平安证券

计算机

2025 年 02 月 06 日

AI 动态跟踪系列（四）

DeepSeek 引发广泛关注，大模型应用落地将加速

强于大市（维持）

行情走势图

相关研究报告

【平安证券】行业动态跟踪报告*计算机*AI 动态跟踪系列（三）：复杂推理大模型 OpenAI o1 亮相，数学与代码能力飞跃*强于大市 20240914

【平安证券】行业动态跟踪报告*计算机*AI 动态跟踪系列（二）：英伟达 GTC 2024 AI 软件与应用有哪些看点？*强于大市 20240327

【平安证券】行业动态跟踪报告*计算机*AI 动态跟踪系列（一）：Duolingo4Q23 业绩超预期，持续关注 AI+教育应用前景*强于大市 20240305

证券分析师

闫磊 投资咨询资格编号 S1060517070006 YANLEI511@pingan.com.cn

付强 投资咨询资格编号 S1060520070001 FUQIANG021@pingan.com.cn

黄韦涵 投资咨询资格编号 S1060523070003 HUANGWEIHAN235@pingan.com.cn

研究助理

王佳一 一般证券从业资格编号 S1060123070023 WANGJIAYI446@pingan.com.cn

平安观点：

- **DeepSeek-V3 和 DeepSeek-R1 陆续发布，国产大模型能力已可比肩海外领军大模型。**2024 年 12 月 26 日，杭州 AI 公司深度求索（DeepSeek）正式发布 DeepSeek-V3 大模型首个版本并同步开源。根据 DeepSeek 网站信息，DeepSeek-V3 为自研 MoE 模型，671B 参数，激活 37B，在 14.8T tokens 上进行了预训练。DeepSeek-V3 多项评测成绩超越了 Qwen2.5-72B 和 Llama-3.1-405B 等其他开源模型，并在性能上和世界顶尖的闭源模型 GPT-4o 以及 Claude-3.5-Sonnet 相当。在训练成本方面，根据 DeepSeek 发布的技术文档论文信息，DeepSeek-V3 的训练时长为 2788K 个 H800 GPU 小时，训练花费约为 557.6 万美元。2025 年 1 月 20 日，DeepSeek 正式发布复杂推理类大模型 DeepSeek-R1，性能对齐 OpenAI o1 正式版。以 DeepSeek 系列大模型为代表的国产大模型性能已可比肩海外领军大模型，且成本更低。

- **DeepSeek 系列大模型引发全球广泛关注，海内外巨头科技公司及云服务平台厂商已相继接入。**2025 年 1 月 15 日，DeepSeek 推出 AI 助手 DeepSeek App。2025 年春节期间，DeepSeek 系列大模型火爆出圈，引发全球广泛关注。根据新浪财经 2 月 1 日引用彭博社信息，DeepSeek 的 AI 助手在 140 个市场中成为下载量最多的移动应用。根据 Appfigures 的数据，DeepSeek 的推理人工智能聊天机器人在 1 月 26 日登上苹果公司 App Store 的榜首，并自那时以来一直保持全球第一的位置。同时，全球也开始了对 DeepSeek 大模型的复刻。以港科大团队为例，港科大助理教授何俊贤的团队，只用了 8K 个样本，就在 7B 模型上复刻出了 DeepSeek-R1-Zero 和 DeepSeek-R1 的训练。当前，海内外巨头科技公司及云服务平台厂商已相继接入了 DeepSeek 大模型，部分 AI 应用领域相关企业也已开始了 DeepSeek 大模型的部署和应用。DeepSeek 大模型获得了全球的广泛关注，认可度持续提升。我们认为，DeepSeek 大模型的开源、低成本和高性能将大幅降低大模型的获得、部署和应用成本，将加快大模型在 B 端和 C 端应用场景的落地。另外，DeepSeek 大模型的出圈将对全球大模型产业的竞争格局产生重要影响，将对海外领军大模型厂商的领先性产生冲击，并同时将对算力的未来发展产生重要影响。

- **DeepSeek 大模型的出圈预计不改算力整体需求向上的态势，但推理和端侧算力有望增长更快。**DeepSeek 在算法效率和计算成本方面有着较大的优势，短期内可能对训练算力的增长有一定的平抑效应，但是不改 AI 算力整体需求长期上升的态势。AI 作为全球智能化发展的主要抓手，大模型当前已应用于端侧、教育、金融、办公、传媒、医疗、智能汽车、企业服务等多个应用场景，应用领域广阔。DeepSeek 低成本而且开源的解决方案，大幅降低了 AI 在各行各业应用的技术和成本门槛，为 AI 的产业化落

地提供了更快的路径。推理和端侧的算力需求增长潜力非常大。同时，较低训练成本以及开源的 DeepSeek，有望带来更低的大模型开发和使用门槛，基于该大模型开发的主体可能更多，也一定程度上为训练算力需求提供了支撑。DeepSeek 并不是压缩了算力市场，反而为算力市场增加了更多的想象空间。DeepSeek 也在积极与国产 AI 算力平台合作。DeepSeek 大模型与国产 AI 芯片适配的逐步成熟，将加快推动国产 AI 芯片在国内大模型训练端和推理端的应用，加快国产 AI 芯片产业链的成熟，为国产 AI 芯片产业带来发展机遇，同时加快我国大模型产业的发展。

- **投资建议：**DeepSeek-V3 和 DeepSeek-R1 等 DeepSeek 系列大模型的陆续发布，表明国产大模型能力已可比肩海外领军大模型。我们认为，DeepSeek 大模型的开源、低成本和高性能将大幅降低大模型的获得、部署和应用成本，将加快大模型在 B 端和 C 端应用场景的落地。DeepSeek 大模型的出圈，短期内可能对训练算力的增长有一定的平抑效应，但预计不改算力整体需求向上的态势，而且推理和端侧算力有望增长更快。DeepSeek 大模型与国产 AI 芯片适配的逐步成熟，将加快推动国产 AI 芯片在国内大模型训练端和推理端的应用，加快国产 AI 芯片产业链的成熟，为国产 AI 芯片产业带来发展机遇，同时加快我国大模型产业的发展。我们坚定看好 AI 主题的投资机会，标的方面：1）国产算力基础设施方面，推荐浪潮信息、中科曙光、紫光股份、神州数码、海光信息、龙芯中科，建议关注寒武纪、景嘉微、软通动力、华勤技术；2）端侧算力方面，推荐恒玄科技、兆易创新，关注乐鑫科技、瑞芯微；3）算法方面，推荐科大讯飞；4）应用场景方面，强烈推荐中科创达、恒生电子、盛视科技，推荐金山办公、德赛西威、万兴科技、福昕软件，建议关注同花顺、拓尔思、彩讯股份、卫宁健康。

- **风险提示：**1）AI 算力供应链风险上升。2）国产大模型算法发展可能不及预期。3）大模型产品的应用落地低于预期。

一、 DeepSeek-V3 和 DeepSeek-R1 陆续发布，国产大模型能力已可比肩海外领军大模型

■ DeepSeek-V3

2024 年 12 月 26 日，杭州 AI 公司深度求索(DeepSeek)正式发布 DeepSeek-V3 大模型首个版本并同步开源。根据 DeepSeek 网站信息，DeepSeek-V3 为自研 MoE 模型，671B 参数，激活 37B，在 14.8T tokens 上进行了预训练。DeepSeek-V3 多项评测成绩超越了 Qwen2.5-72B 和 Llama-3.1-405B 等其他开源模型，并在性能上和世界顶尖的闭源模型 GPT-4o 以及 Claude-3.5-Sonnet 相当。具体而言：

- 百科知识方面：DeepSeek-V3 在知识类任务（MMLU，MMLU-Pro，GPQA，SimpleQA）上的水平相比前代 DeepSeek-V2.5 明显提升，接近当前表现最好的模型 Claude-3.5-Sonnet-1022。
- 长文本方面：在长文本测评中，DROP、FRAMES 和 LongBench v2 上，DeepSeek-V3 平均表现超越其他模型。
- 代码方面：DeepSeek-V3 在算法类代码场景（Codeforces），远远领先于市面上已有的全部非 o1 类模型；并在工程类代码场景（SWE-Bench Verified）逼近 Claude-3.5-Sonnet-1022。
- 数学：在美国数学竞赛（AIME 2024，MATH）和全国高中数学联赛（CNMO 2024）上，DeepSeek-V3 大幅超过了所有开源闭源模型。
- 中文能力方面：DeepSeek-V3 与 Qwen2.5-72B 在教育类测评 C-Eval 和代词消歧等评测集上表现相近，但在事实知识 C-SimpleQA 上更为领先。

图表1 DeepSeek-V3 大模型与海外领军开闭源大模型的测试比较

测试集		DeepSeek-V3	Qwen2.5 72B-Inst.	Llama3.1 405B-Inst.	Claude-3.5-Sonnet-1022	GPT-4o 0513
	模型架构	MoE	Dense	Dense	-	-
	# 激活参数	37B	72B	405B	-	-
	# 总参数	671B	72B	405B	-	-
英文	MMLU (EM)	88.5	85.3	**88.6**	88.3	87.2
	MMLU-Redux (EM)	**89.1**	85.6	86.2	88.9	88
	MMLU-Pro (EM)	75.9	71.6	73.3	**78**	72.6
	DROP (3-shot F1)	**91.6**	76.7	88.7	88.3	83.7
	IF-Eval (Prompt Strict)	86.1	84.1	86	**86.5**	84.3
	GPQA-Diamond (Pass@1)	59.1	49	51.1	**65**	49.9
	SimpleQA (Correct)	24.9	9.1	17.1	28.4	**38.2**
	FRAMES (Acc.)	73.3	69.8	70	72.5	**80.5**
	LongBench v2 (Acc.)	**48.7**	39.4	36.1	41	48.1
代码	HumanEval-Mul (Pass@1)	**82.6**	77.3	77.2	81.7	80.5
	LiveCodeBench(Pass@1-COT)	**40.5**	31.1	28.4	36.3	33.4
	LiveCodeBench (Pass@1)	**37.6**	28.7	30.1	32.8	34.2
	Codeforces (Percentile)	**51.6**	24.8	25.3	20.3	23.6
	SWE Verified (Resolved)	42	23.8	24.5	**50.8**	38.8
	Aider-Edit (Acc.)	79.7	65.4	63.9	**84.2**	72.9
	Aider-Polyglot (Acc.)	**49.6**	7.6	5.8	45.3	16
数学	AIME 2024 (Pass@1)	**39.2**	23.3	23.3	16	9.3
	MATH-500 (EM)	**90.2**	80	73.8	78.3	74.6
	CNMO 2024 (Pass@1)	**43.2**	15.9	6.8	13.1	10.8
中文	CLUEWSC (EM)	90.9	**91.4**	84.7	85.4	87.9
	C-Eval (EM)	**86.5**	86.1	61.5	76.7	76
	C-SimpleQA (Correct)	**64.1**	48.4	50.4	51.3	59.3

资料来源：DeepSeek 官网，平安证券研究所

DeepSeek-V3 生成 Tokens 的速度跳跃式提升。根据 DeepSeek 官网信息，通过算法和工程上的创新，DeepSeek-V3 的生成吐字速度从 20 TPS（Tokens Per Second，每秒生成 Tokens） 大幅提高至 60 TPS，相比 V2.5 模型实现了 3 倍的提升，可以为用户带来更加迅速流畅的使用体验。

在训练成本方面，根据 DeepSeek 发布的技术文档论文信息，DeepSeek-V3 的训练时长为 2788K 个 H800 GPU 小时，训练花费约为 557.6 万美元。根据 IT 之家消息，Meta 训练其 4050 亿参数模型 Llama 3，在 16384 块 H100 GPU 训练集群上训练了 54 天，训练时长约为 21,234K 个 H100 GPU 小时。以此估算，即使不考虑 Llama 3.1 相比 Llama 3 因为模型能力更强而可能需要的更长的训练时间，也不考虑 H100 相比 H800 有更高的性能，DeepSeek-V3 的训练时长仅为同水平大模型（Llama 3.1）的约 13%，训练成本大幅下降。

图表2 DeepSeek-V3 的训练时长为 2788K 个 H800 GPU 小时

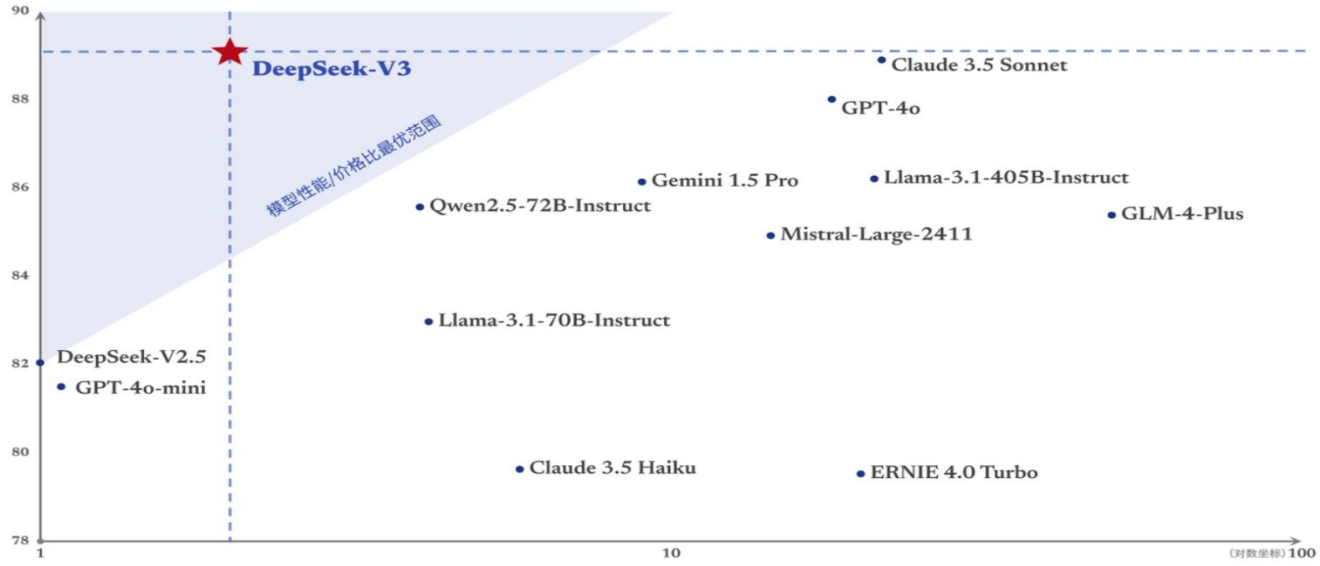
Training Costs	Pre-Training	Context Extension	Post-Training	Total
in H800 GPU Hours	2664K	119K	5K	2788K
in USD	$5.328M	$0.238M	$0.01M	$5.576M

资料来源：DeepSeek 官网，《DeepSeek-V3 Technical Report》，平安证券研究所

注：图表中训练花费金额的计算是假设 H800 的租赁价格是每小时 2 美元

DeepSeek-V3 的 API 服务价格更具性价比。随着性能更强、速度更快的 DeepSeek-V3 更新上线，DeepSeek 将 V2.5 的模型 API 服务定价调整提高为每百万输入 tokens 0.5 元（缓存命中）/ 2 元（缓存未命中），每百万输出 tokens 8 元。这个价格的提升将在 2025 年 2 月 8 日 24:00 之后生效。在此之前，属于新模型的优惠价格体验期，价格仍会是之前的每百万输入 tokens 0.1 元（缓存命中）/ 1 元（缓存未命中），每百万输出 tokens 2 元。在 2025 年 2 月 8 日之前，已经注册的老用户和在此期间内注册的新用户均可享受这个优惠价格。相比而言，GPT-4o 的 API 价格为每百万输入 tokens 1.25 美元（缓存命中）/ 2.50 美元（缓存未命中），每百万输出 tokens 10 美元。即使是调整之后的价格，DeepSeek-v3 的 API 服务价格也大幅低于 GPT-4o 的价格。在相当的性能下，DeepSeek-V3 的 API 服务价格更具性价比。

图表3 DeepSeek-V3 的 API 服务价格更具性价比



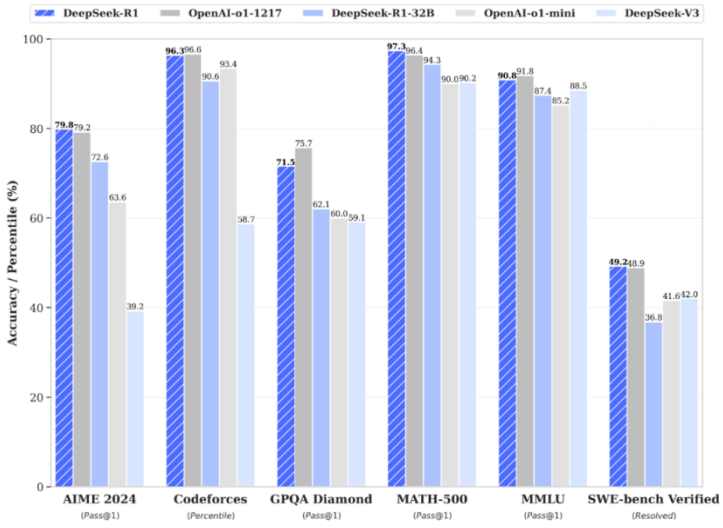
资料来源：DeepSeek 官网，平安证券研究所

注：图表是表示大模型 MMLU Redux ZeroEval 得分 vs 输入 API 价格（元/1M Tokens）

■ DeepSeek-R1

2025 年 1 月 20 日，即 DeepSeek-V3 正式发布之后不到 1 个月，也即 OpenAI o1 正式版推出（推出时间为 2024 年 12 月 6 日）之后不到 2 个月，DeepSeek 正式发布复杂推理模型 DeepSeek-R1 并同步开源模型权重。DeepSeek-R1 在后训练阶段大规模使用了强化学习技术，在仅有极少标注数据的情况下，极大提升了模型推理能力。在数学、代码、自然语言推理等任务上，性能比肩 OpenAI o1 正式版。

图表4 DeepSeek-R1 性能比肩 OpenAI o1 正式版



资料来源：DeepSeek 官网，平安证券研究所

另外，DeepSeek 在开源 DeepSeek-R1-Zero 和 DeepSeek-R1 这两个 660B 模型的同时，通过 DeepSeek-R1 的输出，蒸馏了 6 个小模型开源给社区，其中 32B 和 70B 模型在多项能力上实现了对标 OpenAI o1-mini 的效果。

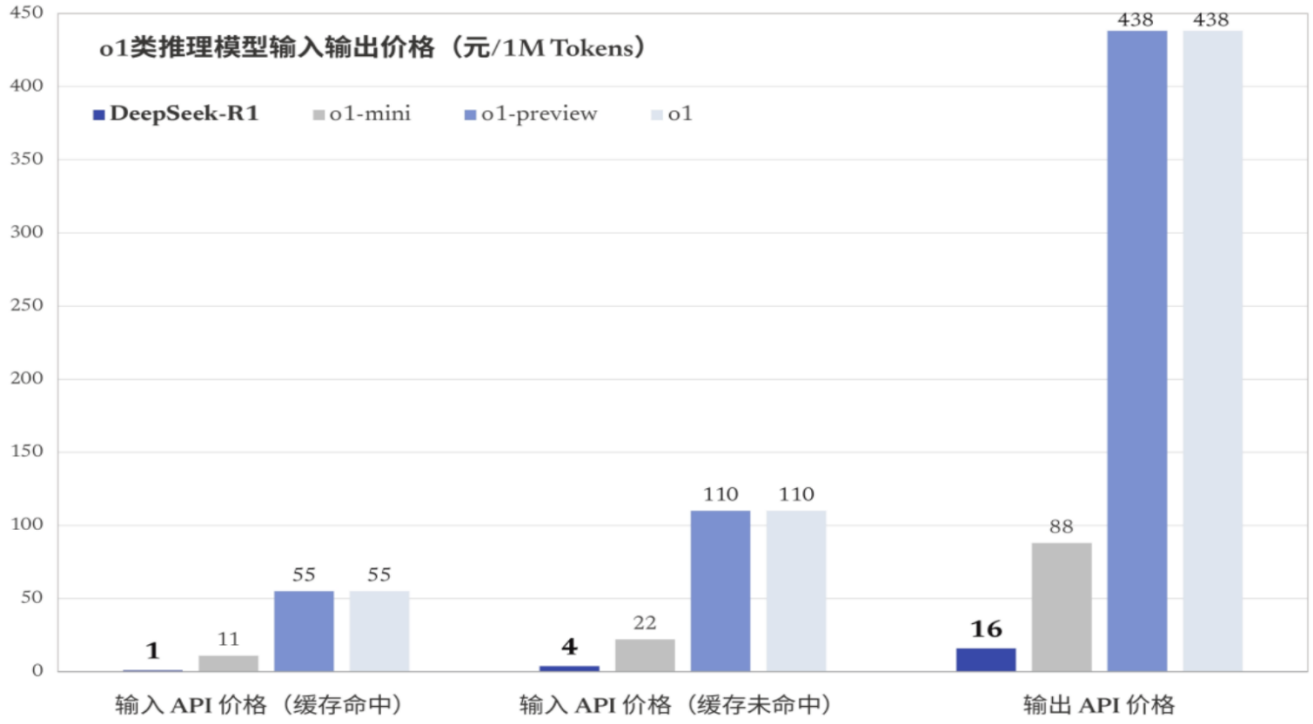
图表5 DeepSeek-R1 蒸馏的 32B 和 70B 模型在多项能力得分上超过了 OpenAI o1-mini

	AIME 2024 pass@1	AIME 2024 cons@64	MATH-500 pass@1	GPQA Diamond pass@1	LiveCodeBench pass@1	CodeForces rating
GPT-4o-0513	9.3	13.4	74.6	49.9	32.9	759.0
Claude-3.5-Sonnet-1022	16.0	26.7	78.3	65.0	38.9	717.0
o1-mini	63.6	80.0	90.0	60.0	53.8	**1820.0**
QwQ-32B	44.0	60.0	90.6	54.5	41.9	1316.0
DeepSeek-R1-Distill-Qwen-1.5B	28.9	52.7	83.9	33.8	16.9	954.0
DeepSeek-R1-Distill-Qwen-7B	55.5	83.3	92.8	49.1	37.6	1189.0
DeepSeek-R1-Distill-Qwen-14B	69.7	80.0	93.9	59.1	53.1	1481.0
DeepSeek-R1-Distill-Qwen-32B	**72.6**	83.3	94.3	62.1	57.2	1691.0
DeepSeek-R1-Distill-Llama-8B	50.4	80.0	89.1	49.0	39.6	1205.0
DeepSeek-R1-Distill-Llama-70B	70.0	**86.7**	**94.5**	**65.2**	**57.5**	1633.0

资料来源：DeepSeek 官网，平安证券研究所

DeepSeek-R1 在性能比肩 OpenAI o1 正式版的同时，API 服务价格相比 OpenAI o1 正式版大幅下降。根据 DeepSeek 官网信息，DeepSeek-R1 API 服务定价为每百万输入 tokens 1 元（缓存命中）/ 4 元（缓存未命中），每百万输出 tokens 16 元，分别为 OpenAI o1 正式版定价的约 1/55、2/55、8/219。相比 OpenAI o1，DeepSeek-R1 极具性价比。综合 DeepSeek-V3 和 DeepSeek-R1 大模型的表现，以 DeepSeek 系列大模型为代表的国产大模型性能已可比肩海外领军大模型，且成本（包括训练成本及 API 服务成本）更低。

图表6 DeepSeek-R1 API 服务价格相比 OpenAI o1 系列大模型大幅下降



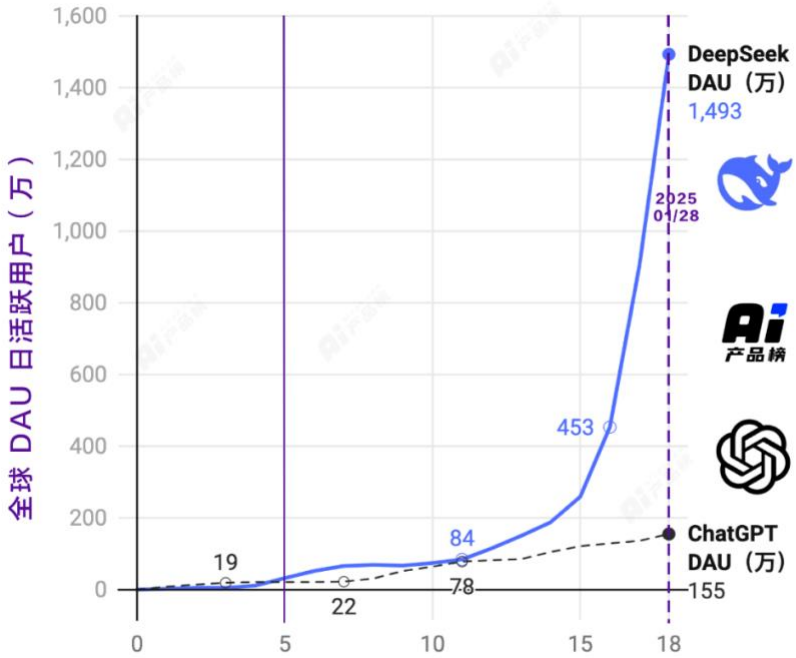
资料来源：DeepSeek 官网，平安证券研究所

DeepSeek 大模型的多模态能力也在持续提升。根据中国新闻网信息，1 月 28 日，人工智能社区 Hugging Face 显示，DeepSeek 发布了开源多模态 AI 模型 Janus-Pro。Janus-Pro 是 Janus 的高级版本，其拥有优化的训练策略，扩展的训练数据以及更大的模型规模，这些改进使得 Janus-Pro 在多模态理解和文本到图像的指令跟踪能力方面都取得了重大进步，同时还增强了文本到图像生成的稳定性。Janus-Pro 系列包括了参数量分别为 7B 和 1.5B 的两个型号。报告公开的测试结果显示，Janus-Pro-7B 在 GenEval 和 DPG-Bench 基准测试中击败了 OpenAI 的 DALL-E 3 和 Stable Diffusion。

二、 DeepSeek 系列大模型引发全球广泛关注，海内外巨头科技公司及云服务平台厂商已相继接入

2025 年 1 月 15 日，DeepSeek 推出 AI 助手 DeepSeek App。2025 年春节期间，DeepSeek 系列大模型火爆出圈，引发全球广泛关注。根据新浪财经 2 月 1 日引用彭博社信息，DeepSeek 的 AI 助手在 140 个市场中成为下载量最多的移动应用。根据 Appfigures 的数据，DeepSeek 的推理人工智能聊天机器人在 1 月 26 日登上苹果公司 App Store 的榜首，并自那时以来一直保持全球第一的位置。这款应用还在美国的 Android Play Store 中占据了榜首位置，并自 1 月 28 日以来一直保持该位置。根据 Sensor Tower 的数据，DeepSeek 在发布后的前 18 天内获得了 1600 万次下载，几乎是 OpenAI 的 ChatGPT 发布时 900 万下载量的两倍。根据 AI 产品榜数据，DeepSeek App 是全球增速最快的 AI 应用，仅上线 18 天，DAU（日活用户）就达到 1500 万。而 ChatGPT 的 DAU 过 1500 万花了 244 天，DeepSeek App DAU 达到 1500 万的速度是 ChatGPT 的约 13 到 14 倍。

图表7 DeepSeek App 仅上线 18 天 DAU（日活用户）就达到 1500 万



资料来源：AI 产品榜，平安证券研究所

同时，全球也开始了对 DeepSeek 大模型的复刻。以港科大团队为例，港科大助理教授何俊贤的团队，只用了 8K 个样本，就在 7B 模型上复刻出了 DeepSeek-R1-Zero 和 DeepSeek-R1 的训练。他们以 Qwen2.5-Math-7B（基础模型）为起点，直接对其进行强化学习。在整个过程中，他们没有进行监督微调（SFT），也没有使用奖励模型。最终，团队复刻出的模型在 AIME 基准上实现了 33.3%的准确率，在 AMC 上为 62.5%，在 MATH 上为 77.2%。这一表现不仅超越了 Qwen2.5-Math-7B-Instruct，并且还可以和使用超过 50 倍数据量和更复杂组件的 PRIME 和 rStar-MATH 相媲美，复刻结果表现亮眼。除了港科大和加利福尼亚大学伯克利分校等学校研究人员对 DeepSeek 大模型的复刻，全球人工智能领域知名社区 Hugging Face 也宣布复刻 DeepSeek-R1，并将这个项目命名为 Open R1。截至目前，研究机构对 DeepSeek-R1 复刻结果的亮眼表现，表明了 DeepSeek 大模型技术理念的稳定性和可靠性，也彰显了 DeepSeek 大模型的发展潜力。

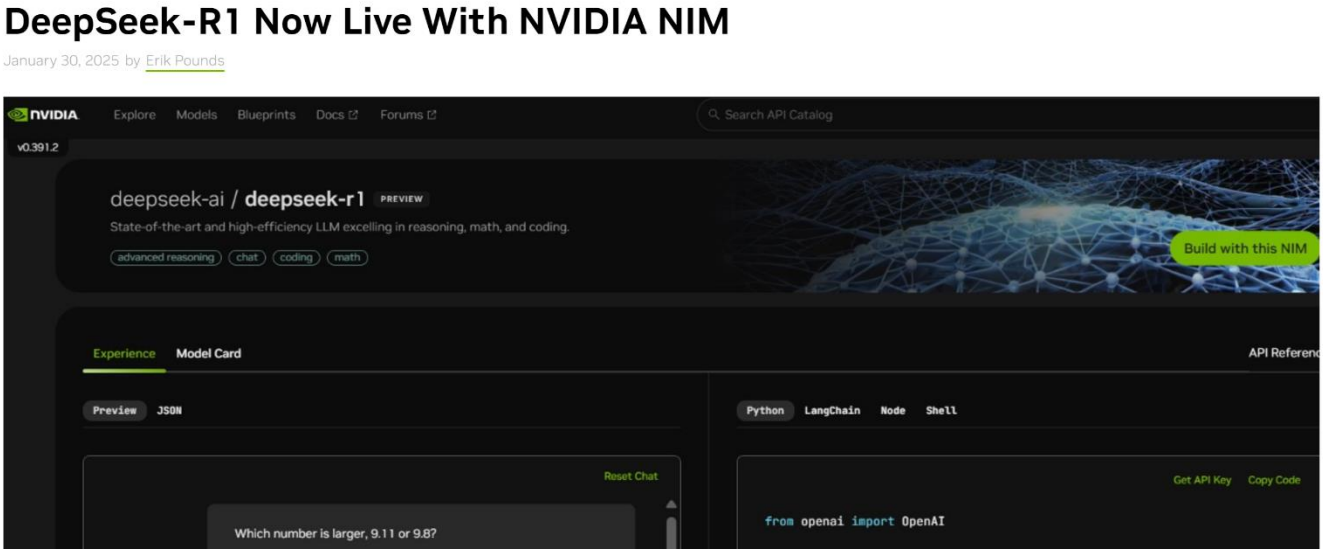
图表8 港科大团队对 DeepSeek-R1 技术理念的复刻成果表现亮眼

	AIME 2024	MATH 500	AMC	Minerva Math	Olympia dBench	Avg.
Qwen2.5-Math-7B-Base	16.7	52.4	52.5	12.9	16.4	30.2
Qwen2.5-Math-7B-Base + 8K MATH SFT	3.3	54.6	22.5	32.7	19.6	26.5
Qwen-2.5-Math-7B-Instruct	13.3	79.8	50.6	34.6	40.7	43.8
Llama-3.1-70B-Instruct	16.7	64.6	30.1	35.3	31.9	35.7
rStar-Math-7B	26.7	78.4	47.5	-	47.1	-
Eurus-2-7B-PRIME	26.7	79.2	57.8	38.6	42.1	48.9
Qwen2.5-7B-SimpleRL-Zero	33.3	77.2	62.5	33.5	37.6	48.8
Qwen2.5-7B-SimpleRL	26.7	82.4	62.5	39.7	43.3	50.9

资料来源：新智元微信公众号，平安证券研究所

英伟达、微软等国际巨头科技公司相继接入 DeepSeek 大模型。根据北京商报信息，1 月 31 日，英伟达宣布，NVIDIA NIM（一种云原生微服务技术）已经可以使用 DeepSeek-R1。微软同日称已将 DeepSeek-R1 正式纳入 Azure AI Foundry，成为该企业级 AI 服务平台的一部分。亚马逊云科技（AWS）也宣布：企业和开发者可以在 Amazon Bedrock 和 Amazon SageMaker AI 中部署 DeepSeek-R1 模型，还可以使用 AWS Trainium 等以经济高效的方式部署 DeepSeek-R1-Distill 模型。

图表9 NVIDIA NIM 已经可以使用 DeepSeek 大模型



资料来源：英伟达网站，平安证券研究所

我国腾讯云、百度云、阿里云等领先云服务平台厂商也相继部署了 DeepSeek 大模型。根据北京商报信息：1）2 月 2 日，腾讯云宣布将 DeepSeek-R1 大模型一键部署至腾讯云“HAI”上，开发者仅需 3 分钟就能接入调用。2 月 4 日，腾讯云又在腾讯云 TI 平台推出“开发者大礼包”：DeepSeek 全系模型一键部署，部分模型限免体验。2）2 月 3 日，百度智能云千帆平台正式上架 DeepSeek-R1 和 DeepSeek-V3 模型，并推出超低价格方案，用户还可享受限时免费服务。3）2 月 3 日，

阿里云宣布阿里云 PAI ModelGallery 支持云上一键部署 DeepSeek-V3、DeepSeek-R1。除了海内外巨头科技公司及云服务平台厂商，根据公司微信公众号信息，美格智能、三六零、江苏银行、万兴科技等部分 AI 应用领域相关企业也已开始了 DeepSeek 大模型的部署和应用。

图表10 部分 AI 应用领域相关企业也已开始了 DeepSeek 大模型的部署和应用

公司	应用 DeepSeek 大模型简介
美格智能	公司 1 月 26 日公众号文章，公司结合美格智能自研的 AIMO 智能体及 DeepSeek-R1 模型的基础能力，开发面向工业智能化、座舱智能体、智能无人机、机器人等领域的 AI Agent 应用。
三六零	公司 2 月 2 日公众号文章，继 360 集团创始人周鸿祎提出无偿为国产大模型 DeepSeek 提供全方位网络安全防护之后，近日，360 数字安全集团宣布其安全大模型正式接入 DeepSeek，将以 DeepSeek 为安全大模型基座，发挥 360 安全大数据优势，通过继续强化学习等技术手段，训练出“DeepSeek 版”安全大模型，让安全真正做到“自动驾驶”。
江苏银行	公司 2 月 3 日公众号文章，江苏银行主动融入数字经济发展浪潮，依托“智慧小苏”大语言模型服务平台，成功本地化部署微调 DeepSeek-VL2 多模态模型、轻量 DeepSeek-R1 推理模型，分别运用于智能合同质检和自动化估值对账场景中，通过对海量金融数据的挖掘与分析，重塑金融服务模式，实现金融语义理解准确率与业务效率双突破，为业务发展注入强劲动力。
万兴科技	公司 2 月 4 日公众号文章，万兴科技率先完成深度求索（DeepSeek）最新推理大模型 DeepSeek-R1 的深入适配，涵盖旗下视频创意、绘图创意及文档创意软件业务多款产品。

资料来源：各公司微信公众号，平安证券研究所

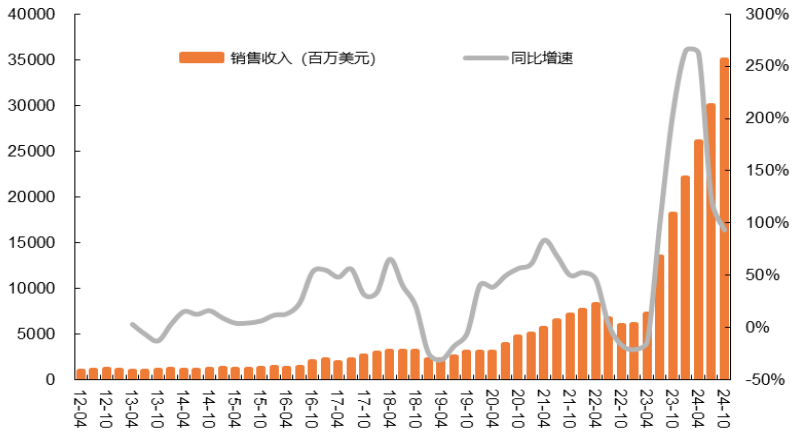
当前，海内外巨头科技公司及云服务平台厂商相继接入了 DeepSeek 大模型，AI 应用领域相关企业也已开始了 DeepSeek 大模型的部署和应用。DeepSeek 大模型获得了全球的广泛关注，认可度持续提升。我们认为，DeepSeek 大模型的开源、低成本和高性能将大幅降低大模型的获得、部署和应用成本，将加快大模型在 B 端和 C 端应用场景的落地。另外，DeepSeek 大模型的出圈将对全球大模型产业的竞争格局产生重要影响，将对海外领军大模型厂商的领先性产生冲击，并同时将对算力的未来发展产生重要影响。

三、 DeepSeek 大模型的出圈预计不改算力整体需求向上的态势，但推理和端侧算力有望增长更快

■ **DeepSeek 大模型预计不会改变 AIGC 算力整体需求向上态势**

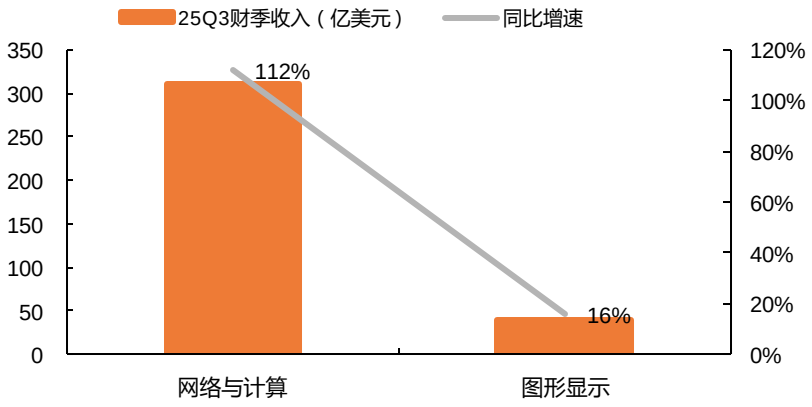
DeepSeek 大模型较低的算力消耗以及较为优秀的模型性能，将有望重塑整个 AI 算力市场。一直以来，训练端的算力争夺成为各国人工智能竞争的焦点，各国持续加大算力投入，主要 CSP 或者大模型厂商都在积极加大对算力端的投入。作为训练芯片的领军供应商，英伟达 2025 各财季收入均保持高增长状态。特朗普上台之后，更是宣布了“星际之门”计划，由 OpenAI、软银等厂商牵头，预计未来 4 年投入 5000 亿美金加强美国本土算力基础设施建设。DeepSeek 系列低成本大模型推出之后，或将改变后续整个 AI 算力投入的结构，模型训练端的需求增长可能在一定程度上受到平抑，而作为后端应用的推理环节的算力需求有望接续增长。

图表11 英伟达各财季收入及同比增速



数据来源：WIND、英伟达公司财报、平安证券研究所

图表12 英伟达 2025Q3 财季网络与计算业务收入快速增长



数据来源：WIND、英伟达公司财报、平安证券研究所

AI 作为全球智能化发展的主要抓手，其应用前景正在为各国所认可。但从发展阶段来看，仍处在早期，杀手级应用还在探索和孕育之中，一旦有了海量的应用，作为底层基础设施的算力，将获得更大的市场空间。当前，大模型已应用于端侧、教育、金融、办公、传媒、医疗、智能汽车、企业服务等多个应用场景，应用领域广阔。DeepSeek 低成本而且开源的解决方案，大幅降低了 AI 在各行各业应用的技术和成本门槛，为 AI 的产业化落地提供了更快的路径。推理和端侧的算力需求增长潜力非常大。同时，较低训练成本以及开源的 DeepSeek，有望带来更低的大模型开发和使用门槛，基于该大模型开发的主体可能更多，也一定程度上为训练算力需求提供了支撑。DeepSeek 并不是压缩了算力市场，反而为算力市场增加了更多的想象空间。

■ **DeepSeek 大模型有望带动端侧应用及推理算力需求**

如前所述，随着 DeepSeek 大模型的持续推出并开源使用，将极大推动整个应用推理端的落地，应用推理端的算力需求将迎来快速增长期。21 世纪报援引巴克莱研报指出，目前，AI 推理计算需求将快速提升，预计其将占通用人工智能总计算需求的 70%以上。

相比训练算力市场，推理算力竞争更为充分，参与者也更多。除了英伟达、AMD 等芯片设计大厂之外，AWS、谷歌、微软、阿里巴巴以及字节跳动等，各家的 ASIC 芯片在不断迭代之中，重点在优化特定负载下的工作性能；Groq、SambaNova、Positron AI 等初创企业也在积极参与，推出更为经济高效的解决方案。

大模型厂商除了参与芯片设计之外，部分厂商甚至开始参与 AI 硬件生态的建设，OpenAI CEO 奥特曼近期在接受采访时表示要开发能够替代手机的 AI 硬件，并称语音是非常关键的交互方式；DeepSeek 的低成本方案也会给这个市场增添新的动力。目前看，机器人、耳机、眼镜、手机、音箱、玩具等已经开始成为 AI 大模型搭载的载体，以后会更多，相关的主控芯片、模组和整机产品厂商，均会受益。

图表13 豆包发布的 AI 智能体耳机

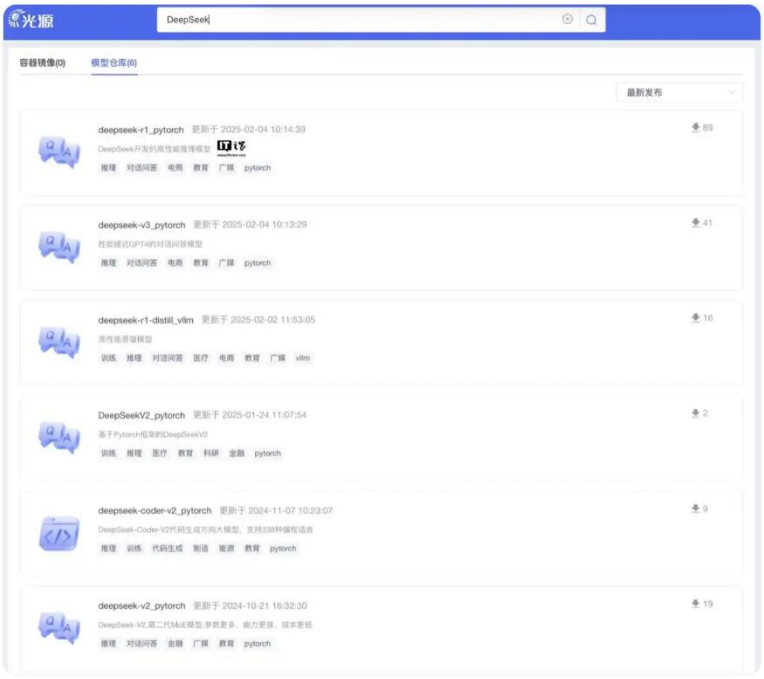
资料来源：公司官网、平安证券

■ **DeepSeek 大模型将加快国产 AI 芯片产业链的成熟**

对国内大模型厂商来讲，供应链风险、网络安全风险依然较高，能够搭建安全、可信以及可持续的算力基础设施，是其稳定运营和发展的基础。中美之间的科技博弈还在持续，而且存在进一步升级的风险，GPU 等核心元器件供应稳定性会受到挑战。同时，DeepSeek 一定程度上影响了现有人工智能的发展模式，基于国外平台的基础设施，网络安全上也存在较大隐患。

国内 AI 算力芯片发展较快，与 DeepSeek 合作的潜力凸显。在 DeepSeek-V3、DeepSeek-R1 开源并引发广泛关注后，国产 AI 算力平台厂商迅速跟进了与 DeepSeek 的合作。根据 IT 之家信息，中科曙光国家先进计算产业创新中心有限公司 2 月 3 日发文宣布，海光信息技术团队成功完成 DeepSeek V3 和 R1 模型与海光 DCU（深度计算单元）国产化适配，并正式上线。用户在“光合开发者社区”中的“光源”板块访问并下载相关模型，或登录光源官网搜索“DeepSeek”，即可基于 DCU 平台部署和使用相关模型。

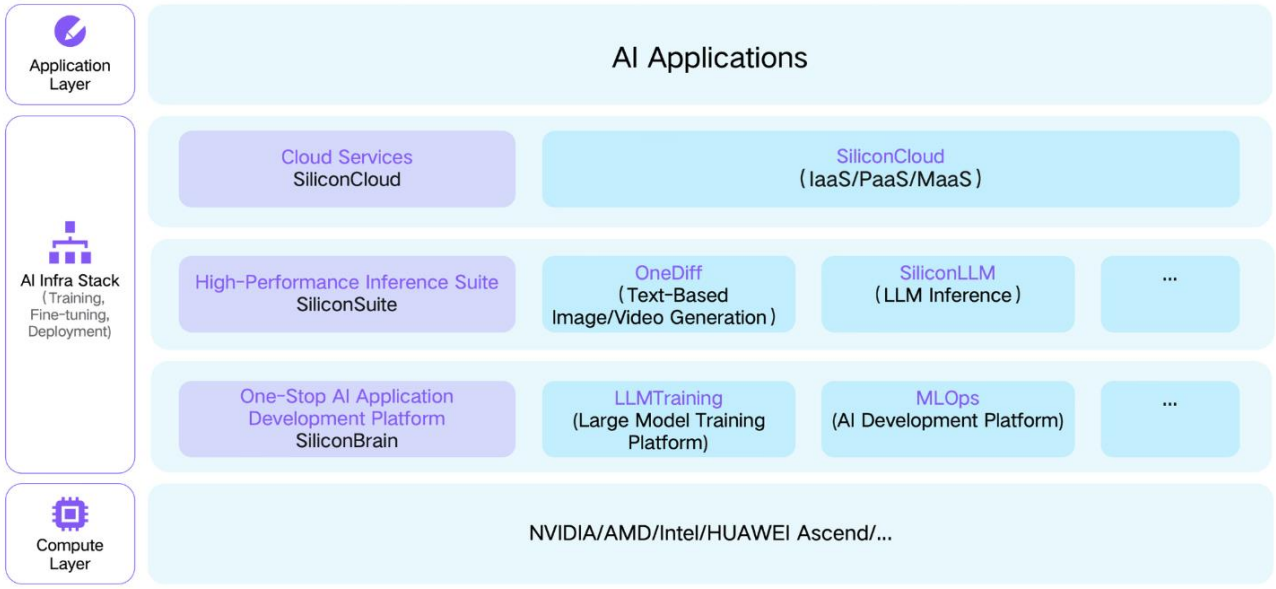
图表14 用户可在“光合开发者社区”中的“光源”板块访问并下载 DeepSeek 相关模型



资料来源：IT 之家，平安证券研究所

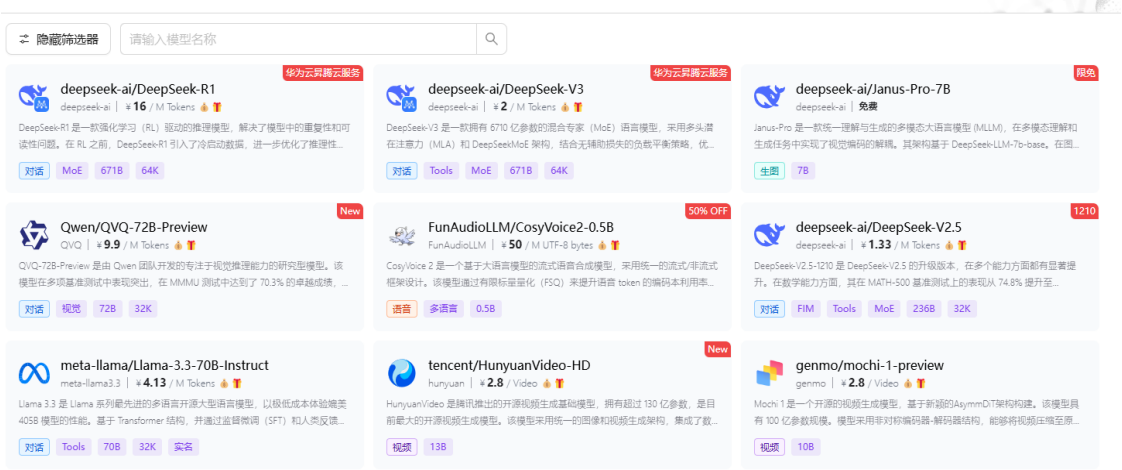
根据证券时报网信息，2 月 1 日，硅基流动和华为云双方联合首发并上线基于华为云昇腾云服务的 DeepSeek R1/V3 推理服务。得益于自研推理加速引擎加持，硅基流动和华为云昇腾云服务支持部署的 DeepSeek 模型可获得持平全球高端 GPU 部署模型的效果。2 月 2 日，云轴科技 ZStack 宣布 AI Infra 平台 ZStack 智塔全面支持企业私有化部署 DeepSeek V3/R1/ Janus Pro 三种模型，并可基于海光、昇腾、英伟达、英特尔等多种国内外 CPU/GPU 适配，将充分发挥 DeepSeek 开源模型和低成本高性能特点，助力企业级 AI 应用进一步落地。根据智通财经信息，2 月 5 日，联想集团与沐曦股份联合发布基于 DeepSeek 大模型的一体机解决方案。该方案以“联想服务器 / 工作站 + 沐曦训推一体 GPU + 自主算法”为核心架构，配合联想 AI force 智能体开发平台，推出智能体一体机与训推一体服务器双产品形态，率先实现从千亿参数大模型训练到场景化推理落地的全链条覆盖，为企业破解算力部署复杂、技术门槛高、安全可控难三大核心痛点提供了新路径。

图表15 硅基流动搭建的针对 AIGC 的计算架构



资料来源：硅基流动官网、平安证券研究所

图表16 硅基流动云平台支持的主要大模型



资料来源：硅基流动官网、平安证券研究所

我们认为，低成本高性能的 DeepSeek 大模型与国产 AI 芯片适配的逐步成熟，将能够对冲国产 AI 芯片与英伟达等全球 AI 芯片巨头的产品在训练端算力性能的差距，将加快推动国产 AI 芯片在国内大模型训练端和推理端的应用，加快国产 AI 芯片产业链的成熟，为国产 AI 芯片产业带来发展机遇，同时加快我国大模型产业的发展。

四、 投资建议

DeepSeek-V3 和 DeepSeek-R1 等 DeepSeek 系列大模型的陆续发布，表明国产大模型能力已可比肩海外领军大模型。我们认为，DeepSeek 大模型的开源、低成本和高性能将大幅降低大模型的获得、部署和应用成本，将加快大模型在 B 端和 C

端应用场景的落地。DeepSeek 大模型的出圈，短期内可能对训练算力的增长有一定的平抑效应，但预计不改算力整体需求向上的态势，而且推理和端侧算力有望增长更快。DeepSeek 大模型与国产 AI 芯片适配的逐步成熟，将加快推动国产 AI 芯片在国内大模型训练端和推理端的应用，加快国产 AI 芯片产业链的成熟，为国产 AI 芯片产业带来发展机遇，同时加快我国大模型产业的发展。我们坚定看好 AI 主题的投资机会，标的方面：1）国产算力基础设施方面，推荐浪潮信息、中科曙光、紫光股份、神州数码、海光信息、龙芯中科，建议关注寒武纪、景嘉微、软通动力、华勤技术；2）端侧算力方面，推荐恒玄科技、兆易创新，关注乐鑫科技、瑞芯微；3）算法方面，推荐科大讯飞；4）应用场景方面，强烈推荐中科创达、恒生电子、盛视科技，推荐金山办公、德赛西威、万兴科技、福昕软件，建议关注同花顺、拓尔思、彩讯股份、卫宁健康。

五、 风险提示

1）AI 算力供应链风险上升。美国对华半导体出口管制升级，将倒逼我国国产 AI 芯片产业链加快成熟。但如果我国国产 AI 芯片的迭代速度不达预期，将影响我国 AI 算力的发展，进而制约大模型的突破。

2）国产大模型算法发展可能不及预期。当前，虽然 DeepSeek 等国产大模型能力已可比肩海外领军大模型，但海外领军大模型厂商在算力储备等方面仍有优势，如果国产大模型后续算法迭代不及预期，则国产大模型厂商的追赶进度存在不达预期的风险。

3）大模型产品的应用落地低于预期。当前，我国国产大模型已经开始在教育、医疗、汽车、办公、智能硬件等 B 端和 C 端应用场景持续落地。DeepSeek 低成本而且开源的解决方案，大幅降低了 AI 在各行各业应用的技术和成本门槛，为 AI 的产业化落地提供了更快的路径。但如果大模型产品的市场拓展不及预期，我国大模型产品的应用落地将存在低于预期的风险。

平安证券研究所投资评级：

股票投资评级:

强烈推荐 （预计 6 个月内，股价表现强于市场表现 20%以上）

推　　荐 （预计 6 个月内，股价表现强于市场表现 10%至 20%之间）

中　　性 （预计 6 个月内，股价表现相对市场表现在 ± 10%之间）

回　　避 （预计 6 个月内，股价表现弱于市场表现 10%以上）

行业投资评级:

强于大市 （预计 6 个月内，行业指数表现强于市场表现 5%以上）

中　　性 （预计 6 个月内，行业指数表现相对市场表现在 ± 5%之间）

弱于大市 （预计 6 个月内，行业指数表现弱于市场表现 5%以上）

公司声明及风险提示：

负责撰写此报告的分析师（一人或多人）就本研究报告确认：本人具有中国证券业协会授予的证券投资咨询执业资格。

平安证券股份有限公司具备证券投资咨询业务资格。本公司研究报告是针对与公司签署服务协议的签约客户的专属研究产品，为该类客户进行投资决策时提供辅助和参考，双方对权利与义务均有严格约定。本公司研究报告仅提供给上述特定客户，并不面向公众发布。未经书面授权刊载或者转发的，本公司将采取维权措施追究其侵权责任。

证券市场是一个风险无时不在的市场。您在进行证券交易时存在赢利的可能，也存在亏损的风险。请您务必对此有清醒的认识，认真考虑是否进行证券交易。

市场有风险，投资需谨慎。

免责条款：

此报告旨为发给平安证券股份有限公司（以下简称“平安证券”）的特定客户及其他专业人士。未经平安证券事先书面明文批准，不得更改或以任何方式传送、复印或派发此报告的材料、内容及其复印本予任何其他人。

此报告所载资料的来源及观点的出处皆被平安证券认为可靠，但平安证券不能担保其准确性或完整性，报告中的信息或所表达观点不构成所述证券买卖的出价或询价，报告内容仅供参考。平安证券不对因使用此报告的材料而引致的损失而负上任何责任，除非法律法规有明确规定。客户并不能仅依靠此报告而取代行使独立判断。

平安证券可发出其它与本报告所载资料不一致及有不同结论的报告。本报告及该等报告反映编写分析员的不同设想、见解及分析方法。报告所载资料、意见及推测仅反映分析员于发出此报告日期当日的判断，可随时更改。此报告所指的证券价格、价值及收入可跌可升。为免生疑问，此报告所载观点并不代表平安证券的立场。

平安证券在法律许可的情况下可能参与此报告所提及的发行商的投资银行业务或投资其发行的证券。

平安证券

平安证券研究所　　电话：4008866338

深圳	上海	北京
深圳市福田区益田路 5023 号平安金融中心 B 座 25 层	上海市陆家嘴环路 1333 号平安金融大厦 26 楼	北京市丰台区金泽西路 4 号院 1 号楼丽泽平安金融中心 B 座 25 层