计算机

国内模型开启世界级竞赛

开源 DeepSeek-R1 低成本对标 o1，震撼海外科技界。2025 年 1 月 20 日，DeepSeek 开源 DeepSeek-R1 模型，在数学、代码、自然语言推理等任务上，性能比肩 OpenAI o1 正式版。同时 DeepSeek 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。DeepSeek-R1 API 服务定价远低于 OpenAI o1。OpenAI CEO 1 月 24 日称将向 ChatGPT 免费用户提供 o3-min，我们认为这体现了 DeepSeek 给到 OpenAI 的竞争压力。海外微软、亚马逊、英伟达、AMD 纷纷将 DeepSeek 模型适配到自己的云服务或硬件，美国总统特朗普称 DeepSeek 给美国的科技行业敲响警钟，彰显了业界对 DeepSeek 技术实力的认可。

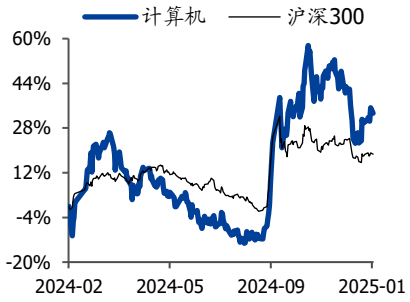
DeepSeek 技术路径解析：算法层面多维度优化。DeepSeek 团队在算法上的创新和工程上的极致优化。据 DeepSeek-R1 论文 DeepSeek-V3 技术报告，其优化包括以下方向：1）不需要监督微调，纯强化学习驱动。2）强化学习算法的创新：包括开发 GRPO 算法节省了强化学习的训练成本；没有应用结果或过程奖励模型，而采用了基于规则的奖励系统。3）多头潜在注意力机制（MLA）和专家混合架构(MOE)的结合。4)FP8 混合精度框架。5)使用 PTX 从更底层调用硬件。DeepSeek 证明了在大模型的发展进程中除了算力，软件层面的优化同样占据着举足轻重的地位，中国拥有大量高素质的算法设计、模型优化等方面的专业人才，为中国在大模型领域追赶和超越世界前沿水平提供了重要的支撑。

字节大模型进展不断，应用落地加速。1）1 月 20 日豆包实时语音大模型在豆包 APP 全量开放，在情绪理解和情感表达方面与 GPT-4O 相比优势明显。情商层面，模型在情感理解、情感承接以及情感表达等方面也取得显著进展，能较为准确地捕捉、回应人类情感信息。2）1 月 22 日，豆包全新基础模型 Doubao-1.5-pro 发布，能力全面升级，并进一步提升了多模态能力。Doubao-1.5-pro 使用 MoE 架构，仅用较小激活参数，即可比肩一流超大稠密预训练模型的性能，探索模型性能和推理性能之间的极致平衡。豆包团队还通过 RL 算法的突破和工程优化研发了深度思考模式，在 AIME 上已经超过 O1-preview，O1 等推理模型。

国产模型进步影响深远，打开广阔投资机遇。国产大模型技术的不断进步带来的变革令人期待。1.更低的成本让企业在开发 AI 应用时，能够以、更高的效率进行，有望加速国内 AI 应用从概念走向实际落地。DeepSeek 开源的蒸馏小模型超越 OpenAI o1-mini 也有望为模型加速在端侧落地。2.算力效率提高，AGI 有望来临。我们认为算力利用效率的提高一方面有望加速大模型的进步，另一方面也降低了大模型的训练和部署门槛，有望激励更多玩家入局大模型产业。微软 CEO 引用“杰文斯悖论”，表示随着 AI 的效率和可访问性越来越高，我们将看到它的使用量猛增。大模型应用对算力的需求为国产算力产业链带来了巨大的发展机遇。3.投资内容更加丰富，包括 1）互联网大厂合作生态如软件服务商；2）AI Agent 如各领域 SAAS；3）其他细分领域如目前 AI

增持（维持）

行业走势



相关研究

1、《计算机：AI Agent 由硬及软》 2025-01-26
2、《计算机：国产化需求更加凸显》 2025-01-18
3、《计算机：海外巨头竞相发布 ADAS 更新，FSD 有望三个月后表现超过人类》 2025-01-18

技术应用于军事领域，特种云建设有望加速；AI 编程提升效率，计算机行业公司深度受益。

建议关注：

AI Agent 软件：萤石网络、汉得信息、中科创达、鼎捷数智、海天瑞声、新致软件、云天励飞、焦点科技、泛微网络、致远互联、金山办公、润达医疗、星环科技、协创数据、创业黑马、恒生电子、迈富时、小商品城、金证股份、卫宁健康、创业慧康、晶泰控股、佳发教育、嘉和美康、金桥信息、新大陆等。

互联网大厂 AI 链：寒武纪、恒玄科技、天键股份、润欣科技、实丰文化、乐鑫科技、萤石网络、中芯国际、孩子王、润泽科技、欧陆通、华懋科技、浪潮信息、中兴通讯、中科曙光、兆易创新、国光电器、法本信息、新致软件、亚康股份、申菱环境、兆龙互连等。

军工 AI：能科科技、品高股份、海格通信、振芯科技、道通科技。

风险提示：AI 技术迭代不及预期风险；经济下行超预期风险；行业竞争加剧风险。

内容目录

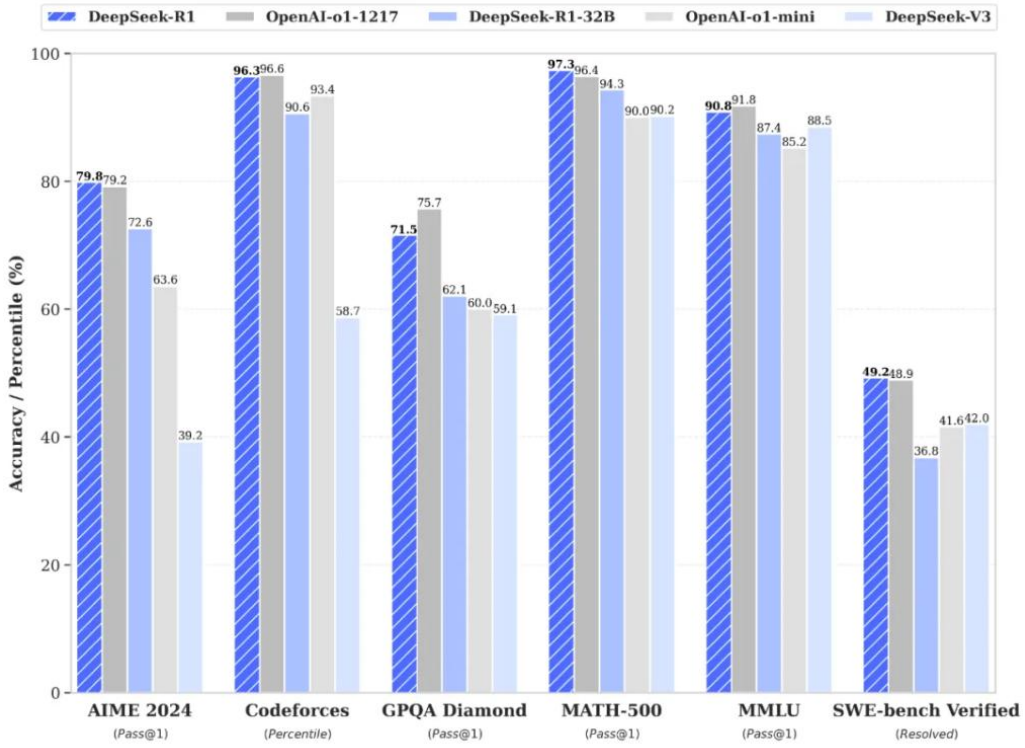
图表目录

开源 DeepSeek-R1 低成本对标 o1，震撼海外科技界

2025 年 1 月 20 日，DeepSeek 正式发布 DeepSeek-R1，并同步开源模型权重。DeepSeek-R1 遵循 MIT License，允许用户通过蒸馏技术借助 R1 训练其他模型。同时上线 API，对用户开放思维链输出。

DeepSeek-R1 性能对齐 OpenAI-o1 正式版。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1：DeepSeek-R1 在多项评测基准上得分



资料来源：DeepSeek，国盛证券研究所

开源蒸馏小模型超越 OpenAI o1-mini。

DeepSeek 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。

图表2：DeepSeek 蒸馏得到的小模型在多项能力得分优秀

	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，国盛证券研究所

价格远低于 OpenAI o1。
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 tokens16 元。相比之下，OpenAI 的 o1 模型价格为每百万输入 tokens55 元（缓存命中）/110 元（缓存未命中），每百万输出 tokens 438 元。

1 月 24 日，OpenAI CEO Sam Altman 发帖称将向 ChatGPT 免费用户提供 o3-mini，plus 套餐将获得大量 o3-mini 使用量。1 月 28 日，Sam Altman 再次发帖称赞 DeepSeek-R1 并表示将会有一些版本发布。我们认为这体现了 DeepSeek 给到 OpenAI 的竞争压力。

海外各大云厂商及硬件厂商迅速适配：
据财联社 1 月 25 日电，AMD 宣布已将新的 DeepSeek-V3 模型集成到 Instinct MI300X GPU 上，旨在与 SGLang 一起实现最佳性能。

英伟达官网 1 月 30 日宣布 DeepSeek-R1 模型现已作为 NVIDIA NIM 微服务预览版提供。DeepSeek-R1 NIM 微服务可以在单个 NVIDIA HGX H200 系统上每秒提供多达 3872 个 token。DeepSeek-R1 NIM 微服务通过支持行业标准 API 简化了部署。企业可以通过在其首选的加速计算基础设施上运行 NIM 微服务来最大限度地提高安全性和数据隐私。企业还可以为专门的 AI 代理创建定制的 DeepSeek-R1 NIM 微服务。

据 21 世纪经济报道，微软现在已经将 DeepSeek-R1 模型添加到其 Azure AI Foundry，开发者可以用新模型进行测试和构建基于云的应用程序和服务。同时，微软还将 R1 的精炼版本引入“Copilot+PC”，率先提供给搭载骁龙 X 芯片、英特尔酷睿 Ultra 200V 处理器的 PC 设备，然后是搭载 AMD Ryzen AI 9 的设备。

亚马逊云科技也宣布，用户可以在 Amazon Bedrock 和 Amazon SageMaker AI 两大 AI 服务平台上部署 DeepSeek-R1 模型。

据 NBC 新闻，美国总统特朗普在佛罗里达旅行时表示：“中国公司发布 DeepSeek AI 应该给我们的行业敲响警钟，我们需要集中精力进行竞争。”

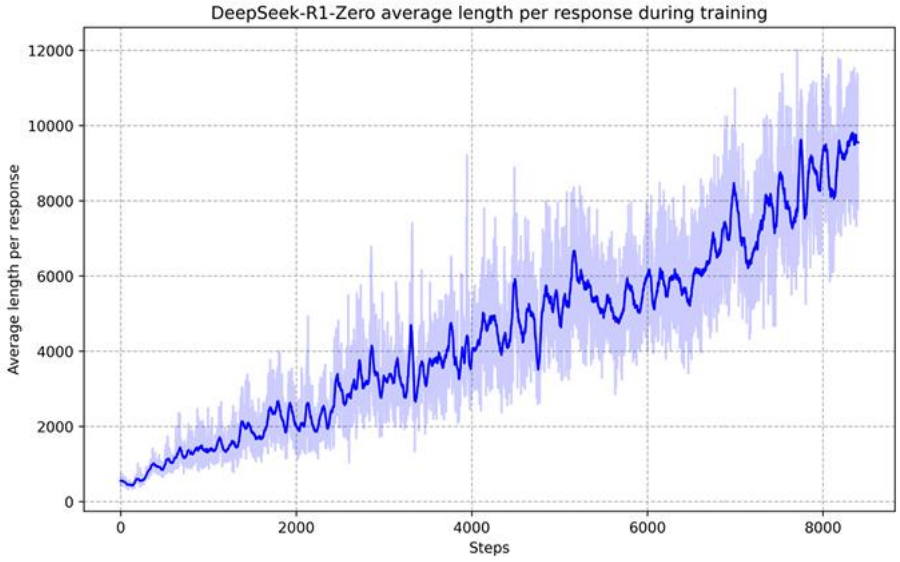
我们认为，海外科技厂商迅速适配 DeepSeek，以及特朗普的重视都充分体现了 DeepSeek 技术实力备受认可。

DeepSeek 技术路径解析：算法层面多维度优化

DeepSeek 能以极低成本实现对标 OpenAI o1 模型的能力，来源于团队在算法上的创新和工程上的极致优化。据 DeepSeek-R1 论文《DeepSeek-R1:Incentivizing Reasoning Capability in LLMs via Reinforcement Learning》和 DeepSeek-V3 技术报告，其优化包括以下方向：

不需要监督微调，纯强化学习驱动：
将强化学习用于基础模型能推理任务上获得显著提升，但以前的工作严重依赖大量的监督数据来提高模型性能。DeepSeek 证明了即使不使用监督微调（SFT）作为冷启动，直接通过大规模强化学习（RL）也可以显著提高推理能力。

图表3：RL 过程中 DeepSeek-R1-Zero 在训练集上的平均响应长度。DeepSeek-R1-Zero 自然而然地学会了用更多的思考时间来解决推理任务



资料来源：《DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning》，国盛证券研究所

强化学习算法的创新：
1）开发了 GRPO 算法(Group Relative Policy Optimization)节省了强化学习的训练成本。

2）没有应用结果或过程奖励模型，而采用了基于规则的奖励系统，主要由两种类型的奖励组成：
准确率奖励：准确率奖励模型评估响应是否正确。例如，对于具有确定性结果的数学问题，模型需要以指定格式提供最终答案，从而实现可靠的基于规则的正确性验证。对于 LeetCode 问题，可以使用编译器根据预定义的测试用例生成反馈。
格式奖励：如强制模型将其思考过程置于'<think>'和'</think>'标签之间。

多头潜在注意力机制和专家混合架构：
在架构方面，DeepSeek-V3 采用多头潜在注意力（MLA）进行高效推理，并采用 DeepSeekMoE 进行经济高效的训练。这两种架构已在 DeepSeekV2 中得到验证，证明了它们在实现高效训练和推理的同时保持稳健模型性能的能力。

图表4：DeepSeek-V3 的基本架构示意图

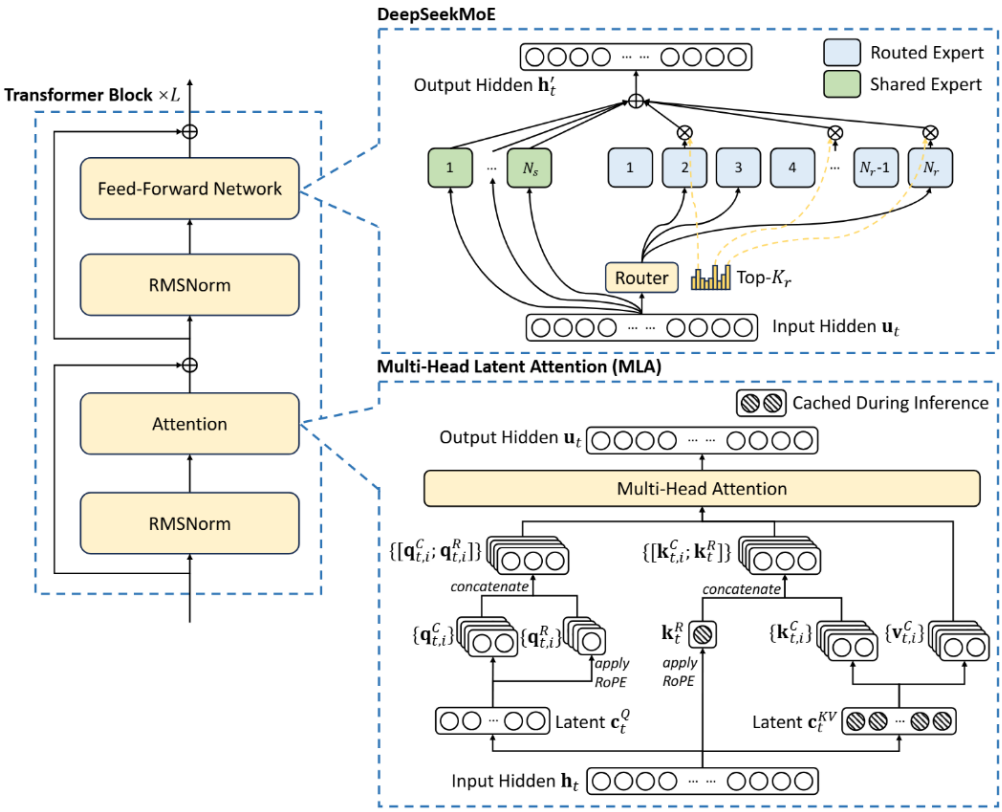


Figure 2 | Illustration of the basic architecture of DeepSeek-V3. Following DeepSeek-V2, we adopt MLA and DeepSeekMoE for efficient inference and economical training.

资料来源：DeepSeek-V3 技术报告，国盛证券研究所

混合精度框架：

DeepSeek 提出了 FP8 训练的混合精度框架。在这个框架中，大多数计算密集操作都是在 FP8 精度中进行的，而一些关键操作则以其原始数据格式进行计算，以平衡训练效率和数值稳定性。

图表5：FP8 数据格式的混合精度框架

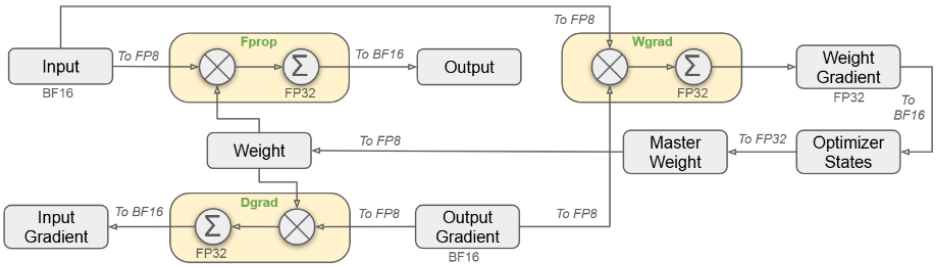


Figure 6 | The overall mixed precision framework with FP8 data format. For clarification, only the `Linear` operator is illustrated.

资料来源：DeepSeek-V3 技术报告，国盛证券研究所

使用 PTX 从更底层调用硬件：

据 Deepseek-v3 技术报告，团队采用了定制的 PTX（Parallel Thread Execution）指令并自动调整通信块大小。据英伟达官网，PTX 为通用并行编程提供了稳定的编程模型和指令集。它旨在在支持 NVIDIA Tesla 架构定义的计算功能的 NVIDIA GPU 上高效运行。CUDA 和 C/C++等语言的高级语言编译器会生成 PTX 指令，这些指令针对本机目标架构指令进行了优化并被转换为本机目标架构指令。

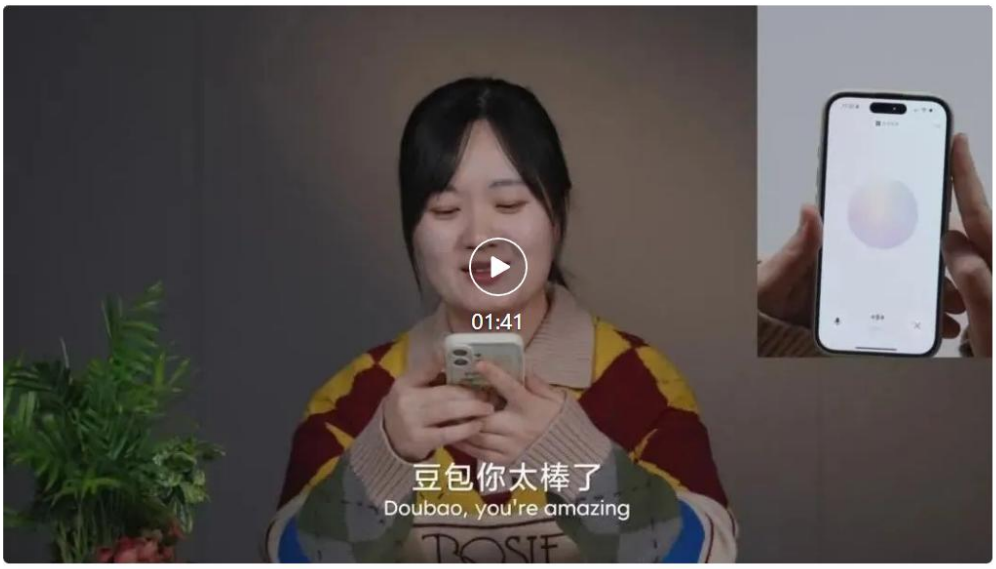
我们认为 DeepSeek 的成功是大模型发展道路上的一座里程碑，它有力地证明了在大模型的发展进程中，除了依靠算力的堆砌，软件层面的优化同样占据着举足轻重的地位，包括算法的创新、模型架构的改良、训练策略的优化等。在软件领域，中国拥有大量高素质的算法设计、模型优化等方面的专业人才，为中国在大模型领域追赶和超越世界前沿水平提供了重要的支撑。我们相信中国的大模型产业将在全球舞台上继续发挥更优秀表现，为推动人类科技进步做出更大的贡献。

字节大模型进展不断，应用落地加速

豆包实时语音大模型开放，情商智商双高

2025 年 1 月 20 日，豆包实时语音大模型正式推出，并在豆包 APP 全量开放，豆包实时语音大模型是一款语音理解和生成一体化的模型，实现了端到端语音对话。相比传统级联模式，在语音表现力、控制力、情绪承接方面表现惊艳，并具备低时延、对话中可随时打断等特性。

图表6：豆包实时语音大模型高情商回应用户呼唤，并能准确模仿经典文艺作品

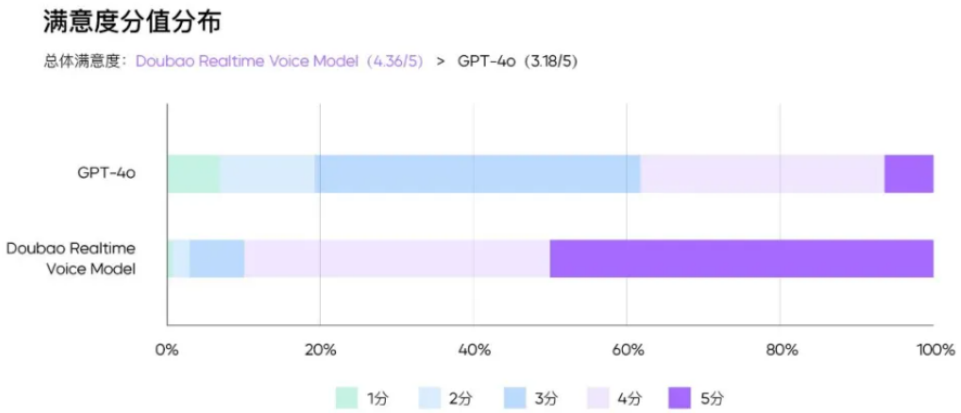


注：高情商回应用户呼唤，并能准确模仿经典文艺作品。

资料来源：豆包大模型团队，国盛证券研究所

豆包实时语音大模型模型情绪理解和表达能力突出。豆包团队围绕拟人度、有用性、情商、通话稳定性、对话流畅度等多个维度进行考评。整体满意度（以 5 分为满分）方面，豆包实时语音大模型评分为 4.36，GPT-4o 为 3.18。其中，50%的测试者对豆包实时语音大模型表现打出满分。在模型优点评测中豆包实时语音大模型在情绪理解和情感表达方面与 GPT-4O 相比优势明显。尤其是“一听就是 AI 与否”评测中，超过 30%的反馈表示 GPT-4o “过于 AI”，而豆包实时语音大模型相应比例仅为 2%以内。

图表7：豆包团队模型评测满意度分值分布



资料来源：豆包大模型团队，国盛证券研究所

Doubao-1.5-pro 基础模型更新，能力全面升级

1 月 22 日，豆包全新基础模型 Doubao-1.5-pro 正式发布，本次更新 Doubao-1.5-pro 基础模型能力全面提升，在多个公开评测基准上表现优异。

图表8：Doubao-1.5-pro 在多个基准上的测评结果

		Doubao-1.5-pro	Llama3.1-405B	GPT4o-0806	Gemini-exp-1205	Claude-3.5-Sonnet-latest	Qwen2.5	DeepseekV3
Knowledge	MMLU	88.6	88.6	**88.7**	86.8	88.5	85.6	88.5
	MMLU_PRO	**80.1**	73.3	74.9	76.4	78.0	71.1	75.9
	GPQA	**65.0**	51.1	53.1	62.1	**65.0**	49.0	59.1
MATH	Math	88.6	73.8	75.9	**89.7**	78.3	83.1	87.8
	OlympiadBench	59.8	34.1	40.7	**64.7**	43.5	50.0	59.1
Code	MBPP+	78.0	72.8	78.3	78.6	76.5	76.9	**79.3**
	McEval	**70.2**	58.7	68.2	67.0	68.2	61.7	69.4
	FullStackBench	**65.1**	53.6	61.8	62.6	60.3	56.9	63.3
Reasoning	BBH	91.6	89.2	91.7	**92.6**	**92.6**	88.3	92.3
	DROP	**93.0**	91.2	79.8	89.7	88.3	87.4	91.6
Instruction Following	IFEVal	89.5	86.0	85.7	**89.8**	89.3	84.1	86.1
	SysBench	67.6	58.9	62.2	**69.0**	**69.0**	47.2	66.3
Chinese	CMMLU	**90.9**	75.4	77.3	84.3	81.2	84.3	83.5
	C-Eval	**91.8**	72.7	76.0	83.9	80.0	84.1	86.5

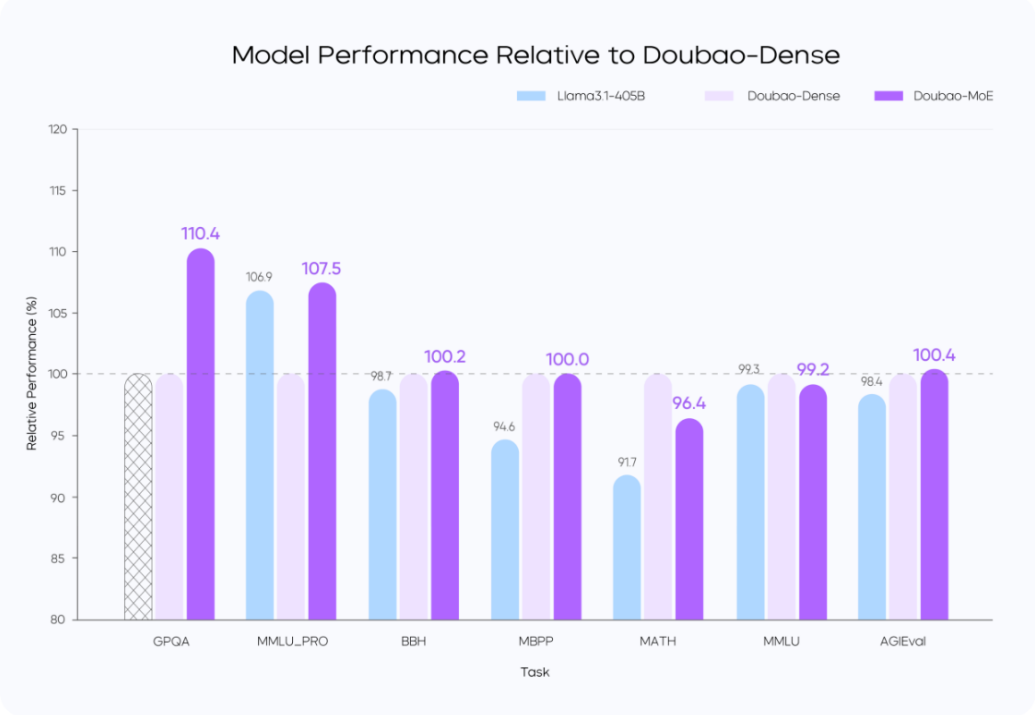
Doubao-1.5-pro 在多个基准上的测评结果

资料来源：豆包大模型团队，国盛证券研究所

模型性能与推理性能的极致平衡：Doubao-1.5-pro 使用稀疏 MoE 架构，在预训练阶段，仅用较小参数激活的 MoE 模型，性能即可超过 Llama3.1-405B 等超大稠密预训练模型。团队通过对稀疏度 Scaling Law 的研究，确定了性能和效率比较平衡的稀疏比例，并根据 MoE Scaling Law 确定了小参数量激活的模型即可达到世界一流模型的性能。

MoE 模型的性能通常可以用表现相同的稠密模型的总参数量和 MoE 模型的激活参数量的比值来确定，此前业界在这一性能杠杆上的普遍水平为不到 3 倍。豆包团队通过模型结构和训练算法优化，在完全相同的部分训练数据（9T tokens）对比验证下，用激活参数仅为稠密模型参数量 1/7 的 MoE 模型，超过了稠密模型的性能，将性能杠杆提升至 7 倍。

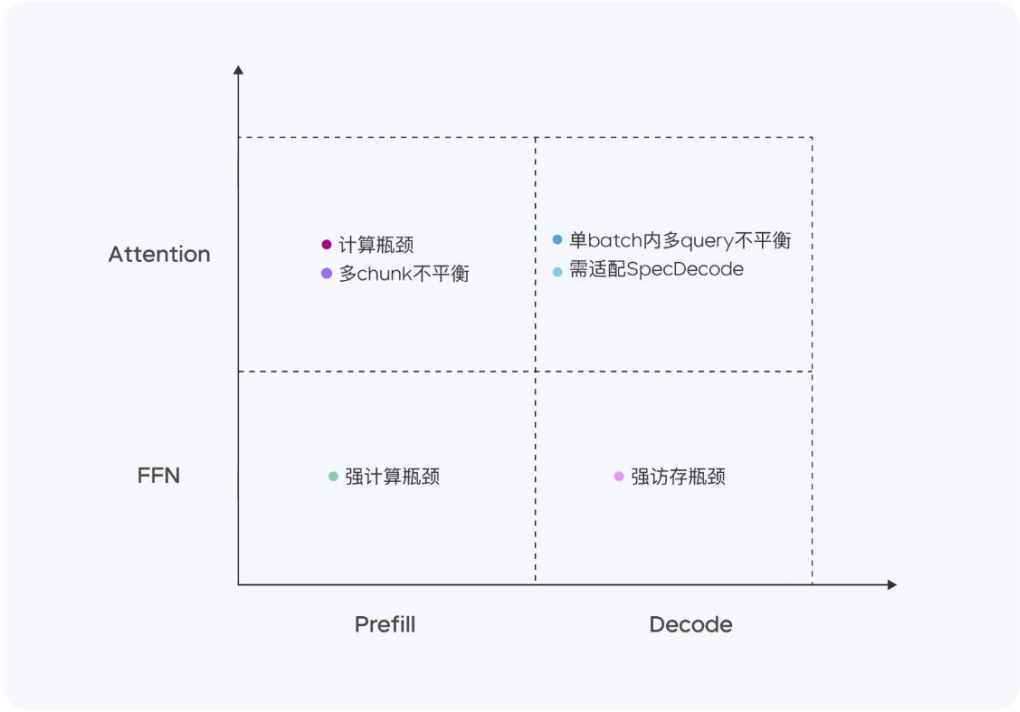
图表9：Doubao-Dense、Doubao-MoE 和 Llama3-405B 的性能对比



资料来源：豆包大模型团队，国盛证券研究所

高性能推理系统：Doubao-1.5-pro 是一个高度稀疏的 MoE 模型，在 Prefill/Decode 与 Attention/FFN 构成的四个计算象限中，表现出显著不同的计算与访存特征。针对 Prefill/Decode 与 Attention/FFN 构成的四个计算象限中，表现出显著不同的计算与访存特征四个不同象限，豆包采用异构硬件结合不同的低精度优化策略，在确保低延迟的同时大幅提升吞吐量，在降低总成本的同时兼顾 TTFT 和 TPOT 的最优化目标。

图表10：不同阶段的计算和访存特征



资料来源：豆包大模型团队，国盛证券研究所

豆包凭借自研服务器集群方案，灵活支持低成本芯片，硬件成本比行业方案大幅度降低。还通过定制化网卡和自主研发的网络协议，显著优化了小包通信的效率。在算子层面，实现了计算与通信的高效重叠，从而保证了多机分布式推理的稳定性和高效性。

多模态方面，Doubao-1.5-pro 的视觉推理能力表现优越，在各类评测基准上均取得了优异表现。

图表11：Doubao-1.5-pro 在多个视觉基准上的测评结果

	Benchmark	Doubao-1.5-pro	GPT4o-1120	Claude3.5-Sonnet	Gemini-2-flash	Qwen2-VL-72B	InternVL-2.5-78B
College-level Problems	MMMU(val)	**73.8**	70.7	70.4	70.7	64.5	70.1
	MMMU-Pro	**59.3**	54.5	54.7	57.0	46.2	48.6
Mathematical Reasoning	MathVision	**48.6**	30.4	38.3	41.3	25.9	32.2
	OlympiadBench	**48.5**	25.9	27.8	43.6	11.2	25.1
	MathVista	**78.8**	63.8	65.4	73.1	70.5	76.6
Document and Diagrams Reading	TextVQA(val)	84.7	81.4	76.5	75.6	**85.5**	83.4
	ChartQA(test avg.)	88.0	86.7	**90.8**	85.2	88.3	88.3
	InfoVQA(test)	**88.0**	80.7	74.3	77.8	84.5	84.1
	DocVQA(test)	**96.7**	91.1	95.2	92.1	96.5	95.1
	Charxiv(RQ/DQ)	54.4 / 84.3	52.0 / **86.5**	**60.2**/ 84.3	55.2/81.8	43.0 / 81.3	42.4 / 82.3
General Visual Question Answering	RealWorldQA	**78.9**	75.4	66.6	74.5	77.8	78.7
	MMStar	71.9	63.9	65.1	69.4	68.6	**73.1**
	MMBench-en	87.5	83.5	81.7	83.0	85.9	**88.3**
	MMBench-cn	86.0	82.1	83.4	82.9	83.4	**88.5**
Spatial and Counting Understanding	Blink	**68.4**	68.0	59.6	62.6	61.1	63.8
	CountBench	**89.6**	85.1	86.8	88.2	88.6	84.1
Video Understanding	Video-MME	74.1	73.4	61.7	**78.2**	71.2	72.1
	EgoSchema-subest	75.4	74.8	64.4	71.8	**80.6**	78.2

资料来源：豆包大模型团队，国盛证券研究所

Doubao 深度思考模式探索智能的边界。豆包团队致力于使用大规模 RL 的方法不断提升模型的推理能力，拓宽当前模型的智能边界。通过 RL 算法的突破和工程优化，充分发挥 test time scaling 的算力优势，完成了 RL scaling，研发了 Doubao 深度思考模式

阶段性进展 Doubao-1.5-pro-AS1-Preview 在 AIME 上已经超过 O1-preview，O1 等推理模型。并且随着 RL 的持续，模型能力还在不断提升中。

国产模型进步影响深远，投资领域更加丰富

我们认为 DeepSeek 和豆包等国产大模型技术的不断进步，带来的变革令人期待。在国内 AI 应用落地、算力产业发展以及投资领域都产生了深远而广泛的影响。

1.国内 AI 应用落地加速

豆包和 DeepSeek 等国内大模型的进步，促使企业在开发 AI 应用时，能够以更低的成本、更高的效率进行，有望加速国内 AI 应用从概念走向实际落地。据 AI 产品榜，2024 年 12 月豆包 MAU 为 7116 万，月增速达 18.64%。字节旗下虚拟角色 APP 猫箱 MAU 为 688 万，月增速达 50.18%。我们认为豆包实时语音大模型的推出有望进一步改善字节旗下应用体验，加速用户增长。

我们认为 DeepSeek 开源的蒸馏小模型超越 OpenAI o1-mini 也为模型加速在端侧落地做出极大贡献。DeepSeek 开源的 32B 和 70B 模型在多项能力上实现了对标 OpenAI o1-mini 的效果，有望部署在 AI 手机、AI 眼镜、AI 玩具等各类智能终端。

2.算力效率提高，长期利好相关产业链

DeepSeek 采用了创新的算法和技术，显著提高了算力的利用效率。我们认为算力利用效率的提高一方面有望加速大模型的进步，另一方面也降低了大模型的训练和部署门槛，有望激励更多玩家入局大模型产业。

1 月 27 日微软 CEO 在 X 上发帖引用“杰文斯悖论”，表示随着 AI 的效率和可访问性越来越高，我们将看到它的使用量猛增，将其变成我们无法满足的商品。我们认为随着大模型应用的落地速度加快，对算力的需求也将日益增长，也为国产算力产业链带来了巨大的发展机遇。

图表12：微软 CEO 引用“杰文斯悖论”

资料来源：X，国盛证券研究所

3.投资内容更加丰富

（一）互联网大厂合作生态

以字节为首的互联网大厂为推广其大模型，需要积极与众多企业展开合作，推动生态的繁荣发展。例如目前以互联网大厂和独角兽创业公司为主的大模型厂商不一定在各领域具备深耕的行业 know-how，另外大模型厂商的研发人员通常在软件行业内属于较高端人才，人力成本相对较高，所以大模型厂商将精力聚焦于提升基座模型框架能力，由其他软件服务商去对接具体的行业客户做定制化开发是性价比更高的选择。因此我们认为在垂类深耕的公司以及具备长期软件服务经验的公司有望把握机遇，在对接大模型和具体行业的过程中深度受益。

（二）AI Agent

AI Agent 作为人工智能发展的重要方向，DeepSeek 和豆包大模型的进展为其提供了更强大的技术支持。AI Agent 落地的典型场景在 C 端可赋能各种硬件智能终端，如 AI 眼镜、AI 玩具、智能家居等，B 端则可推动 SaaS 平台从简单的业务管理工具转变为驱动智能化业务的引擎。英伟达 CEO 近期表示 SaaS 企业正坐拥金矿，将诞生数百万 AI 智能体推动企业在特定任务上实现更高效的智能化管理。

字节旗下的扣子作为新一代 AI 应用开发平台，无论是否有编程基础，都可以在扣子上快速搭建基于大模型的各类 AI 应用，并将 AI 应用发布到各个社交平台、通讯软件，也可以通过 API 或 SDK 将 AI 应用集成到业务系统中。借助扣子提供的可视化设计与编排工具，可以通过零代码或低代码的方式，快速搭建出基于大模型的各类 AI 项目，满足个性化需求、实现商业价值。

图表13：扣子界面

资料来源：火山引擎官网，国盛证券研究所

（三）细分领域

目前 AI 技术大量应用于军事领域，下游终端目前主要表现为无人机、机器狗等硬件设备，但 AI 技术大幅应用的前提是上游大量云基础设施建设以及中游大数据软件分析，我们认为我国特种云建设有望加速。

图表14：AI 应用于军事依赖上游云建设和大数据分析帮助

资料来源：百度图片，国盛证券研究所

AI 编程提升效率，计算机行业公司深度受益。在 AI 编码方面，大模型能够帮助程序员快速生成代码、查找代码错误、进行代码优化等，提高软件开发效率。当前大模型编程能力进步迅速，AI 代码工具不断涌现。集成大模型的 IDE Cursor 到 2024 年 8 月已有超过 40000 名企业客户；IDE 插件 GitHub Copilot 自全面推出以来有超过 77000 个组织采用；国内也有阿里通义灵码 AI 程序员等应用。我们认为编程问题有比较准确、快迭代的评判标准，同时 Github 等社区拥有海量高质量数据，是大模型能较快取得进步的方向。AI 编程提效显著，计算机行业有望高度受益。众多科技企业已广泛应用 AI 编程，例如 AI 生成了 Google 超过 25%的代码。据埃森哲报告分析，软件平台的工作在生成式 AI 的影响范围内的工作时间占比达到 68%，我们认为计算机行业是可深度受益于 AI 发展进行提效的行业。AI 编程提升人效，有望在今年下半年起为业内公司的业绩增长带来动力，成为行业成长的超级逻辑。

建议关注

AI Agent 软件：萤石网络、汉得信息、中科创达、鼎捷数智、海天瑞声、新致软件、云天励飞、焦点科技、泛微网络、致远互联、金山办公、润达医疗、星环科技、协创数据、创业黑马、恒生电子、迈富时、小商品城、金证股份、卫宁健康、创业慧康、晶泰控股、佳发教育、嘉和美康、金桥信息、新大陆等。

互联网大厂 AI 链：寒武纪、恒玄科技、天键股份、润欣科技、实丰文化、乐鑫科技、萤石网络、中芯国际、孩子王、润泽科技、欧陆通、华懋科技、浪潮信息、中兴通讯、中科曙光、兆易创新、国光电器、法本信息、新致软件、亚康股份、申菱环境、兆龙互连等。

军工 AI：能科科技、品高股份、海格通信、振芯科技、道通科技。

风险提示

AI 技术迭代不及预期风险：若 AI 技术迭代不及预期，则对产业链相关公司会造成一定不利影响。

经济下行超预期风险：若宏观经济景气度下行，固定资产投资额放缓，影响企业再投资意愿，从而影响消费者消费意愿和产业链生产意愿，对整个行业将会造成不利影响。

行业竞争加剧风险：若相关企业加快技术迭代和应用布局，整体行业竞争程度加剧，将会对目前行业内企业的增长产生威胁。