证券研究报告/行业研究

报告日期：2025 年 02 月 14 日

"低成本、高性能、强推理"三位一体，DeepSeek 驱动高质量模型平价化

——DeepSeek 专题研究

投资要点

- **DeepSeek模型密集更新，用户数将持续高速增长**

自 2024 年起，DeepSeek在AI领域迅速崛起并不断迭代。2024年12月底至2025年1月底，更新尤为密集，发布了参数众多且性能提升的 V3、支持思维链输出和模型训练的 R1，以及深耕图像领域的视觉和多模态模型。2024年12月底到2025年1月底，全球用户数从34.7万激增至1.19亿。与ChatGPT相比，DeepSeek仅用一年多就达到ChatGPT两年的用户规模，在国内1月跃居月均活跃用户数榜首，APP下载量也大幅增长。

- **DeepSeek具备低成本、高性能、强推理三大特点**

DeepSeek-V3通过算法创新和工程优化大幅提升模型效率，从而降低成本，提高性价比。DeepSeek-V3 训练成本仅为 557 万美元，耗时不到两个月。DeepSeek通用及推理模型成本相较于OpenAI等同类模型大幅下降。DeepSeek-R1在继承了V3的创新架构的基础上，在后训练阶段大规模使用了强化学习技术，自动选择有价值的数据进行标注和训练，减少数据标注量和计算资源浪费，并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， DeepSeek在 AIME 2024 测评中上获得 79.8% 的 pass@1 得分，略微超过 OpenAI-o1；在 MATH-500 上，获得了 97.3% 的得分，与 OpenAI-o1性能相当，并且显著优于其他模型。

- **DeepSeek驱动模型平价化，建议关注算力、AI应用和端侧的投资机会**

1）算力：随着更多用户对 DeepSeek 的使用，以及未来更多AI 应用的不断涌现，对算力的需求呈现出几何级增长趋势。AI 技术的进步，虽然模型效率提高了，但不断增长的用户和应用数量，却对算力资源提出了更高要求，消耗也随之剧增。2）B 端应用：AIAgent 正在对传统 SaaS 应用进行全面重构。与传统知识库结构化管理模式相比，AIAgent 的向量数据库具备强大的自主学习能力，能够自动理解文档内容，实现更加高效的知识管理，为企业的数字化转型提供了有力支持。C 端应用：作为生成式 AI 的重要商业化应用，AIAgent 在电商、教育、旅游、酒店以及客服等多个行业得到了广泛应用。3）端侧：AI正在内容、应用、硬件、生态上影响世界，AIAgent已从“数字”走向“具身”；随着市场发展，大模型更广泛地接入硬件产品，做好软硬件协同发展是未来竞争的关键。

- **投资建议**

1）建议关注以国产算力和AI推理需求为核心的算力环节，尤其是IDC、服务器、国产芯片等算力配套产业，推荐海光信息、浪潮信息。2）DeepSeek迅速集成进各云厂商的平台中，直接拉高模型能力下限，AI应用开发提速升级。建议关注：B端：鼎捷数智、用友网络；C端：金山办公。3）小模型能力提升促进了端侧模型部署，我们看好AI终端作为新一代计算平台爆发可能。建议关注：科大讯飞、立讯精密、歌尔股份。

- **风险提示**

AI产业商业化落地不及预期的风险、市场竞争加剧风险、政策不确定性风险。

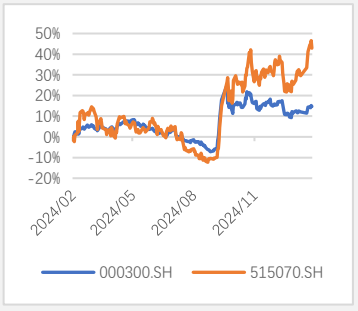
投资评级： 看好

分析师：吴起涤

执业登记编号：A0190523020001

wuqidi@yd.com.cn

人工智能 AI ETF 与沪深 300 指数走势对比



资料来源：同花顺 iFinD，源达信息证券研究所

目录

图表目录

一、"低成本、高性能、强推理"三位一体，DeepSeek 模型持续迭代升级

1.DeepSeek 模型密集更新，用户数将持续高速增长

自 2024 年起，DeepSeek 在 AI 领域迅速崛起并不断迭代。从年初发布初始版本，到后续融入数学、视觉语言技术的版本，技术实力稳步提升。2024 年 12 月底至 2025 年 1 月底，更新尤为密集，发布了参数众多且性能提升的 V3、支持思维链输出和模型训练的 R1，以及深耕图像领域的视觉和多模态模型。

表 1：DeepSeek 持续迭代升级

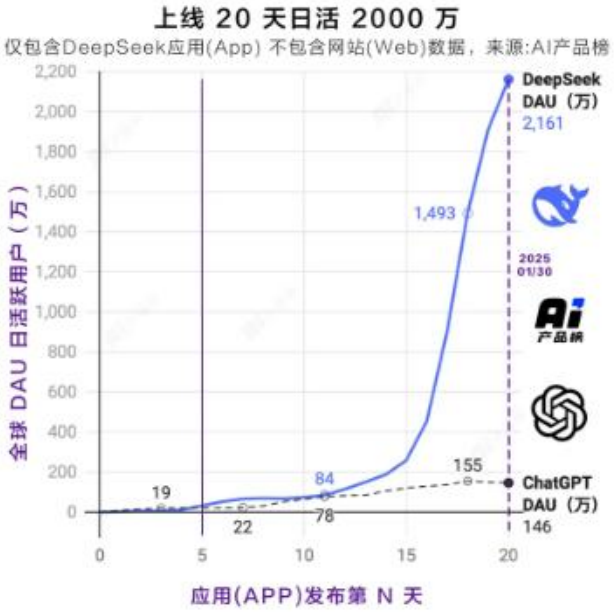
时间	模型名称	模型类型	主要特点
2024/1/5	DeepSeek LLM	通用语言模型	首次亮相的大型语言模型，标志着 DeepSeek 在自然语言处理领域的初步探索。
2024/2/5	DeepSeek-Math	数学专用模型	专注于数学问题解决能力，提升逻辑推理和复杂数学任务处理性能。
2024/3/11	DeepSeek-VL	多模态模型	引入视觉语言融合技术，支持图像与文本的联合理解，拓展多模态应用场景。
2024/5/7	DeepSeek-V2	语言生成模型	优化语言生成流畅度与准确性，显著提升文本输出的自然度和逻辑性。
2024/6/17	DeepSeek-Coder-DeepSeek-VL2	代码生成/多模态模型	强化代码生成能力与多模态交互功能，支持编程任务和跨模态内容生成。
2024/10/17	DeepSeek-Janus	多语言跨领域模型	支持多语言处理与跨领域任务，增强模型的泛化能力。
2024/12/26	DeepSeek-V3	综合性能优化模型	提升模型综合性能，优化训练策略与架构设计，为后续版本奠定基础。
2025/1/20	DeepSeek-R1	推理优化模型	采用混合专家（MoE）架构，动态路由技术使推理成本仅为 GPT-4 Turbo 的 17%，在多个基准测试中超越 OpenAI o1 模型。
2025/1/27	DeepSeek-Janus-Pro	多模态专业模型	高级版本多模态模型，优化训练策略与数据规模，击败 DALL-E 3 和 Stable Diffusion，支持文本到图像的稳定生成。

资料来源：DeepSeek 官网，源达信息证券研究所

DeepSeek 更新效果显著，用户数量爆发式增长。2024 年 12 月底到 2025 年 1 月底，全球用户数从 34.7 万激增至 1.19 亿。2 月 8 日，国内 APP 端日活 3494 万，海外 3685 万，全球 Web 端 4800 万。与 ChatGPT 相比，DeepSeek 仅用一年多就达到 ChatGPT 两年的用户规模，在国内 1 月跃居月均活跃用户数榜首，APP 下载量也大幅增长。

基于当前发展态势，DeepSeek 未来用户数还会高速增长。它技术实力强，R1 模型媲美 ChatGPT-o1；技术路径巧妙，推理成本仅为 GPT-4 Turbo 的 17%；开源与闭源双轨战略，满足不同用户需求；云服务厂商上线其大模型，芯片厂商完成适配，手机、汽车企业接入，应用场景不断拓展，有望在 AI 领域占据重要地位。

图 1：DeepSeek：全球增速最快的 AI 应用



资料来源：AI 产品榜，源达信息证券研究所

图 2：DeepSeek：增长 1 亿用户所用时间最短的 AI 应用

资料来源：AI 产品榜，源达信息证券研究所

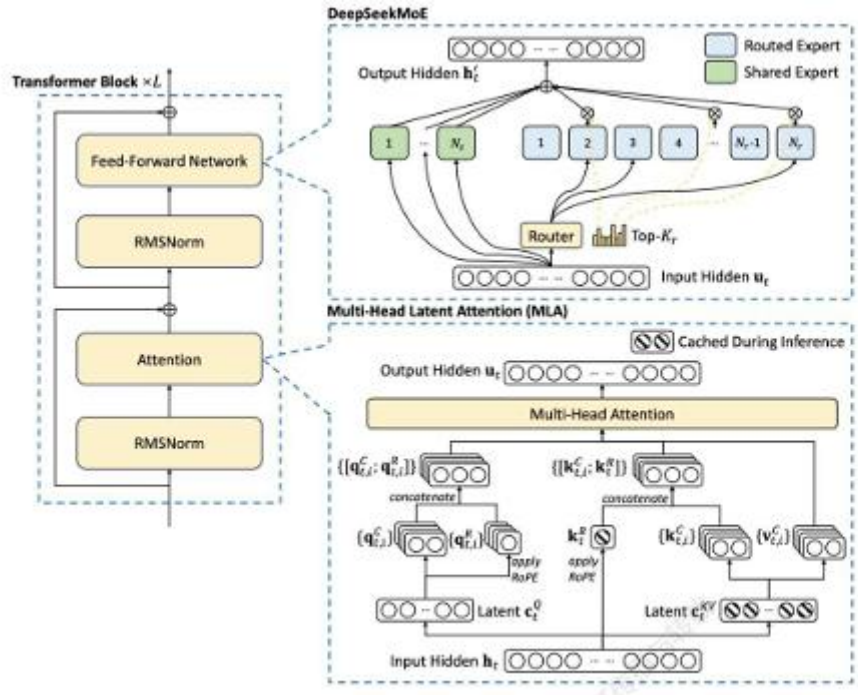
2.低成本：DeepSeek 位于模型性价比最优范围，较 OpenAI 等同类模型大幅下降

DeepSeek-V3 通过算法创新和工程优化大幅提升模型效率，从而降低成本，提高性价比。1）从算法创新层面来看，DeepSeek-V3 采用了自主研发的 MoE 架构，总参数量达 671B，每个 token 激活 37B 参数，实现多维度对标 GPT-4o。其稀疏专家模型 MoE，拓展至 256 个路由专家加 1 个共享专家，每个 token 激活 8 个路由专家、最多被发送到 4 个节点，并引入冗余专家部署策略，实现推理阶段 MoE 不同专家间的负载均衡，还提出无辅助损失的负载均衡策略，减少性能下降。此外，多头注意力机制 MLA 围绕推理阶段的显存、带宽和计算效率展开，通过创新底层软件架构，引入数学变换减少 kv cache 内存占用，缓解

transformer 推理时的显存和带宽瓶颈，优化注意力计算方式，进一步提高效率。同时，采用创新训练目标 MTP，让模型训练时一次性预测多个未来令牌，扩展预测范围，增强对上下文的理解能力，优化训练信号密度，将推理速度提升 1.8 倍。2）在工程优化方面，DeepSeek-V3 创新性地大范围落地 FP8 + 混合精度策略，计算精度从主流的 FP16 降到 FP8，保留混合精度策略，在重要算子模块保留 FP16/32 保证准确度和收敛性，兼顾模型稳定性和降低算力成本。3) 在解决通信瓶颈问题上，采用 DualPipe 高效流水线并行算法，实现接近于 0 的通信开销。

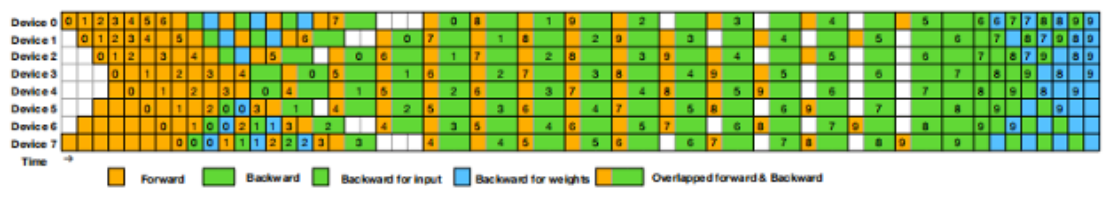
一系列的创新与优化，使得 DeepSeek-V3 训练成本仅为 557 万美元，耗时不到两个月。根据论文，DeepSeek-V3 正式训练成本仅为 557.6 万美元。在预训练阶段，每训练一万亿个标记的 DeepSeek-V3 仅需 18 万 H800 GPU 小时，即在 DeepSeek 拥有的 2048 块 H800 GPU 集群上仅需 3.7 天。加上 266.4 万 GPU 小时预训练、119 万 GPU 小时上下文长度扩展、5000 GPU 小时后期训练，得出 DeepSeek-V3 的完整训练仅需 278.8 万 GPU 小时。假设 H800GPU 的租赁价格为 GPU 小时 2 美元，总训练成本仅为 557.6 万美元。

图 3：DeepSeek-V3 模型架构



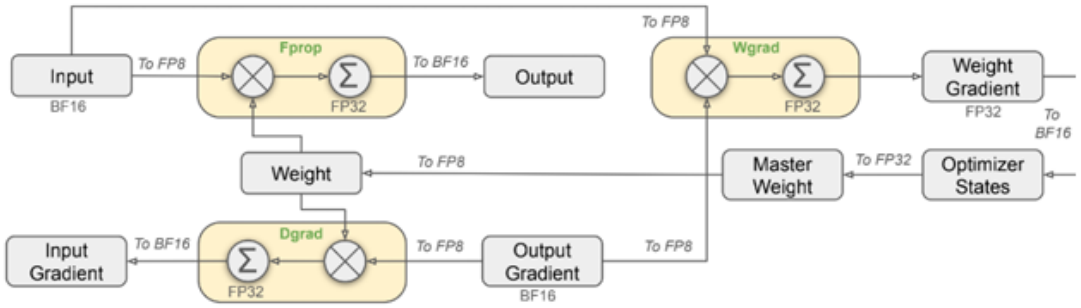
资料来源：DeepSeek 官网，源达信息证券研究所

图 4：DeepSeek-V3 DualPipe 调度策略



资料来源：DeepSeek 官网，源达信息证券研究所

图 5：DeepSeek-V3 混合精度框架



资料来源：DeepSeek 官网，源达信息证券研究所

DeepSeek 通用及推理模型成本相较于 OpenAI 等同类模型大幅下降。1）通用模型：DeepSeek-V3 模型 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。此外，V3 模型设置长达 45 天的优惠价格体验期：2025 年 2 月 8 日前，V3 的 API 服务价格仍保持每百万输入 tokens 0.1 元（缓存命中）/ 1 元（缓存未命中），每百万输出 tokens 2 元。2）推理模型：DeepSeek-R1 模型 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。

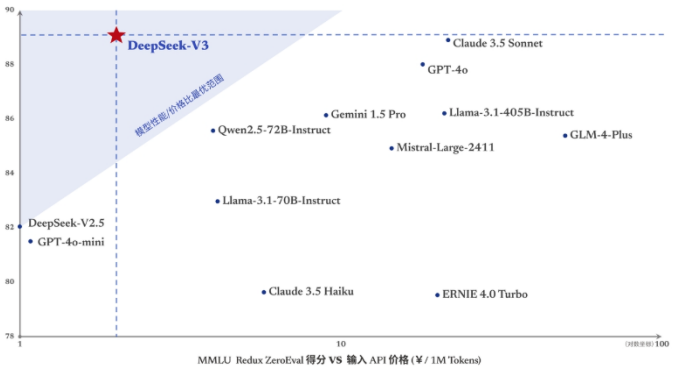
表 2：不同大模型 API 服务定价对比

大模型名称	API 服务定价
DeepSeek	DeepSeek-V3：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 DeepSeek-R1：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（2 月 9 日前优惠期减半）
ChatGPT	GPT-4 Turbo：输入每百万 tokens 约 70 元 o3-mini：价格较前代降低 63%
通义千问	文本模型： qwen2-72b-instruct：输入价格为 0.005 元 / 1,000tokens，输出价格为 0.01 元 / 1,000 tokens qwen1.5-110b-chat：输入价格为 0.007 元 / 1,000 tokens，输出价格为 0.014 元 / 1,000 tokens qwen-72b-chat：输入和输出价格均为 0.02 元 / 1,000 tokens 视觉理解模型： Qwen-VL-Plus：输入价格为 0.0015 元 / 千 tokens Qwen-VL-Max：输入价格为 0.003 元 / 千 tokens
文心一言	将于 4 月 1 日 0 时起全面免费

豆包	后付费模式：以豆包通用模型 pro-32k 为例，推理输入 0.0008 元 / 千 Tokens、推理输出 0.002 元 / 千 Tokens，模型推理的综合价格为 0.001 元 / 千 Tokens 预付费模式：以豆包通用模型 pro-32k 为例，10K TPM 的包月价格为 2000 元，平均价格为 0.0046 元 / 千 Tokens
kimi	开放平台多模态图片理解模型： moonshot-v1-8k-vision-preview：每 1M tokens 价格 12 元 moonshot-v1-32k-vision-preview：每 1M tokens 价格 24 元 moonshot-v1-128k-vision-preview：每 1M tokens 价格 60 元 上下文缓存： Cache 创建费用：24 元 / M token Cache 存储费用：5 元 / M token / 分钟 Cache 调用费用：0.02 元 / 次

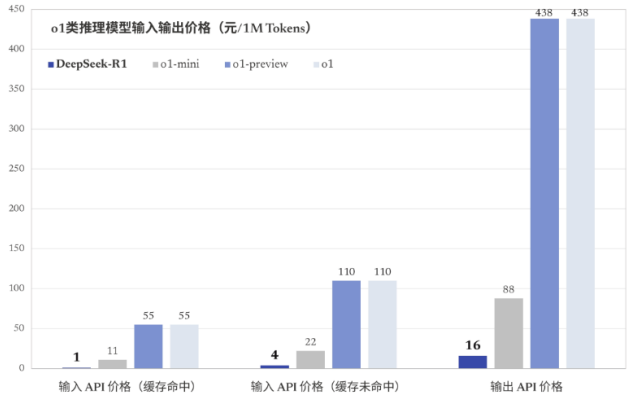
资料来源：源达信息证券研究所

图 6：DeepSeek-V3 位于模型性能/性价比最优范围



资料来源：DeepSeek 官网，源达信息证券研究所

图 7：o1 类推理模型输入输出价格

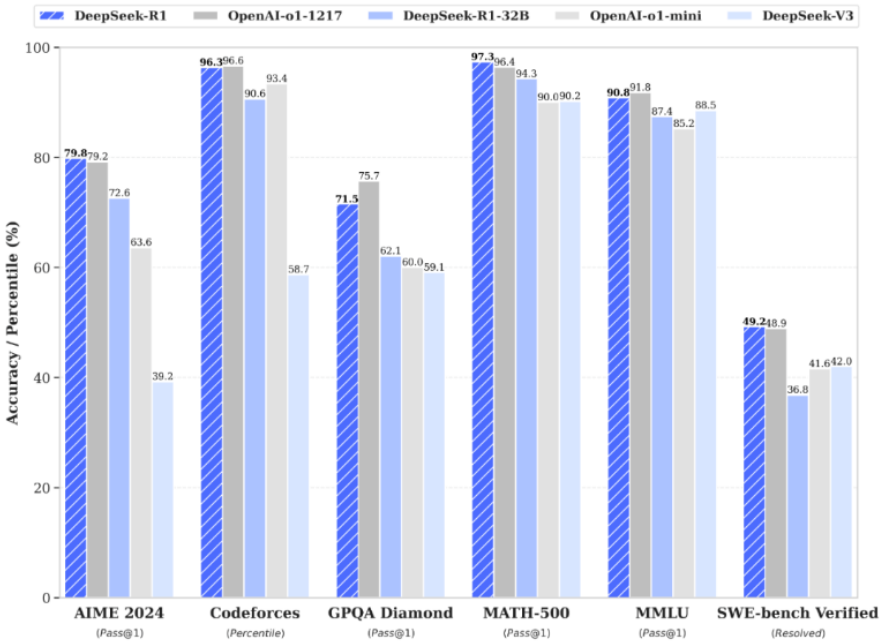


资料来源：DeepSeek 官网，源达信息证券研究所

3.高性能&强推理：Deepseek 算法能力突出，模型性能跻身世界前列

DeepSeek-R1 在继承了 V3 的创新架构的基础上，在后训练阶段大规模使用了强化学习技术，自动选择有价值的数据进行标注和训练，减少数据标注量和计算资源浪费，并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，DeepSeek 在 AIME 2024 测评中上获得 79.8% 的 pass@1 得分，略微超过 OpenAI-o1；在 MATH-500 上，获得了 97.3% 的得分，与 OpenAI-o1 性能相当，并且显著优于其他模型。

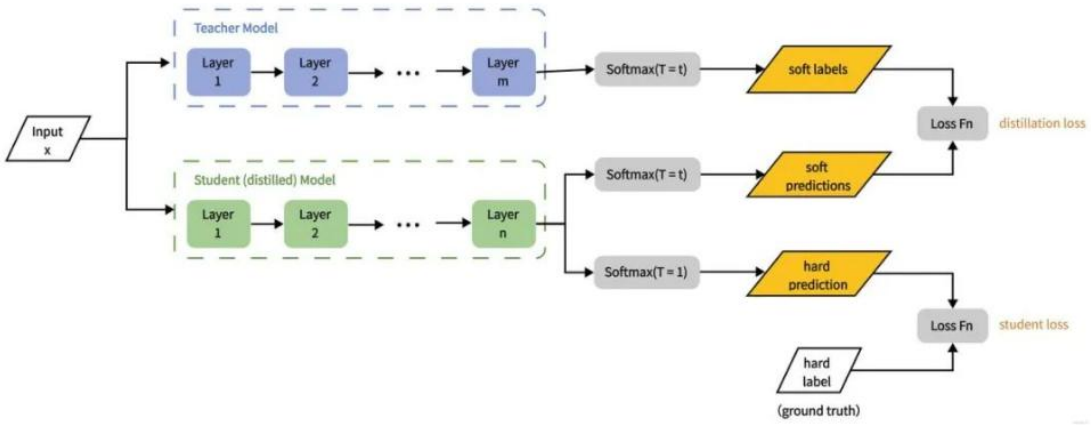
图 8：DeepSeek 模型性能优异



资料来源：DeepSeek 官网，源达信息证券研究所

DeepSeek 的蒸馏技术显著提升小模型推理能力。据 DeepSeek-V3 的技术文档，该模型使用数据蒸馏技术生成的高质量数据提升了训练效率。通过已有的高质量模型来合成少量高质量数据，作为新模型的训练数据，从而达到接近于在原始数据上训练的效果。DeepSeek 发布了从 15 亿到 700 亿参数的 R1 蒸馏版本。这些模型基于 Qwen 和 Llama 等架构，表明复杂的推理能力可以被封装在更小、更高效的模型中。蒸馏过程包括使用由完整 DeepSeek-R1 生成的合成推理数据对这些较小的模型进行微调，从而在降低计算成本的同时保持高性能。让规模更大的模型先学到高水平推理模式，再把这些成果移植给更小的模包。

图 9：蒸馏技术原理



资料来源：CSDN，源达信息证券研究所

图 10：DeepSeek 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

资料来源：DeepSeek 官网，源达信息证券研究所

二、DeepSeek 驱动模型平价化，建议关注算力、AIAgent 和端侧的投资机会

1.DeepSeek 驱动模型平价化，算力需求大幅增长

DeepSeek 的爆火使得“杰文斯悖论”这一经济学名词受到关注。“杰文斯悖论” 由经济学家威廉·斯坦利·杰文斯于 1865 年提出。当时，英国面临煤炭资源可能耗尽的担忧，人们认为提高煤炭使用效率能缓解资源短缺。但杰文斯却指出，技术进步带来的效率提升，反而会导致资源消耗的增加。例如，当煤炭动力技术效率提高，意味着能以更低成本获取更多能量，这会促使更多依赖煤炭能源的产业兴起，如工厂、火车、轮船等，进而刺激煤炭需求大幅增长，加速煤炭资源的消耗。

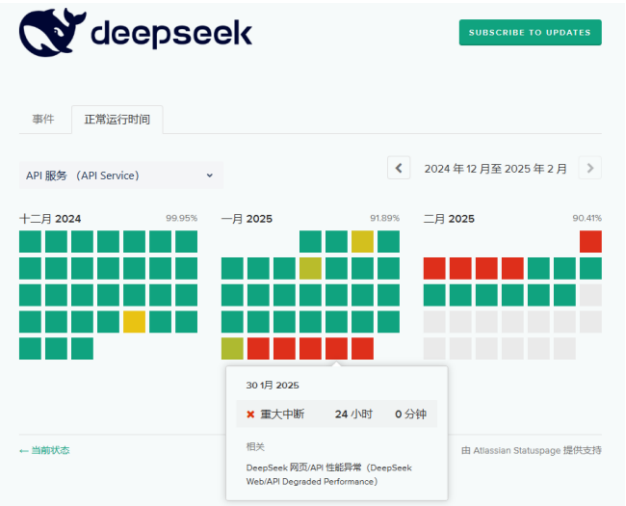
DeepSeek 爆火使得更多用户开始使用 AI 服务，如同打开了 AI 应用需求的 “潘多拉魔盒”。随着更多用户对 DeepSeek 的使用，以及未来更多 AI 应用的不断涌现，对算力的需求呈现出几何级增长趋势。AI 技术的进步，就像曾经煤炭动力技术的提升，虽然模型效率提高了，但不断增长的用户和应用数量，却对算力资源提出了更高要求，消耗也随之剧增。

DeepSeek 用户量大幅攀升，面临服务器资源不足的问题。目前网站已经暂停 API 充值，显示“当前服务器资源紧张，为避免对您造成业务影响，我们已暂停 API 服务充值。存量充值金额可继续调用，敬请谅解”。我们在使用时也发现，网页 deepseek 问答经常反馈“服务器繁忙，请稍后再试”。

图 11：DeepSeek 显著提升了算力利用效率

资料来源：甲子光年智库，源达信息证券研究所

图 12：DeepSeek 面临服务器资源不足的问题



资料来源：DeepSeek 官网，源达信息证券研究所

建议关注以国产算力和 AI 推理需求为核心的算力环节，尤其是 IDC、服务器、国产芯片等算力配套产业，推荐海光信息、浪潮信息。

2.AIAgent 市场空间广阔，B 端、C 端应用大有可为

DeepSeek 迅速集成进各云厂商的平台中，直接拉高模型能力下限，AI 应用开发提速升级。在 AI 领域蓬勃发展的当下，1 月 20 日深度求索推出的大模型 DeepSeek-R1 凭借其在数学、代码、自然语言推理等任务上比肩 OpenAI o1 模型正式版的出色性能，以及 MIT 许可协议下支持免费商用、任意修改和衍生开发的优势，迅速成为了海内外各大云厂商的宠儿。截至 2 月 5 日，华为云、腾讯云、阿里云、百度智能云等国内主流云平台，以及亚马逊 AWS、微软 Azure 等国际云巨头纷纷宣布将 DeepSeek 迅速集成进各自的平台中。比如腾讯云将 R1 大模型一键部署至高性能应用服务 HAI 上，开发者仅需 3 分钟就能接入调用，还推出“开发者大礼包”，实现 DeepSeek 全系模型一键部署；阿里云 PAI Model Gallery 支持云上一键部署 DeepSeek-V3、DeepSeek-R1。这种广泛且迅速的集成，直接拉高了模型能力下限。以往一些受限于基础模型能力不足而难以实现的复杂功能，在集成

DeepSeek 后得以轻松达成。众多企业借助这些集成了 DeepSeek 的云平台，在开发 AI 应用时，从模型搭建到功能实现的周期大幅缩短，开发效率显著提升，实现了 AI 应用开发的提速升级。

表 3：国内外云厂商接入 DeepSeek 模型

厂商名称	接入时间	相关信息
华为云	2025/2/1	联合硅基流动首发并上线基于华为云昇腾云服务的 DeepSeek R1/V3 推理服务
腾讯云	2025/2/2	支持一键部署 DeepSeek-R1 模型，开发者仅需 3 分钟即可完成模型的启动和配置
阿里云	2025/2/3	PAI Model Gallery 支持云上一键部署 DeepSeek-V3 和 DeepSeek-R1 模型
百度智能云	2025/2/3	千帆平台正式上架 DeepSeek-R1 和 DeepSeek-V3 模型，并推出超低价格方案及限时免费服务
火山引擎	2025/2/4	全面支持 DeepSeek 系列大模型，包括 V3 和 R1 等不同尺寸的模型
京东云	2025/2/4	正式上线 DeepSeek-R1 和 DeepSeek-V3 模型，支持公有云在线部署、专混私有化实例部署两种模式
天翼云	2025/2/5	在其智算产品体系中全面接入 DeepSeek-R1 模型
微软 Azure	2025/1/30	用户可以在 Azure AI Foundry 和 GitHub 上部署 DeepSeek-R1 模型
亚马逊 AWS	2025/1/30	用户可以在 Amazon Bedrock 和 Amazon SageMaker AI 中部署 DeepSeek-R1 模型
英伟达	2025/1/31	DeepSeek-R1 模型登陆 NVIDIA NIM，在单个英伟达 HGX H200 系统上，完整版 DeepSeek-R1 671B 的处理速度可达每秒 3872 Token

资料来源：源达信息证券研究所

AIAgent 市场空间广阔，B 端、C 端大有可为。根据头豹研究院，2023 年我国 AIAgent 市场规模达到了 554 亿元，预计到 2028 年，这一数字将攀升至 8520 亿元，年均复合增长率高达 72.7%。其中，垂直领域的 AIAgent 更是异军突起，迅速成为科技行业的焦点。业内人士预测，垂直领域的 AI 代理市场规模有望达到 SaaS 的十倍，甚至可能催生出超过 3000 亿美元估值的独角兽企业。

从市场应用层面来看，AIAgent 的价值主要体现在 ToC 端和 ToB 端两个方面。1）B 端场景：AIAgent 正在对传统 SaaS 应用进行全面重构。与传统知识库结构化管理模式相比，AIAgent 的向量数据库具备强大的自主学习能力，能够自动理解文档内容，实现更加高效的知识管理，为企业的数字化转型提供了有力支持。2）C 端场景：作为生成式 AI 的重要商业化应用，AIAgent 在电商、教育、旅游、酒店以及客服等多个行业得到了广泛应用。通过

智能化的交互服务，AIAgent 不仅提升了用户体验，还推动了传统行业的升级转型，为消费者带来了更加便捷、个性化的服务。

DeepSeek 迅速集成进各云厂商的平台中，直接拉高模型能力下限，AI 应用开发提速升级。建议关注：B 端：鼎捷数智、用友网络；C 端：金山办公。

3.DeepSeek 引领开源技术生态，低成本高性能利好端侧 AI 爆发

AI 正在内容、应用、硬件、生态上影响世界，AI 智能体已从“数字”走向“具身”；随着市场发展，大模型更广泛地接入硬件产品，做好软硬件协同发展是未来竞争的关键。

图 13：AI 智能体“数字化”走向“具身化”示意

资料来源：QuestMobile，源达信息证券研究所

- AIPC

AIPC 凭借高性能硬件和生产力属性有望成为端侧模型落地首站。AIPC 的关键特性在于，通过在本地运行大模型，以更具定制性、高效性和安全性的方式，满足用户个性化需求。早期，AIPC 已能实现文生文、文生图、自动报告生成以及 AI 本地知识库等功能，但鉴于当时模型能力存在一定局限，普遍采用云 + 端混合模型方案。随着 DeepSeek 模型性能的提升，其应用范围也在不断拓展。联想、华为等知名品牌厂商敏锐捕捉到这一技术优势，纷纷将 Deepseek 接入自家系统。这一举措，使得用户在使用这些品牌的设备时，能够获得更加智能、便捷的 AI 交互体验。无论是日常办公中的文档处理，还是休闲娱乐时的创意激发，用户都能感受到 AIPC 带来的高效与便利 ，享受到更加流畅自然的人机交互，真正体验到科技为生活带来的变革。

图 14：2025 年全球 AI PC 出货量能占到 PC 出货量的 35%

资料来源：Canalys，源达信息证券研究所

- **AI 眼镜**

眼镜成端侧 AI 落地绝佳载体，贴合现代生活应用场景广泛。眼镜是人类穿戴设备中最靠近嘴巴、耳朵和眼睛这三大感官的物体，技术进步使其成为端侧 AI 落地绝佳载体：AI 眼镜集成相机、眼镜、麦克风和蓝牙耳机等组件的多重功能，能直接自然地实现声音、语言、视觉的多模态输入输出，完美契合 AI 复杂功能使用条件。

图 15：AI 智能交互眼镜构成及功能

资料来源：艾瑞咨询，源达信息证券研究所

科技企业纷纷布局，涉及产品数量超 50 款。Ray-Ban Meta 的成功充分证明 AI 眼镜作为一款创新产品的可行性，科技企业们纷纷布局，刺激整个产业蓬勃发展，据 VR 陀螺，目前已有超 40 家国内外厂商入局 AI 眼镜，其中包括互联网大厂、手机巨头、AR 明星企业，涉及产品数量预计超过 50 款。

表 4：科技厂商 AI 眼睛布局

厂商名称	产品名称	发布时间	售价	卖点
华为	华为 Vision Glass	2024 年	1349 元起	轻巧舒适，等效 4 米前 120 英寸巨幕，3D 模式下可呈现超 200 英寸巨幕，支持多种 3D 片源
华为	华为智能眼镜 2	2024 年	1399 元起	搭载鸿蒙系统，支持多种智能交互功能
华为	华为 X GENTLE MONSTER Eyewear II LANG-01	2024 年	889 元起	时尚设计与智能功能结合
小米	小米 AI 眼镜	预计 2025 年 4 月	暂未公布	搭载 AI 功能、音频耳机模块、摄像头模块，全面对标 Ray-Ban Meta
Rokid	Rokid Glasses	2024 年 11 月 18 日	2499 元	搭载高通骁龙 AR1 平台，1200 万像素摄像头，支持高清摄影摄像
雷鸟创新	雷鸟 V3	2025 年 1 月 7 日	1799 元	搭载猎鹰影像系统、通义独家定制大模型、第一代骁龙®AR1 旗舰级芯片
雷鸟创新	雷鸟 X3 Pro	预计 2025 年 Q2	暂未公布	搭载萤火光引擎、RayNeo 波导，全彩 MicroLED 光引擎
Meta	Ray-Ban Meta	2023 年 9 月	299 美元起	内置定向扬声器、麦克风、摄像头等组件，可用于 FPV 拍摄 / 视频录制、通话、听音乐等
三星	三星 AI 智能眼镜	预计 2025 年 9 月	暂未公布	搭载 AR1、支持 Gemini 模型
苹果	苹果 AI 眼镜	预计 2025 年	暂未公布	据报道正在研发，可能具备 AR 功能
李未可科技	李未可 Meta Lens Chat	预计 2024 年 Q4 或 2025 年 Q1	暂未公布	与博士眼镜达成战略合作，进驻博士眼镜全国线下门店
蜂巢科技	界环 AI 音频眼镜	2024 年 9 月	600-800 元	提供 8 框 14 色，快拆结构设计，重量约 30 克，续航 11 小时
闪极科技	闪极 AI“拍拍镜”	2024 年 12 月 19 日	999 元起	国内首款实现量产的 AI 眼镜，搭载闪极自研的全球首款 AI 记忆系统 Loomo OS
XREAL	XREAL One	2024 年	3299 元起	智能 AR 眼镜，具备高清显示和智能交互功能

INMO	INMO GO 2	2024 年 11 月 29 日	3999 元起	一体式 AI+AR 智能眼镜，具备实时同声翻译、便携提词等功能
看见科技	Looktech AI Glasses	2024 年	暂未公布	具备 AI 拍照、识别等功能
星宸科技	星宸 AI 眼镜	2024 年	暂未公布	搭载星宸科技的 AI 技术，具备智能交互功能

资料来源：源达信息证券研究所

三、投资建议

1）建议关注以国产算力和 AI 推理需求为核心的算力环节，尤其是 IDC、服务器、国产芯片等算力配套产业，推荐海光信息、浪潮信息。

2）DeepSeek 迅速集成进各云厂商的平台中，直接拉高模型能力下限，AI 应用开发提速升级。建议关注：B 端：鼎捷数智、用友网络；C 端：金山办公。

3）小模型能力提升促进了端侧模型部署，我们看好 AI 终端作为新一代计算平台爆发可能。建议关注：科大讯飞、立讯精密、歌尔股份。

表 5：相关公司万得一致盈利预测

公司	代码	PB	归母净利润（亿元）			PE			总市值（亿元）
			2024E	2025E	2026E	2024E	2025E	2026E	
海光信息	688041.SH	15.9	19.5	28.7	39.0	162.3	110.2	81.1	3159
浪潮信息	000977.SZ	4.9	23.0	28.7	34.3	40.5	32.5	27.2	931
鼎捷数智	300378.SZ	5.6	1.8	2.2	2.7	66.1	53.5	43.1	118
用友网络	600588.SH	7.2	-0.4	3.2	6.2	-1553.2	194.3	100.4	625
金山办公	688111.SH	16.6	15.2	19.1	24.2	116.5	93.0	73.5	1776
科大讯飞	002230.SZ	7.7	6.0	9.6	13.3	211.0	130.9	94.7	1262
立讯精密	002475.SZ	4.9	135.9	171.8	209.1	23.2	18.3	15.1	3151
歌尔股份	002241.SZ	3.0	27.3	36.4	45.3	35.6	26.7	21.5	972

资料来源：Wind，源达信息证券研究所

四、风险提示

AI 产业商业化落地不及预期的风险。目前各环节 AI 产品的商业化模式尚处于探索阶段，如果各环节产品的推进节奏不及预期，或对相关企业业绩造成不利影响。

市场竞争加剧风险。海外 AI 厂商凭借先发优势，以及较强的技术积累，在竞争中处于优势地位，如果国内 AI 厂商技术迭代不及预期，经营状况或将受到影响；同时，目前国内已有众多企业投入 AI 产品研发，后续可能存在同质化竞争风险，进而影响相关企业的收入。

政策不确定性风险。AI 技术的发展直接受各国政策和监管影响。随着 AI 在各个领域的渗透，政府可能会进一步出台相应的监管政策以规范其发展。如果企业未能及时适应和遵守相关政策，可能面临相应处罚，甚至被迫调整业务策略。此外，政策的不确定性也可能导致企业战略规划和投资决策的错误，增加运营的不确定性。

投资评级说明

行业评级	以报告日后的 6 个月内，证券相对于沪深 300 指数的涨跌幅为标准，投资建议的评级标准为：		
	看	好：	行业指数相对于沪深 300 指数表现＋10%以上
	中	性：	行业指数相对于沪深 300 指数表现－10%~＋10%以上
	看	淡：	行业指数相对于沪深 300 指数表现－10%以下
公司评级	以报告日后的 6 个月内，行业指数相对于沪深 300 指数的涨跌幅为标准，投资建议的评级标准为：		
	买	入：	相对于沪深 300 指数表现＋20%以上
	增	持：	相对于沪深 300 指数表现＋10%~＋20%
	中	性：	相对于沪深 300 指数表现－10%~＋10%之间波动
	减	持：	相对于沪深 300 指数表现－10%以下

办公地址

石家庄

河北省石家庄市长安区跃进路 167 号源达办公楼

上海

上海市浦东新区峨山路 91 弄 100 号陆家嘴软件园 2 号楼 701 室

分析师声明

作者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，以勤勉的职业态度，独立、客观地出具本报告。分析逻辑基于作者的职业理解，本报告清晰准确地反映了作者的研究观点。作者所得报酬的任何部分不曾与，不与，也不将与本报告中的具体推荐意见或观点而有直接或间接联系，特此声明。

重要声明

河北源达信息技术股份有限公司具有证券投资咨询业务资格，经营证券业务许可证编号：911301001043661976。

本报告仅限中国大陆地区发行，仅供河北源达信息技术股份有限公司（以下简称：本公司）的客户使用。本公司不会因接收人收到本报告而视其为客户。本报告的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，也不保证所包含信息和建议不发生任何变更。本公司已力求报告内容的客观、公正，但文中的观点、结论和建议仅供参考，不包含作者对证券价格涨跌或市场走势的确定性判断。本报告中的信息或所表述的意见均不构成对任何人的投资建议，投资者应当对本报告中的信息和意见进行独立评估。

本报告仅反映本公司于发布报告当日的判断，在不同时期，本公司可以发出其他与本报告所载信息不一致及有不同结论的报告；本报告所反映研究人员的不同观点、见解及分析方法，并不代表本公司或其他附属机构的立场。同时，本公司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。

本公司及作者在自身所知情范围内，与本报告中所评价或推荐的证券不存在法律法规要求披露或采取限制、静默措施的利益冲突。