下载日志已记录，仅供内部参考

GPT-6 Astra 亮相，ASIC 放量重塑算力格局

——全球产业趋势跟踪周报（0908）

定期报告

本周产业集中在人工智能、半导体。本周海外大模型竞争再度明显升温：OpenAI、Anthropic、谷歌三家头部厂商在极短时间内完成新一轮旗舰或主力模型更新。博通最新财报进一步确认，AI 算力正在加速重构，定制芯片成为 AI 算力扩张的重要环节。

【主题与产业趋势】英伟达业绩指引大幅超预期。本周海外大模型竞争再度明显升温：OpenAI 于 9 月 4 日发布 GPT-6 Astra，Anthropic 于 9 月 2 日推出 Claude Fable 5.1 与 Claude Mythos 5.1，谷歌在同日发布 Gemini 3.8 Flash，三家头部厂商在极短时间内完成新一轮旗舰或主力模型更新。从三个模型的共性趋势看，实际生产效率成为衡量前沿模型价值的重要标准，科研能力是旗舰模型最关注的细分方向之一。与此同时，不同厂商的产品路线也进一步分化：OpenAI 强化计算机操作能力，Anthropic 继续深耕 Agent，谷歌则依靠 Flash 系列强化性能、速度与成本之间的平衡。**ASIC 放量重塑算力格局。**博通最新财报进一步确认，AI 算力正在加速重构，定制芯片成为 AI 算力扩张的重要环节。FY2026 Q3 公司营收创历史新高，并给出未来两个财年的 AI 半导体收入指引，定制 ASIC 进入规模化部署阶段，并逐步成为头部模型厂商和云厂商优化推理成本、能耗和系统效率的重要选择。随着谷歌、Anthropic、OpenAI、Meta 等客户推进多代 XPU 项目，AI 芯片市场正在形成 GPU 与定制 ASIC 并行发展的格局。

【重要资讯】集成电路：财联社 9 月 2 日电，据央行官网，8 月 31 日至 9 月 1 日，二十国集团（G20）在美国阿什维尔召开今年第二次财长和央行行长会议。会议讨论了全球经济形势与前景、促进经济增长、全球失衡、发展中国家主权债务等议题。中国人民银行行长潘功胜出席会议并发言。潘功胜指出，贸易摩擦和保护主义通过影响供应链、推升通胀、扰乱市场预期等因素拖累全球经济。各方应坚定维护多边主义，充分用好 G20 财金合作平台，强化宏观经济政策协同，携手应对全球性风险挑战。解决主权债务问题应坚持“共同行动、公平负担”原则。G20 应发挥引领作用，确保各方在共同框架下有效落实“待遇可比”原则。经济增长是化解发展中国家债务问题、恢复债务可持续性的根本出路。**非农就业：**财联社 9 月 4 日电，美国 8 月非农就业人数增加 16.2 万人，预估为增加 5.5 万人，前值为减少 2.3 万人。劳动力参与率为 61.6%，前一个月为 61.4%。平均每小时工资较前一个月增加 0.3%，预估为增加 0.3%，前一个月为增加 0.2%。其中，私营部门非农就业人数增加 12.7 万人，前一个月为增加 7.1 万人，市场预估为增加 5 万人。制造业非农就业人数增加 1.6 万人，前一个月为增加 1.4 万人；预估为增加 0.5 万人。失业率为 4.1%，前一个月为 4.1%。

【全球股市表现】上周全球股市涨少跌多，其中金融、电信服务、能源表现较好，信息技术、材料、工业表现一般。

风险提示：经济数据不及预期，政策理解不全面，海外政策超预期收紧。

相关报告

1.《英伟达业绩指引大幅超预期，国内大模型密集发布——全球产业趋势跟踪周报（0831）》

2.《mRNA 癌症疫苗实现关键突破，英伟达服务器涨价——全球产业趋势跟踪周报（0824）》

3.《AI 迈向物理世界的载体，国内外厂商发布最新大模型——全球产业趋势跟踪周报（0817）》

4.《创新药出海超预期，国产半导体设备迈向破局之路——全球产业趋势跟踪周报（0811）》

5.《海外厂商云收入保持高增，国产大模型下压价格中枢——全球产业趋势跟踪周报（0803）》

6.《海外科技巨头财报对于 AI 产业指引——全球产业趋势跟踪周报（0727）》

7.《Kimi K3 登顶开源模型第一，超节点建设新机遇——全球产业趋势跟踪周报（0720）》

8.《世界人工智能大会即将召开，关注国产算力投资机会——全球产业趋势跟踪周报（0713）》

李昊阳 S1090524120005
lihaoyang1@cmschina.com.cn
杨晨 S1090526060002
yangchen4@cmschina.com.cn
王德健 S1090526080002
wangdejian@cmschina.com.cn

下载日志已记录，仅供内部参考

正文目录

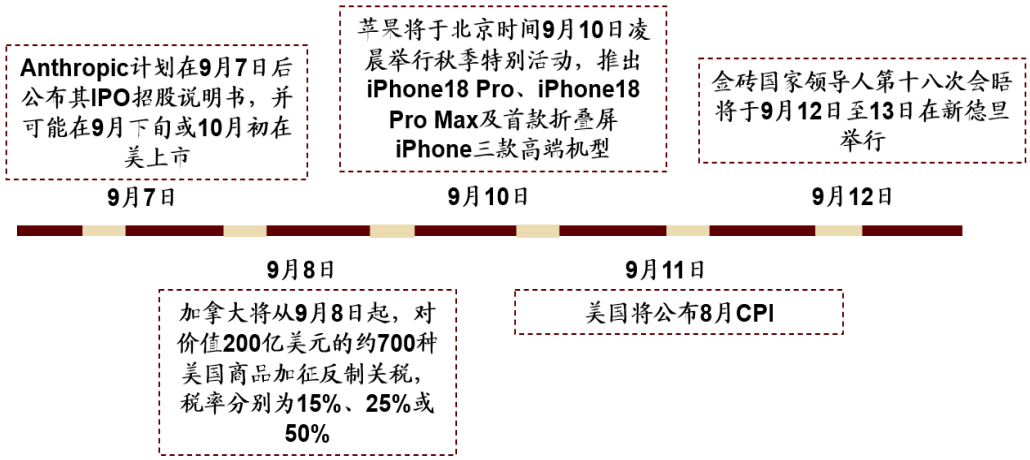
一、核心关注与投资建议 3
1、本周重要事项前瞻 3
2、市场交易热点 3
3、主题与产业趋势变化：OpenAI、Anthropic、谷歌接连发布最新模型，ASIC 放量重塑算力格局 4
（1）OpenAI、Anthropic、谷歌接连发布最新模型 4
（2）ASIC 放量重塑算力格局 10
二、全球观察：全球市场跟踪 14
1、全球股市行业表现 14
2、全球强势股与异动股简析 14
三、重要资讯速递 17
1、国内重要资讯与产业政策 17
2、国外重要资讯与产业政策 19
3、全球重点科技公司跟踪 20

下载日志已记录，仅供内部参考

一、核心关注与投资建议

1、本周重要事项前瞻

图 1：本周重要事件前瞻

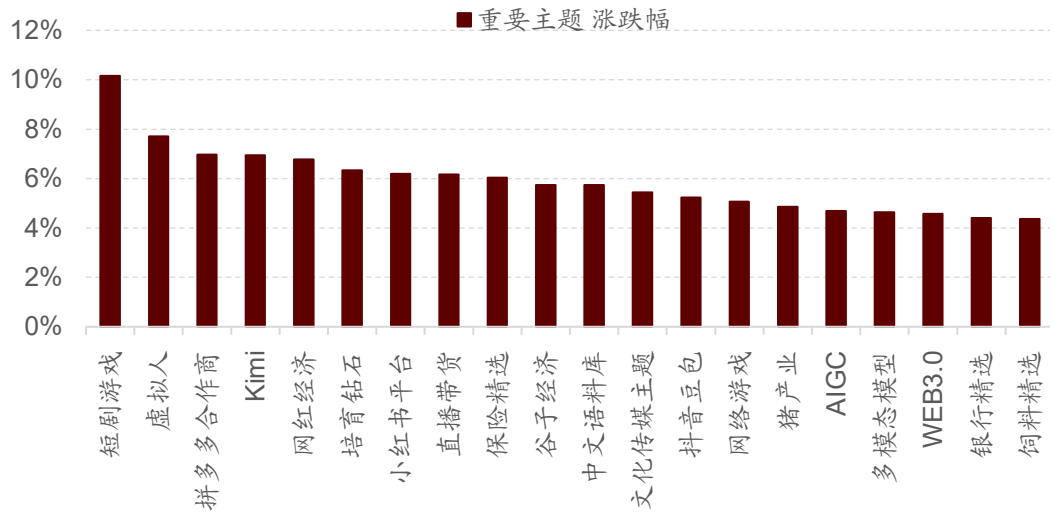


资料来源：Wind，财联社，招商证券

2、市场交易热点

上周，A 股涨幅居前的重要主题指数为短剧游戏、虚拟人、拼多多合作商。

图 2：上周涨幅居前的重要主题指数为短剧游戏、虚拟人、拼多多合作商



资料来源：Wind，招商证券

注：本文所涉及个股均为列示或者代表与相关产业链有关联，不代表个股投资建议，也不代表招商行业研究观点，具体个股推荐请以招商各行业研究组推荐为准。

下载日志已记录，仅供内部参考

3、主题与产业趋势变化：OpenAI、Anthropic、谷歌接连发布最新模型，ASIC 放量重塑算力格局

（1）OpenAI、Anthropic、谷歌接连发布最新模型

本周海外大模型竞争再度明显升温：OpenAI 于 9 月 4 日发布 GPT-6 Astra，Anthropic 于 9 月 2 日推出 Claude Fable 5.1 与 Claude Mythos 5.1，谷歌在同日发布 Gemini 3.8 Flash，三家头部厂商在极短时间内完成新一轮旗舰或主力模型更新。从三个模型的共性趋势看，实际生产效率成为衡量前沿模型价值的重要标准，科研能力是旗舰模型最关注的细分方向之一。与此同时，不同厂商的产品路线也进一步分化：OpenAI 强化计算机操作能力，Anthropic 继续深耕 Agent，谷歌则依靠 Flash 系列强化性能、速度与成本之间的平衡。

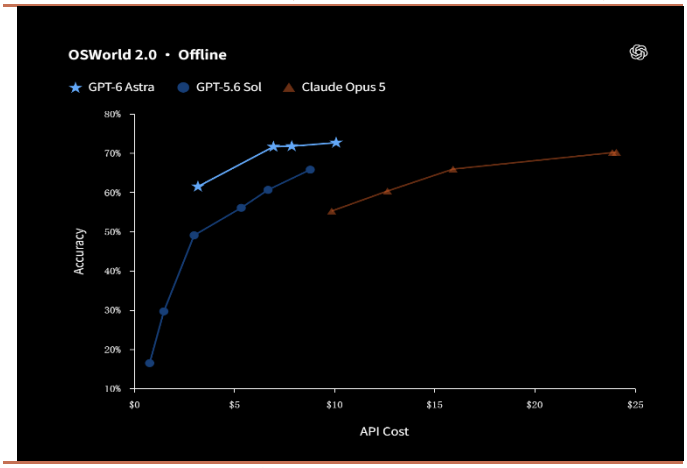
➢ GPT-6 Astra 实现全方位提升

9 月 4 日，OpenAI 正式发布 GPT-6 Astra。OpenAI 总裁发布时直言“欢迎来到 AGI 时代”，这是 OpenAI 历史上规模最大的训练任务，在德克萨斯州园区动用超过 10 万块 GPU 完成了预训练，也是 OpenAI 首款由前代模型深度参与训练监督的旗舰产品。总体看，GPT-6 Astra 在多项基准测试中大幅刷新此前自家旗舰模型 GPT-5.6 Sol 的同时，更超越了其竞争对手 Anthropic 的旗舰模型。

GPT-6 Astra 计算机操作能力出现明显跃升，Agent 开始具备更强的完整任务执行能力。在 OSWorld 2.0 中，GPT-6 Astra 准确率达到 72.6%，高于 GPT-5.6 Sol 的 65.7%，同时单任务模拟执行时间由约 75 分钟降至约 40 分钟，缩短约 47%。配合新版 Codex harness，在 Mind2Web 中任务完成速度达到 GPT-5.6 Sol 当前体验约 1.9 倍。

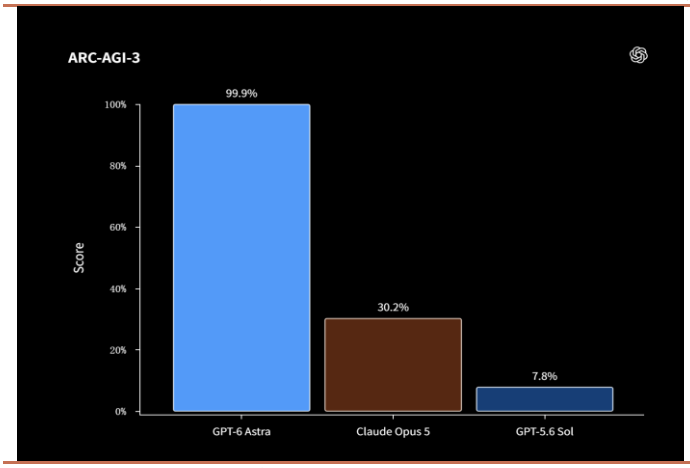
抽象推理与未知环境适应能力方面，GPT-6 Astra 呈现较大的代际提升。衡量陌生环境中通用推理适应能力的 ARC-AGI-3 测试中，GPT-6 Astra 取得 99.9%，而 GPT-5.6 Sol 只有 7.8%，同时远高于 Claude Opus 5，成为此次发布中最醒目的单项能力变化之一。部分测试者称 Astra 在 96%的关卡中超过了人类的行动效率基线，模型在面对新环境时快速理解规则、探索并形成有效行动策略的能力得到大幅增强。

图 3：GPT-6 Astra 计算机操作能力出现明显跃升



资料来源：OpenAI，招商证券

图 4：GPT-6 Astra 未知环境适应能力呈现较大代际提升

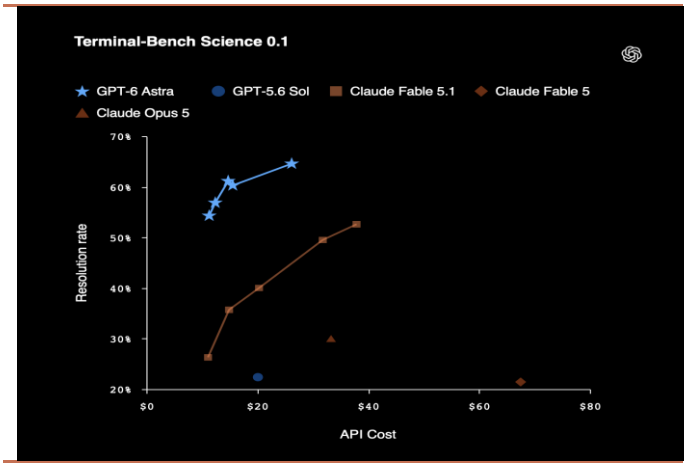


资料来源：OpenAI，招商证券

此外，**GPT-6 Astra 在科研和专业工作自动化方面的提升同样不容小觑。**智能体科研能力测试 Terminal-Bench Science 0.1 中，GPT-6 Astra 在该测试中达到 64.6%，明显高于 GPT-5.6 Sol 的 22.4%，最高水平接近前者的 3 倍，也高于 Claude Fable 5.1 的 52.6%；更重要的是，GPT-6 不仅成功率更高，API 成本还低于 GPT-5.6 和 Claude Fable 5.1，单位成本效率的改善有助于模型在复杂科研任务中的落地。AutomationBench 测试下，GPT-6 的端到端自动化业务能力的优势同样明显，其准确率最高达到 41.4%，相比 GPT-5.6 Sol 最高 18.1%提高 23.3 个百分点，完成率达到此前的 2.3 倍，同时超过 Claude Fable 5.1 的 31.4%。

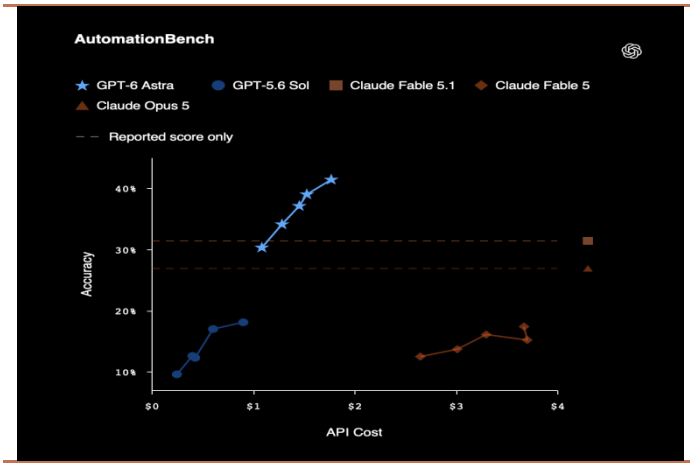
下载日志已记录，仅供内部参考

图 5：GPT-6 Astra 的科研单位成本效率显著改善



资料来源：OpenAI，招商证券

图 6：GPT-6 Astra 端到端自动化业务能力优势同样明显



资料来源：OpenAI，招商证券

价格方面，GPT-6 Astra 的输入、缓存输入和输出单价均为 GPT-5.6 Sol 当前价格的 2.5 倍，OpenAI 此次并未延续前几代旗舰模型持续降价的思路，而是对最高端能力重新建立明显的价格溢价。官方称在部分评测中，Astra 使用更少输出 token 获得更好结果，估算每项任务的 API 成本反而更低。OpenAI 的长远意图不止于提高 token 单价，从按量收费转向按任务完成效果收费是未来大势所趋，商业模式有望演变为基于项目复杂度、耗时或产出价值的定价体系。

表 1：GPT-6 Astra 的输入、缓存输入和输出单价均为 GPT-5.6 Sol 当前价格的 2.5 倍（单位：美元/百万 tokens）

模型	输入	缓存输入	输出
GPT-6 Astra	10	1	50
GPT-5.6 Sol	4	0.4	20

资料来源：OpenAI，招商证券

此次 GPT-6 Astra 依托于 OpenAI 迄今为止规模最大的算力投入，在得克萨斯州 Stargate 基地使用超过 10 万块 GPU 完成训练。这一数字不仅远超 GPT-5.6 Sol，也反映出规模化训练仍是提升模型复杂推理和长任务执行能力的关键。大规模并行计算使得 Astra 能够处理更长的上下文、更多步骤的连锁操作，并在真实软件环境中保持稳定表现，高昂的硬件成本与能源消耗，也直接推高了后续 API 定价，反映头部 AI 企业在基础设施上的军备竞赛仍在延续。

此外，随着 GPT-6 Astra 自主推理能力增强，其安全监控面临新挑战。其书面推理过程变得更加简洁和隐蔽，部分任务中可见的推理步骤显著减少，导致人类难以追踪模型的决策逻辑，OpenAI 虽然已引入更严格的拒答机制、实时监控系统和任务中止功能，但同时承认模型越强，完全监控的难度越高。这一矛盾引发深层忧虑：不仅要防范模型被恶意用于网络攻击，还要解决如何理解模型为何做出特定行为的问题，自主性与透明度之间的张力，将成为下一代 AI 治理的核心议题之一。

➢ Fable 5.1 提升聚焦 Agent 和科研

相比 Fable 5，Fable 5.1 的主要增量集中在三个方面：复杂 Agent 任务完成率明显提高，科研任务能力出现跨级提升；通过缓存价格下降改善高频 Agent 工作负载的经济性；提高了安全机制的精确度，使原本受到较多限制的网络安全和生命科学能力能够在安全边界内释放更多实际效用。

从官方 Benchmark 看，Agent 能力依旧为本轮升级最明确的主线，任务完成率明显提高。Terminal-Bench 4.0 主要考察模型在终端环境中持续操作、编写和调试代码的能力，Fable 5.1 得分达到 55.8%，较 Fable 5 提高 13.8 个百分点；AutomationBench 则模拟真实企业中的多步骤业务流程，成绩从 17.1%提高到 31.4%，接近翻倍，且两个测试结果均大幅超越 GPT-5.6 Sol。此类测试共同涉及任务拆解、工具调用及执行反馈，相比单轮问答 Benchmark，更能反映 Agent 进入真实生产环境后的可用性，通过早期真实用户测试进一步体现了这种提升。

下载日志已记录，仅供内部参考

图 7：Agent 能力和科研任务是本次 Fable 5.1 最大的升级点

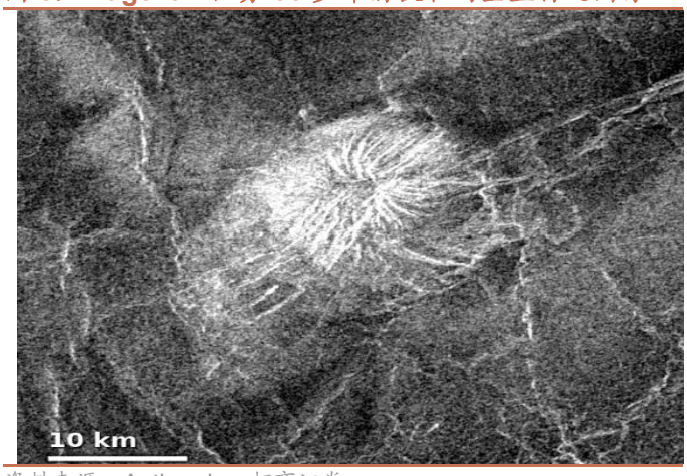
	Fable 5.1	Fable 5	Opus 5	GPT-5.6 Sol
Agentic scientific research Terminal-Bench-Science 0.1 [1]	**52.6%**	24.7%	29.0%	22.4%
Agentic coding Terminal-Bench 4.0	**55.8%** 60.9% (Mythos 5.1)	42.0%	52.3%	37.3%
Knowledge work GDPval-AA v2	**1853**	1723	1824	1711
Computer use OSWorld 2.0 [2]	**77.9%** partial	72.9% partial	75.4% partial	— partial
Computer use OSWorld 2.0	**41.7%** strict	36.1% strict	39.6% strict	— strict
Multidisciplinary reasoning Humanity's Last Exam	**60.9%** no tools	57.8% no tools	56.6% no tools	— no tools
	65.0% with tools	63.8% with tools	63.6% with tools	— with tools
Business workflows AutomationBench	**31.4%**	17.1%	26.9%	19.6%
Agentic coding CursorBench 3.2.0	**73.4%**	70.5%	70.0%	67.2%

资料来源：Anthropic，招商证券

科研任务是 Fable 5.1 相较 Fable 5 提升同样显著的领域。Terminal-Bench-Science 0.1 主要测试模型能否在科研环境中理解问题后完成连续分析，Fable 5.1 成绩由上一代的 24.7%提升至 52.6%，超过翻倍，Opus 5 为 29.0%、GPT-5.6 Sol 仅为 22.4%。结合 Humanity's Last Exam 等偏知识和通用推理测试的提升幅度远小于科研 Agent 测试，表明此次 Anthropic 模型的关注点更多是把已有知识、推理和工具组合起来解决现实科研任务的能力。

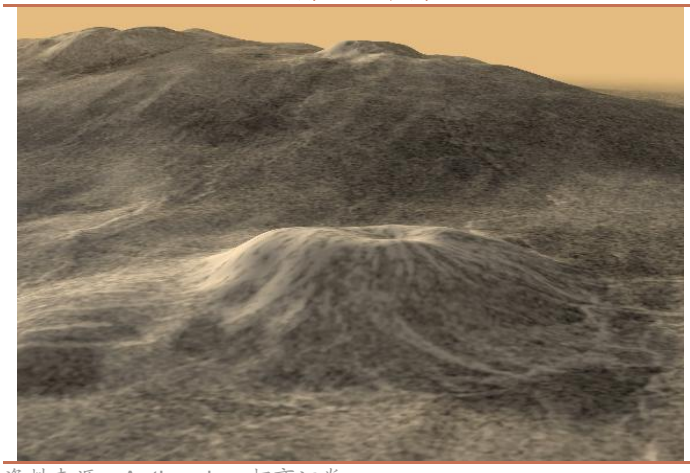
Anthropic 此次还首次大规模展示了新模型直接形成科研结果的案例。在生命科学领域，Mythos 5.1 调用开源蛋白设计和折叠工具进行蛋白 binder 设计，在 12 个靶点上的有效设计命中率接近 50%，而当前蛋白设计典型命中率约为 10%-15%；其中三个靶点的结合亲和力达到 Adaptyv Bio 蛋白设计竞赛最佳方案约 10 倍。在计算分析和建模方面，Anthropic 使用 Fable 5.1 基于 NASA Magellan 任务 30 多年前获取的金星雷达图像，以及此前只覆盖约五分之一金星表面的既有地图，自主训练神经网络并构建新的数字高程模型，最终模型生成了覆盖约三分之一金星表面的高分辨率高程图，将原先约 10-20 公里的有效细节尺度提高到 2-3 公里，高程准确度最高改善约 25%。

图 8：Magellan 任务 30 多年前获取的金星雷达图像



资料来源：Anthropic，招商证券

图 9：Fable 5.1 生成的高分辨率高程图



资料来源：Anthropic，招商证券

此外，Anthropic 进一步提高网络安全防护机制的精确度，Fable 5.1 允许模型用于软件漏洞识别后，Claude Code 单次会话中由网络安全防护机制触发的干预次数平均较 Fable 5 原有机制减少约 60%，在维持高风险任务限制的同时提高了防御性网络安全任务的可用性；针对基础生物学和医疗等无害请求的安全机制触发率，相比 Fable 5 发布时下降约 85%。

相比 Fable 5，Fable 5.1 并未降低基础 API 单价，输入和输出价格仍分别为 10 美元/百万 tokens 和 50 美元/百万 tokens；本轮价格优化主要集中在缓存输入，缓存读取价格降至 0.25 美元/百万 tokens，较 Fable 5 下降 75%，Anthropic 预计，由此可以使典型工作负载总体成本下降约 25%，高度 Agent 化的任务成本最高下降约 45%。尽管

下载日志已记录，仅供内部参考

Anthropic 将其旗舰模型价格下调，但整体价格相比其竞争对手 OpenAI 的旗舰模型 GPT-5.6 Sol 仍然偏高，与 GPT-6 Astra 接近。

表 2：Claude Fable 5.1 缓存输入相较 Claude Fable 5 发布时下降 75%（单位：美元/百万 tokens）

模型	输入	输出	缓存输入
Claude Mythos 5.1	10	50	-
Claude Fable 5.1	10	50	0.25
Claude Fable 5	10	50	1
Claude Opus 5	5	25	0.5
GPT-5.6 Sol	4	20	0.4
GPT-6 Astra	10	50	1

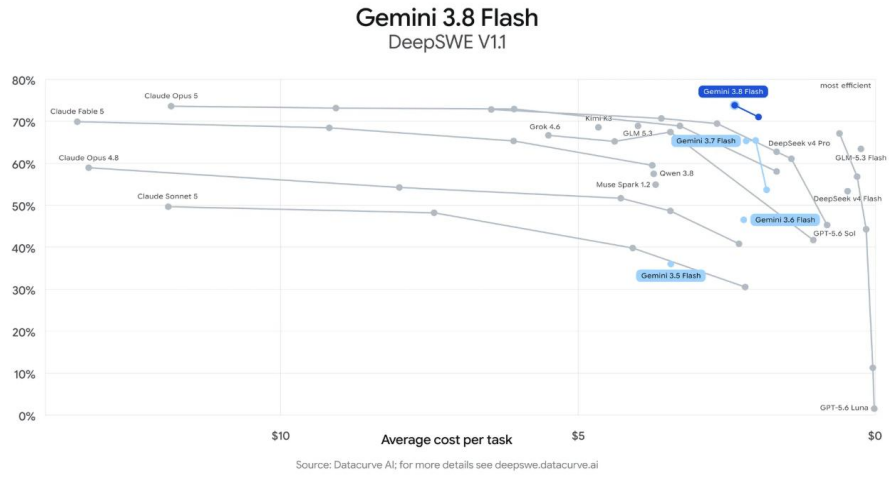
资料来源：Anthropic，OpenAI，招商证券

➢ 长程软件工程成为 Gemini 3.8 Flash 最大升级方向

Gemini 3.8 Flash 最突出的提升在于长程软件工程任务。在 DeepSWE v1.1 测试中，Gemini 3.8 Flash high 取得 73.7%的成绩，较 Gemini 3.7 Flash 此前约 65.3%的水平明显提高，与 Claude Opus 5 基本处于同一水平，略高于 GPT-5.6 Sol。DeepSWE 测试的重点是让模型自主解决完整工程问题，需要经历理解代码库、修改文件、调用工具、测试结果以及继续修正等过程。根据 DeepSWE 数据，Gemini 3.8 Flash 平均单任务成本仅约 2.36 美元，Claude Opus 5 为 11.84 美元、GPT-5.6 Sol 为 6.46 美元，说明谷歌此次提升并未建立在显著提高调用成本。

此外，Gemini 3.8 Flash 专业知识工作能力继续提高，金融和法律方面的解决能力已经进入前列。在 Vals Finance Agent v2 中，Gemini 3.8 Flash 得分达到 61.4%，高于 Claude Opus 5 的 58.6%、GPT-5.6 Sol 的 53.8%；衡量长流程法律工作能力的 Harvey Legal Agent Benchmark 中达到 10.0%，高于 Gemini 3.7 Flash 的 8.8%、Claude Opus 5 的 6.7%。

图 10：在 DeepSWE V1.1 测试中，Gemini 3.8 Flash 性价比步入前沿水平

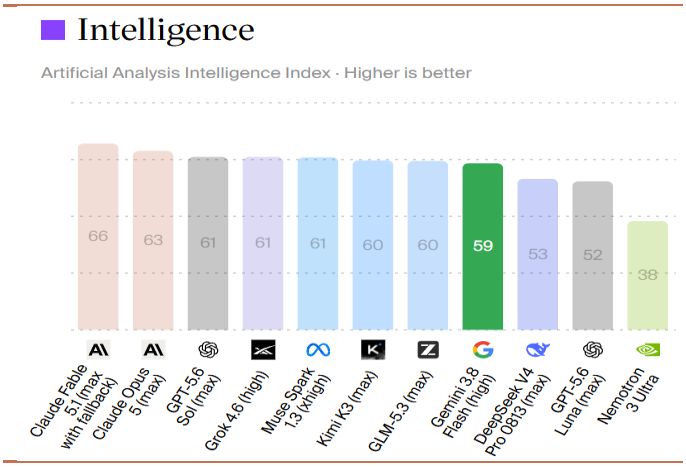


资料来源：谷歌，招商证券

Gemini 3.8 Flash 的核心优势是在高智能水平下，仍然维持极高输出速度。根据 Artificial Analysis 实测数据，Gemini 3.8 Flash 得分为 59，虽然低于 Claude Fable 5、GPT-5.6 Sol 等最前沿模型，但已经进入第一梯队，其综合推理、知识和任务处理能力并未因 Flash 定位而明显牺牲。更突出的优势体现在速度端，其输出速度达到 302 tokens/s，位列所有模型第一，较第二名 Muse Spark 的 209 tokens/s 快约 45%，约为 GPT-5.6 Luna 的 2.5 倍。谷歌当下正在显著强化其性能、速度兼顾的模型路线，Gemini 3.8 Flash 正是最好的证明，其虽然并非智能水平最高的模型，但在接近前沿能力的同时显著提高 token 生成吞吐速度，更适合实时交互及大规模高频调用场景。

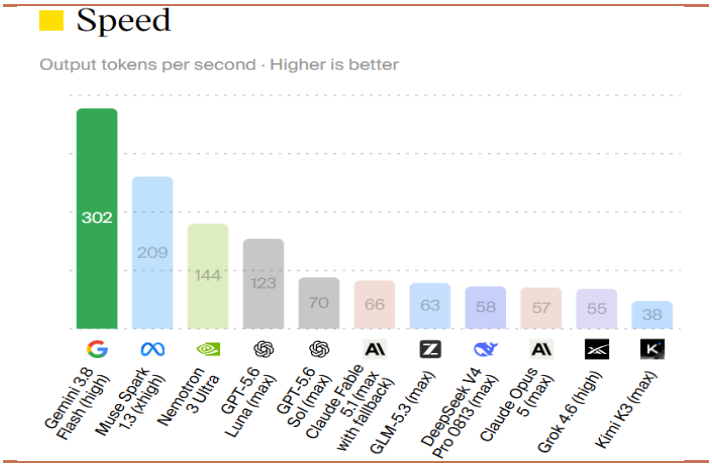
下载日志已记录，仅供内部参考

图 11：Gemini 3.8 Flash 进入高智能行列



资料来源：Artificial Analysis，招商证券

图 12：Gemini 3.8 Flash 输出速度维持优势



资料来源：Artificial Analysis，招商证券

Gemini 3.8 Flash 当前定价为输入 0.75 美元/百万 tokens、输出 3.75 美元/百万 tokens，整体处于海外主流高性能模型的较低价格区间。横向比较，Claude Sonnet 5 为输入 2 美元/百万 tokens、输出 10 美元/百万 tokens，GPT-5.6 Terra 输入价格 2 美元/百万 tokens、输出价格 12 美元/百万 tokens；与国产开源模型相比，DeepSeek V4 Flash、GLM-5.3-Flash 的 token 价格仍更低。Gemini 3.8 Flash 价格定位介于海外旗舰模型 Flash 版本和国内开源模型 Flah 版本中介区间，其优势在于以相对克制的定价覆盖更广泛的调用需求，在海外主流模型中形成较突出的性价比位置，有利于降低企业大规模 API 调用的成本压力。

表 3：Gemini 3.8 Flash 与同定位模型价格对比（单位：美元/百万 tokens）

厂商	模型	输入	缓存输入	输出
谷歌	Gemini 3.8 Flash	0.75	0.075	3.75
OpenAI	GPT-5.6 Luna	0.2	0.02	1.2
OpenAI	GPT-5.6 Terra	2	0.2	12
Anthropic	Claude Sonnet 5	2	-	10
DeepSeek	DeepSeek V4 Flash	0.22（空闲）、0.44（高峰）	0.007（空闲）、0.014（高峰）	0.66（空闲）、1.32（高峰）
智谱	GLM-5.3-Flash	0.075	0.015	0.25

资料来源：谷歌，Anthropic，OpenAI，DeepSeek，DeepSeek，智谱，招商证券

➢ 主流海外模型出现降价趋势，开源模型维持 token 调用重要主力

近期海内外大模型定价策略逐步趋同，模型价差逐渐缩小：OpenAI 在 7 月底至 8 月连续调整 GPT-5.6 系列价格，先对 Terra 和 Luna 进行降价，随后进一步下调 Sol 价格，覆盖高中低不同模型档位；谷歌在 8 月中旬下调 Gemini 3.6 Flash 价格，并以阶段性优惠方式强化 Flash 系列的性价比优势。相比之下，DeepSeek 于 8 月 17 日上调 V4 Flash 和 V4 Pro 价格，同时引入峰谷定价机制，通过时段差异改善算力利用效率。

整体来看，海外头部厂商当前更倾向于通过主动降价扩大 API 调用规模，以此应对国内模型厂商的低价策略，提高开发者和企业用户渗透率；DeepSeek 的涨价则反映出高性能国产模型在需求增长后，开始更加重视算力供需与商业化效率。当下模型能力持续提升的同时，企业、科研机构对成本上升愈发敏感，因此在需求端尚未出现大规模应用放量的背景下，延续此前的高定价并不明智，海外模型厂商高价定价策略今年以来受到国内低价模型厂商的冲击逐步显现，迫使海外厂商主动降价。

表 4：近期部分大模型执行价格调整（单位：美元/百万 tokens）

调整时间	厂商	模型	调整前价格	调整后价格	调整幅度
2026/7/30	OpenAI	GPT-5.6 Terra	输入 2.50；缓存输入 0.25；输出 15.00	输入 2.00；缓存输入 0.20；输出 12.00	整体降价 20%
2026/7/30	OpenAI	GPT-5.6 Luna	输入 1.00；缓存输入 0.10；输出 6.00	输入 0.20；缓存输入 0.02；输出 1.20	降价 80%

下载日志已记录，仅供内部参考

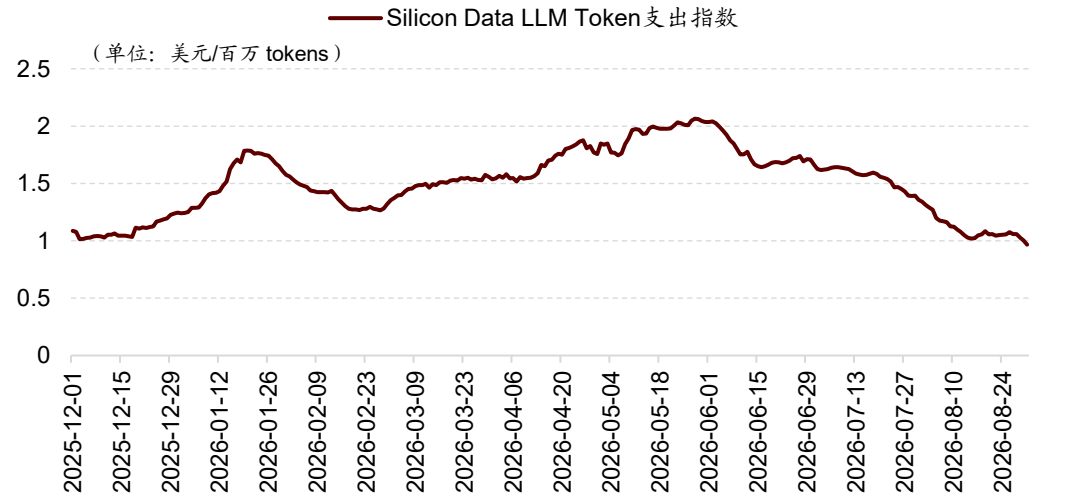
2026/8/13	谷歌	Gemini 3.6 Flash	输入 1.50；输出 7.50	输入 0.75；输出 3.75	降价 50%，作为阶段性优惠价格执行至 2026 年底
2026/8/17	DeepSeek	DeepSeek V4 Flash	输入 0.14；缓存输入 0.0028；输出 0.28	闲时 0.22 /0.007 /0.66；高峰：0.44 /0.014 /1.32	整体涨价并引入峰谷定价；闲时为高峰价格一半
2026/8/17	DeepSeek	DeepSeek V4 Pro	输入 0.435；缓存输入 0.0036；输出 0.87	闲时：0.66 /0.022 /1.98；高峰：1.32 /0.044 /3.96	整体涨价并引入峰谷定价；闲时为高峰价格一半
2026/8/21	OpenAI	GPT-5.6 Sol	输入 5.00；输出 30.00	输入 4.00;输出 20.00	输入降 20%，输出降 33%，阶段性优惠持续 3 个月

资料来源：OpenAI，谷歌，DeepSeek，招商证券

Token 价格方面，衡量市场实际使用结构下每百万 tokens 的有效支出水平的 Silicon Data LLM Token 支出指数，自 5 月底高点以来已回落超 50%。该指数经历“去年年底起上升-今年 5 月见顶-6 月以来快速下行”三个阶段：2025 年 12 月初指数约为 1.09 美元/百万 tokens，此后明显抬升，2026 年 1 月已经进入 1.5 美元/百万 tokens 附近；2 月阶段性回落后，3-5 月重新上行，并于 5 月 28 日达到 2.07 美元/百万 tokens 的高点；此后持续回落，截至 8 月 31 日已经降至 0.97 美元/百万 Token，较 5 月高点累计下降约 53.2%。

根据指数的走势，可以观察出实际使用场景下模型支出已然形成较清晰的下降趋势，这一走势主要受到模型定价和实际调用结构共同变化影响。由于 Silicon Data LLM Token 支出指数本身按照真实市场使用量加权，因此上半年指数上升，可以理解为用户对高性能推理模型和 Agent 任务的使用增加，高价模型及更长输出在实际消费中的权重提高，带动每百万 tokens 平均支出上行；进入下半年后，供给端的降价明显加速，低价且能力更强的模型进入市场并获得青睐，模型使用组合逐渐向低成本方案倾斜。

图 13：Silicon Data LLM Token 支出指数自 5 月底高点回落超 50%



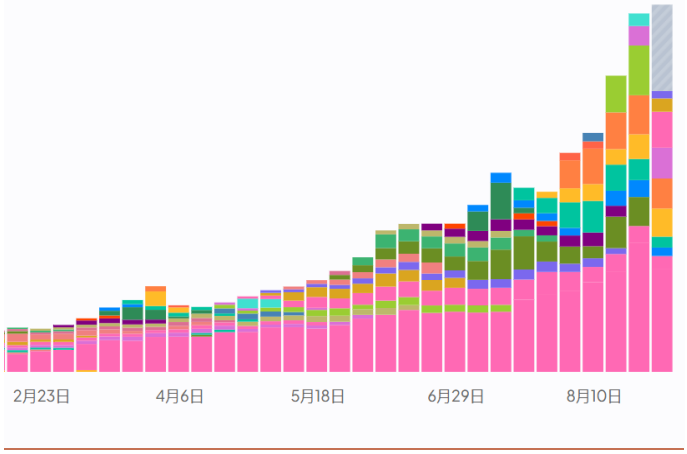
资料来源：Bloomberg，招商证券

Token 调用量层面，OpenRouter 平台总 token 调用量自 2025 年底以来仍高速扩张，当下 token 调用量已经呈现指数级增长。截至 2026 年 8 月 30 日的数据显示，全球周调用量已经达到 113 万亿 token，相比去年同期增长约 22 倍，相比年初增长约 16 倍。模型价格下降趋势，推动 token 消费进一步增长，更多企业能够将 AI 嵌入日常业务流程，闭源模型利润率可能受到压缩，但整个 AI 产业链的价值空间仍有望扩大。

开源及开放权重模型已经成为 OpenRouter 月度 token 调用的重要主力。其中以 GLM、DeepSeek、混元、MiniMax 为代表的开源大模型凭借极致的性价比，在 OpenRouter 平台的周调用量多次领跑。国产模型份额显著提升，最新周度数据 Top10 中，中国厂商模型占 7 席，对应 Token 量约 52.22 万亿，占 Top10 合计约 73%，国内模型与海外前沿模型的能力差距缩小后，竞争已经开始体现在真实开发者调用量上，高性价比模型放量速度明显更快，价格与性能兼顾的特性更容易承接大规模推理需求。

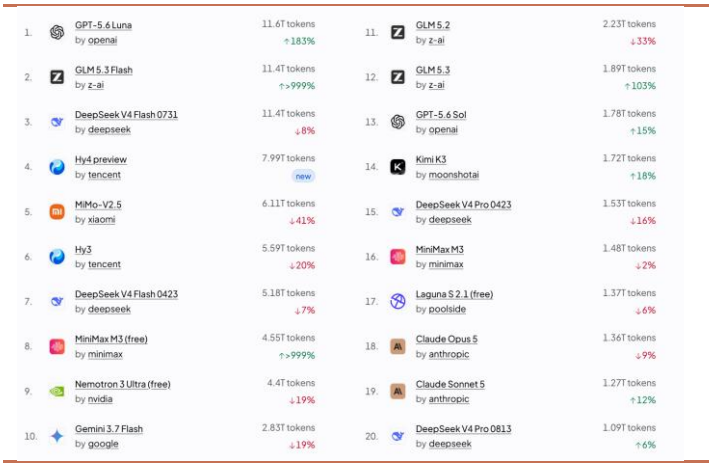
下载日志已记录，仅供内部参考

图 14：全球周 token 调用量已经达到 113 万亿



资料来源：OpenRouter，招商证券

图 15：开源模型成为月度 token 调用的主力



资料来源：OpenRouter，招商证券

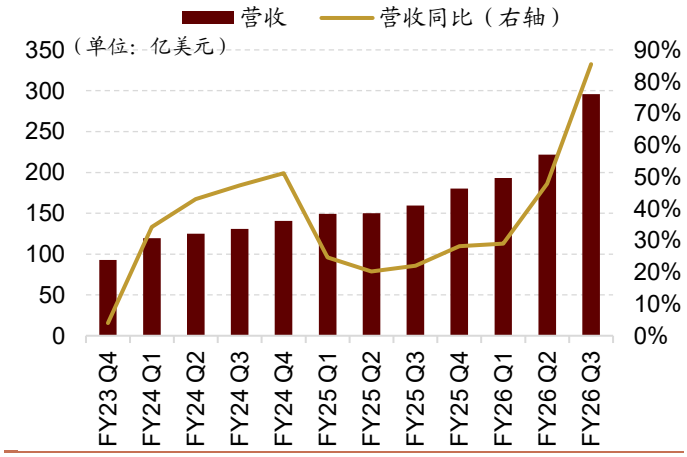
（2）ASIC 放量重塑算力格局

博通最新财报进一步确认，AI 算力正在加速重构，定制芯片成为 AI 算力扩张的重要环节。FY2026 Q3 公司营收创历史新高，并给出未来两个财年的 AI 半导体收入指引，定制 ASIC 进入规模化部署阶段，并逐步成为头部模型厂商和云厂商优化推理成本、能耗和系统效率的重要选择。随着谷歌、Anthropic、OpenAI、Meta 等客户推进多代 XPU 项目，AI 芯片市场正在形成 GPU 与定制 ASIC 并行发展的格局。

➢ 博通 FY2026 Q3 营收创历史新高，毛利率环比小幅回落

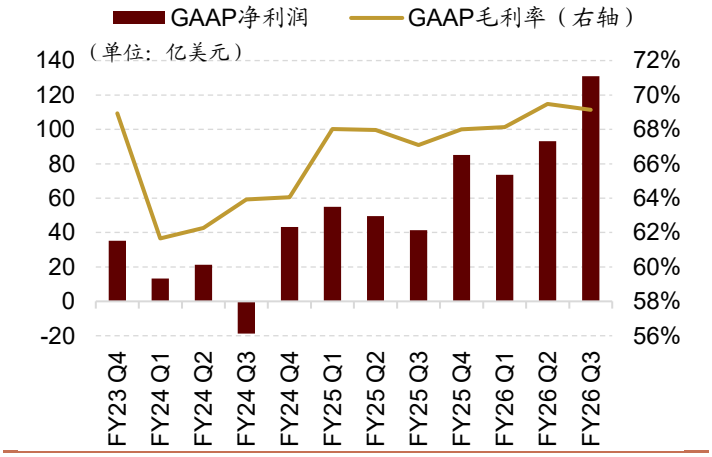
博通 FY2026 Q3 实现营收 295.91 亿美元，同比增长 86%，环比增长 33%，再创单季历史新高；GAAP 净利润 130.88 亿美元，同比增长 216%；Non-GAAP 净利润 163.72 亿美元，同比增长 95%，盈利增速明显快于收入。公司 Q3 财季 GAAP 毛利率约 69.13%，较 FY2026 Q2 的 69.48%小幅下降，环比回落主要源于 AI 半导体收入占比提升，毛利率相对较低的定制 XPU 占比持续提升，而高毛利的软件业务占比相应下降，对综合毛利率形成结构性稀释，但收入大幅增长仍支撑利润维持较高水平。

图 16：博通 FY26 Q3 实现营收 295.91 亿美元



资料来源：博通，iFind，招商证券

图 17：博通 FY26 Q3 GAAP 净利润 130.88 亿美元



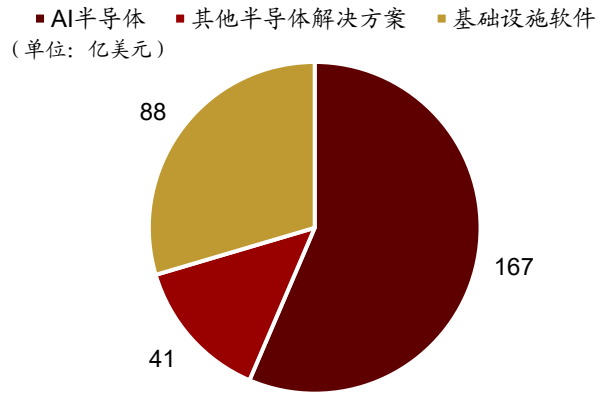
资料来源：博通，iFind，招商证券

业务结构上，博通最新收入结构已经进一步向 AI 半导体业务倾斜。FY2026 Q3 半导体解决方案收入 208.39 亿美元，同比增长 127%，占总收入 70%；基础设施软件收入 87.52 亿美元，同比增长 29%，占比 30%。半导体内部的分化更加明显：AI 半导体收入 167 亿美元，同比增长 221%、环比增长 54%，已经占公司总收入约 56%；公司披露其 XPU 出货量同比增长超过 3.5 倍，占 AI 收入约 73%，AI 网络收入同比增长超过 2.5 倍，AI 加速器与网络同步快速扩张。相比之下，非 AI 半导体收入约 41 亿美元，同比增长 5%、环比基本持平，宽带和服务器存储有所增长，无线业务形

下载日志已记录，仅供内部参考

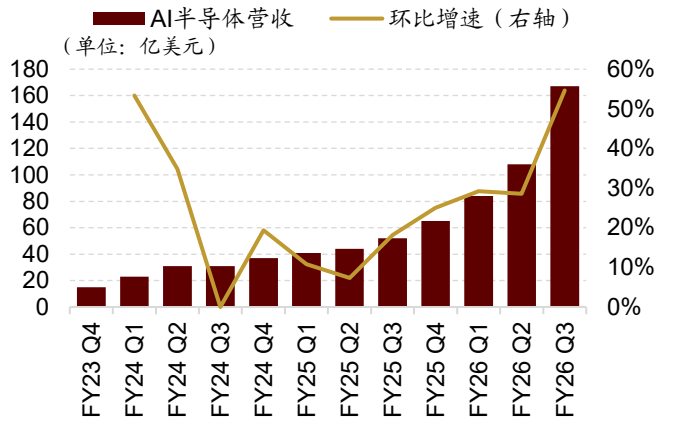
成部分拖累。软件业务继续保持稳健增长，收入达 88 亿美元，同比增长 29%，ARR 同比增长 15%，VMware 整合后的收入基础已经趋于稳定。

图 18：博通最新收入结构已经进一步向 AI 半导体倾斜



资料来源：博通，招商证券

图 19：FY26 Q3 博通 AI 半导体收入 167 亿美元



资料来源：博通，招商证券

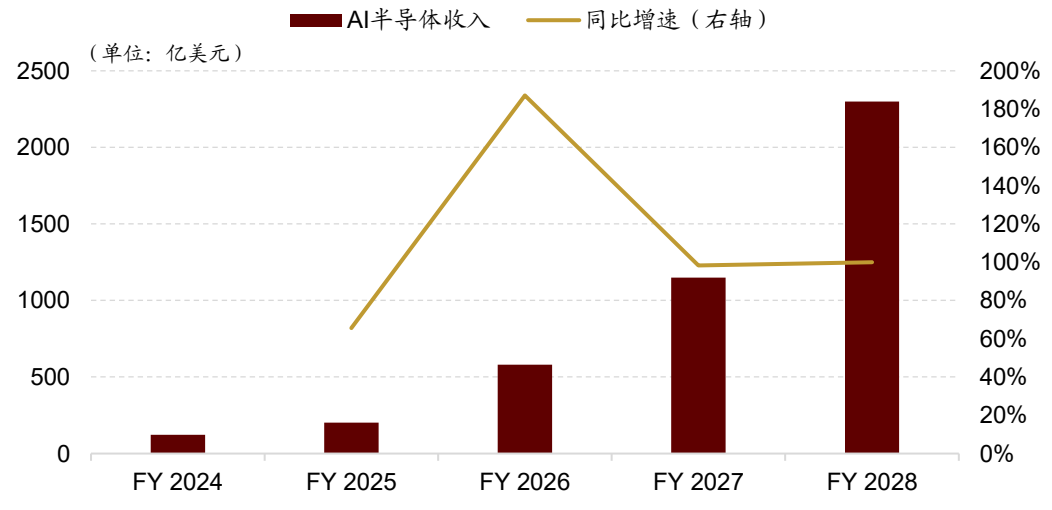
➢ 博通罕见给出未来两个财年的 AI 半导体收入指引

博通预计下一季度总营收达到约 348 亿美元，同比增长 93%，收入增长速度较 Q3 的 86%进一步提高。其中半导体解决方案收入预计约 261 亿美元，同比增长 136%，仍然是整体增长的核心；基础设施软件收入预计约 87 亿美元，同比增长 25%，环比基本稳定。AI 半导体收入将由 Q3 的 167 亿美元进一步升至 217 亿美元，同比增长 236%，XPU 和 AI networking 收入均预计同比增长约 3 倍；非 AI 半导体预计约 43 亿美元，整体维持平稳。

盈利能力方面，由于低于软件业务毛利率的 XPU 收入占比继续快速提高，同时新一代 XPU 包含更多 HBM 等存储器件，公司预计 Non-GAAP 综合毛利率由 Q3 的 75%下降至约 73%，但依靠收入规模带来的经营扩张，Non-GAAP 营业利润率仍预计维持约 66%。此外，公司预计 Q4 资本开支增加至 14 亿美元，主要用于扩大半导体相关产能。

根据 Q4 指引，博通预计 FY2026 全年 AI 半导体收入达到约 580 亿美元，同比增长 186%，高于此前 560 亿美元的预测。管理层进一步给出较为罕见的两年期明确指引，预计 FY2027 AI 半导体收入约 1150 亿美元，FY2028 进一步达到约 2300 亿美元，即未来两年均接近翻倍增长。公司表示 1150 亿美元的 FY2027 目标已经获得相应供应保障，当前客户实际需求甚至高于这一指引，主要由 Google、Anthropic、OpenAI 和 Meta 等客户的 XPU 部署以及配套 AI networking 需求支撑。

图 20：博通预计 FY27 AI 半导体收入约 1150 亿美元，FY28 进一步达到约 2300 亿美元



资料来源：博通，招商证券

下载日志已记录，仅供内部参考

➢ ASIC 业务从单一大客户驱动转向多客户，AI 网络和光互联成为经营核心抓手

XPU 客户结构进一步清晰，ASIC 业务从单一大客户驱动转向多客户、多代际项目推进。博通管理层披露公司已有 **6 个 XPU 客户**，并说明了部分客户的部署进展：博通已向 Anthropic 和谷歌大规模交付 Ironwood TPU v7、向谷歌出货 TPU v8i，并交付 OpenAI 第一代定制加速器 Jalapeño，Meta MTIA 也进入后续出货节奏。在头部云厂商和前沿模型公司正在用自研芯片获取更好的性能功耗比和成本结构背景下，博通 AI 半导体业务已经围绕核心 AI 客户建立多代 XPU 开发、供应和网络配套。

AI 网络和光互联成为公司经营扩张的核心抓手之一。公司管理层表示，AI networking 收入未来几年预计保持与 XPU 相当的增长速度，产品上覆盖 Tomahawk 交换芯片、PCIe switching、optical DSP、EML、VCSEL、CW laser 等环节，围绕 AI 集群形成“XPU 计算芯片+以太网交换+高速光互联”的系统级供应能力，经营重心拓展至 AI 数据中心互联架构。

此外，博通开始参与 AI 基础设施融资，经营角色从芯片供应商延伸到算力项目推动方。CFO 披露，公司在 6 月与 Apollo、Blackstone 建立 AI XPV 平台，目标支持 OpenAI 和 Anthropic 到 2028 年底超过 20GW 的计算基础设施，首期融资规模为 350 亿**美元**，相关基础设施预计从 2026 年中开始部署在 Fluidstack 相关站点。博通正在帮助部分战略客户解决 AI 算力建设中的资金错配问题，但融资安排也带来一定风险，融资平台有助于提高 AI 收入可见度的同时，使博通在部分客户基础设施项目中承担了表外或有责任。

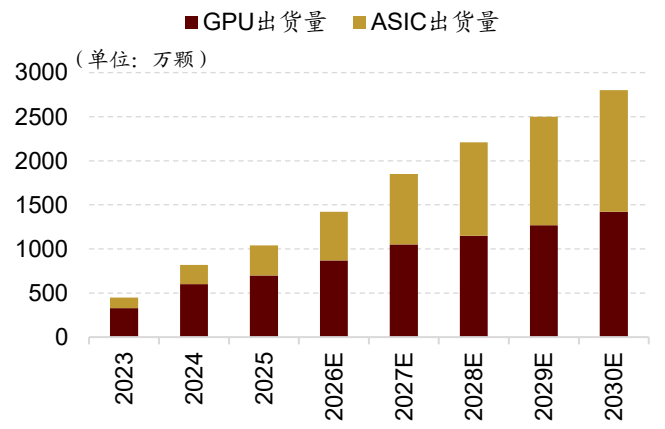
➢ ASIC 有望显著放量

当下 AI 加速器供给正在形成 GPU 之外的定制 ASIC 主线。AI 加速器市场上 ASIC 的出现并非为了全面替代 GPU，GPU 会继续向高价值训练场景升级，ASIC 在高频推理和云厂商自有负载中快速放量。全球 AI 加速器逐渐呈现出两个趋势：出货量层面，ASIC 数量快速提升；份额方面，GPU 仍保持绝对主导。

出货量层面，全球 AI 加速器出货量的核心趋势是总量持续扩张、ASIC 增速更快，出货结构有望更加均衡：根据 Bloomberg Intelligence 预测，GPU 和 AI ASIC 合计出货量从 2024 年约 800 万颗量级，提升到 2030 约 2800 万颗；GPU 出货占比从 2023 年的 74%降至 2030 年约 51%，并在此后维持 50%附近，2025-2030 年 CAGR 约为 15%；ASIC 成为核心增量，出货占比从 2023 年的 26%提升到 2030 的 49%，2025-2030 年 CAGR 也明显高于 GPU，达到 32%。

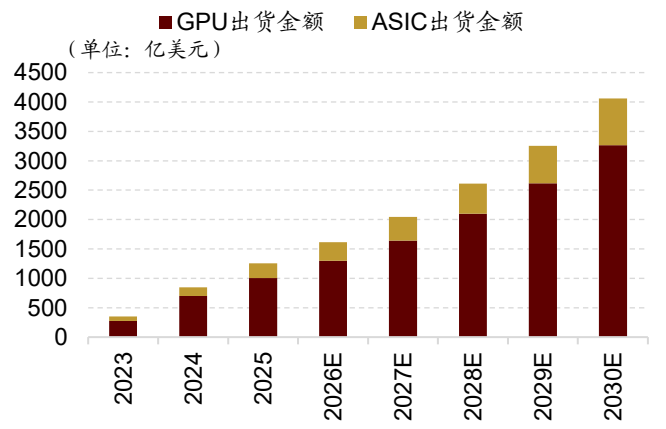
份额方面，根据 IDC 预测，GPU 出货金额从 2023 年的 277 亿美元增至 2030 年的 3263 亿美元，ASIC 从 2023 年的 74 亿美元增至 2030 年的 795 亿美元，GPU 金额占比长期维持在约 80%，ASIC 虽然在颗数上快速追赶，但单颗价值量仍低于 GPU。

图 21：ASIC 出货结构有望更加均衡



资料来源：Bloomberg Intelligence，招商证券

图 22：GPU 市场份额占比将长期维持在约 80%



资料来源：IDC，招商证券

ASIC 放量显著，核心原因是 AI 推理需求规模化以后，头部云厂商更重视单位算力的成本、能耗以及对自身架构的适配程度。GPU 通用性强，适合训练、预训练等多类负载复用；ASIC 则更适合谷歌、亚马逊、Meta、OpenAI 等超大客户的稳定、高频的重复性推理任务，公司可围绕特定模型、内存带宽、网络互联和功耗约束做定制优化。但先进

下载日志已记录，仅供内部参考

制程 ASIC 需要高额前期投入和庞大的芯片设计团队，Bloomberg Intelligence 提到先进节点 ASIC 项目前期投入通常可达 10 亿-30 亿美元，从设计到流片需要 2-3 年，因此只有少数拥有大规模内部算力需求、资本开支能力和研发资源的头部厂商能够长期推进。

下载日志已记录，仅供内部参考

二、全球观察：全球市场跟踪

1、全球股市行业表现

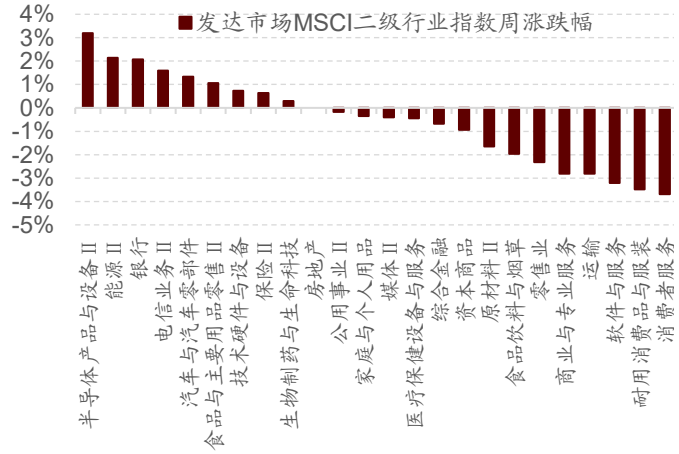
上周全球股市涨少跌多，其中金融、电信服务、能源表现较好，信息技术、材料、工业表现一般。具体来看，**美股涨少跌多**，能源、信息技术、公共事业表现较好，可选消费、材料、房地产表现一般；**欧股涨少跌多**，能源、金融表现较好，可选消费、工业、公共事业表现一般；**英国股市涨少跌多**，能源、金融、电信服务表现较好，工业、公共事业、信息技术表现一般；**日本涨少跌多**，公共事业、电信服务、金融表现较好，信息技术、材料、工业表现一般；**港股涨少跌多**，金融、电信服务表现较好，材料、工业、信息技术表现一般；**A 股涨跌参半**，电信服务、金融、日常消费表现较好，信息技术、材料、能源表现一般。

表 5：全球股市行业表现（2026/08/31-2026/09/04）

	美国	欧洲	英国	日本	中国香港	A 股
能源	2.3%	2.0%	3.4%	-0.4%	-0.6%	-2.3%
信息技术	1.1%	-1.9%	-2.0%	-4.2%	-2.4%	-4.2%
公共事业	0.7%	-2.1%	-2.4%	4.9%	-0.3%	0.4%
医疗保健	0.2%	-0.2%	-0.4%	-2.0%	-2.1%	-1.9%
金融	0.0%	0.8%	1.6%	2.5%	4.6%	2.8%
电信服务	-0.5%	-0.4%	1.5%	3.0%	0.4%	4.8%
日常消费	-0.9%	-0.5%	-0.2%	-1.6%	-0.5%	1.9%
工业	-1.1%	-2.9%	-3.1%	-2.0%	-2.7%	-1.4%
房地产	-1.3%				-2.0%	-0.9%
材料	-1.6%	-0.2%	-0.8%	-3.4%	-3.4%	-4.2%
可选消费	-2.1%	-3.1%	-1.1%	-1.9%	-2.0%	0.3%

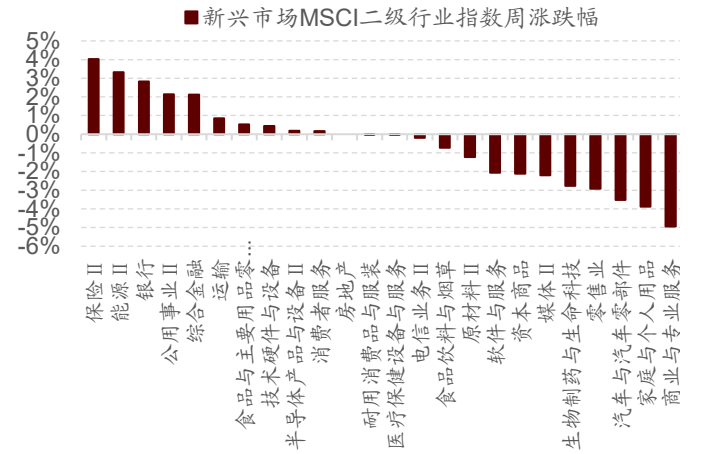
资料来源：Wind，招商证券

图 23：发达市场 MSCI 二级行业指数周涨跌幅（2026/08/31-2026/09/04）



资料来源：Wind，招商证券

图 24：新兴市场 MSCI 二级行业指数周涨跌幅（2026/08/31-2026/09/04）



资料来源：Wind，招商证券

2、全球强势股与异动股简析

过去一周领涨的 200 亿美元以上市值公司中，信息技术、金融居多；过去一周领跌的 200 亿美元以上市值公司中，信息技术居多。

异动股简析

【DELL.N】本周戴尔科技股价大涨的主要原因是最新财报把“AI 服务器需求强劲”从主题交易兑现为订单、收入和

下载日志已记录，仅供内部参考

积压订单的实质增长。公司 FY2027 Q2 披露，AI 服务器业务单季获得 609 亿美元订单、确认 164 亿美元收入，期末 AI 积压订单达到 950 亿美元，电话会进一步说明过去 12 个月已将 1317 亿美元需求转化为订单，客户数超过 6500 家，需求从早期大型云厂商继续扩散到 neocloud、主权 AI 和企业客户。市场对戴尔的定价逻辑因此发生变化：公司不再只是传统 PC 和通用服务器厂商，而是机柜级系统、存储、网络的核心集成商。此外，AI 相关需求同时带动传统服务器、网络和存储，企业本地和边缘 AI 基础设施因数据控制、知识产权保护和 token economics 具备吸引力，进一步打开戴尔在企业 AI 落地中的空间。公司称部分 AI 部署需要多达 50 种独特设计，围绕性能、功耗、冷却和数据中心环境做工程适配，并已率先交付基于 NVIDIA Vera Rubin 平台的 rack systems，强化其在下一代 AI 工厂建设中的卡位。

【FICO.N】本周 FAIR ISAAC 股价大跌的主要原因在于美国住房金融监管政策对 FICO 信用评分垄断地位的冲击。FICO 长期在美国住房抵押贷款评分中的标准化地位和定价权，而近期 FHFA 负责人指示 Fannie Mae 和 Freddie Mac 允许所有抵押贷款机构使用其竞争对手的 VantageScore 系统，VantageScore 方面也称 VantageScore 4.0 已获准用于面向 Fannie Mae 和 Freddie Mac 的所有抵押贷款发起机构，并强调该模型可覆盖更多传统 FICO 体系难以评分的成年人。市场担忧在于，FICO 长期依赖美国按揭信用评分中的标准化地位和较强定价权，一旦 GSE 渠道全面接受竞争模型，贷款机构可能拥有议价和替代选择，FICO 评分业务的增长假设、收费能力和估值溢价都会受到重新审视。此前市场对 FICO 的担忧更多集中在价格上调和监管审查，本周政策落地强化了竞争实质化的预期，因此引发集中抛压。

表 6：全球 200 亿美元市值以上公司本周涨幅 TOP20（2026/08/31-2026/09/04）

代码	公司	上市地	一级行业	二级行业	总市值	周涨跌幅	2026 年涨跌幅
DELL.N	戴尔科技	纽约证券交易所	信息技术	技术硬件与设备	3397	14.9%	320.2%
2884.TW	玉山金控	台湾证券交易所	金融	银行	241.51	12.4%	29.0%
ALNY.O	阿里拉姆制药	纳斯达克	医疗保健	制药、生物科技与生命科学	356	12.2%	-33.1%
601319.SH	中国人保	上海	金融	保险 II	488	11.9%	-9.0%
2454.TW	联发科	台湾证券交易所	信息技术	半导体与半导体生产设备	2217	10.8%	214.9%
DE.N	迪尔(DEERE)	纽约证券交易所	工业	资本货物	1870	10.0%	49.8%
2328.HK	中国财险	香港联交所	金融	保险 II	498	9.8%	11.2%
6505.TW	台塑石化	台湾证券交易所	能源	能源 II	233	9.4%	67.1%
ULTA.O	ULTA 美妆	纳斯达克	可选消费	零售业	241.20	9.0%	-6.8%
2882.TW	国泰金	台湾证券交易所	金融	保险 II	528	9.0%	51.7%
MU.O	美光科技	纳斯达克	信息技术	半导体与半导体生产设备	11481	9.0%	256.4%
VST.N	VISTRA	纽约证券交易所	公用事业	公用事业 II	501.10	8.9%	-7.2%
ITUB.N	艾涛巴西联合银行	纽约证券交易所	金融	银行	458	8.7%	16.3%
601658.SH	邮储银行	上海	金融	银行	940	8.6%	3.8%
PBR.N	巴西石油	纽约证券交易所	能源	能源 II	749	8.6%	78.7%
DG.N	达乐(DOLLAR GENERAL)	纽约证券交易所	可选消费	零售业	294	8.4%	1.8%
9984.T	软银	东京证券交易所	电信服务	电信服务 II	2032	8.3%	27.2%
0992.HK	联想集团	香港联交所	信息技术	技术硬件与设备	513.03	8.1%	254.6%
NTR.N	NUTRIEN	纽约证券交易所	材料	材料 II	379	8.1%	30.9%
0267.HK	中信股份	香港联交所	工业	资本货物	516	7.7%	20.0%

资料来源：Wind，招商证券

下载日志已记录，仅供内部参考

表 7：全球 200 亿美元市值以上公司本周跌幅 TOP20（2026/08/31-2026/09/04）

代码	公司	上市地	一级行业	二级行业	总市值	周涨跌幅	2026 年涨跌幅
FICO.N	FAIR ISAAC	纽约证券交易所	信息技术	软件与服务	201.35	-19.2%	-44.9%
EIX.N	爱迪生国际(EDISON INTL)	纽约证券交易所	公用事业	公用事业Ⅱ	218	-19.1%	-1.7%
MDB.O	MONGODB	纳斯达克	信息技术	软件与服务	297	-17.4%	-12.1%
ADSK.O	欧特克(AUTODESK)	纳斯达克	信息技术	软件与服务	455	-16.4%	-26.4%
CDNS.O	铿腾电子	纳斯达克	信息技术	软件与服务	806	-14.0%	-6.4%
PCG.N	太平洋煤气与电力(PG&E)	纽约证券交易所	公用事业	公用事业Ⅱ	383	-13.9%	-10.5%
688008.SH	澜起科技	上海	信息技术	半导体与半导体生产设备	342.26	-11.9%	59.6%
SNPS.O	新思科技	纳斯达克	信息技术	软件与服务	755	-11.0%	-16.2%
PANW.O	PALO ALTO NETWORKS	纳斯达克	信息技术	技术硬件与设备	2716	-10.3%	80.9%
DDOG.O	DATADOG	纳斯达克	信息技术	软件与服务	765	-10.1%	56.6%
EXPE.O	亿客行	纳斯达克	可选消费	零售业	358	-9.5%	5.8%
TCOM.O	携程网	纳斯达克	可选消费	零售业	297	-9.5%	-42.9%
300274.SZ	阳光电源	深圳	工业	资本货物	270	-9.5%	-48.0%
AON.N	怡安保险(AON)	纽约证券交易所	金融	保险Ⅱ	685	-9.1%	-7.8%
DSY.PA	达索系统	巴黎证券交易所	信息技术	软件与服务	326	-9.0%	-11.9%
RMS.PA	爱马仕国际	巴黎证券交易所	可选消费	耐用消费品与服装	1765	-8.8%	-31.2%
TTWO.O	TAKE-TWO 互动软件	纳斯达克	信息技术	软件与服务	401	-8.8%	-16.1%
6098.T	RECRUIT	东京证券交易所	工业	商业和专业服务	1521	-8.8%	83.9%
EFX.N	艾可菲(EQUIFAX)	纽约证券交易所	工业	商业和专业服务	208	-8.6%	-17.7%
688012.SH	中微公司	上海	信息技术	半导体与半导体生产设备	473.95	-8.6%	83.4%

资料来源：Wind，招商证券

下载日志已记录，仅供内部参考

三、重要资讯速递

1、国内重要资讯与产业政策

（1）货币政策—央行行长潘功胜出席 G20 财长和央行行长会议 强调继续实施好适度宽松的货币政策

财联社 9 月 2 日电，据央行官网，8 月 31 日至 9 月 1 日，二十国集团（G20）在美国阿什维尔召开今年第二次财长和央行行长会议。会议讨论了全球经济形势与前景、促进经济增长、全球失衡、发展中国家主权债务等议题。中国人民银行行长潘功胜出席会议并发言。

潘功胜指出，贸易摩擦和保护主义通过影响供应链、推升通胀、扰乱市场预期等因素拖累全球经济。各方应坚定维护多边主义，充分用好 G20 财金合作平台，强化宏观经济政策协同，携手应对全球性风险挑战。解决主权债务问题应坚持“共同行动、公平负担”原则。G20 应发挥引领作用，确保各方在共同框架下有效落实“待遇可比”原则。经济增长是化解发展中国家债务问题、恢复债务可持续性的根本出路。国际上，贸易保护主义上升、国家安全问题泛化、政策环境不可预期是近年来全球失衡问题加剧的重要原因。解决全球失衡，需要各国推动自身结构性改革。逆差国削减财政赤字，提高本国储蓄率，顺差国适度推动消费和投资增长。同时，要坚持多边主义，加强国际合作。解决全球失衡应着眼于中期承诺。各国都应制定中长期政策方案，做出明确承诺并坚定执行，避免来回“翻烧饼”。

潘功胜表示，中国经济运行总体平稳向好，结构持续优化，高质量发展不断取得新进展，金融市场平稳运行，经济长期向好的基本面与支撑条件没有改变。中国政府坚持并持续推进经济结构的转型升级，并将此作为“十五五”规划的战略重点。中方从不刻意追求贸易顺差，坚持扩大内需、坚持高水平对外开放，以中国大市场为各方提供新机遇，为新一轮全球经济动态平衡贡献力量。中国人民银行将持续推进货币政策框架转型，完善利率体系，继续实施好适度宽松的货币政策，为中国经济稳定增长和金融市场的平稳运行营造良好的货币金融环境。

新闻来源： https://www.cls.cn/detail/2472193

（2）人工智能—工信部：探索通过首购首用、风险补偿等模式 加大大模型、智能体、Token 等服务采购力度

财联社 8 月 31 日电，工业和信息化部办公厅发布关于开展人工智能应用服务商培育专项行动的通知。其中提到，推动规模化应用。引导服务商围绕高频、刚需、可复用的业务，封装构建模块化、标准化解决方案产品包，形成一批“小快轻准”的人工智能产品服务。依托产业联盟、行业协会、开源社区等平台，分行业开展针对性供需对接活动，对服务商进行宣介推广。组织开放一批真实场景，积极引入服务商开展落地实践，推动服务商的技术研发和成果转化，打造标杆方案，形成带动效应。探索通过首购首用、风险补偿等模式，加大大模型、智能体、Token 等服务采购力度，提升行业用模用智质效。有条件的地区建立完善“出海”综合服务体系，依托一带一路、金砖、中国-东盟等合作机制平台，推动优质人工智能应用项目出海落地。

新闻来源： https://www.cls.cn/detail/2469446

（3）PMI—8 月中国制造业采购经理指数为 49.8% 较上月回升

财联社 8 月 31 日电，国家统计局服务业调查中心、中国物流与采购联合会今天公布 8 月中国采购经理指数。其中，制造业采购经理指数较上月有较大幅度回升，行业景气度明显改善。8 月中国制造业采购经理指数为 49.8%，较上月上升 0.6 个百分点。国家统计局服务业调查中心首席统计师霍丽慧介绍，在调查的 21 个行业中，有 16 个行业采购经理指数较上月上升，制造业行业的景气度还是明显改善的。具体来看，制造业产需两端同步扩张，生产指数和新订单指数分别为 50.4%和 50.6%，较上月分别上升 0.5 和 2.1 个百分点。国内需求明显回升，国外市场需求也有所改善。

新闻来源： https://www.cls.cn/detail/2469103

（4）AI 影视—芒果超媒董事长：AI 长剧《后西游记》成本仅为同类 S 级真人剧几分之一

财联社 9 月 2 日电，芒果超媒董事长蔡怀军表示，AI 长剧《后西游记》总成本约为同类 S 级真人剧的几分之一；30 集、单集 40 分钟的体量，从策划到上线，时间较传统三维动画压缩了 70%。首批 5 集播出后，神话题材和电影级 AI 视觉呈现带来了新的用户，“特别是男性用户群体拉新率超出了我们的预期”。

下载日志已记录，仅供内部参考

新闻来源：https://www.cls.cn/detail/2471962

（5）集成电路—工信部：前 7 个月我国集成电路设计收入同比增长 21.5%

财联社 8 月 31 日电，工信部发布数据，前 7 个月，信息技术服务收入同比增长 10.0%，占全行业收入的 68.7%。其中，云计算、大数据服务共实现收入同比增长 12.3%，占信息技术服务收入的 16.7%；集成电路设计收入同比增长 21.5%；电子商务平台技术服务收入同比增长 5.1%。

新闻来源：https://www.cls.cn/detail/2469466

（6）半导体—华为半导体负责人何庭波更新韬定律论文

《科创板日报》记者注意到，关于韬定律，华为半导体负责人何庭波今日在中科院科技论文预发布平台 ChinaXiv，再次发布相关论文《Huawei's τ Chip Was Supposed to Melt?》，提到热量是韬定律最棘手的隐忧，而麒麟芯片实测结果，颠覆了这一质疑。麒麟 2026 的芯片晶体管密度从约每平方毫米 1.55 亿提升到 2.38 亿个，跃升 55%；在相同性能下，NPU 功耗下降 66%，GPU 下降 58%，CPU 性能核心下降 41%。原因是在芯片内部，能量的主要消耗并非来自计算本身，而是来自数据传输。

新闻来源：https://www.cls.cn/detail/2474741

（7）私募—私募个人合格投资者门槛升格 要求家庭金融资产不低于 500 万或家庭金融净资产不低于 300 万

财联社 9 月 4 日电，证监会正就《私募投资基金募集监督管理办法》征求意见，旨在进一步加强对私募投资基金募集业务活动监管。新规规定，个人认定为合格投资者需具有二年以上证券、基金、期货、股权等投资经历，且家庭金融资产不低于 500 万元，或家庭金融净资产不低于 300 万元，或近三年本人年均收入不低于 50 万元，并具备相应风险识别和承担能力。对于主要财产投向不动产项目公司股权、单一标的、境外资产、场外衍生品等的私募基金，自然人投资者应具有四年以上投资经历，家庭金融资产不低于 1000 万元，家庭金融净资产不低于 600 万元。

新闻来源：https://www.cls.cn/detail/2474402

（8）新型电网—国家能源局召开新型电网建设工作部署会 加快输电通道等重大电网工程核准建设

财联社 9 月 2 日电，国家能源局 9 月 1 日召开新型电网建设工作部署会。会议指出，新型电网建设一头连着新能源发电等上游项目建设，一头连着下游工商业、居民等用户用电，带动关联效应强、投资规模巨大，对于当前促投资和长远调结构具有重要的支撑作用。一是加快输电通道等重大电网工程核准建设。电网企业要精益化推进项目建设，充分发挥重大项目投资带动作用，地方能源主管部门要加快各项支持要件的办理。二是做好新型电网建设的总体统筹推进。地方能源主管部门要梳理省内新型电网项目和任务清单，层层压实责任，指导属地电网企业加大省内主干电网、配电网工程等建设力度。三是加大政策支持和落实工作力度。国家能源局将会同有关部委和地方能源主管部门，发挥好重点能源项目定期监测调度和对接工作机制作用，协调解决工程实施中的重大问题。支持广大民营企业积极投资新型电网建设，共享新机遇、新红利。

新闻来源：https://www.cls.cn/detail/2471458

（9）IPO—月之暗面据悉计划今年在香港进行 IPO

财联社 9 月 4 日电，月之暗面据悉计划今年在香港进行 IPO，寻求募资至多 50 亿美元。

新闻来源：https://www.cls.cn/detail/2474583

（10）人工智能—智谱上半年营收 9.539 亿元 同比增长约 400%

财联社 8 月 31 日电，智谱上半年营收 9.539 亿元，同比增长约 400%；毛利 2.516 亿元；经营亏损 21.5 亿元，预估亏损 29.3 亿元；研发支出 21.3 亿元。智谱上半年开放平台及 API 业务收入约 8.25 亿元（约 1.23 亿美元），同比增长 2,735.7%。

新闻来源：https://www.cls.cn/detail/2469870

下载日志已记录，仅供内部参考

2、国外重要资讯与产业政策

（1）非农就业—美国 8 月非农就业人数意外大增 16.2 万人 失业率报 4.1%

财联社 9 月 4 日电，美国 8 月非农就业人数增加 16.2 万人，预估为增加 5.5 万人，前值为减少 2.3 万人。劳动力参与率为 61.6%，前一个月为 61.4%。平均每小时工资较前一个月增加 0.3%，预估为增加 0.3%，前一个月为增加 0.2%。其中，私营部门非农就业人数增加 12.7 万人，前一个月为增加 7.1 万人，市场预估为增加 5 万人。制造业非农就业人数增加 1.6 万人，前一个月为增加 1.4 万人；预估为增加 0.5 万人。失业率为 4.1%，前一个月为 4.1%。

新闻来源：https://www.cls.cn/detail/2474650

（2）美联储—美联储沃勒"鸽派"表态后，美债收益率全线下行

美联储理事沃勒表示，只要通胀继续放缓，他倾向于维持利率不变，此番言论推动美国国债上涨，使各期限国债收益率下降了 3 至 5 个基点，其中两年期国债收益率领跌，因为两年期国债对美联储利率变动的敏感度高于长期国债。由于市场预期美联储本月将加息，两年期国债收益率本周一度突破 4.40%，创下自 2025 年 1 月以来的新高，随后下跌多达 7 个基点至 4.30%。美元下跌多达 0.5%，兑所有 G10 货币均走弱。

新闻来源：https://baijiahao.baidu.com/s?id=1875320227163886750&wfr=spider&for=pc

（3）债市—欧美日债市遭遇抛售潮

财联社 9 月 1 日电，受能源价格上涨、通胀预期升温以及主要央行货币政策预期变化影响，全球主要国债收益率普遍走高。日本 30 年期国债收益率升至 4.18%以上，创历史新高；英国 30 年期国债收益率升至 5.88%，续创 1998 年以来最高；德国 10 年期国债收益率升至 3.339%，为 2011 年以来最高。美国 30 年期国债收益率升至 5.28%，日内上涨 4 个基点，逼近近期新高。

新闻来源：https://www.cls.cn/detail/2470868

（4）美伊冲突—美军称完成一波针对伊朗军事目标的打击行动

财联社 9 月 2 日电，美军中央司令部当地时间 9 月 1 日称，中央司令部部队于 9 月 1 日成功完成了一波针对伊朗军事目标的打击行动。美军打击了伊朗伊斯兰革命卫队目标，包括防空阵地、雷达系统、海上资产和设施、布雷能力以及通信站点。美军中央司令部称，"此次打击行动是在革命卫队最近试图袭击霍尔木兹海峡的商业船只以及美军人员之后进行的"。

新闻来源：https://www.cls.cn/detail/2471388

（5）人工智能—OpenAI 上调 GPT-6 Astra 的 API 定价

财联社 9 月 4 日电，OpenAI 表示，在价格上，GPT-6 Astra 的 API 定价相比上一代 GPT-5.6 Sol 更高：按每百万 Token 计费，输入 10 美元，输出 50 美元，且缓存读取和写入另行计费。相比之下，GPT-5.6 Sol 价格为输入 5 美元，输出 30 美元。

新闻来源：https://www.cls.cn/detail/2474162

（6）业绩—博通 2026 财年 Q3 营收 295.91 亿美元 同比增长 86%

财联社 9 月 3 日电，博通 2026 财年 Q3 营收 295.91 亿美元，同比增长 86%，去年同期 159.52 亿美元，市场预期 293.62 亿美元；调整后 EPS3.32 美元，同比增长 96%，分析师预期 3.23 美元。Q4 展望营收为 348 亿美元，市场预估为 350.5 亿美元。在财报电话会议上，博通 CEO 表示，将 2026 财年人工智能业务营收指引从 560 亿美元上调至 580 亿美元，同比增长 186%；到 2027 年公司已确保供应将人工智能营收翻倍至约 1150 亿美元；预计 2028 财年人工智能半导体营收将再次翻倍，达到 2300 亿美元。博通 CEO 表示，本季度已开始向谷歌量产交付下一代 TPU 8I 版本产品；全新谷歌 TPU 性能与英伟达 Vera Rubin 持平或更优。

新闻来源：https://www.cls.cn/detail/2472562

（7）日本央行—日本央行行长称在制定政策时会比以往更强烈地对外汇波动作出反应

财联社 9 月 2 日电，日本央行行长植田和男称，不会单独关注外汇，在制定政策时会比以往更强烈地对外汇波动作

下载日志已记录，仅供内部参考

出反应。密切关注外汇波动，外汇波动是价格前景的风险因素之一。

新闻来源：https://www.cls.cn/detail/2471445

（8）业绩—戴尔预计本财年营收 1920 亿美元

财联社 9 月 2 日电，戴尔预计本财年营收 1920 亿美元，分析师预期 1737.6 亿美元，公司原本预计 1650 亿-1690 亿美元；AI 优化服务器全年营收预期由 600 亿美元上调至 740 亿美元。

新闻来源：https://www.cls.cn/detail/2471349

（9）存储—报道称三星存储器业务部门将 2031 年前约 70%的产能分配给长约订单 HBM 抢货情况相当激烈

《科创板日报》1 日讯，AI 建设热潮引爆 HBM 需求，并排挤传统 DRAM 供给、加剧内存缺货。韩媒报道，三星的存储器业务部门已将 2031 年前约 70%的产能分配给长约订单（long-term agreements，简称 LTAs），主要签约客户包括英伟达，微软及谷歌等大型科技公司。然而，HBM 抢货情况相当激烈，就连这些大型科技公司都无法取得合约约定的全部供货量，使现货价格进一步飙升。

新闻来源：https://www.cls.cn/detail/2470629

（10）云计算—Anthropic 与英伟达支持的 Lambda 签署 350 亿美元云计算协议

财联社 9 月 1 日电，据报道，Anthropic 与英伟达支持的云服务提供商 Lambda 签署一项价值 350 亿美元的云计算协议，这个数据中心的租赁由英伟达持有。

新闻来源：https://www.cls.cn/detail/2470224

3、全球重点科技公司跟踪

表 8：全球重点科技公司跟踪

日期	公司	信息	信息来源
1、互联网科技公司——新动向			
2026.8.31	字节跳动	《科创板日报》8 月 31 日讯，字节跳动将豆包股价格从 6 月的 14.85 美元/股提升至 17.02 美元/股，并增加豆包股数量发放，员工可自行调整总包中的豆包股比例。	https://www.chinastarmarket.cn/detail/2469475
2026.9.1	华为	《科创板日报》9 月 1 日讯，华为 2026 年上半年实现营收 4678.19 亿元，同比增长 9.54%；归母净利润 234.28 亿元，同比下降 36.8%；研发投入 1213.82 亿元，继续投向 AI、芯片、云计算、智能驾驶等方向。	https://www.chinastarmarket.cn/detail/2469686
2026.9.3	英伟达、Hugging Face	财联社 9 月 3 日电，英伟达将以 129.3 亿美元收购开源 AI 平台 Hugging Face，二者将共同扩展 Hugging Face 平台，强化其基础设施，并扩大全球开发者和机构对 AI 的访问。Hugging Face 将继续作为整个 AI 生态系统的开放平台。开发者将选择他们想要的模型、框架、云和推理服务提供商以及计算平台。NVIDIA 计算技术无需在 Hugging Face 上构建或部署。Hugging Face 将继续支持整个生态系统的开源和开放权重模型，涵盖所有模型构建器。它将继续支持多云和多加速器的开发与部署，使开发者能够使用最适合其工作的硬件和基础设施。	https://www.chinastarmarket.cn/detail/2473463
2、消费电子——新技术、新单品、新观念			
2026.9.3	华为、小米、荣耀	《科创板日报》9 月 3 日讯，华为、小米、荣耀近日集中上调多款手机售价，部分机型单次涨幅达千元；报道指出存储芯片等上游元器件涨价向终端产品传导。	https://www.chinastarmarket.cn/detail/2473470
2026.9.4	英伟达	财联社 9 月 4 日电，英伟达宣布英伟达 Rtx Spark Windows 个人电脑将于 2026 年 10 月上市。	https://www.cls.cn/detail/2473685
3、新能源汽车			
2026.9.3	宁德时代	财联社 9 月 3 日电，宁德时代董事长曾毓群通过视频为	https://www.chinastarmark

下载日志已记录，仅供内部参考

		“2026 世界动力电池大会”致辞时表示，今年上半年国内新能源汽车销量超过 740 万辆，上市新车超过 600 款，平均每天接近三辆，大家比速度、比参数、比价格，开发周期越来越短。他表示，动力电池精度极高，容错量极低，在量产规模中所有小的问题都会被成倍放大，在设计阶段就要严控每一道工序和每一个参数等，提前识别异常，对待品质问题零容忍。	et.cn/detail/2473227
2026.9.4	特斯拉	《科创板日报》9 月 4 日讯，特斯拉在美国得克萨斯州奥斯汀发布量产版 Cybercab 无人驾驶出租车，该车取消方向盘及油门刹车踏板，搭载 FSD 完全自动驾驶系统。	https://www.chinastarmarket.cn/detail/2473791
4、工业及人形机器人			
2026.9.4	OpenAI	《科创板日报》9 月 4 日讯，OpenAI CEO 奥尔特曼证实公司推进机器人研发，布局人形机器人与数据中心场景特种机器人两条技术路线；基础设施、制造业场景是当前优先方向。	https://www.chinastarmarket.cn/detail/2473365
5、半导体芯片			
2026.8.31	燧原科技	财联社 8 月 31 日电，燧原科技(688801.SH)公告称，中信证券、国泰海通、广发证券发布关于公司首次公开发行股票并在科创板上市参与战略配售投资者的专项核查报告。本次发行股票 4303.5173 万股，初始战略配售 860.7034 万股，占比 20%。战略配售对象包括中信证券投资、员工资管计划、启善投资、深圳外滩、通富微电、中兴通讯等 10 名投资者，其中启善投资拟认购金额上限 2.62 亿元，通富微电及中兴通讯各 8000 万元。	https://www.chinastarmarket.cn/detail/2469882
2026.9.1	三星	《科创板日报》1 日讯，AI 建设热潮引爆 HBM 需求，并排挤传统 DRAM 供给、加剧内存缺货。韩媒报道，三星的存储器业务部门已将 2031 年前约 70%的产能分配给长约订单（long-term agreements，简称 LTAs），主要签约客户包括英伟达，微软及谷歌等大型科技公司。然而，HBM 抢货情况相当激烈，就连这些大型科技公司都无法取得合约约定的全部供货量，使现货价格进一步飙升。	https://www.cls.cn/detail/2470629
2026.9.2	台积电	财联社 9 月 2 日电，据报道，台积电的 CoWoS 订单已排满至 2027 年，部分客户甚至需要等待一年以上才能拿到订单。在台积电供应瓶颈持续的情况下，英特尔的新技术作为替代方案正受到关注。据悉，联发科在近期财报电话会议上正式宣布，计划同时采用 CoWoS 和 EMIB-T 技术，大规模开发人工智能专用集成电路（ASIC）；博通和 Meta 也在考虑从 CoWoS 过渡到 EMIB 以降低成本，而谷歌和英伟达也对 EMIB-T 表现出浓厚的兴趣。	https://www.chinastarmarket.cn/detail/2471880
2026.9.3	台积电	财联社 9 月 3 日讯，台积电全球在建晶圆厂已接近 20 座；公司高管称，对芯片制造工具的需求自去年年底以来几乎翻倍，主要受 AI 带来的先进产能需求推动。	https://www.chinastarmarket.cn/detail/2472703
2026.9.3	博通	财联社 9 月 3 日电，博通 2026 财年 Q3 营收 295.91 亿美元，同比增长 86%，去年同期 159.52 亿美元，市场预期 293.62 亿美元；调整后 EPS3.32 美元，同比增长 96%，分析师预期 3.23 美元。Q4 展望营收为 348 亿美元，市场预估为 350.5 亿美元。在财报电话会议上，博通 CEO 表示，将 2026 财年人工智能业务营收指引从 560 亿美元上调至 580 亿美元，同比增长 186%；到 2027 年公司已确保供应将人工智能营收翻倍至约 1150 亿美元；预计 2028 财年人工智能半导体营收将再次翻倍，达到 2300 亿美元。博通 CEO 表示，本季度已开始向谷歌量产交付下一代 TPU 8I 版本产品；全新谷歌 TPU 性能与英伟达	https://www.chinastarmarket.cn/detail/2472562

下载日志已记录，仅供内部参考

		Vera Rubin 持平或更优。	
2026.9.3	澜起科技、三星、SK 海力士	财联社 9 月 3 日电，澜起科技(688008.SH)发布投资者关系活动记录表公告，公司 7 月在业界率先试产 CXL3.2MXC 芯片并成功导入三星、SK 海力士下一代 CXL 产品，预计 2027 年进入 CXL 规模商用元年。PCIeSwitch 芯片预计年内完成工程样片流片，PCIe6.x/CXL3.xRetimer 芯片已完成量产版本流片并列入 PCI-SIG 供应商名录，计划今年完成 PCIe7.0Retimer 及以太网 PHYRetimer 芯片工程样片流片。DDR5 第六子代 RCD 芯片已送样，支持速率达 9200MT/s，第五子代开始规模出货，第三四子代合计出货占比突破 50%。第二子代 MRCD/MDB 芯片处于规模试用阶段，大规模上量需等待新一代 CPU 明年量产。时钟芯片已收到量产订单，高精度 RTC 芯片预计下半年完成工程样片流片。AI 服务器 CPU 配置比例提升将带动内存模组及互连芯片需求增长。	https://www.chinastarmarket.cn/detail/2473108
2026.9.3	长江存储控股	《科创板日报》9 月 3 日讯，长江存储控股股份有限公司首次公开发行股票并在科创板上市的 IPO 审核状态变更为“已问询”。	https://www.chinastarmarket.cn/detail/2473639
6、人工智能			
2026.8.31	智谱	《科创板日报》8 月 31 日讯，智谱披露上市后首份中报，上半年收入 9.54 亿元；开放平台及 API 业务收入 8.25 亿元，占总收入 86.5%，Token 调用量较年初增长超 40 倍。	https://www.chinastarmarket.cn/detail/2469870
2026.9.2	Anthropic	财联社 9 月 2 日电，Anthropic 推出 Claude Fable 5.1 与 Claude Mythos 5.1。	https://www.chinastarmarket.cn/detail/2471321
2026.9.3	字节跳动	财联社 9 月 3 日电，字节跳动将获得 296 亿美元银团贷款。据了解，字节跳动最初计划筹集 200 亿美元，因银行认购踊跃，最终将规模扩大至 296 亿美元。资金主要用于一般企业用途。交易尚未正式签约，银行仍在确认分配额度。完成后，该笔贷款将成为今年亚洲第二大美元计价银团贷款，仅次于软银集团 3 月的 400 亿美元过桥贷款。字节跳动正加速 AI 投入，考虑将 2026 年资本支出提升至最高 700 亿美元，约为去年两倍以上，主要用于扩建数据中心及其他 AI 基础设施。字节跳动上一次大规模境外贷款发生在 2024 年，筹集约 108 亿美元。截至发稿，字节跳动方面暂未回应。	https://www.chinastarmarket.cn/detail/2473052
2026.9.4	OpenAI	《科创板日报》9 月 4 日讯，OpenAI 正式发布 GPT-6 Astra，称其在计算机操作、软件工程、网络安全、科学研究等领域达到先进水平，将先向 Daybreak 项目客户开放。	https://www.chinastarmarket.cn/detail/2474162
2026.9.4	Anthropic	财联社 9 月 4 日电，据报道，知情人士透露，Anthropic PBC 接近敲定将其循环信贷额度扩大至 150 亿美元，为这家人工智能公司在公开提交备受瞩目的 IPO 文件之前扫清了一道障碍。知情人士表示，摩根士丹利将负责牵头，高盛、摩根大通和花旗也将在信贷安排中担任重要角色。	https://www.chinastarmarket.cn/detail/2473766
7、传媒和游戏			
2026.9.2	芒果超媒	财联社 9 月 2 日电，芒果超媒董事长蔡怀军表示，AI 长剧《后西游记》总成本约为同类 S 级真人剧的几分之一；30 集、单集 40 分钟的体量，从策划到上线，时间较传统三维动画压缩了 70%。首批 5 集播出后，神话题材和电影级 AI 视觉呈现带来了新的用户，“特别是男性用户群体拉新率超出了我们的预期”。	https://www.cls.cn/detail/2471962

下载日志已记录，仅供内部参考

8、商业航天			
2026.9.1	星河动力	2026 年 9 月 1 日 10 时，星河动力航天公司自主研发的中大型可重复使用液体运载火箭智神星一号在酒泉东风商业航天创新试验区智神星系列专用发射工位点火升空，火箭按程序顺利进入预定轨道。该火箭为两级构型，全长 52 米，芯级直径 3.35 米，起飞质量约 283 吨，起飞推力 350 吨。LEO 运载能力可充分覆盖当前低轨星座组网单次发射 5—7 吨的主流载荷需求，火箭设计重复使用次数不少于 25 次。	https://www.chinastarmarket.cn/detail/2470383

资料来源：财联社，科创板日报，第一电动，彭博，招商证券

下载日志已记录，仅供内部参考

分析师承诺

负责本研究报告的每一位证券分析师，在此申明，本报告清晰、准确地反映了分析师本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与，未来也将不会与本报告中的具体推荐或观点直接或间接相关。

评级说明

报告中所涉及的投资评级采用相对评级体系，基于报告发布日后 6-12 个月内公司股价（或行业指数）相对同期当地市场基准指数的市场表现预期。其中，A 股市场以沪深 300 指数为基准；香港市场以恒生指数为基准；美国市场以标普 500 指数为基准。具体标准如下：

股票评级

强烈推荐：预期公司股价涨幅超越基准指数 20%以上

增持：预期公司股价涨幅超越基准指数 5-20%之间

中性：预期公司股价变动幅度相对基准指数介于±5%之间

减持：预期公司股价表现弱于基准指数 5%以上

行业评级

推荐：行业基本面向好，预期行业指数超越基准指数

中性：行业基本面稳定，预期行业指数跟随基准指数

回避：行业基本面转弱，预期行业指数弱于基准指数



重要声明

本报告由招商证券股份有限公司（以下简称“本公司”）编制。本公司具有中国证监会许可的证券投资咨询业务资格。本报告基于合法取得的信息，但本公司对这些信息的准确性和完整性不作任何保证。本报告所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价，在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。除法律或规则规定必须承担的责任外，本公司及其雇员不对使用本报告及其内容所引发的任何直接或间接损失负任何责任。本公司或关联机构可能会持有报告中所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业务服务。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突。

本报告版权归本公司所有。本公司保留所有权利。未经本公司事先书面许可，任何机构和个人均不得以任何形式翻版、复制、引用或转载，否则，本公司将保留随时追究其法律责任的权利。