

传媒

报告日期：2025年02月10日

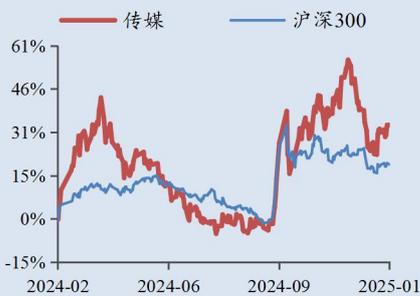
DeepSeek 重塑 AI 产业格局，助力 AI 端侧落地

——AI 产业系列跟踪专题研究报告

华龙证券研究所

投资评级：推荐（维持）

最近一年走势



分析师：孙伯文

执业证书编号：S0230523080004

邮箱：sunbw@hlzq.com

相关阅读

《智谱发布 Realtime、4V、Air 新模型，商业化应用场景持续拓宽——传媒行业周报》2025.01.20

《CES 2025 盛大开幕，“AR+AI”应用改变移动办公范式——传媒行业周报》2025.01.14

《AI+文化出海+IP 衍生，打开传媒成长新空间——传媒行业 2025 年年度投资策略报告》2024.12.06

摘要：

- **DeepSeek-V3 和 R1 模型凭借卓越的性能和创新的技术，迅速在全球 AI 领域崭露头角。**2024 年 12 月 26 日，公司发布全新系列模型 DeepSeek-V3 并同步开源，DeepSeek-V3 为自研 MoE 模型，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 不分伯仲。2025 年 1 月 20 日，DeepSeek 发布了推理模型 DeepSeek-R1，该模型在后训练阶段大规模使用了强化学习技术，在仅仅使用极少标注数据的情况下，极大提升了模型推理能力，在数据、代码、自然语言推理等任务上，性能比肩 OpenAI o1 正式版。
- **DeepSeek-V3 技术创新：架构优化与训练效率提升。**DeepSeek-V3 在继承 DeepSeek-V2 高效架构的基础上，通过创新的负载均衡策略和多标记预测（MTP）目标，进一步优化了模型性能。该模型率先采用无需辅助损失的负载均衡策略，有效降低了因负载均衡导致的性能下降。同时，MTP 目标不仅提升了模型性能，还支持推理加速中的推测性解码。在预训练阶段，DeepSeek-V3 通过 FP8 混合精度训练框架和算法、框架、硬件的协同设计，克服了跨节点 MoE 训练中的通信瓶颈，实现了计算与通信的高效重叠，显著提升了训练效率并降低了成本。最终，以仅 266.4 万 H800 GPU 小时的经济成本完成了在 14.8 万亿标记上的预训练，成为目前最强的开源基础模型。在后训练阶段，DeepSeek-V3 通过知识蒸馏技术，将长链推理（CoT）模型的推理能力融入其中，显著提升了推理性能，同时保持了输出风格和长度的可控性。
- **科技巨头纷纷接入 DeepSeek，推动全球 AI 生态发展。**2025 年 1 月 30 日，英伟达宣布 DeepSeek-R1 模型作为 NVIDIA NIM 微服务预览版上线，为企业提供了高效推理服务。紧接着，1 月 29 日，微软将 DeepSeek-R1 纳入 Azure AI Foundry 平台，支持企业通过无服务器端点访问该模型。此外，DeepSeek-R1 还上线了 GitHub Models，开发者可免费试用。2 月 1 日，硅基流动与华为云联合发布了基于昇腾云服务的 DeepSeek R1/V3 推理服务，凭借自研推理加速引擎，该服务在性能上与全球高端 GPU 部署模型相当，同时提供稳定、生产级的服务能力，满足大规模商用需求，并依托华为云昇腾云服务的强大算力支持。
- **投资建议：**DeepSeek-V3/R1 的推出代表中国在全球 AI 领域的技

术引领，极大地提振了全球 AI 产业的信心，为 AI 产业的长期增长提供了强大动力。借助 AI 算法，公司可以精准分析用户偏好，实现个性化内容推荐，进而提高用户参与度和留存率，增强其市场竞争力。同时，AI 能够根据企业的需求生成创意内容，如剧本创作、角色设计、故事背景等，为影视、动漫、游戏等领域提供丰富的创意灵感，推动行业内容的多元化发展。因此，维持传媒行业“推荐”评级。建议重点关注：三七互娱、姚记科技、昆仑万维、世纪天鸿、上海电影、博纳影业、光线传媒、万达电影、视觉中国、汤姆猫、实丰文化。

- **风险提示：**市场需求增长不及预期、技术研发进展不及预期、竞争加剧导致盈利空间受限风险、外部环境风险、政策支持风险。

表：重点关注公司及盈利预测

重点公司	股票	2025/02/05	EPS (元)				PE				投资评级
代码	名称	股价(元)	2023A	2024E	2025E	2026E	2023A	2024E	2025E	2026E	
002555.SZ	三七互娱	15.36	1.20	--	--	--	12.8	12.6	11.1	10.0	未评级
002605.SZ	姚记科技	28.92	1.38	1.42	1.68	1.94	21.3	19.4	16.4	14.5	未评级
300418.SZ	昆仑万维	43.51	1.05	--	--	0.22	42.6	--	207.2	111.9	未评级
300654.SZ	世纪天鸿	12.86	0.12	--	--	--	111.2	87.2	72.4	61.1	未评级
601595.SH	上海电影	25.89	0.28	--	--	--	91.4	83.5	43.9	31.8	未评级
001330.SZ	博纳影业	5.36	-0.40	--	0.31	0.36	--	--	17.3	14.5	未评级
300251.SZ	光线传媒	11.44	0.14	0.17	0.41	0.44	80.3	58.9	31.4	28.0	未评级
002739.SZ	万达电影	10.61	0.42	--	0.55	0.69	24.6	--	19.4	15.8	未评级
000681.SZ	视觉中国	25.35	0.21	--	--	--	122.0	125.4	105.9	93.7	未评级
300459.SZ	汤姆猫	6.05	-0.25	--	--	--	--	--	--	--	未评级
002862.SZ	实丰文化	35.64	-0.54	--	--	--	--	--	--	--	未评级

数据来源：iFind，华龙证券研究所；注：标的盈利预测均来自 iFind 一致预期

内容目录

1 DeepSeek-V3、R1 模型性能强劲“火出圈”	1
1.1 DeepSeek-V3 性能卓越，比肩全球顶尖模型	1
1.2 R1 模型性能对标 OpenAI o1 正式版	3
2 DeepSeek 模型技术创新	5
3 科技巨头纷纷接入 DeepSeek，推动全球 AI 生态发展	6
4 投资建议	7
5 风险提示	7

图目录

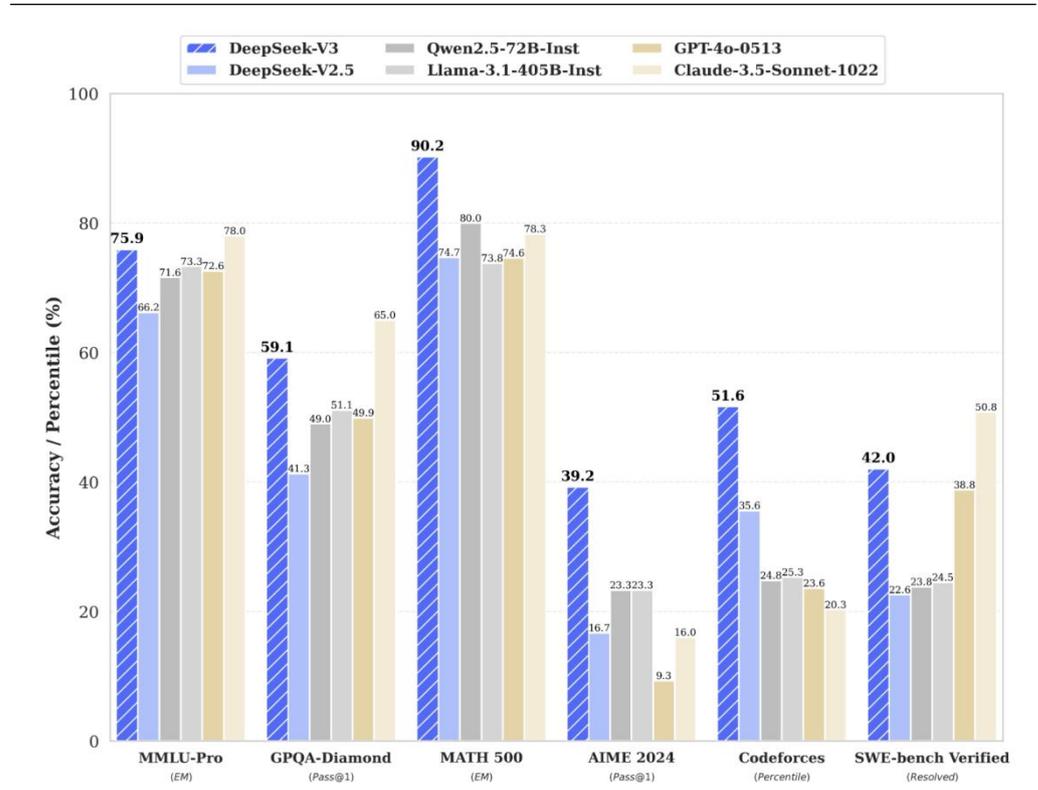
图 1: DeepSeek-V3 实现多项评测成绩领先	1
图 2: DeepSeek-V3 在四类测试集的评分	2
图 3: DeepSeek-V3 的 API 服务价格	3
图 4: DeepSeek-R1 性能比肩 OpenAI o1 正式版	4
图 5: 蒸馏小模型超越 OpenAI o1-mini	4
图 6: DeepSeek-V3 的基本架构	5
图 7: Multi-Token Prediction (MTP)示意图	6

1 DeepSeek-V3、R1 模型性能强劲“火出圈”

1.1 DeepSeek-V3 性能卓越，比肩全球顶尖模型

2024 年 12 月 26 日，公司发布全新系列模型 DeepSeek-V3 并同步开源，DeepSeek-V3 为自研 MoE 模型，671B 参数，激活 37B，在 14.8T token 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 不分伯仲。DeepSeek-V3 通过算法和工程上的创新，将生成吐字速度从 20TPS 大幅提高至 60TPS，相比 V2.5 模型实现了 3 倍的提升，带来更加流畅的用户使用体验。

图 1: DeepSeek-V3 实现多项评测成绩领先



数据来源：DeepSeek 官网，华龙证券研究所

DeepSeek-V3 的百科知识、长文本分析处理、代码、数学和中文能力都遥遥领先。分项来看，DeepSeek-V3 在知识类任务 (MMLU, MMLU-Pro, GPQA, SimpleQA) 上的水平相比前代 DeepSeek-V2.5 显著提升，接近当前表现最好的模型 Claude-3.5-Sonnet-1022。在长文本测评中，DeepSeek-V3 平均表现超越其他模型。在算法类代码场景 (Codeforces)，DeepSeek-V3 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。在美国数学竞赛 (AIME2024, 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。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。

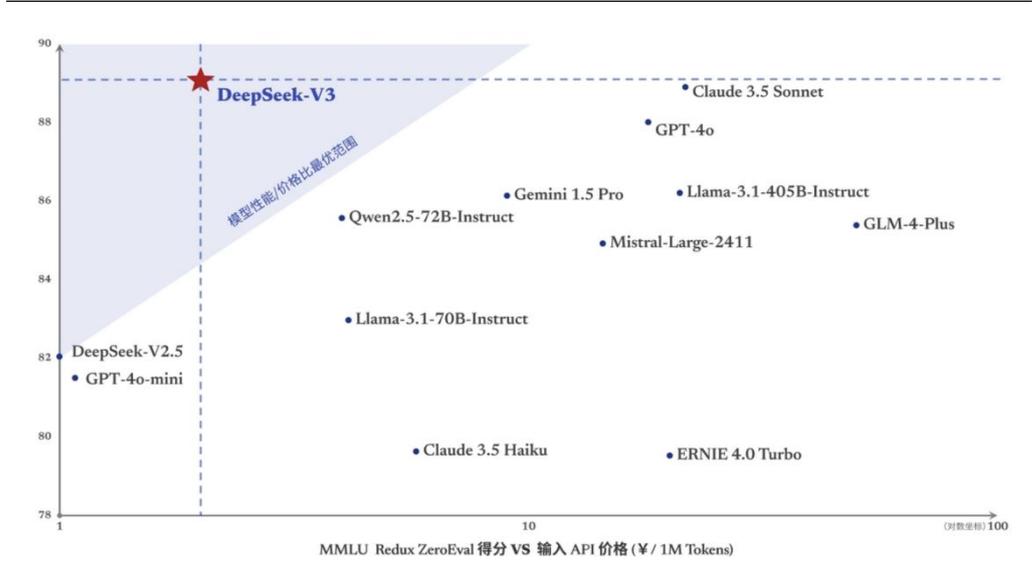
图 2：DeepSeek-V3 在四类测试集的评分

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
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

数据来源：DeepSeek 官网，华龙证券研究所

DeepSeek-V3 模型性价比极为突出。在性能上实现显著领先的同时，定价远低于市场上其他同类产品，其 API 服务定价调整为每百万输入 tokens 0.5 元（缓存命中）/ 2 元（缓存未命中），每百万输出 tokens 8 元。

图 3: DeepSeek-V3 的 API 服务价格

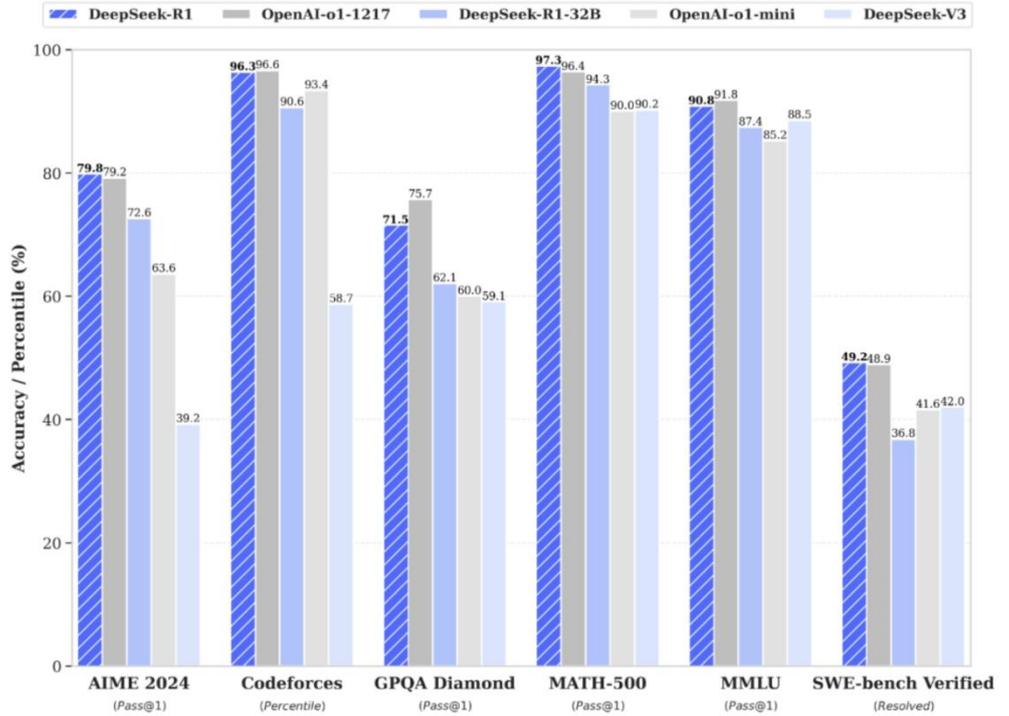


数据来源: DeepSeek 官网, 华龙证券研究所

1.2 R1 模型性能对标 OpenAI o1 正式版

2025 年 1 月 20 日, DeepSeek 发布了推理模型 DeepSeek-R1, 该模型在后训练阶段大规模使用了强化学习技术, 在仅仅使用极少标注数据的情况下, 极大提升了模型推理能力, 在数据、代码、自然语言推理等任务上, 性能比肩 OpenAI o1 正式版。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。

图 4：DeepSeek-R1 性能比肩 OpenAI o1 正式版



数据来源：DeepSeek 官网，华龙证券研究所

公司在开源 DeepSeek-R1-Zero 和 DeepSeek-R1 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 5：蒸馏小模型超越 OpenAI o1-mini

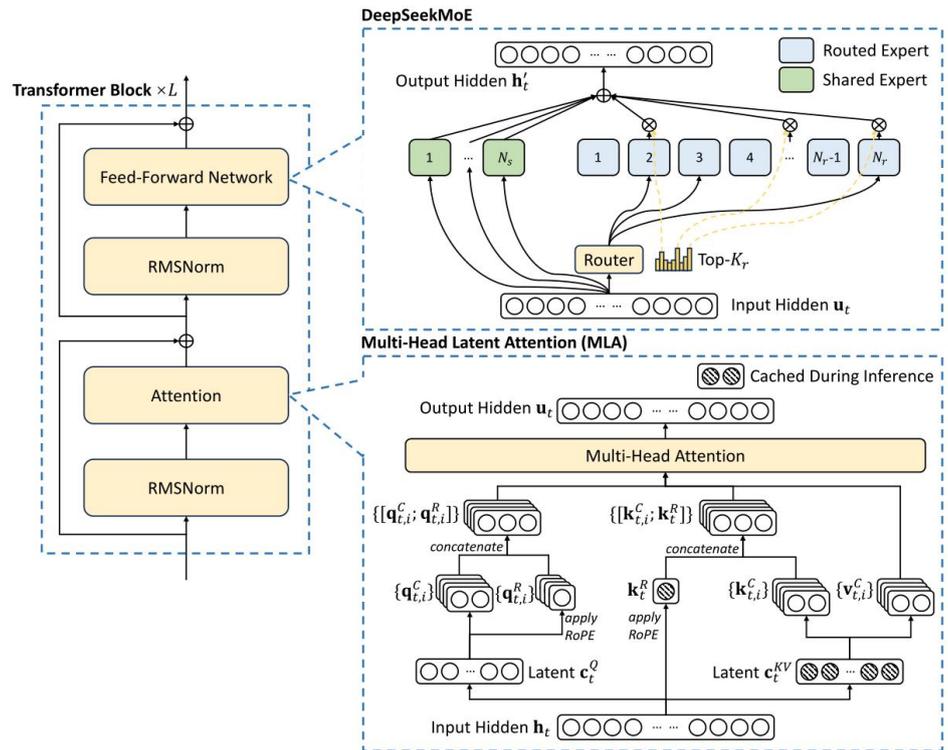
	AIME 2024 pass@1	AIME 2024 cons@64	MATH-500 pass@1	GPQA Diamond pass@1	LiveCodeBench pass@1	CodeForces rating
GPT-4o-0513	9.3	13.4	74.6	49.9	32.9	759.0
Claude-3.5-Sonnet-1022	16.0	26.7	78.3	65.0	38.9	717.0
o1-mini	63.6	80.0	90.0	60.0	53.8	1820.0
QwQ-32B	44.0	60.0	90.6	54.5	41.9	1316.0
DeepSeek-R1-Distill-Qwen-1.5B	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

数据来源：DeepSeek 官网，华龙证券研究所

2 DeepSeek 模型技术创新

DeepSeek-V3 在 DeepSeek-V2 高效架构的基础上，率先采用了一种无需辅助损失的负载均衡策略，该策略将因鼓励负载均衡而导致的性能下降降至最低。同时，公司研究了一种多标记预测(Multi-Token Prediction, MTP)目标，证明其对模型性能有益，并可用于推理加速中的推测性解码。

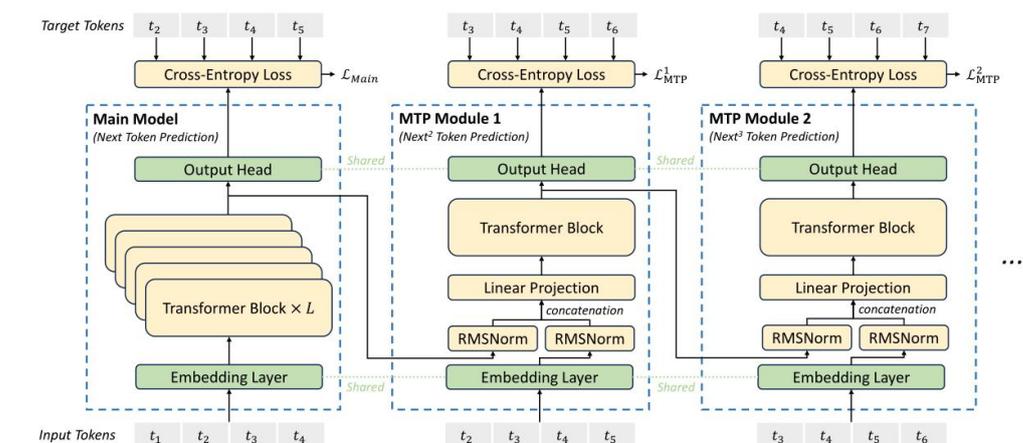
图 6: DeepSeek-V3 的基本架构



数据来源: arXiv, DeepSeek-AI, Aixin Liu, et al. : 《DeepSeek-V3 Technical Report》, 华龙证券研究所

在预训练阶段，公司追求极致的训练效率。DeepSeek 设计了一个 FP8 混合精度训练框架，并首次在超大规模模型上验证了 FP8 训练的可行性和有效性。通过算法、框架和硬件的协同设计，克服了跨节点 MoE (Mixture of Experts) 训练中的通信瓶颈，实现了近乎完全的计算与通信重叠。显著提升了训练效率，降低了训练成本。同时，公司以仅 266.4 万 H800 GPU 小时的经济成本，完成了 DeepSeek-V3 在 14.8 万亿标记上的预训练，打造了目前最强的开源基础模型。

图 7：Multi-Token Prediction (MTP)示意图



数据来源：arXiv, DeepSeek-AI, Aixin Liu, et al. : 《DeepSeek-V3 Technical Report》, 华龙证券研究所

在后训练阶段，DeepSeek 引入了一种创新方法，将长链推理 (Chain-of-Thought, CoT) 模型中的推理能力蒸馏到标准大语言模型中。该模型将 R1 的验证和反思模式融入 DeepSeek-V3，并显著提升了其推理性能。同时控制了 DeepSeek-V3 的输出风格和长度。

3 科技巨头纷纷接入 DeepSeek，推动全球 AI 生态发展

英伟达于 2025 年 1 月 30 日正式宣布，DeepSeek-R1 模型已作为 NVIDIA NIM 微服务预览版上线；2025 年 1 月 29 日，微软正式宣布将 DeepSeek-R1 模型纳入 Azure AI Foundry 平台。企业可以在 Azure AI Foundry 上通过模型目录以无服务器端点的形式访问 DeepSeek-R1。此外，DeepSeek-R1 还上线了 GitHub Models，开发者可以在 GitHub 上免费试用该模型；2025 年 2 月 1 日，硅基流动和华为云团队双方联合首发并上线基于华为云昇腾云服务的 DeepSeekR1/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。提供稳定的、生产级服务能力，让模型能够在大规模生产环境中稳定运行，并满足业务商用部署需求。华为云昇腾云服务可以提供澎湃、弹性、充足的算力。

4 投资建议

DeepSeek-V3/R1 的推出代表中国在全球 AI 领域的技术引领，极大地提振了全球 AI 产业的信心，为 AI 产业的长期增长提供了强大动力。借助 AI 算法公司可以精准分析用户偏好，实现个性化内容推荐，进而提高用户参与度和留存率，增强其市场竞争力。同时，AI 能够根据企业的需求生成创意内容，如剧本创作、角色设计、故事背景等，为影视、动漫、游戏等领域提供丰富的创意灵感，推动行业内容的多元化发展。因此，维持传媒行业“推荐”评级。建议重点关注：三七互娱、姚记科技、昆仑万维、世纪天鸿、上海电影、博纳影业、光线传媒、万达电影、视觉中国、汤姆猫、实丰文化。

5 风险提示

市场需求增长不及预期：市场需求的不足可能导致企业对 AI 技术的投资减少，进而延缓 AI 应用的普及和商业化进程。

技术研发进展不及预期：如果 DeepSeek 大模型或相关应用的研究进度未能达到预期目标，可能会对算力需求和应用落地产生负面影响。技术突破的延迟可能导致市场推广受阻，进而影响整个 AI 产业的发展节奏。

竞争加剧导致盈利空间受限风险：AI 市场竞争激烈，各厂商为争夺份额可能降低产品定价，这将压缩企业利润空间，影响盈利能力和对算力的投入。

外部环境风险：若美国持续强化对芯片出口的管控力度，国内大模型训练所需的芯片供应稳定性将受到冲击，进而对 Deepseek 等企业的技术研发与业务拓展产生不利影响。

政策支持风险：政策对于 AI 产业的发展具有关键的引导与推动作用。若相关政策的落地实施进度滞后于预期，或者支持力度未能达到行业发展的实际需求，Deepseek 等企业可能面临资源获取不足、市场拓展受限等问题，从而对企业的长期发展带来一定的阻碍。

免责及评级说明部分

分析师声明：

本人具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉尽责的职业态度，独立、客观、公正地出具本报告。不受本公司相关业务部门、证券发行人、上市公司、基金管理公司、资产管理公司等利益相关者的干涉和影响。本报告清晰准确地反映了本人的研究观点。本人在预测证券品种的走势或对投资证券的可行性提出建议时，已按要求进行相应的信息披露，在自己所知情的范围内本公司、本人以及财产上的利害关系人与所评价或推荐的证券不存在利害关系。本人不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。据此入市，风险自担。

投资评级说明：

投资建议的评级标准	类别	评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后的6-12个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅。其中：A股市场以沪深300指数为基准。	股票评级	买入	股票价格变动相对沪深300指数涨幅在10%以上
		增持	股票价格变动相对沪深300指数涨幅在5%至10%之间
		中性	股票价格变动相对沪深300指数涨跌幅在-5%至5%之间
		减持	股票价格变动相对沪深300指数跌幅在-10%至-5%之间
	行业评级	卖出	股票价格变动相对沪深300指数跌幅在-10%以上
		推荐	基本面向好，行业指数领先沪深300指数
		中性	基本面稳定，行业指数跟随沪深300指数
		回避	基本面向淡，行业指数落后沪深300指数

免责声明：

本报告的风险等级评定为R4，仅供符合华龙证券股份有限公司（以下简称“本公司”）投资者适当性管理要求的客户（C4及以上风险等级）参考使用。本公司不会因为任何机构或个人接收到报告而视其为当然客户。

本报告信息均来源于公开资料，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，本公司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，但本公司没有义务和责任及时更新本报告所涉及的内容并通知客户。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。

本报告仅为参考之用，并不构成对具体证券或金融工具在具体价位、具体时点、具体市场表现的投资建议，也不构成对所述金融产品、产品发行或管理人作出任何形式的保证。在任何情况下，本公司仅承诺以勤勉的职业态度，独立、客观地出具本报告以供投资者参考，但不就本报告中的任何内容对任何投资作出任何形式的承诺或担保。据此投资所造成的任何一切后果或损失，本公司及相关研究人员均不承担任何形式的法律责任。

在法律许可的情况下，本公司及所属关联机构可能会持有报告中提及的公司所发行证券的头寸并进行证券交易，也可能为这些公司提供或正在争取提供投资银行、财务顾问或金融产品等相关服务。本公司的资产管理部、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。本公司的员工可能担任本报告所提及的公司的董事。客户应充分考虑可能存在的利益冲突，勿将本报告作为投资决策的唯一参考依据。

版权声明：

本报告版权归华龙证券股份有限公司所有，本公司对本报告保留一切权利。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。

华龙证券研究所

北京	兰州	上海	深圳
地址：北京市东城区安定门外大街189号天鸿宝景大厦西配楼F4层 邮编：100033	地址：兰州市城关区东岗西路638号文化大厦21楼 邮编：730030 电话：0931-4635761	地址：上海市浦东新区浦东大道720号11楼 邮编：200000	地址：深圳市福田区民田路178号华融大厦辅楼2层 邮编：518046