

DeepSeek 十问十答

华泰研究

2025 年 2 月 04 日 | 中国内地

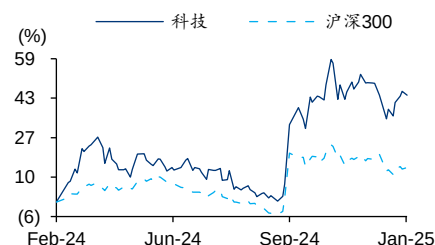
动态点评

科技

增持（维持）

研究员	黄乐平, PhD
SAC No. S0570521050001	huangleping@htsc.com
SFC No. AUZ066	+(852) 3658 6000
研究员	陈旭东
SAC No. S0570521070004	chenxudong@htsc.com
SFC No. BPH392	+(86) 21 2897 2228
联系人	于可熠
SAC No. S0570122120079	yukeyi@htsc.com
SFC No. BVF938	+(86) 21 2897 2228

行业走势图



资料来源：Wind，华泰研究

1. DeepSeek 是否会抑制算力增长？短期抑制 VS 长期爆发

我们认为 DeepSeek 本质是对现有 AI 算法的效率优化，短期可能导致训练需求下降，但随着模型普及和应用场景扩展，长期推理需求将显著增长，这种趋势类似于“蒸汽机降低煤耗但提升煤炭总用量”的历史现象。

2. DeepSeek 是否会改变 AI 算力的增长范式：目前不会

AI 算力大致分为 1) 用于研发前沿模型相关的探索性算力 (AGI 方向)，2) 面向消费者的应用性算力 (现有模型推理)。目前北美四大 AI 公司主要通过扩大 GPU 集群规模的方式探索下一代大模型。只要这个探索工作还在继续产生正向回报，AI 算力的增长范式短期或不会发生变化。

3. DeepSeek 会改变市场投资逻辑吗？软件有望跑赢硬件

我们认识到，1) 未来大模型竞争中，“算法效率”的重要性或上升，投资重点可能从“算力军备”转向“算法效率”。2) 开源协议使中小开发者能基于前沿模型二次开发，创造更多创新机会。投资角度，看好美股软件表现好于硬件。

4. 芯片市场格局是否会改变？高端 GPU 用途受限，ASIC 占比或提升

DeepSeek 的成功显示即使不使用最先进的 GPU，也可以开发满足一般消费者需求的大模型。这可能意味着，英伟达的 Blackwell/Rubin 等最先进的 GPU 的用途，可能会局限在探索下一代超大规模模型 (Frontier Model) 上。

5. DeepSeek 真的那么便宜吗？可能没有

DeepSeek 在其 V3 技术报告中估计其训练成本只有 557.6 万美元，根据 SemiAnalysis 分析，这只包含预训练阶段的部分成本，而加上 GPU 算力投资，研发、数据收集等其他重要成本或远大于这个数字。

6. DeepSeek 到底有哪些创新？混合专家 MoE，强化学习，蒸馏等

主要创新包括模型架构 (混合专家 MoE, MLA)、训练方法 (纯强化学习)、蒸馏优化和推理效率提升等，显著提升了 AI 算法效率和性能。

7. DeepSeek 会带动中国科技资产价值重估吗？有可能

目前 (2025/2/3)，恒生科技指数 12 月前向 PE 20.0 倍，远低于纳斯达克的 35.4 倍。DeepSeek 的成功可能提高中国 AI 公司估值预期。

8. DeepSeek 会推动端侧智能发展吗？有可能，但需要时间

DeepSeek 的高性价比模型有利于模型在智能手机和汽车等智能设备上落地，我们认为智能硬件迭代不会一蹴而就，模型能力提升只是其中一环。

9. DeepSeek 会导致美国提升出口管制压力吗？可能会

我们注意到 DeepSeek 发布后，美国媒体进一步限制中国发展 AI 声音抬头。建议关注后续：1) 高端 AI 芯片出口管制，2) 前沿模型的开源限制，3) 模型回传限制，4) 数据获取限制等风险。

10. DeepSeek 会改变开源软件生态吗？会

目前基础大模型的开发，主要集中在 OpenAI、Google、Anthropic、阿里，字节、百度的科技巨头手中。DeepSeek 这次的成功丰富了 AI 开源生态，也为中小开发者依托开源生态实现快速发展提供了一条发展路径。

风险提示：中美贸易摩擦升级风险，宏观下行风险，创新品渗透不及预期风险。本研报中涉及到未上市公司或未覆盖个股内容，均系对其客观公开信息的整理，并不代表本研究团队对该公司、该股票的推荐或覆盖。

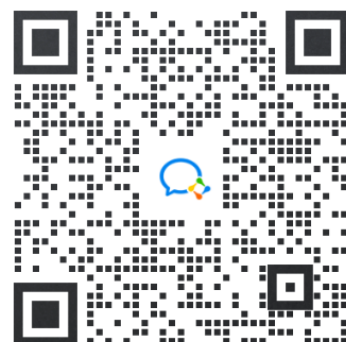
免责声明：

1. 本资料来源于网络公开渠道，版权归属版权方；
2. 本资料仅限会员学习使用，如他用请联系版权方；
3. 会员费用作为信息收集整理及运营之必须费用；
4. 如侵犯您的合法权益，请联系客服微信将及时删除



行业报告资源群

1. 进群福利：进群即领万份行业研究、管理方案及其他学习资源，直接打包下载
2. 每日分享：6份行研精选报告、3个行业主题
3. 报告查找：群里直接咨询，免费协助查找
4. 严禁广告：仅限行业报告交流，禁止一切无关信息



微信扫码，长期有效

知识星球 行业与管理资源

专业知识社群：每月分享8000+份行业研究报告、商业计划、市场研究、企业运营及咨询管理方案等，涵盖科技、金融、教育、互联网、房地产、生物制药、医疗健康等；已成为投资、产业研究、企业运营、价值传播等工作助手。



微信扫码，行研无忧

正文目录

DeepSeek 十问十答.....	3
问题#1: DeepSeek 是否会抑制算力增长?	3
问题#2: DeepSeek 是否会改变 AI 算力的增长范式.....	4
问题#3: DeepSeek 会改变市场投资逻辑吗?	5
问题#4: DeepSeek 是否会改变芯片市场格局?	6
问题#5: DeepSeek 真的那么便宜吗?	7
问题#6: DeepSeek 到底有哪些创新?	8
问题#7: DeepSeek 会带动中国科技资产价值重估吗?	9
问题#8: DeepSeek 会推动端侧智能发展吗?	10
问题#9: DeepSeek 会导致美国提升出口管制压力吗?	11
问题#10: DeepSeek 会改变开源软件生态吗?	12
风险提示.....	12

图表目录

图表 1: MAAMG 季度资本开支.....	3
图表 2: 大模型对半导体需求增长快于摩尔定律.....	4
图表 3: 美股软/硬件板块股价涨跌幅.....	5
图表 4: 硬件/软件 ETF 主要权重股估值和股价表现	5
图表 5: Nvidia、AMD、Broadcom、Marvell、Intel 数据中心及 AI 相关收入趋势	6
图表 6: DeepSeek AI TCO	7
图表 7: DeepSeek 技术原理.....	8
图表 8: Nasdaq vs 恒生科技一年前向 PE	9
图表 9: Nasdaq vs 恒生科技一年前向 PE	9
图表 10: DeepSeek-R1 模型已作为 NVIDIA NIM 微服务发布	10
图表 11: DeepSeek 或催化美国提升出口限制压力	11
图表 12: 开源模型性能进步速度快于闭源模型	12
图表 13: 国内外主流 AI 大模型对比.....	12

DeepSeek 十问十答

DeepSeek 的横空出世引发了 AI 产业的深刻讨论。本文就 DeepSeek 对算力需求、芯片市场、技术创新、端侧智能、出口管制、投资逻辑、软件生态、开源生态以及中国科技资产估值等十个关键问题进行了深入分析。

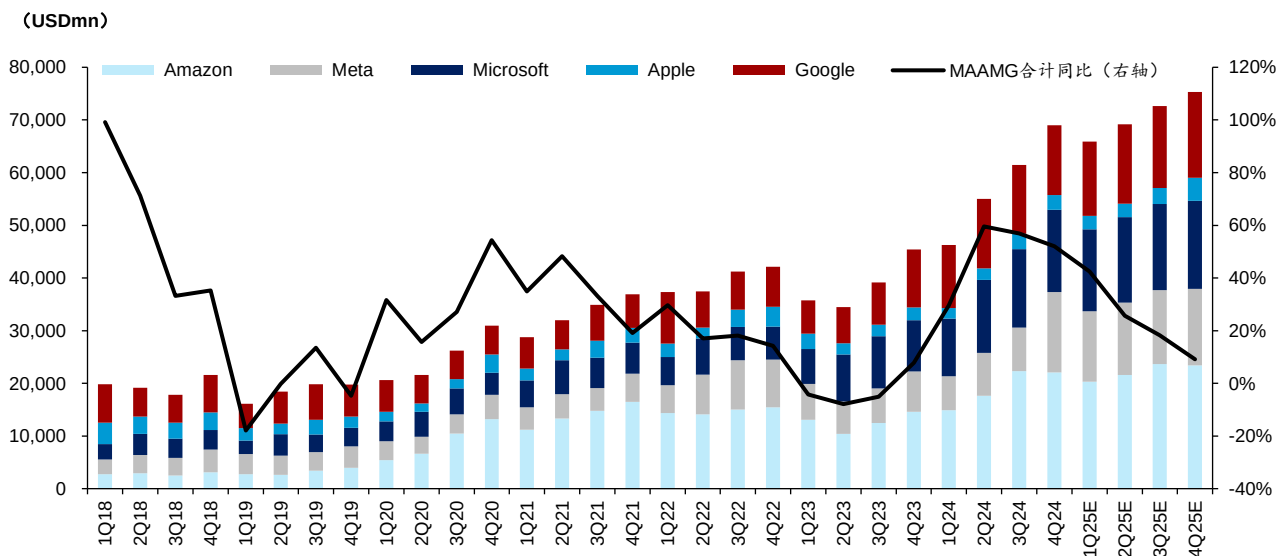
问题#1: DeepSeek 是否会抑制算力增长?

据 DeepSeek V3 技术报告,V3 模型的训练总计只需要 278.8 万 GPU 小时,相当于在 2048 卡的 H800GPU 集群上训练约 2 个月,合计成本约 557.6 万美金,相较而言, Llama 3 系列模型的计算预算则多达 3930 万 H100 GPU 小时, DeepSeek 训练成本约相当于 Llama 3 系列模型的 7%。

我们认为, DeepSeek 对算力需求的影响呈现出短期抑制、长期增长的复杂趋势。短期内, DeepSeek 的低成本高效训练方法可能导致训练需求下降。然而,从长远来看,随着模型的普及和应用场景的扩展,推理需求将显著增长。这种趋势类似于“蒸汽机降低煤耗但提升煤炭总用量”的历史现象。

随着 AI 技术成本的下降,其应用范围将大幅扩展,最终可能导致算力需求的大幅增长。微软 CEO 纳德拉认为 AI 效率提升将激发指数级需求。

图表1: MAAMG 季度资本开支



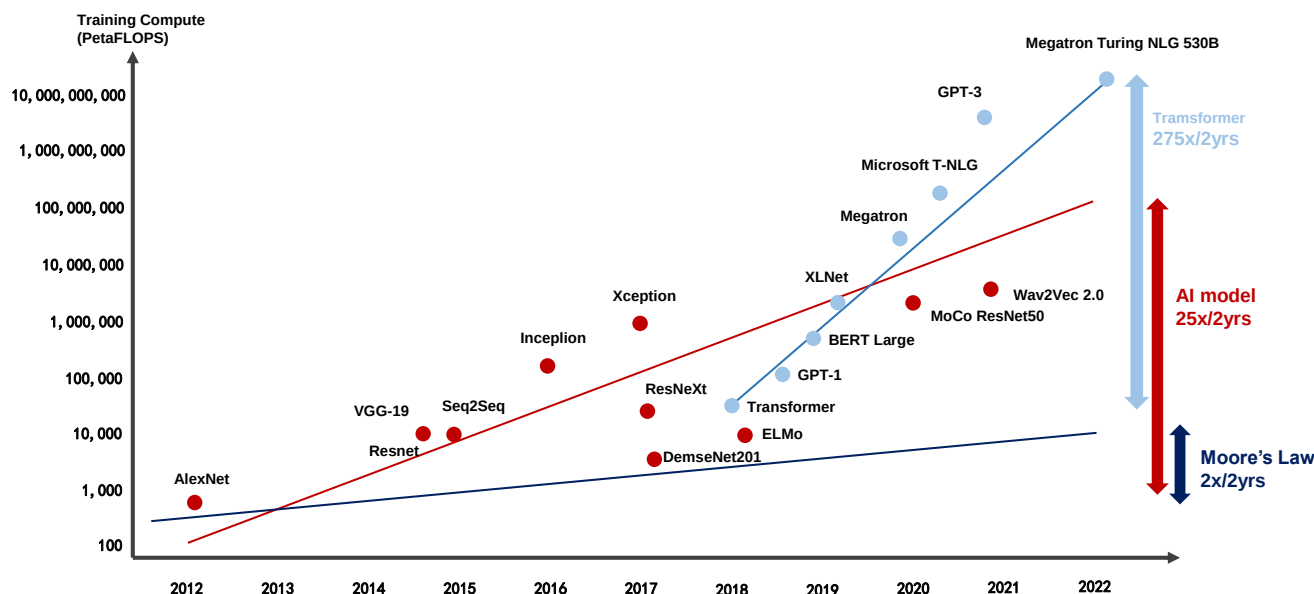
注: 预测数据来自 Factset 一致预期

资料来源: Factset, 华泰研究

问题#2: DeepSeek 是否会改变 AI 算力的增长范式

AI 算力大致分为 1) 用于研发通向通用人工智能 (AGI) 的前沿模型相关的探索性算力 (AGI 方向), 2) 面向一般消费者的应用性算力 (现有模型推理)。过去两年, 推动算力增长的主要动力是, 探索性算力增速 (25x/2 年) 远高于摩尔定律 (2 倍/年), 导致 GPU 需求激增。只要这个探索工作还在继续产生正向回报, AI 算力的增长范式短期或不会发生变化。

图表2: 大模型对半导体需求增长快于摩尔定律



资料来源: NVIDIA Ian Buck statement (2018), 华泰研究

在通用人工智能 (AGI) 愿景的驱动下, 我们看到, 主要科技巨头仍然在加大投入, 例如:

- 1/24, Meta 宣布计划 2025 年资本支出达 600-650 亿美元, 主要用于 AI 基础设施 (训练集群和数据中心建设)。
- 1/21, OpenAI 宣布和软银、Oracle 启动的 “Project Stargate” 计划投资超 1000 亿美元建设 AI 基础设施, 显示资本仍集中流向需要海量算力的前沿探索。

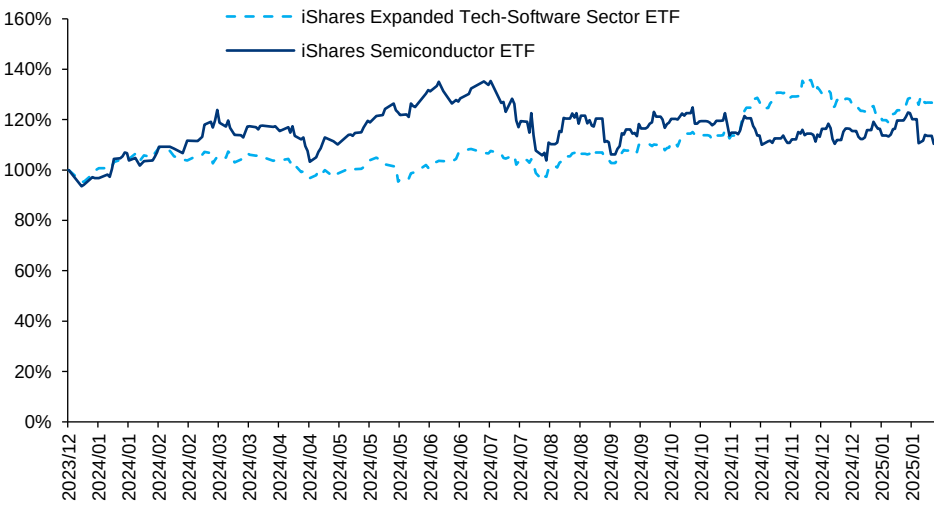
根据 Factset 一致预期, 微软、谷歌、亚马逊、Meta、苹果等北美五大科技公司合计资本开支 2025 年有望继续增长 19.6%。其中很大部分投入是用于包括 GPT-5、Llama4 等在内下一代模型的算力投资。另一方面, Agent 等应用目前仍处于探索的初期阶段, 大规模商用的时间点仍存在不确定性, 所以我们认为目前 AI 算力的增长范式没有变化。

问题#3: DeepSeek 会改变市场投资逻辑吗?

DeepSeek 这次的“惊喜”让我们认识到,

- 1) 未来大模型公司之间的竞争中,"算法效率"的重要性可能上升,投资重点可能从"算力军备"转向"算法效率"。AI 竞赛正从"算力军备"转向"算法效率"。未来的竞争重点或将更多地集中在算法优化和生态活力上。
- 2) 开源协议能使中小开发者基于前沿模型二次开发,推动大模型创新从少数科技巨头向分布式社区转移,为中小软件企业创造更多创新机会。从投资角度,我们认为 2025 年是 AI 发展进入商业化落地的一年。以 Agent AI 为代表的企业软件有望迅速落地,提升企业工作效率,带动美股软件板块表现或好于硬件。

图表3: 美股软/硬件板块股价涨跌幅



资料来源: Factset, 华泰研究

图表4: 硬件/软件 ETF 主要权重股估值和股价表现

		市值	收盘价	PE		PB		PS		EPS增速	股价变动 (%)				
公司名称		(百万美元)	当地货币	2025E	2026E	2025E	2026E	2025E	2026E	2025E	1W	30D	90D	180D	1Y
硬件															
NVDA US	Nvidia	2,940,514	120	27.0	22.1	17.8	11.6	14.8	12.2	51%	1.4	-16.9	-14.2	21.4	81.5
AVGO US	Broadcom	1,037,171	221	34.8	28.8	13.4	11.7	16.9	14.6	31%	9.5	-4.9	27.2	62.4	80.7
TSM US	TSMC	840,276	209	22.9	19.2	6.7	5.2	8.0	6.7	33%	8.8	0.3	6.9	35.0	80.8
ASML US	ASML	282,964	739	30.0	24.8	13.7	11.1	8.8	7.8	23%	7.1	3.5	9.3	-11.6	-17.0
QCOM US	Qualcomm	191,155	173	15.4	14.1	6.5	5.7	4.5	4.2	10%	1.0	9.6	4.2	10.8	22.0
AMD US	AMD	188,165	116	23.5	17.0	3.0	2.7	5.9	4.7	49%	0.8	-7.5	-18.1	-9.9	-34.7
TXN US	TI	168,404	185	34.4	26.9	10.1	10.4	9.8	8.7	3%	-1.3	-3.1	-8.8	0.3	16.0
AMAT US	AMAT	146,571	180	19.4	17.5	6.9	6.0	5.0	4.7	8%	3.3	6.1	-3.3	1.0	7.2
ADI US	ADI	105,120	212	29.9	24.1	3.2	3.1	10.3	9.1	11%	-1.6	-1.6	-2.6	5.3	10.3
LRCX US	Lam research	104,286	81	21.7	20.8	10.8	9.5	5.9	5.7	23%	7.2	7.9	6.1	8.3	-3.4
小计				25.9	21.5	9.2	7.7	9.0	7.9	39%	4.1	-8.7	-1.5	25.9	64.7
软件															
MSFT US	Microsoft	3,085,549	415	31.4	27.4	8.9	6.9	11.1	9.8	12%	-4.5	-2.0	0.9	4.2	0.9
ORCL US	Oracle	475,650	170	24.2	20.9	14.8	8.7	7.3	6.4	14%	7.4	2.2	-1.0	35.1	46.9
CRM US	Salesforce	327,007	342	30.4	26.7	4.9	4.5	7.9	7.2	12%	-1.6	2.6	14.9	42.4	19.6
NOW US	ServiceNow	209,786	1,018	62.2	51.3	16.5	12.4	16.1	13.4	18%	-10.7	-5.2	4.9	31.2	30.3
ADBE US	Adobe	190,426	437	21.4	19.0	12.9	11.0	8.1	7.4	11%	-0.3	1.6	-10.1	-14.8	-31.1
PLTR US	Palantir	179,882	82	176.3	145.1	29.9	25.6	53.6	44.0	23%	9.3	3.3	61.3	213.4	384.7
INTU US	Intuit	168,373	602	31.2	27.3	8.4	6.8	9.2	8.2	14%	-0.8	-4.6	-2.8	-1.3	-6.0
PANW US	Palo Alto Networks	121,016	184	58.1	50.8	16.9	12.6	13.2	11.4	12%	-2.6	0.6	0.7	21.2	6.9
APP US	Applovin	110,385	370	60.7	44.2	40.2	21.7	21.9	18.1	50%	7.7	5.4	123.7	450.1	705.9
CRWD US	CrowdStrike	93,089	398	91.0	71.9	23.8	17.1	20.6	16.9	16%	6.5	10.9	29.4	72.6	31.1
小计				58.7	48.5	17.7	12.7	16.9	14.3	14%	-2.1	-0.7	6.7	29.1	36.7

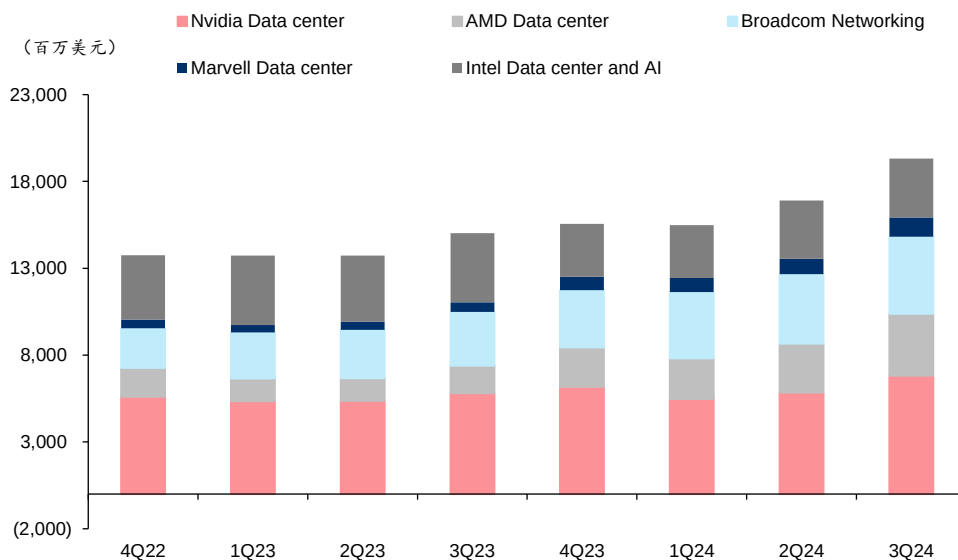
注: 预测数据来自 Factset 一致预期, 数据截至 2025/2/3

资料来源: Factset, 华泰研究

问题#4: DeepSeek 是否会改变芯片市场格局?

根据 Jon Peddie Research, 3Q24 英伟达在全球 GPU 市场份额达到 90%。其中, H100 等高端 GPU 是主要产品之一。DeepSeek 的成果显示, 在面向一般消费者的大模型市场, 企业可以通过使用 A100、H800 等相对低端的芯片实现类似性能。这可能会影响英伟达 2025 年以后, B200 等最先进 GPU 在云计算、主权 AI 等领域的普及。Blackwell/Rubin 等最先进的 GPU 的用途, 初期可能会被局限在探索下一代超大规模模型 (Frontier Model) 上。

图表5: Nvidia、AMD、Broadcom、Marvell、Intel 数据中心及 AI 相关收入趋势



资料来源: Factset, 华泰研究

问题#5: DeepSeek 真的那么便宜吗?

据 DeepSeek V3 技术报告,V3 模型的训练总计只需要 278.8 万 GPU 小时,相当于在 2048 卡的 H800GPU 集群上训练约 2 个月,合计成本约 557.6 万美金,相较而言,Llama 3 系列模型的计算预算则多达 3930 万 H100 GPU 小时,DeepSeek 训练成本约相当于 Llama 3 系列模型的 7%。

但是,SemiAnalysis 在报告中指出,557.6 万美金这个数字主要指的是模型预训练的 GPU 成本,并不包括研发、数据收集、清理等其他重要成本。实际上,DeepSeek 的总体投资规模相当可观。据 SemiAnalysis 估计,其 GPU 投资就超过 5 亿美元。考虑到服务器资本支出、运营成本等因素,DeepSeek 的总拥有成本(TCO)在 4 年内可能达到 25.73 亿美元。DeepSeek 的成本优势主要体现在其高效的训练方法和创新的模型架构上。例如,其推理成本降至 OpenAI 的 1/50,这在实际应用中可以带来显著的成本节约。然而,这种成本优势并不意味着整体 AI 开发和运营成本的大幅降低。

图表6: DeepSeek AI TCO

DeepSeek AI TCO						
Chip	Unit	A100	H20	H800	H100	Total
Years	#	4	4	4	4	
# of GPUs	#	10,000	30,000	10,000	10,000	60,000
NVDA \$ ASP	\$	\$13,500	\$12,500	\$20,000	\$23,000	\$46,000
Server CapEx / GPU	\$	\$23,716	\$24,228	\$31,728	\$34,728	\$79,672
Total Server CapEx	\$m	\$237	\$727	\$317	\$347	\$1,281
Cost to Operation	\$m	\$157	\$387	\$170	\$230	\$715
Total TCO (4y Ownership)	\$m / hr	\$395	\$1,114	\$487	\$577	\$1,996

资料来源: Semianalysis, 华泰研究

DeepSeek 在多个方面展现出技术创新,主要包括模型架构创新、训练方法突破、蒸馏优化、推理效率提升等。其中,混合专家(MoE)架构和多头潜在注意力(MLA)的引入显著提升了模型性能和效率。R1-zero 模型采用纯强化学习(RL)训练,跳过监督微调,验证了 RL 在 AI 训练中的优先级和有效性。这些创新使 DeepSeek 在性能、效率和成本方面都取得了显著进展,为 AI 技术的发展提供了新的方向。特别是在解决复杂数学、物理和推理问题时,其速度是 ChatGPT 的两倍,且在编程问题上提供了迅速而全面的答案。

- 1) 采用混合专家 (MoE) 架构, 动态调用子模型降低计算量;
 - 2) 引入多头潜在注意力 (MLA) 压缩内存, 支持长文本处理;
 - 3) 强化学习 (RL): R1-zero 模型跳过监督微调, 通过纯 RL 直接训练, 验证了 RL 的优先级和有效性;
 - 4) 蒸馏优化: 利用 R1 生成数据微调小模型, 提升特定场景性能 (如数学、代码任务)
- 模型架构创新、训练方法突破、蒸馏优化、推理效率提升等。

The diagram illustrates the architecture of the proposed DeepSeekMoE Transformer Block, which is composed of three main components: the Transformer Block, DeepSeekMoE, and Multi-Head Latent Attention (MLA).

Transformer Block: This block consists of a sequence of L identical sub-layers. Each sub-layer contains an RMS Norm layer, followed by an Attention layer, another RMS Norm layer, and a Feed-Forward Network (FFN). The output of the FFN is added to the input of the Attention layer via a residual connection (indicated by a circle with a plus sign).

DeepSeekMoE: This component is a multi-expert MoE layer. It takes an Input Hidden $\mathbf{u}_t$ and routes it to a set of N_s Shared Experts (green boxes) and N_r Routed Experts (blue boxes). The Router selects the Top- K_r experts based on a gating mechanism. The output of the DeepSeekMoE is the Output Hidden $\mathbf{h}'_t$, which is the sum of the outputs from the selected experts.

Multi-Head Latent Attention (MLA): This component is a multi-head attention mechanism. It takes an Input Hidden $\mathbf{h}_t$ and a Latent $\mathbf{c}_t^{KV}$ as input. The Input Hidden $\mathbf{h}_t$ is projected to $\{\mathbf{q}_{t,i}^C; \mathbf{q}_{t,i}^R\}$ and $\{\mathbf{k}_{t,i}^C; \mathbf{k}_{t,i}^R\}$. The Latent $\mathbf{c}_t^{KV}$ is projected to $\{\mathbf{v}_{t,i}^C\}$. The Q and K vectors are concatenated and passed through a Multi-Head Attention layer. The output is the Output Hidden $\mathbf{u}_t$. The R vectors are used to generate the Latent $\mathbf{c}_t^{KV}$. The diagram also indicates that the R vectors are cached during inference.

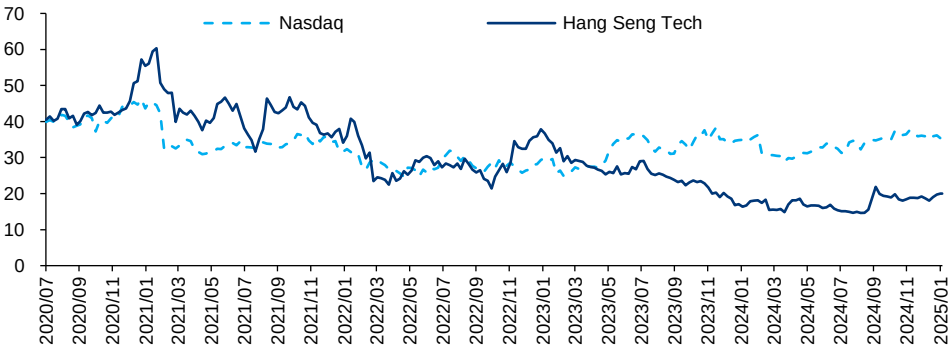
8



问题#7: DeepSeek 会带动中国科技资产价值重估吗?

目前(2025/2/3),恒生科技指数 12 月前向 PE 20.0 倍,远低于纳斯达克的 35.4 倍。DeepSeek 的成功可能提高投资者对中国 AI 公司产业链的估值预期,吸引更多资本投资中国 AI 领域。

图表8: Nasdaq vs 恒生科技一年前向 PE



资料来源: Bloomberg, 华泰研究

图表9: Nasdaq vs 恒生科技一年前向 PE

公司/指数名称		市值 (百万美元)	收盘价 (当地货币)	PE		股价变动 (%)				
				2025E	2026E	1W	30D	90D	180D	1Y
CCMP INDEX	Nasdaq		19,627	28.2	23.7	1.5	0.0	6.4	21.2	25.6
HSTECH US	Hang Seng Tech		4,468	15.9	13.0	1.2	7.5	0.8	39.5	55.4
MSFT US	Microsoft	3,085,549	415	31.4	27.4	-4.5	-2.0	0.9	4.2	0.9
GOOGL US	Google	2,329,879	204	22.7	19.8	6.4	6.4	20.2	28.4	43.3
META US	Meta	1,502,413	689	27.4	24.1	4.4	14.0	20.4	41.0	45.1
AMZN US	Amazon	2,499,208	238	37.9	30.7	1.0	6.0	19.1	46.0	38.3
AAPL US	Apple	3,548,938	236	32.2	28.7	2.7	-3.0	5.6	12.5	27.0
小计				30.3	26.2	2.0	4.3	13.2	26.4	30.9
9888 HK	Baidu	24,712	84	8.8	8.0	-0.4	4.2	-6.5	3.9	-17.2
9988 HK	Alibaba	229,535	94	10.2	9.2	7.7	13.8	-4.5	22.9	33.0
700 HK	Tencent	478,471	404	15.0	13.4	2.2	-2.4	-5.5	11.2	44.7
1810 HK	Xiaomi	100,250	38	34.1	27.7	2.3	4.7	34.6	138.7	211.1
小计				17.0	14.6	3.0	5.1	4.5	44.2	67.9

注: 数据截至 2025/2/3

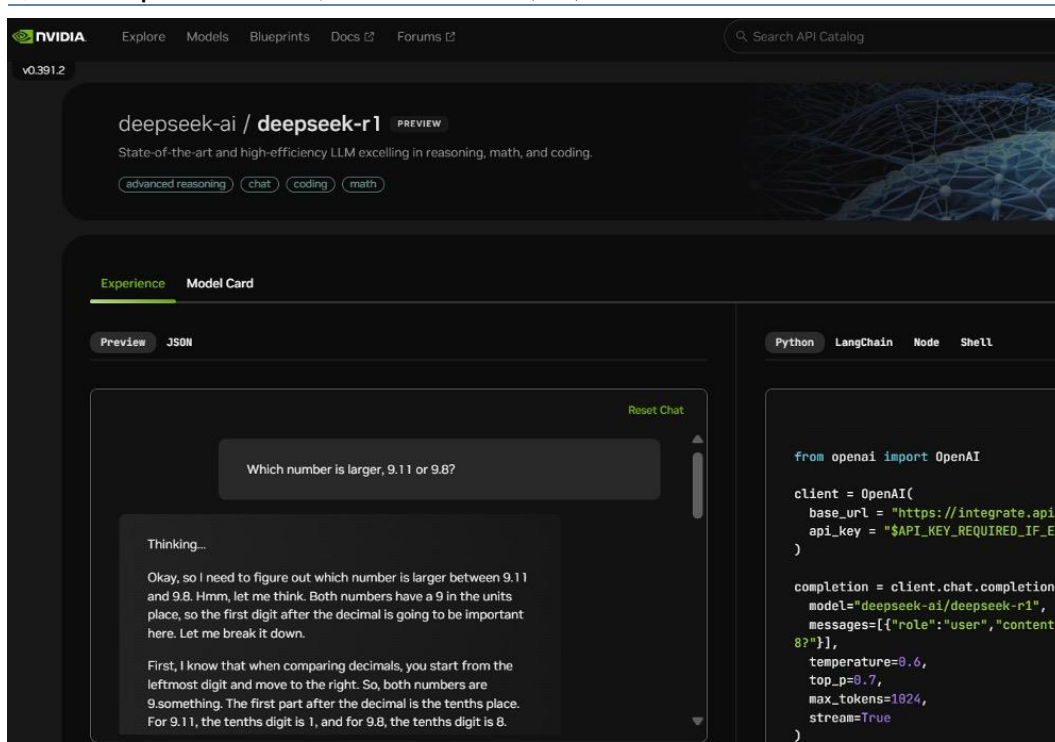
资料来源: Bloomberg, 华泰研究

问题#8: DeepSeek 会推动端侧智能发展吗?

DeepSeek 很可能会显著推动端侧智能的发展,其高性价比模型使得更多企业可能考虑在端侧设备上部署 AI 应用,推动智能设备的智能化进程。微软推出的专为 NPU 设计的 DeepSeek-R1 模型,支持 Copilot+PC 等设备,实现半连续运行的主动智能体验,为智能手机、汽车等端侧设备提供了高效本地化部署方案。DeepSeek 的技术进步可能加速端侧智能的普及和发展,推动 AI 技术向更广泛的终端设备渗透,从而改变现有的计算范式和用户交互方式。这不仅提升了端侧 AI 性能,还有助于保护用户隐私,降低了 AI 应用部署的技术门槛。

但是,从 Apple Intelligence 过去一年的发展历程中,我们看到,智能硬件的迭代是循序渐进过程,不会一蹴而就。模型能力提升只是其中一环,还存在生态链协调等很多挑战,对 2025 年 AI 手机等端侧智能的发展不应抱过高预期。

图表10: DeepSeek-R1 模型已作为 NVIDIA NIM 微服务发布



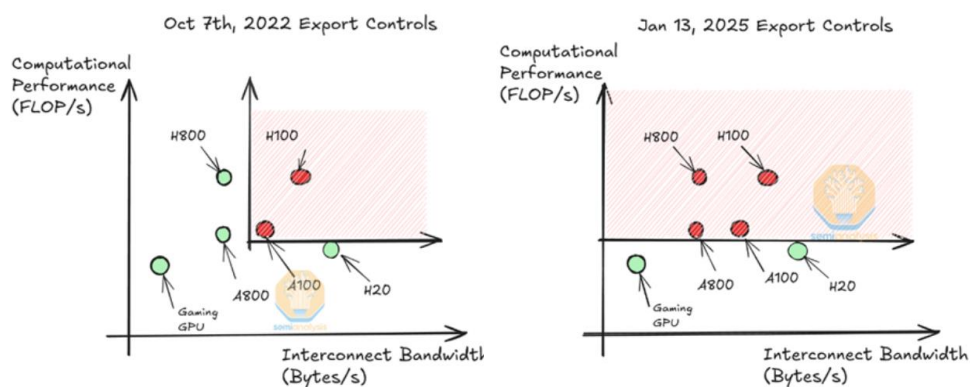
资料来源:英伟达官网,华泰研究

问题#9: DeepSeek 会导致美国提升出口管制压力吗?

DeepSeek 发布后，美国媒体进一步限制中国发展 AI 声音抬头。我们看到以下几个风险：

- 1) 芯片出口管制收紧：美国可能会加强对高端 AI 芯片的出口管制；
- 2) 开源限制：美国政府可能会限制科技公司开源大模型，以防止技术扩散；
- 3) 模型回传限制：在 2024 年年底发布的出口管制政策中，美国政府已经限制在新加坡等第三国训练好的模型回传中国，阻止技术转移；
- 4) 数据获取限制：美国可能会限制用于 AI 训练的大规模数据集的获取；
- 5) 算力资源限制：除了芯片管制，还可能限制美国云服务提供商的高性能计算资源的使用。

图表11：DeepSeek 或催化美国提升出口限制压力

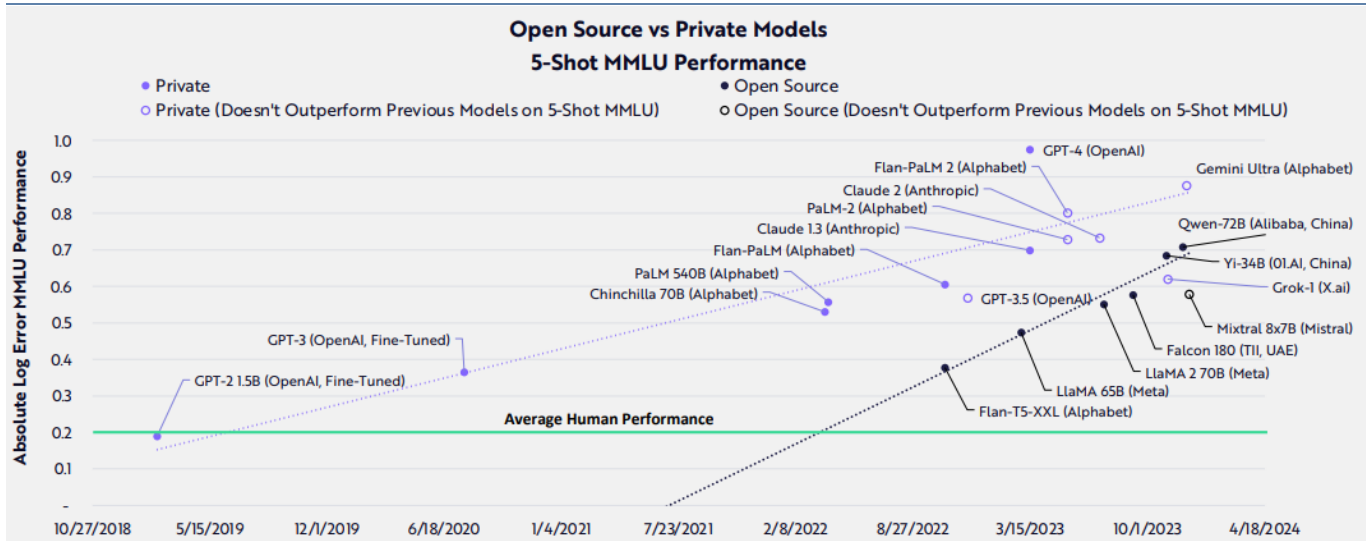


资料来源：Semianalysis，华泰研究

问题#10: DeepSeek 会改变开源软件生态吗?

全球主要大模型公司中, OpenAI、Google、Anthropic、字节、百度等采用闭源模式, Meta、阿里等采用开源模式。如下图所示, 闭源模型一直保持对开源模型的性能优势。

图表12: 开源模型性能进步速度快于闭源模型



资料来源: Big Ideas 2024 (Ark Investment, 2024), 华泰研究

这次 DeepSeek 通过开源接近最先进的闭源的高性能模型, 可能降低 AI 技术的使用门槛。开源模式带来边际成本持续下降的技术红利, 为 AI 技术的真正普及奠定了基础。此外, DeepSeek 的做法也可能推动其它 AI 公司重新思考其商业模式。

图表13: 国内外主流 AI 大模型对比

地区	公司	模型	参数量	模态 (输入; 输出)	开源/闭源	发布时间
美国	OpenAI	GPT-o1	300B	文本, 图像; 文本, 图像	闭源	2024 年 12 月
美国	OpenAI	GPT-4o	未公开	文本, 图像, 音频, 视频 (帧); 文本, 图像, 音频	闭源	2024 年 5 月
美国	OpenAI	GPT-4 turbo	未公开	文本, 图像; 文本, 图像	闭源	2023 年 11 月
美国	OpenAI	Sora	未公开	文本, 图像; 图像, 视频	闭源	2024 年 2 月
美国	谷歌	Gemini 1.5	未公开	文本, 图像, 音频, 视频 (帧); 文本, 图像	闭源	2024 年 2 月
美国	谷歌	Gemini	未公开	文本, 图像, 视频; 文本, 图像	闭源	2023 年 12 月
美国	Anthropic	Claude 3	未公开	文本, 图像; 文本	闭源	2024 年 3 月
中国	百度	ERNIE 4.0	未公开	文本, 图像; 文本, 图像, 视频	闭源	2023 年 10 月
中国	商汤	日日新 5.0	600B (MoE)	文本, 图像; 文本, 图像, 视频	闭源	2024 年 4 月
中国	智谱 AI	GLM-4	未公开	文本, 图像; 文本, 图像	闭源	2024 年 1 月
美国	Meta	Llama 3	70B	文本; 文本	开源	2024 年 4 月
美国	微软	Phi-3	3.8B	文本; 文本	开源	2024 年 4 月
中国	阿里巴巴	Qwen 1.5	72B	文本; 文本	开源	2024 年 2 月
中国	零一万物	Yi-VL	34B	文本; 文本	开源	2024 年 1 月
美国	xAI	Grok-1	314B (MoE)	文本; 文本	开源	2023 年 11 月
中国	深度求索	DeepSeek-R1	671B	文本; 文本	开源	2025 年 1 月

资料来源: Stanford ecosystem graphs, 公司官网, 华泰研究

风险提示

1) 中美贸易摩擦升级, 影响产品供需与公司海外布局风险;

2) 宏观经济下行风险;

3) 创新品渗透不及预期, 导致公司增长不及预期。

4) 本研报中涉及到未上市公司或未覆盖个股内容, 均系对其客观公开信息的整理, 并不代表本研究团队对该公司、该股票的推荐或覆盖

免责声明

分析师声明

本人，黄乐平、陈旭东，兹证明本报告所表达的观点准确地反映了分析师对标的证券或发行人的个人意见；彼以往、现在或未来并无就其研究报告所提供的具体建议或所表达的意见直接或间接收取任何报酬。

一般声明及披露

本报告由华泰证券股份有限公司（已具备中国证监会批准的证券投资咨询业务资格，以下简称“本公司”）制作。本报告所载资料是仅供接收人的严格保密资料。本报告仅供本公司及其客户和其关联机构使用。本公司不因接收人收到本报告而视其为客户。

本报告基于本公司认为可靠的、已公开的信息编制，但本公司及其关联机构（以下统称为“华泰”）对该等信息的准确性及完整性不作任何保证。

本报告所载的意见、评估及预测仅反映报告发布当日的观点和判断。在不同时期，华泰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。同时，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。以往表现并不能指引未来，未来回报并不能得到保证，并存在损失本金的可能。华泰不保证本报告所含信息保持在最新状态。华泰对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本公司不是 FINRA 的注册会员，其研究分析师亦没有注册为 FINRA 的研究分析师/不具有 FINRA 分析师的注册资格。

华泰力求报告内容客观、公正，但本报告所载的观点、结论和建议仅供参考，不构成购买或出售所述证券的要约或招揽。该等观点、建议并未考虑到个别投资者的具体投资目的、财务状况以及特定需求，在任何时候均不构成对客户私人投资建议。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。对依据或者使用本报告所造成的一切后果，华泰及作者均不承担任何法律责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现，过往的业绩表现不应作为日后回报的预示。华泰不承诺也不保证任何预示的回报会得以实现，分析中所做的预测可能是基于相应的假设，任何假设的变化可能会显著影响所预测的回报。

华泰及作者在自身所知情的范围内，与本报告所指的证券或投资标的不存在法律禁止的利害关系。在法律许可的情况下，华泰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，为该公司提供投资银行、财务顾问或者金融产品等相关服务或向该公司招揽业务。

华泰的销售人员、交易人员或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。华泰没有将此意见及建议向报告所有接收者进行更新的义务。华泰的资产管理部门、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。投资者应当考虑到华泰及/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信赖依据。有关该方面的具体披露请参照本报告尾部。

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、发布的机构或人员，也并非意图发送、发布给因可得到、使用本报告的行为而使华泰违反或受制于当地法律或监管规则的机构或人员。

本报告版权仅为本公司所有。未经本公司书面许可，任何机构或个人不得以翻版、复制、发表、引用或再次分发他人（无论整份或部分）等任何形式侵犯本公司版权。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并需在使用前获取独立的法律意见，以确定该引用、刊发符合当地适用法规的要求，同时注明出处为“华泰证券研究所”，且不得对本报告进行任何有悖原意的引用、删节和修改。本公司保留追究相关责任的权利。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

中国香港

本报告由华泰证券股份有限公司制作，在香港由华泰金融控股（香港）有限公司向符合《证券及期货条例》及其附属法律规定的机构投资者和专业投资者的客户进行分发。华泰金融控股（香港）有限公司受香港证券及期货事务监察委员会监管，是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。在香港获得本报告的人员若有任何有关本报告的问题，请与华泰金融控股（香港）有限公司联系。

香港-重要监管披露

- 华泰金融控股（香港）有限公司的雇员或其关联人士没有担任本报告中提及的公司或发行人的高级人员。
- 有关重要的披露信息，请参华泰金融控股（香港）有限公司的网页 https://www.htsc.com.hk/stock_disclosure 其他信息请参见下方 “美国-重要监管披露”。

美国

在美国本报告由华泰证券（美国）有限公司向符合美国监管规定的机构投资者进行发表与分发。华泰证券（美国）有限公司是美国注册经纪商和美国金融业监管局（FINRA）的注册会员。对于其在美国分发的研究报告，华泰证券（美国）有限公司根据《1934 年证券交易法》（修订版）第 15a-6 条规定以及美国证券交易委员会人员解释，对本研究报告内容负责。华泰证券（美国）有限公司联营公司的分析师不具有美国金融监管（FINRA）分析师的注册资格，可能不属于华泰证券（美国）有限公司的关联人员，因此可能不受 FINRA 关于分析师与标的公司沟通、公开露面和所持交易证券的限制。华泰证券（美国）有限公司是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。任何直接从华泰证券（美国）有限公司收到此报告并希望就本报告所述任何证券进行交易的人士，应通过华泰证券（美国）有限公司进行交易。

美国-重要监管披露

- 分析师黄乐平、陈旭东本人及相关人士并不担任本报告所提及的标的证券或发行人的高级人员、董事或顾问。分析师及相关人士与本报告所提及的标的证券或发行人并无任何相关财务利益。本披露中所提及的“相关人士”包括 FINRA 定义下分析师的家庭成员。分析师根据华泰证券的整体收入和盈利能力获得薪酬，包括源自公司投资银行业务的收入。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或不时会以自身或代理形式向客户出售及购买华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或其高级管理层、董事和雇员可能会持有本报告中所提到的任何证券（或任何相关投资）头寸，并可能不时进行增持或减持该证券（或投资）。因此，投资者应该意识到可能存在利益冲突。

新加坡

华泰证券（新加坡）有限公司持有新加坡金融管理局颁发的资本市场服务许可证，可从事资本市场产品交易，包括证券、集体投资计划中的单位、交易所交易的衍生品合约和场外衍生品合约，并且是《财务顾问法》规定的豁免财务顾问，就投资产品向他人提供建议，包括发布或公布研究分析或研究报告。华泰证券（新加坡）有限公司可能会根据《财务顾问条例》第 32C 条的规定分发其在华泰内的外国附属公司各自制作的信息/研究。本报告仅供认可投资者、专家投资者或机构投资者使用，华泰证券（新加坡）有限公司不对本报告内容承担法律责任。如果您是非预期接收者，请您立即通知并直接将本报告返回给华泰证券（新加坡）有限公司。本报告的新加坡接收者应联系您的华泰证券（新加坡）有限公司关系经理或客户主管，了解来自或所分发的信息相关的事宜。

评级说明

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力（含此期间的股息回报）相对基准表现的预期（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数，台湾市场基准为台湾加权指数，日本市场基准为日经 225 指数，新加坡市场基准为海峡时报指数，韩国市场基准为韩国有价证券指数，英国市场基准为富时 100 指数），具体如下：

行业评级

增持：预计行业股票指数超越基准

中性：预计行业股票指数基本与基准持平

减持：预计行业股票指数明显弱于基准

公司评级

买入：预计股价超越基准 15% 以上

增持：预计股价超越基准 5%~15%

持有：预计股价相对基准波动在-15%~5%之间

卖出：预计股价弱于基准 15% 以上

暂停评级：已暂停评级、目标价及预测，以遵守适用法规及/或公司政策

无评级：股票不在常规研究覆盖范围内。投资者不应期待华泰提供该等证券及/或公司相关的持续或补充信息

法律实体披露

中国: 华泰证券股份有限公司具有中国证监会核准的“证券投资咨询”业务资格, 经营许可证编号为: 91320000704041011J

香港: 华泰金融控股(香港)有限公司具有香港证监会核准的“就证券提供意见”业务资格, 经营许可证编号为: AOK809

美国: 华泰证券(美国)有限公司为美国金融业监管局(FINRA)成员, 具有在美国开展经纪交易商业业务的资格, 经营业务许可编号为: CRD#:298809/SEC#:8-70231

新加坡: 华泰证券(新加坡)有限公司具有新加坡金融管理局颁发的资本市场服务许可证, 并且是豁免财务顾问。公司注册号: 202233398E

华泰证券股份有限公司**南京**

南京市建邺区江东中路228号华泰证券广场1号楼/邮政编码: 210019

电话: 86 25 83389999/传真: 86 25 83387521

电子邮件: ht-rd@htsc.com

深圳

深圳市福田区益田路5999号基金大厦10楼/邮政编码: 518017

电话: 86 755 82493932/传真: 86 755 82492062

电子邮件: ht-rd@htsc.com

北京

北京市西城区太平桥大街丰盛胡同28号太平洋保险大厦A座18层/

邮政编码: 100032

电话: 86 10 63211166/传真: 86 10 63211275

电子邮件: ht-rd@htsc.com

上海

上海市浦东新区东方路18号保利广场E栋23楼/邮政编码: 200120

电话: 86 21 28972098/传真: 86 21 28972068

电子邮件: ht-rd@htsc.com

华泰金融控股(香港)有限公司

香港中环皇后大道中99号中环中心53楼

电话: +852-3658-6000/传真: +852-2567-6123

电子邮件: research@htsc.com

http://www.htsc.com.hk

华泰证券(美国)有限公司

美国纽约公园大道280号21楼东(纽约10017)

电话: +212-763-8160/传真: +917-725-9702

电子邮件: Huatai@htsc-us.com

http://www.htsc-us.com

华泰证券(新加坡)有限公司

滨海湾金融中心1号大厦, #08-02, 新加坡 018981

电话: +65 68603600

传真: +65 65091183

©版权所有2025年华泰证券股份有限公司