Deepseek：国产AI应用的“诺曼底时刻”

计算机行业深度

华西计算机团队
2025年2月3日

分析师：刘泽晶
SAC NO：S1120520020002
邮箱：liuzj1@hx168.com.cn

分析师：赵宇阳
SAC NO：S1120523070006
邮箱：zhaoyy1@hx168.com.cn

核心逻辑

Deepseek有望改变AI生态

我们认为DeepSeek的成功有望改变现有AI的产业格局，一方面是中国在全球AI产业的竞争形态，另一方面是大模型开源与闭源的竞争形态：1）对于训练而言，最引人注目的自然是FP8的使用。根据深度学习与NLP公众号，DeepSeek-V3是第一个（至少在开源社区内）成功使用FP8混合精度训练得到的大号MoE模型。2）与OpenAI依赖人工干预的数据训练方式不同，DeepSeek R1采用了R1-Zero路线，直接将强化学习应用于基础模型，无需依赖监督微调（SFT）和已标注数据。3）**低成本模型有望引领AI产业“新路径”：开源+MOE**。4）**开源VS闭源**：开源重构AI生态，与闭源共同繁荣下游。

堆算力的AI“老路径”遭到强力挑战

1）NV、博通等大跌意味着纯算力路径依赖被挑战：DeepSeek在没有最高端算力卡并且以极低的价格建立了一个突破性的AI模型，纯算力路径依赖得到挑战；2）国内外科技巨头持续提升capex指引，剑指NV GPU需求高景气，国产Deepseek模型爆火，高端算力/高集群能力并非唯一解；3）国产算力异军突起，充分受益国产模型deepseek崛起。据华为云2月1日消息，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。

2025：端侧AI爆发元年

1）token成本持续降低，AI agent加速元年：1月27日后，Deepseek-V3发布后英伟达股价大跌，与之相对，苹果、Meta、谷歌等应用提供商股价表现明显更好。**谷歌、OpenAI、Anthropic、字节跳动等国内外领先大模型厂商纷纷剑指智能体开发，发布Project Astra、Operator、Computer Use、UI-TARS等产品，2025年有望成为AI智能体加速元年。**2）相比云端AI，终端AI在成本、能耗、隐私等方面都具有优势。**豆包大模型的成功为字节系AI智能终端的爆发提供了有力支撑。**

- **受益标的：AI终端：**乐鑫科技、恒玄科技、润欣科技、中科蓝讯、翱捷科技、博士眼镜、亿道信息、云天励飞、天键股份、星宸科技；**AI应用：**麦迪科技、能科科技、润达医疗、开普云、新致软件、微盟集团、彩讯股份、汉得信息、拓尔思、同花顺、财富趋势、创业黑马、万兴科技；**国产算力：**中芯国际、海光信息、寒武纪、中科曙光；**算力云：**金山云、品高股份、优刻得、青云科技等。

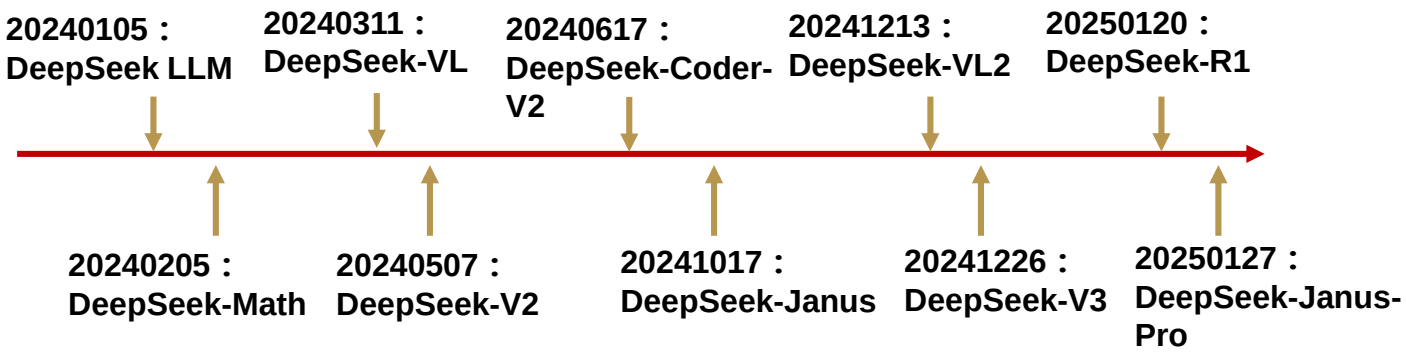
- **风险提示：市场竞争加剧；产品研发不及预期。**

01 Deepseek改变行业生态

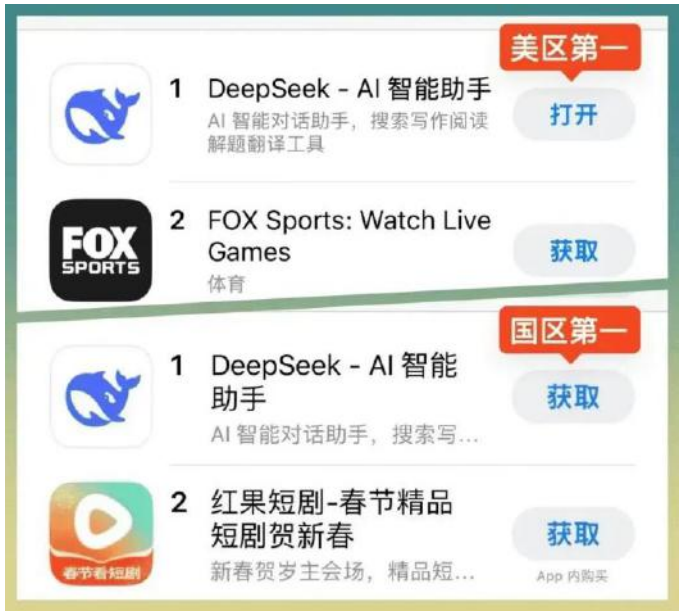
1.1. Deepseek：AI生产函数的根本性改变

- **DeepSeek是一家中国人工智能公司，成立于2023年7月17日，总部位于浙江杭州。它由量化资管巨头幻方量化创立，专注于大语言模型（LLM）及相关AI技术的研发。**
- 简单来说，DeepSeek是一款大语言模型（LLM），主打“极致性价比”。它能写代码、解数学题、做自然语言推理，性能比肩OpenAI的顶尖模型o1，但成本却低到离谱——训练费用仅557.6万美元，是GPT-4o的十分之一，API调用成本更是只有OpenAI的三十分之一。

DeepSeek开源模型时间轴



国内外应用市场下载排名

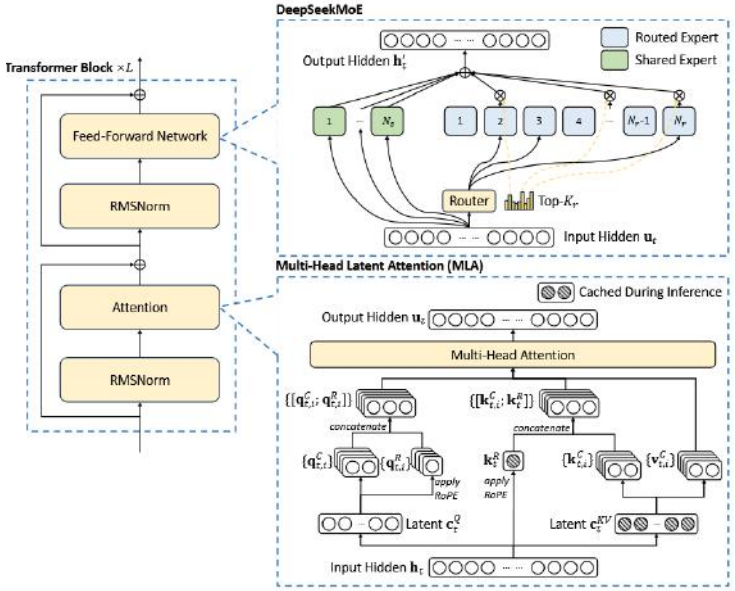


资料来源：算法工程笔记，编程奇点，华西证券研究所

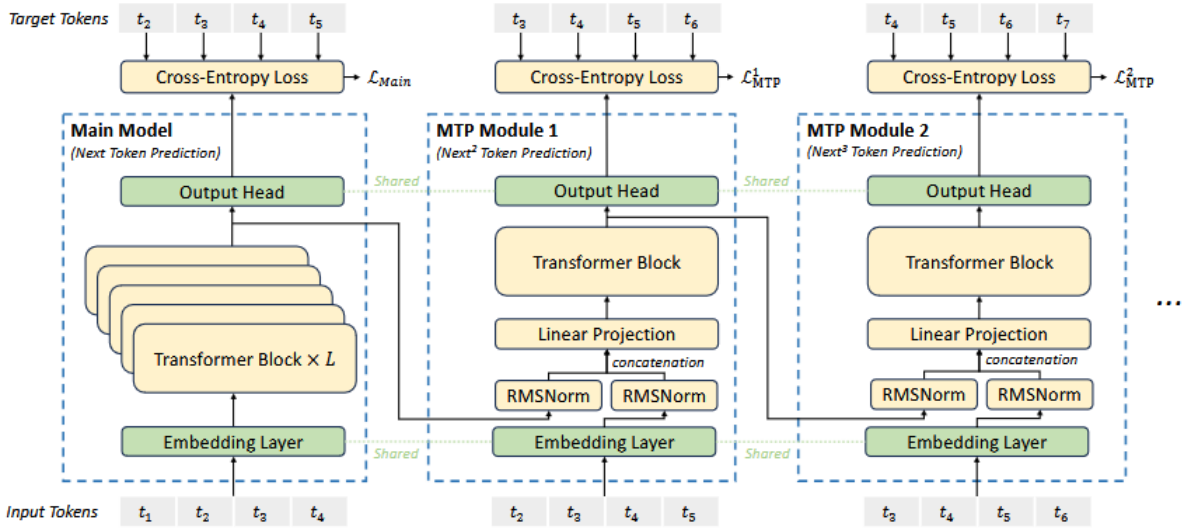
1.2.1 Deepseek：算法能力被突出（ DeepSeek-V3 ）

- **对于训练而言，最引人注目的自然是FP8的使用。根据深度学习与NLP公众号，DeepSeek-V3是第一个（至少在开源社区内）成功使用FP8混合精度训练得到的大号MoE模型。**
- 众所周知，FP8伴随着数值溢出的风险，而MoE的训练又非常不稳定，这导致实际大模型训练中BF16仍旧是主流选择。为了解决以上问题，1）DeepSeek-V3在训练过程中统一使用E4M3格式，并通过细粒度的per-tile（1x128）和per-group（128x128）量化来降低误差。FP8的好处还体现在节省显存上（尤其是激活值）。2）此外，DeepSeek-V3使用BF16来保存优化器状态，以及对部分操作进行选择性重计算（例如RMSNorm, MLA Up-Proj, SwiGLU）。3）在并行策略上，DeepSeek-V3使用64路的专家并行，16路的流水线并行，以及数据并行（ZeRO1）为了降低通信开销。4）在算法层面，DeepSeek-V3使用分组路由的方式，限制每个token只会激活4个节点上的专家，从而减半跨节点的通信流量。5）在系统层面，将节点间通信和节点内通信进行流水，最大化使用网络带宽和NVLink带宽。。

DeepSeek-V3 的基本架构图



DeepSeek-V3的MTP

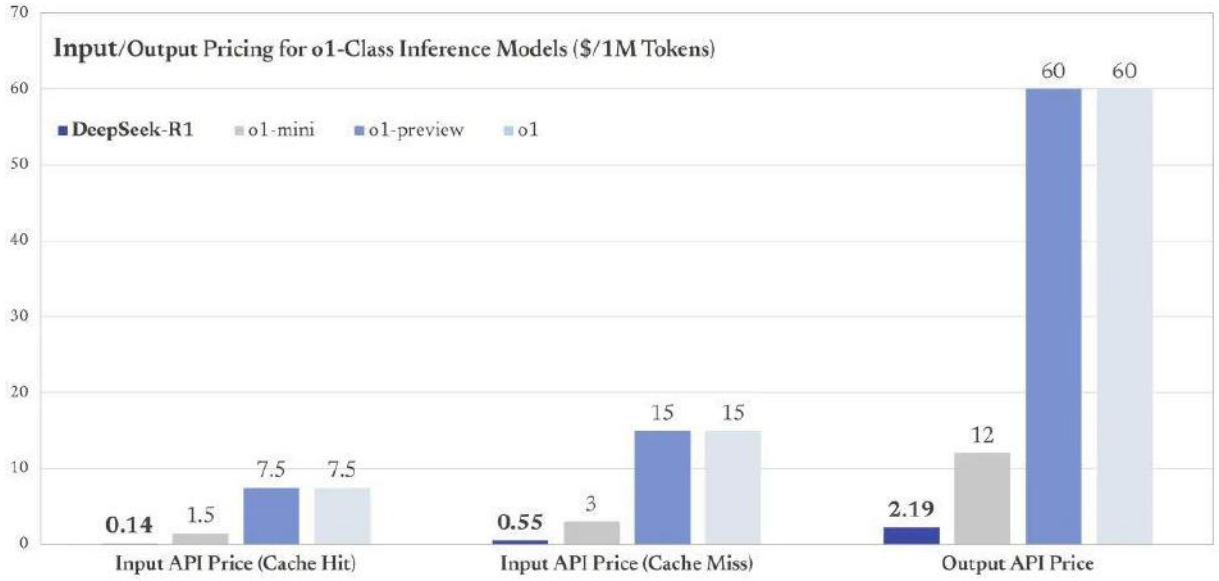


资料来源： 人工智能AB面，华西证券研究所

1.2.2 Deepseek：算法能力被突出（DeepSeek R1）

- **DeepSeek R1的技术关键在于其创新的训练方法。与OpenAI依赖人工干预的数据训练方式不同，DeepSeek R1采用了R1-Zero路线，直接将强化学习应用于基础模型，无需依赖监督微调（SFT）和已标注数据。**
- R1的总体训练过程如下：1）从base模型开始：使用量少、质量高的冷启动数据(cold data)来sft base模型，使得base模型可以有个良好的初始化；使用RL提升模型的推理能力；在RL阶段接近收敛时，用这个时候的checkpoint生成高质量的数据，将它们与现有的sft数据混合，创建新的sft数据集；2）再次从base模型开始：使用新创建的sft数据集做finetune；执行二阶段RL；得到最终的r1。

推理价格对比



蒸馏模型评测

	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：算法工程笔记，华西证券研究所

1.3 低成本模型有望引领AI产业“新路径”：开源+MOE

- **低训练成本+高性能表现，使得DeepSeek-V3成为国产模型之星**
- **DeepSeek-V3性能表现令人惊叹：不仅全面超越了Llama 3.1 405B，还能与GPT-4o、Claude 3.5 Sonnet等顶尖闭源模型正面竞争。**更令人瞩目的是，DeepSeek-V3的API价格仅为Claude 3.5 Sonnet的1/15，堪称“性价比之王”。
- DeepSeek-V3 的预训练阶段在不到两个月内完成，并花费了 2664K GPU 小时。加上 119K GPU 小时的上下文长度扩展和 5K GPU 小时的后训练，DeepSeek-V3 的完整训练成本仅为 2.788M GPU 小时。假设 H800 GPU 的租赁价格为每 GPU 小时 2 美元，**总训练成本仅为5.576M 美元。**

几款主流模型的API价格对比

Model	Input (cache miss)	Output
DeepSeek	$0.27/M tokens	$1.10/M tokens
Claude Haiku 3.5	$0.25/M tokens	$1.25/M tokens
Claude Sonnet 3.5	$3/M tokens	$15/M tokens
GPT-4o	$5/M tokens	$15/M tokens

DeepSeek-V3 的训练成本（假设 H800 GPU 的租赁价格为每小时 2 美元）

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

Table 1 | Training costs of DeepSeek-V3, assuming the rental price of H800 is $2 per GPU hour.

资料来源： 人工智能AB面，华西证券研究所

1.3 低成本模型有望引领AI产业“新路径”：开源+MOE

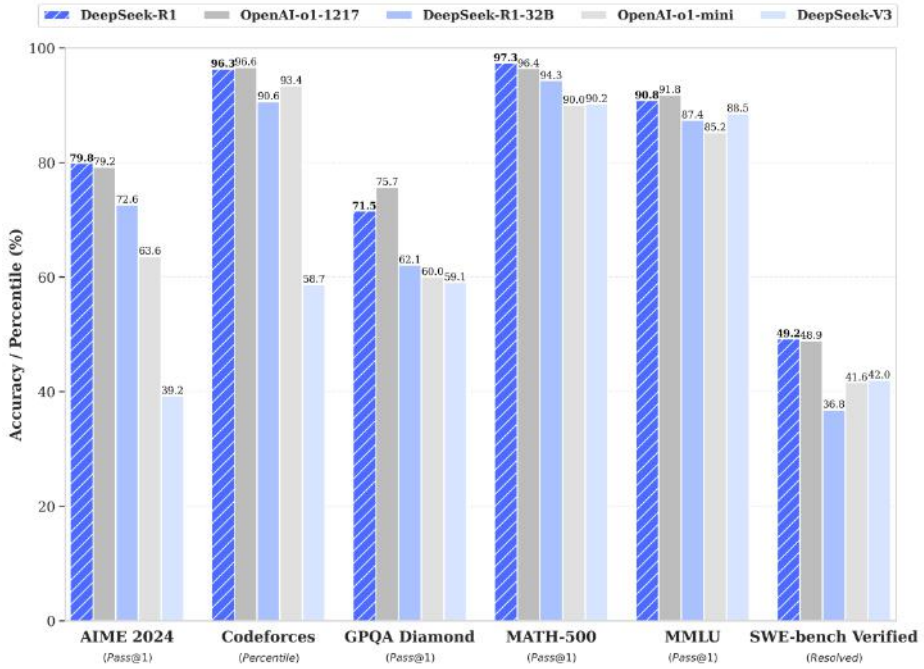
- **DeepSeek-R1：首个开源、媲美OpenAI o1的推理大模型。**DeepSeek-R1展现出了与OpenAI o1相当甚至在某些方面更优的性能。在MATH基准测试上，R1达到了77.5%的准确率，与o1的77.3%相近；在更具挑战性的AIME 2024上，R1的准确率达到71.3%，超过了o1的71.0%。在代码领域，R1在Codeforces评测中达到了2441分的水平，高于96.3%的人类参与者。
- **DeepSeek-R1成功蒸馏出多个小型推理模型，其中 32B & 70B版本 可媲美OpenAI o1-mini。**蒸馏后的小模型也取得了优异成绩，如DeepSeek-R1-Distill-Qwen-7B在AIME 2024上得分55.5%，超过QwQ-32B-Preview（参考P6）。

主流大模型能力对比

	Benchmark (Metric)	Claude-3.5-Sonnet-1022	GPT-4o 0513	DeepSeek V3	OpenAI o1-mini	OpenAI o1-1217	DeepSeek R1
	Architecture	-	-	MoE	-	-	MoE
	# Activated Params	-	-	37B	-	-	37B
	# Total Params	-	-	671B	-	-	671B
English	MMLU (Pass@1)	88.3	87.2	88.5	85.2	**91.8**	90.8
	MMLU-Redux (EM)	88.9	88.0	89.1	86.7	-	**92.9**
	MMLU-Pro (EM)	78.0	72.6	75.9	80.3	-	**84.0**
	DROP (3-shot F1)	88.3	83.7	91.6	83.9	90.2	**92.2**
	IF-Eval (Prompt Strict)	**86.5**	84.3	86.1	84.8	-	83.3
	GPQA Diamond (Pass@1)	65.0	49.9	59.1	60.0	**75.7**	71.5
	SimpleQA (Correct)	28.4	38.2	24.9	7.0	**47.0**	30.1
	FRAMES (Acc.)	72.5	80.5	73.3	76.9	-	**82.5**
	AlpacaEval2.0 (LC-winrate)	52.0	51.1	70.0	57.8	-	**87.6**
	ArenaHard (GPT-4-1106)	85.2	80.4	85.5	92.0	-	**92.3**
Code	LiveCodeBench (Pass@1-COT)	38.9	32.9	36.2	53.8	63.4	**65.9**
	Codeforces (Percentile)	20.3	23.6	58.7	93.4	**96.6**	96.3
	Codeforces (Rating)	717	759	1134	1820	**2061**	2029
	SWE Verified (Resolved)	**50.8**	38.8	42.0	41.6	48.9	49.2
	Aider-Polyglot (Acc.)	45.3	16.0	49.6	32.9	**61.7**	53.3
Math	AIME 2024 (Pass@1)	16.0	9.3	39.2	63.6	79.2	**79.8**
	MATH-500 (Pass@1)	78.3	74.6	90.2	90.0	96.4	**97.3**
	CNMO 2024 (Pass@1)	13.1	10.8	43.2	67.6	-	**78.8**
Chinese	CLUEWSC (EM)	85.4	87.9	90.9	89.9	-	**92.8**
	C-Eval (EM)	76.7	76.0	86.5	68.9	-	**91.8**
	C-SimpleQA (Correct)	55.4	58.7	**68.0**	40.3	-	63.7

Table 4 | Comparison between DeepSeek-R1 and other representative models.

R1与其他开源模型对比效果评测

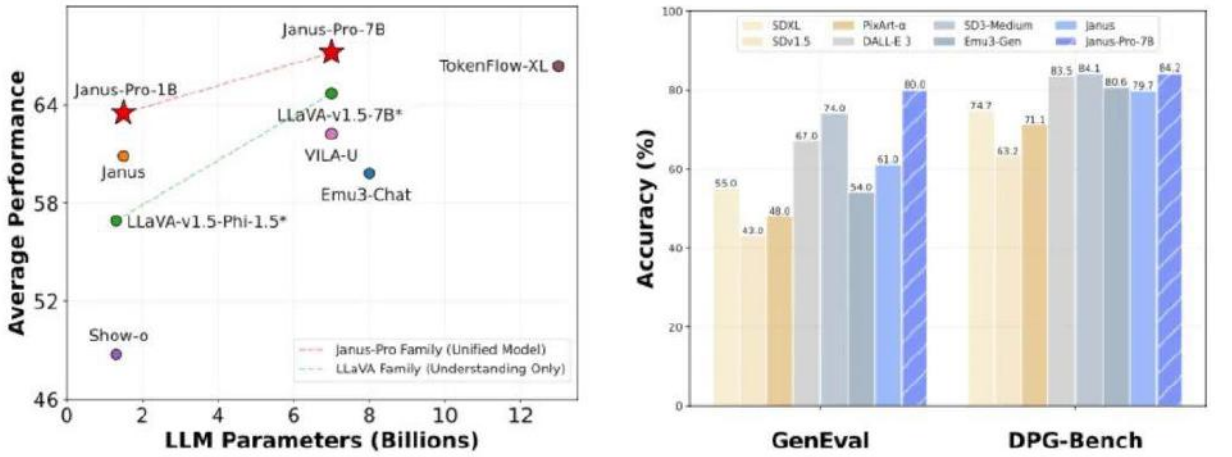


资料来源：腾讯科技，算法工程师笔记，华西证券研究所

1.3 低成本模型有望引领AI产业“新路径”：开源+MOE

- **根据深度学习与NLP公众号，Janus-Pro是一个新颖的自回归框架，统一了多模态理解和生成。通过将视觉编码分离为“理解”和“生成”两条路径，同时仍采用单一的Transformer架构进行处理，解决了以往方法的局限性。这种分离不仅缓解了视觉编码器在理解和生成中的角色冲突，还提升了框架的灵活性。**
- 在视觉生成方面，Janus-Pro通过添加7200万张高质量合成图像，使得在统一预训练阶段真实数据与合成数据的比例达到1：1，实现“更具视觉吸引力和稳定性的图像输出”。在多模态理解的训练数据方面，新模型参考了DeepSeek VL2并增加了大约9000万个样本。作为一个多模态模型，Janus-Pro不仅可以“文生图”，同样也能对图片进行描述，识别地标景点（例如杭州的西湖），识别图像中的文字，并能对图片中的知识（例如下图中的“猫和老鼠”蛋糕）进行介绍。

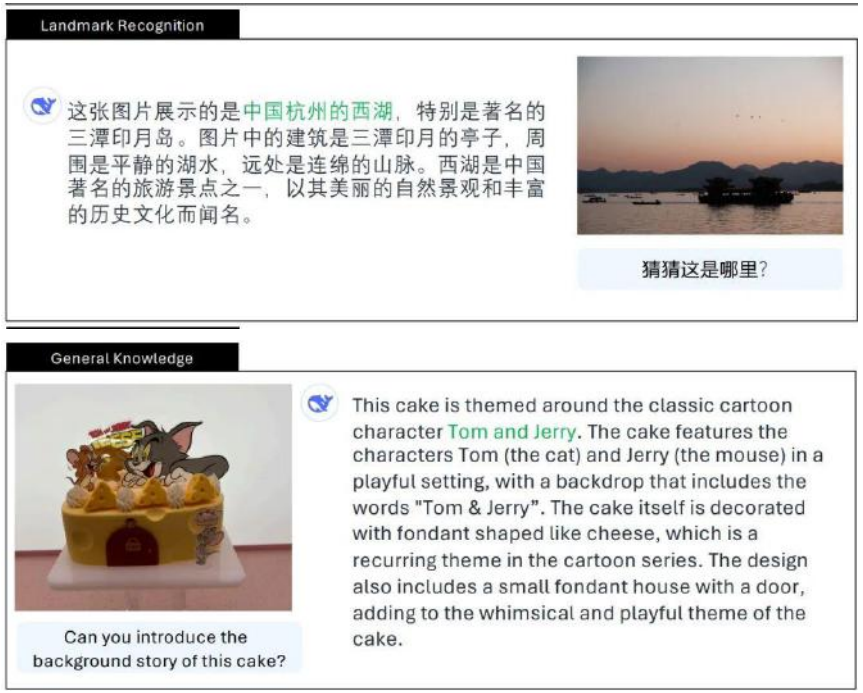
Janus-Pro 70亿参数模型表现好于OpenAI的DALL-E 3



(a) Average performance on four multimodal understanding benchmarks.

(b) Performance on instruction-following benchmarks for text-to-image generation.

Janus-Pro不仅可以“文生图”，同样也能对图片进行描述



资料来源：财联社，华西证券研究所

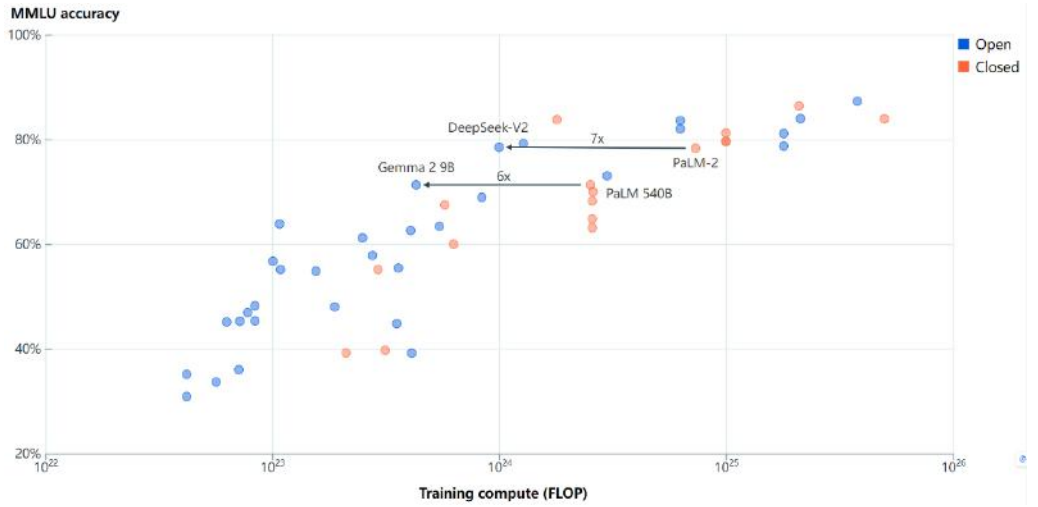
1.4 开源VS闭源：开源重构AI生态，与闭源共同繁荣下游

- 短期博弈：闭源企业通过垄断获取超额利润，但可能陷入“创新者窘境”；长期博弈：开源通过协作降低技术成本，但需解决商业化难题（如红帽的订阅模式）；混合策略：现代企业常采用“核心闭源+外围开源”（如微软的VS Code）或“开源获客+服务收费”（如MongoDB）。
- **开源模型（如DeepSeek）：推动技术民主化，适合需要透明性、定制化的场景；闭源模型（如GPT-4）：提供“开箱即用”体验，适合追求快速落地、无技术团队的企业。趋势：两者界限逐渐模糊，部分闭源厂商开源轻量版模型（如Google的Gemma），而DeepSeek等开源项目通过商业服务扩大影响力，共同推动AI技术普惠。**

开源模型与闭源模型对比

对比维度	开源模型（如 DeepSeek）	闭源模型（如 GPT - 4、Claude）
代码与模型权重	完全公开模型架构、训练代码和部分权重（如 DeepSeek - Coder）	仅提供 API 接口，核心代码和权重保密
可定制性	允许开发者本地部署、微调、二次开发（如基于 DeepSeek - R1 构建垂直领域智能体）	仅支持通过 API 有限定制，无法修改底层模型
透明度与信任	训练数据来源、模型设计可审计（如 DeepSeek 公开技术白皮书）	内部机制不透明，存在“黑箱”风险
商业化模式	开源核心模型 + 增值服务（如 DeepSeek 的 API 服务、企业级支持）	完全依赖 API 订阅付费或企业定制授权
社区协作	全球开发者共同优化模型（如 DeepSeek - Math 数据集被多国研究者用于改进数学推理能力）	仅由内部团队迭代，外部贡献受限
典型代表	DeepSeek 系列、Meta 的 Llama 2、Mistral AI 的 Mixtral	OpenAI 的 GPT - 4、Anthropic 的 Claude、Google 的 Gemini

训练效率对比



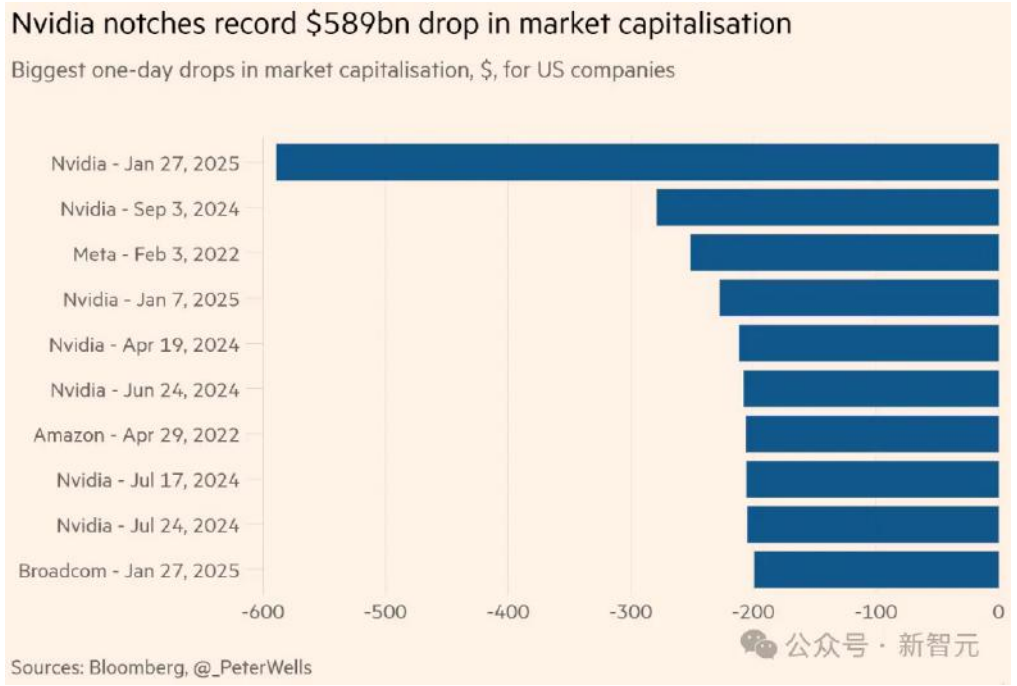
资料来源：两年砍柴，算力魔方，华西证券研究所

02 堆算力的AI“老路径”遭到强力挑战

2.1NV、博通大跌，纯算力路径依赖得到挑战

- **DeepSeek以极低的价格建立了一个突破性的AI模型，而且没有使用尖端芯片，纯算力路径依赖得到挑战。**截止1月27日收盘，AI龙头英伟达创下载入美国金融史有史以来的惨痛大跌，英伟达周一收跌16.97%，市值蒸发近5900亿美元（相当于略超3个AMD或近18个寒武纪），刷新崩盘纪录。
- 除了英伟达外，所有过去两年里与AI芯片关系密切的“卖铲人”们全部遭到严重冲击。据财联社报道，ASIC芯片概念股博通1月27日收跌17.4%，市值蒸发近2000亿美元。芯片代工厂台积电收跌13.3%，市值蒸发逾1500亿美元。在这轮大跌的上周刚刚因为特朗普官宣“星际之门”AI项目大涨的甲骨文，周一收跌13.79%。除英伟达和博通外，美满电子跌19.1%、美光科技跌11.71%，均是两位数跌幅。

NV市值变化



美国前十大市值公司1月27日股价表现

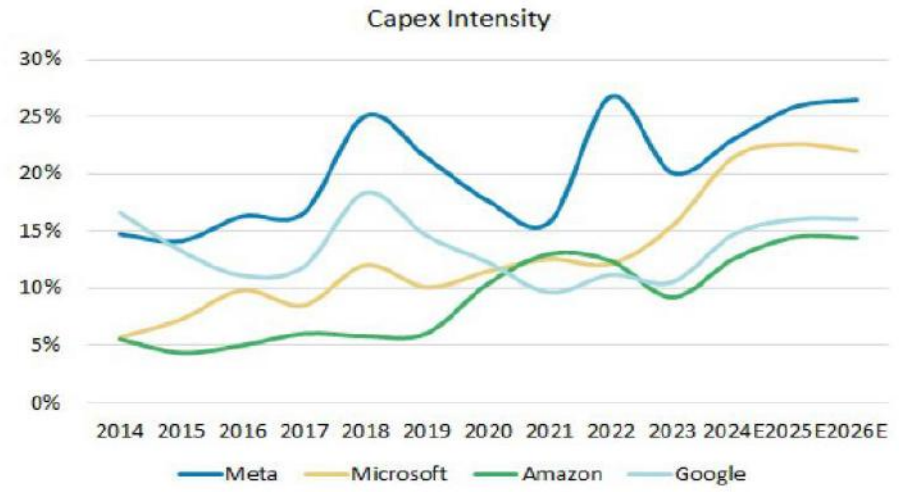
Rank	Name	Market Cap	Price	Today	Price (30 days)
^1 1	Apple AAPL	$3.404 T	$226.41	- 1.63%	
^1 2	Microsoft MSFT	$3.201 T	$430.62	- 3.03%	
v2 3	NVIDIA NVDA	$3.028 T	$123.65	- 13.30%	
4	Amazon AMZN	$2.433 T	$231.45	- 1.45%	
5	Alphabet (Google) GOOG	$2.395 T	$196.62	- 2.62%	
6	Saudi Aramco 2222.SR	$1.806 T	$7.46	- 0.36%	
7	Meta Platforms (Facebook) META	$1.633 T	$647.16	- 0.05%	
8	Tesla TSLA	$1.282 T	$399.66	- 1.70%	
9	TSMC TSM	$1.017 T	$196.13	- 11.61%	
^1 10	Berkshire Hathaway BRK-B	$1.009 T	$468.39	- 1.12%	
v1 11	Broadcom AVGO	$995.25 B	$212.33	- 13.23%	

资料来源：机器学习研究组订阅，小讯AI，华西证券研究所

2.2.1 国内外科技巨头持续提升capex指引，剑指NV GPU需求高景气

- **Capex资本支出逻辑已不足以支撑AI故事，急需推理/应用层“接棒”。**亚马逊、微软以及谷歌的最新财报显示，上个季度它们在不动产和设备上的支出，**达到了惊人的506亿美元，相比之下上年同期是305亿美元。这其中很大一部分资金，都流向了用于支持AI的数据中心。并且这三家公司指引，未来几个月它们的支出将继续走高。另外，Meta也是如此，Meta为自家在Instagram、WhatsApp和Facebook上的AI应用，进行基础设施投资。**
- **新一轮 10 万卡集群竞赛再次证明，AGI 的基建投入仍然如火如荼地进行，AI数据中心成为海外大厂布局重点。**根据海外独角兽24年7月报道，马斯克高调宣布为 xAI 建设 10 万卡集群，OpenAI/Microsoft、Anthropic/AWS、Meta 等大型 AI 公司也在加紧 10 万卡集群建设，每个集群在服务器硬件上的支出已经超过 40 亿美元。还受限于数据中心容量和电力供应能力。一个 10 万 GPU 的集群需要超过 150MW 的数据中心容量，一年的消耗就是 1.59TWh（15.9 亿度电），约等于 15 万个家庭一年的用电量。按$0.078/Kwh 的单价来计算，一个 10 万卡集群每年光在电力这一项上的支出就高达 1.239 亿美元。

科技巨头保持最高的Capex支出强度



Source: Company data, Morgan Stanley Research (E) estimates

10 万卡（H100）集群电量消耗

100k H100 GPU Cluster - Annual Electricity Costs		
Region	Tariff (USD/kWh)	Annual Cost $M
USA - Average	$0.083	$131.9
USA - North Dakota	$0.074	$117.6
USA - Utah	$0.068	$108.0
USA - Arizona	$0.078	$123.9
USA - ND Wind PPA	$0.033	$52.4
USA - Solar PPA (CAISO)	$0.033	$52.4

资料来源：经纬创投，海外独角兽，华西证券研究所

2.2.1 国内外科技巨头持续提升capex指引，剑指NV GPU需求高景气

- 2024年前三季度，百度、阿里巴巴、腾讯三家公司的总资本开支达867.21亿元，同比增长119.80%，这些投资主要用于数据中心、服务器等算力基础设施的采购。根据壹零社，**2025年，字节跳动的资本开支将增至1600亿元，其中900亿元用于算力采购，700亿元用于IDC基建及网络设备，旨在构建自主可控的数据中心集群。Omdia 估计，字节跳动和腾讯2024年各自订购了约 23 万块英伟达芯片，其中包括 H20 型号，这款低配版的 Hopper 经过改动，以满足针对中国客户的美国出口管制条例。**
- 根据半导体行业观察，进入2024年，英伟达的GPU销量依然猛增，英伟达CEO黄仁勋也直言，公司新推出的Blackwell在市场的关注度非常高，也有很多客户在买。根据 Jon Peddie Research 的数据，今年全球 GPU 市场预计将超过 985 亿美元。黄仁勋也认为，数据中心运营商将在未来四年内花费 1 万亿美元升级其基础设施，以满足 AI 开发人员的需求，因此这个机会足以支持多家 GPU 供应商。

海外主流AI视频应用近期访问量及其变化

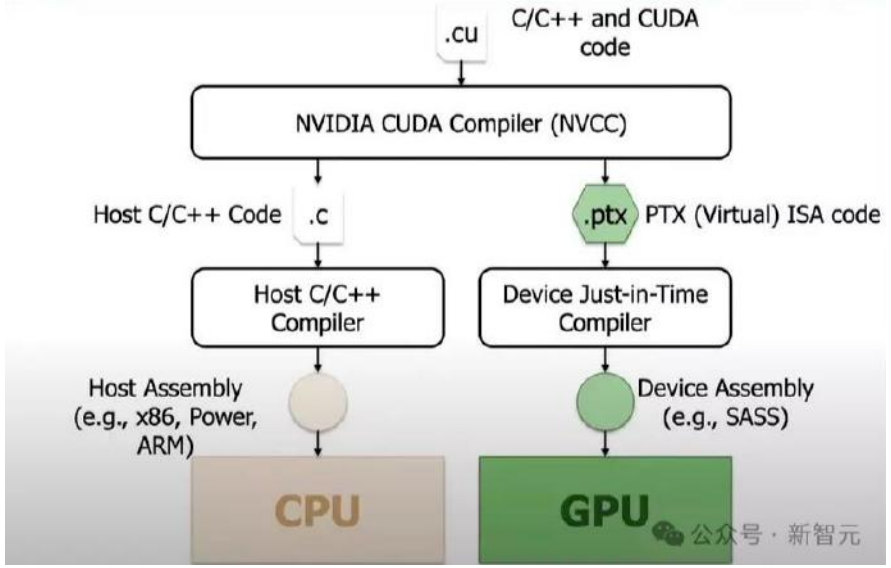
	2024 YE (H100 equivalent)	2025 (GB200)	2025YE (H100 equivalent)
MSFT	750k-900k	800k-1m	2.5m-3.1m
GOOG	1m-1.5m	400k	3.5m-4.2m
META	550k -650k	650k-800k	1.9m-2.5m
AMZN	250k-400k	360k	1.3m-1.6m
XAI	~100k	200k-400k	550k-1m

资料来源：半导体行业观察，华西证券研究所

2.2.2 国产Deepseek模型爆火，高端算力/高集群能力并非唯一解

- R1模型在技术上实现了重要突破——用纯深度学习的方法让AI自发涌现出推理能力，在数学、代码、自然语言推理等任务上，性能比肩美国开放人工智能研究中心（OpenAI）的o1模型正式版，该模型同时延续了该公司高性价比的优势。深度求索公司R1模型训练成本仅为560万美元，远远低于美国开放人工智能研究中心、谷歌、“元”公司等美国科技巨头在人工智能技术上投入的数亿美元乃至数十亿美元。**根据新智元援引外媒报道，他们在短短两个月时间，在2048个H800 GPU集群上，训出6710亿参数的MoE语言模型，比顶尖AI效率高出10倍。**
- **Deepseek突破不是用CUDA实现的，而是通过大量细粒度优化以及使用英伟达的类汇编级别的PTX（并行线程执行）编程。**在训练V3模型时，DeepSeek对英伟达H800 GPU进行了重新配置：为了最大化性能，DeepSeek还通过额外的细粒度线程/线程束级别调整，实现了先进的流水线算法。这些优化远超常规CUDA开发水平，但维护难度极高。然而，这种级别的优化恰恰充分展现DeepSeek团队的卓越技术实力。

当PTX转换为SASS后，就会针对特定代的英伟达GPU进行优化



V3论文中具体提到了关于PTX的细节

In detail, we employ the warp specialization technique (Bauer et al., 2014) and partition 20 SMs into 10 communication channels. During the dispatching process, (1) IB sending, (2) IB-to-NVLink forwarding, and (3) NVLink receiving are handled by respective warps. The number of warps allocated to each communication task is dynamically adjusted according to the actual workload across all SMs. Similarly, during the combining process, (1) NVLink sending, (2) NVLink-to-IB forwarding and accumulation, and (3) IB receiving and accumulation are also handled by dynamically adjusted warps. In addition, both dispatching and combining kernels overlap with the computation stream, so we also consider their impact on other SM computation kernels. Specifically, we employ customized PTX (Parallel Thread Execution) instructions and auto-tune the communication chunk size, which significantly reduces the use of the L2 cache and the interference to other SMs.

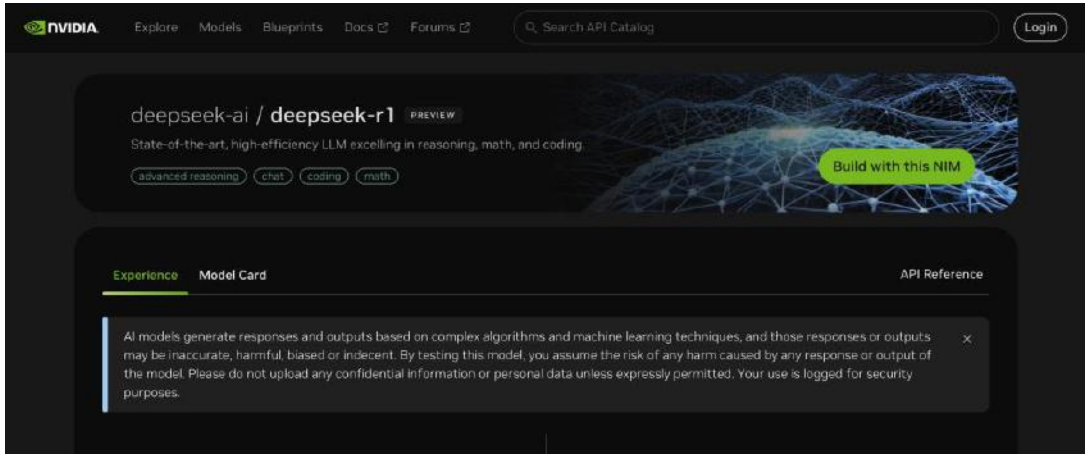
资料来源：新智元，华西证券研究所

2.3 国产算力异军突起，充分受益国产模型deepseek崛起

- **据华为云2月1日消息，DeepSeek-R1开源后引发全球用户和开发者关注。经过硅基流动和华为云团队连日攻坚，现在，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**。据华为云消息，该服务具备以下特点：1）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）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。
- **根据中国经济网，日前，英伟达、微软、亚马逊等AI巨头纷纷宣布，已接入DeepSeek。此外，或许是迫于DeepSeek带来的压力，OpenAI紧急上线新一代推理模型o3-mini，并首次向ChatGPT免费用户开放推理模型**。

硅基流动和华为云首发并上线基于华为云昇腾云服务的DeepSeek R1/V3推理服务

DeepSeek-R1模型已作为NVIDIA NIM微服务预览版



资料来源：中国经济网， 华西证券研究所

2.3 国产算力异军突起，充分受益国产模型deepseek崛起

- **华为于2018年10月发布了AI战略，并于2019年8月在深圳坂田总部正式发布AI处理器昇腾 910、昇腾310和MindSpore全场景AI计算框架。**昇腾系列(HUAWEI Ascend)AI处理器和基础软件构建Atlas人工智能计算解 决方案，包括Atlas系列模块、板卡、小站、服务器、集群等丰富的产品形态，打造面向“端、边、云”的全场景AI基础设施 方案，覆盖深度学习领域推理和训练全流程。
- **海光: DCU产品兼容“类 CUDA”环境，性能快速迭代。**海光 DCU 主要具有三大技术优势。一是强大的计算能力。二是高速并行数据处理能力。海光 DCU 集成片上高带宽内存芯片，可以在大规模数据计算过程中提供优异的数 据处理能力，使海光 DCU 可以适用于广泛的应用场景。三是良好的软件生态环境。海光DCU 采用 GPGPU 架构，兼容“类 CUDA”环境，解决了产品推广过程中的软件生态兼容性问题。深算二号：已经于2023年第三季度发布，并在大数据处理、人工智能和商业计算等领域实现了商用。该产品具有全精度浮点数据和各种常见整型数据计算能力，性能相对于深算一号提升了100%以上。

华为Atlas产品

海光深算二号参数

规格参数		K100	K100_AI
芯片	计算核心	120	-
性能指标	FP64	24.5 TFLOPS	-
	FP32	24.5 TFLOPS	24.5 TFLOPS Tensor 98 TFLOPS
	FP16	100 TFLOPS*	Tensor 200 TFLOPS
	INT8	200 TOPS*	400 TOPS
显存		64GB	64GB
PCIe 接口		PCIe 4.0 x16	PCIe 5.0 x16
TDP		300W	350-400W
尺寸规格		全高全长双宽	全高全长双宽

* 采用稀疏技术

资料来源：AI云原生智能算力架构， AIGC时代的技术服务商，华西证券研究所

2.3 国产算力异军突起，充分受益国产模型deepseek崛起

- 寒武纪：根据AI云原生智能算力架构，公司目前已推出3款加速卡:MLU370-S4、MLU370-X4、MLU370-X8，已与国内主流互联网厂商开展深入的应用适配。而且根据公司官网介绍，**全新一代云端智能训练芯片思元590采用MLUarch05全新架构，实测训练性能较在售产品有了显著提升，它提供了更大的内存容量和更高的内存带宽，其PCIe接口也较上代实现了升级。**
- **国产芯片与NV差距正在逐步缩小，deepseek出现使得国产算力竞争优势进一步凸显。**国产芯片厂商不断加大研发投入，在芯片架构、制程工艺等方面取得了一系列突破。未来，随着技术的不断进步和生态系统的逐步完善，国产 AI 芯片有望缩小与英伟达的差距，在全球 AI 芯片市场中占据更重要的地位。

国产GPU对比

品牌	型号	架构	峰值算力(FP16)	显存	带宽	卡间互联	功耗(TDP)
华为	910B	达芬奇架构	376TFLOPS	64 GB	400GB/s	392GB/S	350W
天数	天垓 100	通用架构	147TFLOPS	32 GB	1.2TB/s	64GB/s	250W
天数	智铠 100	通用架构	200TFLOPS	32 GB	800GB/s	64GB/s	150W
海光	K100 AI 版	通用架构	196TFLOPS	64 GB	896GB/s	N/A	350W
海光	K100	通用架构	100TFLOPS	64 GB	896GB/s	N/A	300W
寒武纪	MLU590	MLUv02 扩展架构	314T	80 GB	2TB/S	318.8GB/5	350W

华为与NV H100等GPU参数对比

	华为 Ascend 910B	L20 (PCIe)	H20 (PCIe/SXM)	H100 (PCIe/SXM)
年份	2023	2023	2023	2022
工艺	7+nm	4nm	4nm	4nm
架构	HUAWEI Da Vinci	Ada Lovelace	Hopper	Hopper
最大功率	400 watt	275W	400W	350/700 watt
GPU 内存	64G HBM2e	48G GDDR6	80G HBM3	80G HBM3
GPU 内存带宽		864GB/s	4.0TB/s	2/3.35 TB/s
二级缓存		96MB	60MB	
GPU 互连（一对一最大带宽）	HCCS 56GB/s	PCIe Gen4 64GB/s	PCIe Gen5 128GB/s, NVLINK 900GB/s	PCIe Gen5 128GB/s, NVLINK 900GB/s
GPU 互连（一对多总带宽）	HCCS 392GB/s	PCIe Gen4 64GB/s	PCIe Gen5 128GB/s, NVLINK 900GB/s	PCIe Gen5 128GB/s, NVLINK 900GB/s
FP32		59.8 TFLOPS	44 TFLOPS	51/67 TFLOPS
TF32 (TensorFloat)		59.8 TFLOPS	74 TFLOPS	756/989 TFLOPS
BFLOAT16 TensorCore		119/119 TFLOPS	148/148 TFLOPS	
FP16 TensorCore	320 TFLOPS			1513/1979 TFLOPS
FP8 TensorCore				3026/3958 TFLOPS
INT8 TensorCore	640 TFLOPS	239/239 TFLOPS	296/296 TFLOPS	3026/3958 TFLOPS

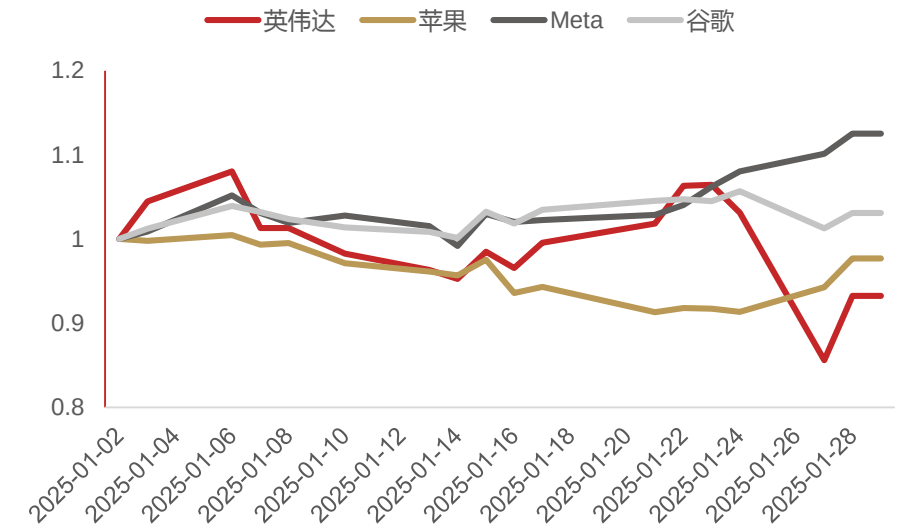
资料来源：北方算网，中科院计算所培训中心，华西证券研究所

03 端侧AI爆发元年

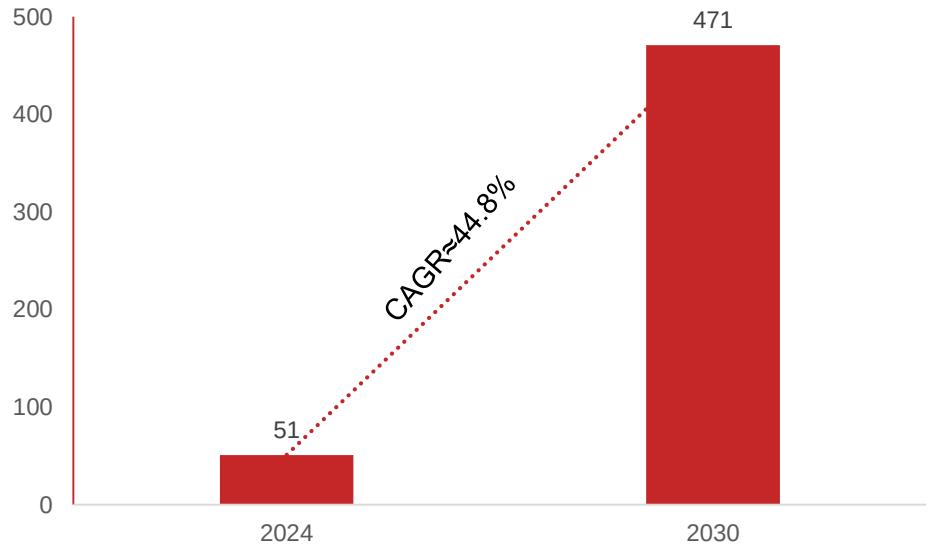
3.1.1 AI智能体加速元年

- **Deepseek-V3发布后英伟达股价大跌。**截至1月27日美股收盘，英伟达下跌近17%，收报118.42美元，单日市值蒸发达5890亿美元，为美国股市历史新高，打破了去年9月初英伟达单日重挫超9%、市值蒸发2790亿美元的纪录。尽管28日英伟达股价收涨近9%，这仍然在一定程度上反映了**市场对英伟达和大模型发展旧模式的信心不足。与之相对，苹果、Meta、谷歌等应用提供商股价表现明显更好。**苹果、Meta分别收涨约3%和2%。
- AI智能体（AI Agent）是指能自主感知环境、做出决策并执行行动的系统，具备自主性、交互性、反应性和适应性等基本特征，能在复杂多变的环境中独立完成任务，包括记忆、规划、工具、行动四个主要模块。**虽然ChatGPT等LLM一直是AI新闻的焦点，但人们逐渐开始意识到其局限性，如幻觉、记忆力短。根据每日经济新闻，OpenAI于2024年底表示GPT模型的改进速度正在放缓，引发业内对未来发展方向的疑问。**Salesforce首席执行官马克·贝尼奥夫更直言AI的未来发展不在于LLM，而在于开发AI智能体。
- 在这一背景下，根据钛媒体、科创板日报和DeepTech深科技，谷歌、OpenAI、Anthropic、字节跳动等国内外领先大模型厂商纷纷剑指智能体开发，发布Project Astra、Operator、Computer Use、UI-TARS等产品，**2025年有望成为AI智能体加速元年。**根据Research and Market、麦肯锡等多份权威报告，在多元化需求驱动下，智能体市场呈爆发式增长态势，2024年全球智能体市场规模约为51亿美元，预计2030年将飙升至471亿美元，复合年增长率高达44.8%。

部分大模型相关企业股价（取2025/01/02股价为1）



全球智能体市场规模（亿美元）

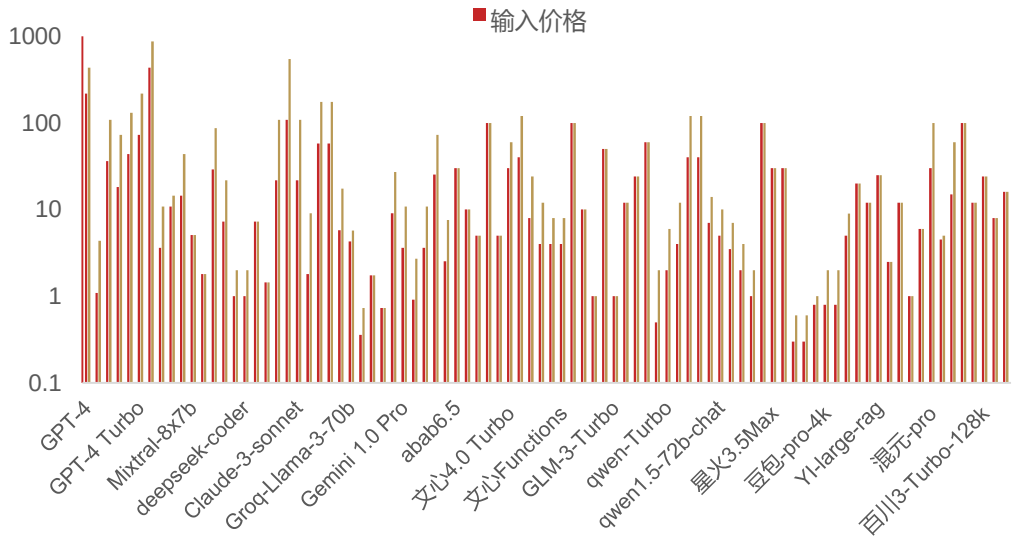


资料来源：华尔街见闻，Wind，我来AI等，华西证券研究所

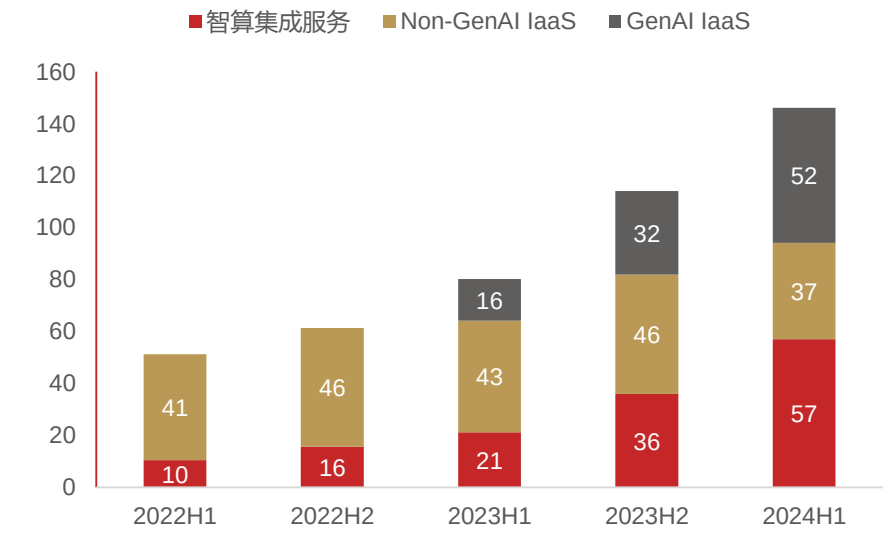
3.1.2 Token成本持续降低推动AI普及

- **主流的模型Token价格持续降低。**根据36Kr今年1月的报道，阿里云已宣布2024年度第3轮大模型降价，通义千问视觉理解模型全线降价超80%。其中，Qwen-VL-Plus直降81%，输入价格仅为0.0015元/千tokens，创下全网最低价格；更高性能的Qwen-VL-Max降至0.003元/千tokens，降幅高达85%。这无疑是对此前字节跳动发布豆包视觉理解模型的回击。在2024年12月18日的火山引擎Force大会上，字节跳动推出的豆包视觉理解模型千tokens输入价格为3厘，1元钱可处理284张720P的图片。当时字节称该价格比业内价低85%。硅谷大模型的价格也出现了下降趋势。根据每日经济新闻，OpenAI的GPT-4o调用API的价格比GPT-4-turbo降低了一半，为5美元/百万Tokens，谷歌Gemini 1.5 Flash的价格降至0.35美元/百万Tokens。预计未来以DeepSeek-V3为代表的模型训练优化和GPU性能增长将继续带动Tokens成本和价格降低。
- **价格下降带动Tokens使用量增长，端侧AI应用初具放量条件。**根据量子位数据，2024下半年国内大模型商用落地日均Tokens消耗量增长近10倍，从1000亿级规模到10000亿规模，月复合增长率达到45%。特别是以低价为卖点的字节跳动，5月份还停留在日均百亿级Tokens水平，不及全行业1/5；8月初突破千亿Tokens大关，并在之后保持迅速增长，12月日均Tokens市场份额占比超50%；月均复合增长率超过60%。IDC数据显示， 2024上半年中国智算服务整体市场同比增长79.6%，市场规模达到146.1亿元人民币。其中，智算集成服务市场同比增长168.4%，市场规模达57.0亿元人民币；GenAI IaaS市场同比增长203.6%，市场规模达52.0亿元人民币；Non-GenAI IaaS市场同比缩减13.7%，市场规模达37.1亿元人民币。

主流大模型Token价格（元/百万Tokens，对数坐标轴）



中国智算服务市场规模（亿元）



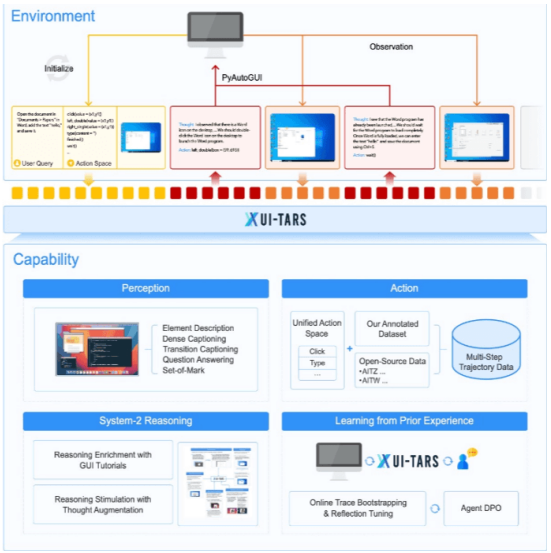
资料来源：AIGCRank，IDC，华西证券研究所

3.1.3 行业巨头剑指智能体开发

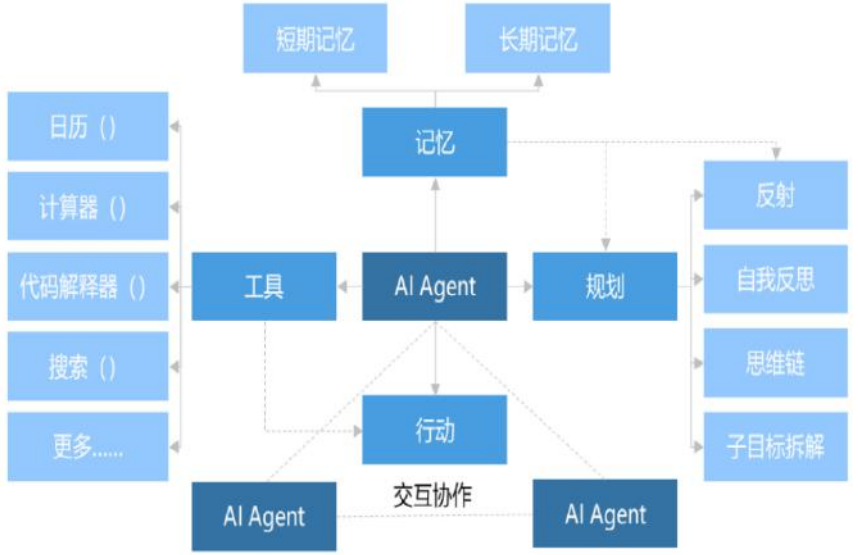
- 2023年10月，**Anthropic率先展示了基于Claude 3.5 Sonnet的计算机操作智能体Computer Use，首次实现让模型移动光标、点击按钮和输入文本**。该功能已向一小部分测试人员开放，包括来自DoorDash、Canva和Asana等公司的第三方开发者，共同构建跨行业智能体系统。
- 1月23日，谷歌发布了首款实现原生多模态输入输出的大模型Gemini2.0，还**推出了3个AI智能体产品：通用大模型助手Project Astra、浏览器助手Project Mariner、编程助手Jules，标志着谷歌AI初步向智能体时代转变**。Astra能进行多模态实时推理，并可调用搜索、地图等工具。Mariner能理解和推理浏览器中的文本、代码、图像、表单等元素，并使用这些信息完成购物、查找航班和酒店等任务。Jules可集成进Github工作流，查看用户已有的代码并修改。
- 1月24日，OpenAI宣布其最新**智能体Operator已向部分用户开放试用**，能自主操作浏览器以完成采购杂货、提交费用报表等任务，旨在通过自动化操作提升用户在日常生活和工作中的效率。OpenAI还宣布与Instacart、Uber、eBay、Priceline、OpenTable、Etsy等公司展开合作，提供更多智能体功能。
- 字节跳动**推出GUI代理模型UI-TARS**，能实时理解动态界面，用户可通过自然语言实现对桌面、移动设备和网页界面的自动化交互。该模型结合了快速直观反应和复杂任务规划的能力，引入了系统化推理机制，**支持多步任务分解、反思思维和里程碑识别等推理模式**；还具备短期和长期记忆功能，能更好地适应动态任务需求；**通过自动收集、筛选和反思新的交互轨迹进行迭代训练，从错误中学习以减少人工干预**。

谷歌智能体用于游戏

UI-TARS智能体



AI智能体基本框架

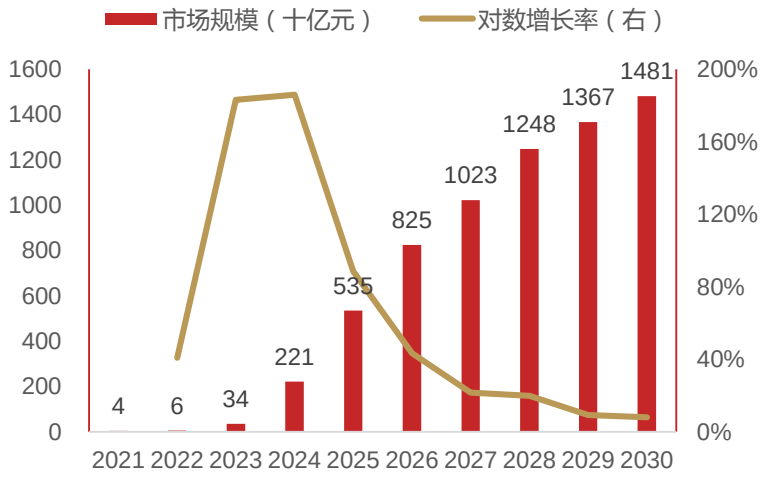


资料来源：雷锋网，深科技，极客网，AI工具集，华西证券研究所

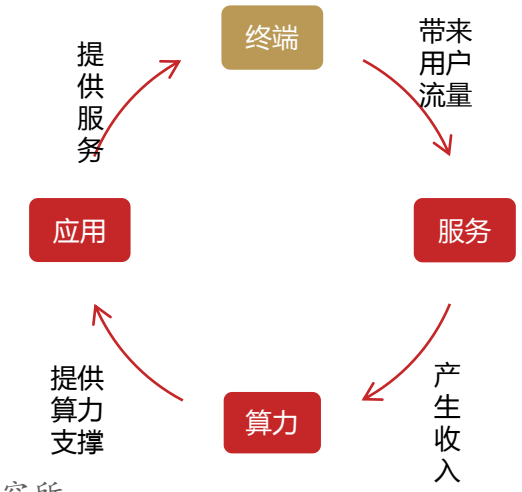
3.2.1 端侧硬件落地加速构建商业闭环

- **智能终端是集成了AI技术，能执行复杂任务、提供智能化服务和交互体验的终端设备**，包括智能手机、AI PC、智能穿戴设备、智能家居设备、车载信息系统等。随着5G商用、AIoT兴起，**智能终端从手机扩展到可穿戴设备、智能家居等领域，多样化和专业化趋势愈发明显**。2022年OpenAI推出GPT-3.5引爆行业热潮，随后国内外科技巨头争相布局大模型，AI技术全面融入智能终端方方面面。
- **伴随AI技术持续迭代和处理器性能进步，AI智能终端市场正处于蓬勃发展期。**根据QYR的数据，2023年中国AI智能终端市场销售收入达到344亿元，预计2030年可以达到14812亿元，年复合增长率约为37.33%；核心厂商包括联想、华为、苹果、荣耀、vivo和小米，2024年合计占有约67.81%的市场份额。从产品类型方面来看，AI PC占有重要地位，预计2030年份额将达到73.88%。根据雷科技和第一财经，英伟达、AMD、英特尔等国际巨头纷纷押注AI终端，发布Thor、锐龙AI、Ultra等产品及软件生态支持；根据财联社和量子之声等，瑞芯微、国芯科技、中科蓝讯等国产厂商也已布局端侧AI芯片，力图抓住机遇追赶英伟达。
- **相比云端AI，终端AI在成本、能耗、隐私等方面都具有优势。**成本方面，将一些处理转移到终端可以减轻日益增加的云基础设施开支。在能耗方面，端侧AI能耗更低，特别是将处理和数据传输相结合时。在可靠性方面，需求高峰期时云端存在大量排队等待和高时延，端侧可规避该问题，甚至可在无网络环境下使用。更重要的是，**端侧AI在用户隐私保护方面具有无可比拟的优势，因为端侧AI的所有信息都保留在终端上，能大大消除用户对隐私泄露的顾虑。基于上述优势，端云协同逐渐成为AI部署的主流模式，我们认为AI智能终端将成为大模型用户入口，引领大模型和AI技术走向商业闭环。**

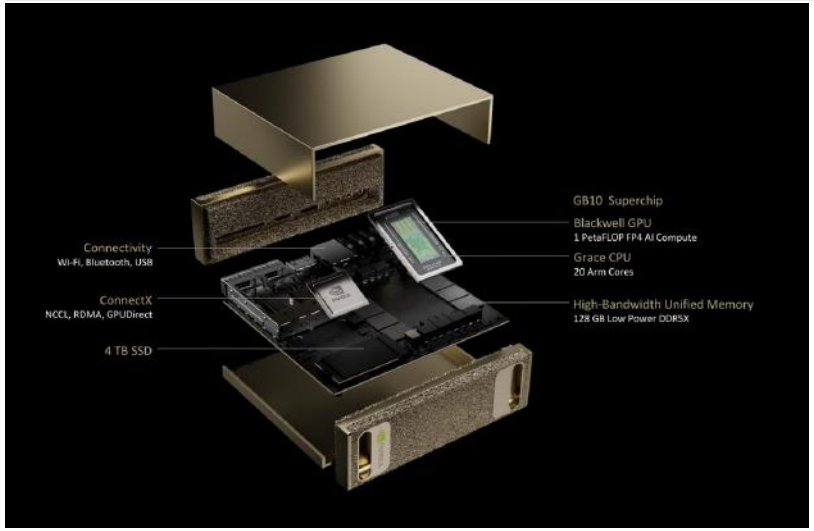
中国AI智能终端市场规模预测



大模型商业闭环



英伟达Project DIGITS端侧AI硬件



资料来源：36氪，QYR，21经济网，电子工程专辑，华西证券研究所

3.2.2 国际巨头争相布局智能终端

谷歌与三星

01

- 谷歌Tensor移动处理器从**初代开始即搭载TPU用于加速机器学习**；最新一代G5的TOPS值增加近40%，实际性能提升14%，**还引入小型嵌入式RISC-V核心，支持设备上训练。**
- 三星Exynos 2400处理器引入AMD RDNA3架构Xclipse 940 GPU，AI性能提升至14.7倍，CPU性能提高70%。与谷歌深度合作，**允许三星手机直接呼出Gemini大模型，提供公式识别、划圈即搜、图文生成等功能。**

Meta

03

- Meta与传统眼镜品牌雷朋合作**推出AI眼镜Ray-Ban Meta**，售价299美元，**推出当季销量即超过30万副。**Ray-Ban Meta搭载高通骁龙AR1 Gen 1芯片；接入Meta AI，提供**音质增强、视觉搜索和实时翻译**等个人助手体验；搭载12MP超广角镜头，能在部分些场景取代手机录制生活「切片」。
- Meta也**为Quest3头显接入Meta AI**，提供聊天功能，**允许头显“看到”用户真实视场中的内容并回答用户提出的问题。**

英伟达

02

- **英伟达GPU为AI提供了主要算力支持。**最新发布的RTX50系显卡AI性能最高达4000TOPS。软件方面，可开启**DLSS4帧生成，首发支持超75款游戏和应用。**
- 终端产品层面，**Project Digits搭载GB10，体积类似Mac Mini，可处理200B参数大模型，满足个人用户运行大模型的需求。**
- Jetson Thor**瞄准机器人、自动驾驶等AI终端，集成Transformer引擎，提供800TFLOPS的FP8算力**，并通过Isaac、Omniverse、GR00T等提供软件支持。

苹果

04

- 苹果于2017年在A11处理器中首次集成NPU，发布了第1代Core ML；到A17Pro和M4算力超35TOPS。
- 软件方面，苹果**为iPhone、Mac等苹果全家桶产品带来Apple Intelligence功能，包括用于创建自定义表情符号的Genmoji、用于生成图像的Image Playground和Image Wand、集成到Siri的ChatGPT**等服务，还能借助相机控制实现视觉智能。
- 苹果还在开发可以识别图像和视频的多模态模型，致力于**增强Vision Pro的计算机视觉能力**，使其可以快速识别周遭环境。

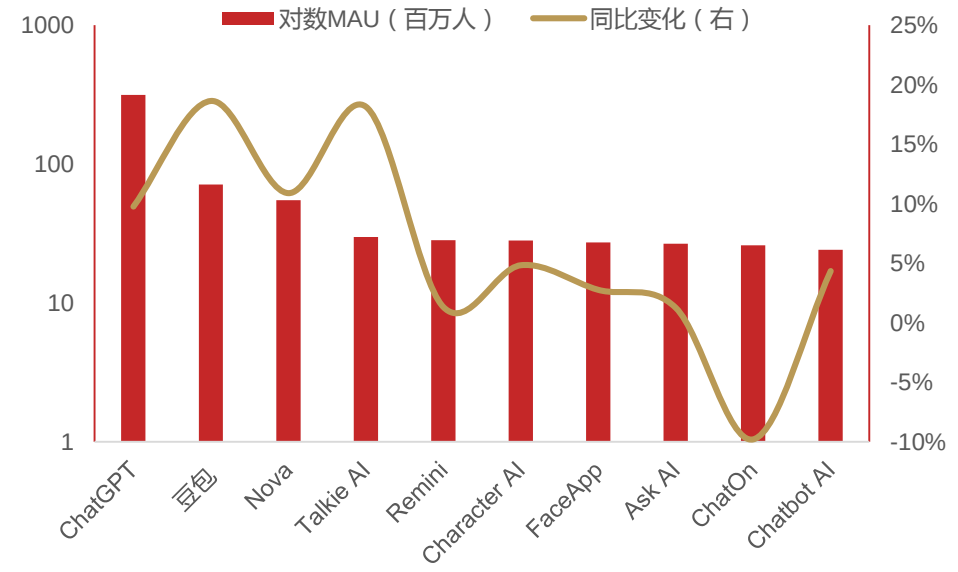
资料来源：根据IT之家等整理，华西证券研究所

3.2.3 字节引领中国智能终端

- 字节跳动AI模型起步较晚，2023年初才开始训练基础模型，但投入巨大。2024年字节跳动采购了约23万张英伟达芯片，成为英伟达第二大客户，仅次于微软。在资金、人力和算力的高投入下，字节大模型后来居上，2024年11月豆包APP累计用户规模即超过1.6亿，MAU达到5998万，仅次于ChatGPT，位列全球第二，全国第一。**豆包大模型的成功为字节系AI智能终端的爆发提供了有力支撑。**字节还充分发挥自身APP工厂的特点，**AI应用如流水线产品一样被快速推向市场，为端侧AI提供了丰富的应用场景。**据a16z全球Top 100 AI应用榜单，字节旗下Cici、Coze、Gauth、Hypic、CapCut五款产品名列前茅，同时字节仍在不断在推出新产品试点，**通过API方式使AI硬件接入Chat Bot产品，让用户无需掏出手机即可与AI助手对话。**
- 硬件AI终端产品可能是字节跳动完善其AI布局的最后一块关键拼图。在2024年12月举行的FORCE原动力大会上，**字节跳动宣布与多家公司共同推出AI+硬件的智跃计划**，合作伙伴包括FoloToy、乐鑫科技、ToyCity和魂伴科技。该计划旨在综合豆包大模型、火山引擎的拟人化语音对话、ToyCity的潮玩设计、乐鑫科技的AI芯片等技术积累，以用户友好的方式将AI技术与智能硬件结合，推动智能玩具、教育工具和互动娱乐产品的普及。**字节还展示了多款基于字节大模型的AI终端产品**，如斐耳声学AI耳机GS Links、AI玩具显眼包、AI耳机Ola Friend、蔚蓝机器狗。**我们认为字节具备领先的大模型技术、丰富的软件生态和海量的用户数据，并通过并购补足了自身缺失的音频和硬件技术，将成为AI智能终端赛道核心玩家。**

基于豆包的AI终端

2024年12月主要AI应用MAU

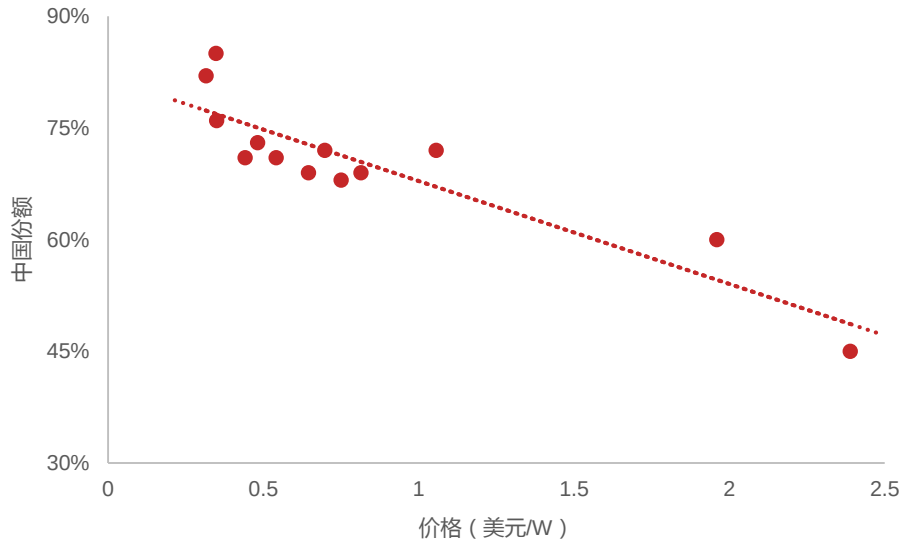


资料来源：AI产品榜，DCD，36氪等，华西证券研究所

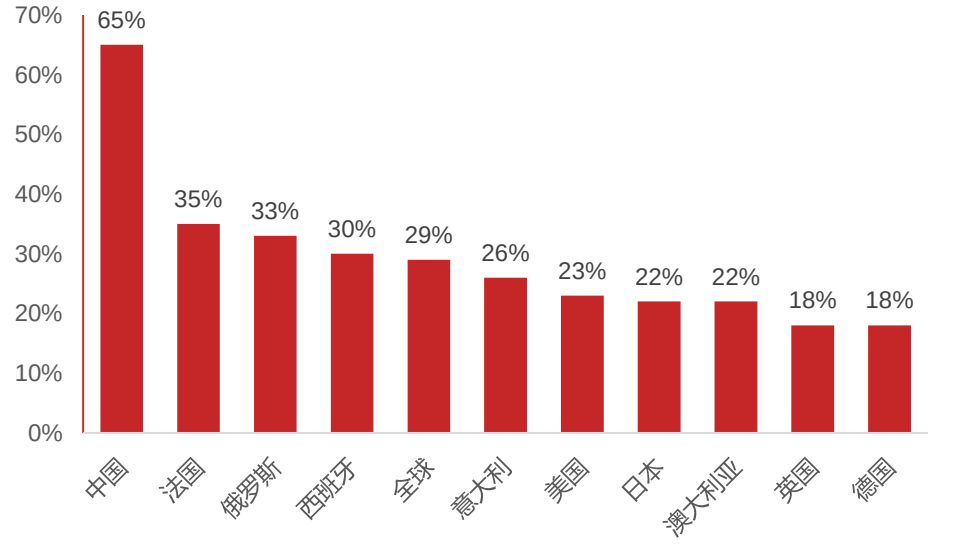
3.2.4 中国厂商供应链优势突出

- **目前智能终端成本较高，阻碍了其普及。**根据Wellsenn XR，以Ray-Ban Meta为例，仅主控芯片高通骁龙AR Gen 1成本即达到55美元，占比超1/3，总成本达164美元，定价则达到299美元。智能终端产业链与手机等传统消费电子产业链重合度较高，**中国供应商在芯片外绝大部分环节都积累了明显优势**。根据半导体产业纵横，如中国屏幕厂商2023年Q2的市场份额已经超过50%，截至2020年末，摄像头模组前三大供应商欧菲光、舜宇光学、丘钛微均为中国企业，2022年韦尔股份（收购豪威）和格科微的CMOS市场份额分列第3、4名，中蓝电子马达出货量2021和2022年进入全球前三。**依托国内完备的产业链，国产智能终端厂商更具成本优势，有望借此扩大市场份额。**如闪极AI拍拍镜A1功能与Ray-ban Meta类似，但最低售价仅999元，极大降低了AI眼镜的门槛。再参考光伏电池，价格随国产厂商份额增长快速下降，我们认为中国智能终端厂商的成本优势也将推动智能终端价格降低和普及化。
- 另一方面，**中国人对AI的态度最积极乐观，接受度较高，也有利于智能终端普及。**根据2018年电通安吉斯的报告，平均有65%的中国人对新兴技术将创造新的就业机会这一观点持乐观态度。与之相对，在美国平均只有23%的人持乐观态度；在全球10个国家的2万名被调查者中，平均只有29%的人认为新兴技术将创造新的就业机会，中国对就业前景的乐观态度远超过全球平均水平。

光伏电池价格随国产厂商份额增长快速下降



中国人对AI的态度最积极乐观



资料来源：Our World in Data，Wellsenn XR，量子位等，华西证券研究所

04 受益标的

4. 受益标的

◆ **受益标的：**

重点公司盈利预测与估值（一）

	代码	公司名称	最新收盘价（元）	市值（亿元）	PE		
					2024E	2025E	2026E
AI终端	688018.SH	乐鑫科技	270.24	303.21	87.53	65.99	50.24
	688608.SH	恒玄科技	397.01	476.59	122.44	81.74	60.31
	300493.SZ	润欣科技	30.66	157.16	142.80	102.27	72.71
	688332.SH	中科蓝讯	143.00	172.04	56.80	43.13	34.50
	688220.SH	翱捷科技	69.52	290.80	-51.85	-98.92	190.05
	300622.SZ	博士眼镜	47.51	83.29	60.55	52.49	45.13
	001314.SZ	亿道信息	48.65	68.83	53.05	47.56	41.92
	688343.SH	云天励飞	48.30	171.53	-	-	-
	301383.SZ	天键股份	60.30	98.21	48.46	36.46	26.74
	301536.SZ	星宸科技	81.40	342.74	132.34	97.42	72.34
算力云	3896.HK	金山云	7.82	297.57	-18.00	-47.17	-139.24
	688227.SH	品高股份	29.01	32.80	-	-	-
	688158.SH	优刻得	13.83	62.66	-	-	-
	688316.SH	青云科技	34.07	16.29	-	-	-

资料来源：Wind，PE根据2025年1月27日收盘价计算，为Wind一致预测结果，华西证券研究所

4. 受益标的

◆ **受益标的：**

重点公司盈利预测与估值（二）

	代码	公司名称	最新收盘价（元）	市值（亿元）	PE		
					2024E	2025E	2026E
AI agent应用	603990.SH	麦迪科技	11.39	34.89	0.00	0.00	0.00
	603859.SH	能科科技	36.03	88.16	35.83	28.07	22.42
	603108.SH	润达医疗	16.35	98.70	32.58	21.95	16.59
	688228.SH	开普云	49.23	33.24	0.00	0.00	0.00
	688590.SH	新致软件	18.51	49.09	48.58	29.11	21.76
	2013.HK	微盟集团	2.29	82.80	-20.73	-79.85	66.71
	300634.SZ	彩讯股份	29.54	133.29	44.42	36.35	30.01
	300170.SZ	汉得信息	19.59	192.93	91.89	76.55	64.21
	300229.SZ	拓尔思	22.11	193.16	102.60	75.56	61.03
	300033.SZ	同花顺	279.13	1,500.60	106.07	87.30	77.52
	688318.SH	财富趋势	152.11	278.27	91.25	81.89	72.54
	300688.SZ	创业黑马	31.13	52.11	-85.66	82.68	58.64
	300624.SZ	万兴科技	68.46	132.36	216.65	127.18	97.63
算力	0981.HK	中芯国际	38.00	4,304.62	65.23	44.35	33.30
	688041.SH	海光信息	128.00	2,975.15	155.34	106.12	78.22
	688256.SH	寒武纪	572.00	2,387.85	-520.71	8,895.80	499.56
	603019.SH	中科曙光	66.99	980.20	45.24	36.87	30.63

资料来源：Wind，PE根据2025年1月27日收盘价计算，为Wind一致预测结果，华西证券研究所

05 风险提示

5. 风险提示

- 市场竞争加剧：近两年来，AI产业竞争者逐渐增多，若竞争加剧可能会造成产品同质化严重，进而对企业创新造成 不良影响，影响产业发展；
- 产品研发不及预期：AI属于技术密集型产业，若产品研发不及预期，会影响产业进程。

免责声明

分析师承诺

作者具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，保证报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求客观、公正，结论不受任何第三方的授意、影响，特此声明。

评级说明

公司评级标准	投资评级	说明
以报告发布日后的6个月内公司股价相对上证指数的涨跌幅为基准。	买入	分析师预测在此期间股价相对强于上证指数达到或超过15%
	增持	分析师预测在此期间股价相对强于上证指数在5%—15%之间
	中性	分析师预测在此期间股价相对上证指数在-5%—5%之间
	减持	分析师预测在此期间股价相对弱于上证指数5%—15%之间
	卖出	分析师预测在此期间股价相对弱于上证指数达到或超过15%
行业评级标准		
以报告发布日后的6个月内行业指数的涨跌幅为基准。	推荐	分析师预测在此期间行业指数相对强于上证指数达到或超过10%
	中性	分析师预测在此期间行业指数相对上证指数在-10%—10%之间
	回避	分析师预测在此期间行业指数相对弱于上证指数达到或超过10%

华西证券研究所：

地址：北京市西城区太平桥大街丰汇园11号丰汇时代大厦南座5层

网址：http://www.hx168.com.cn/hxzq/hxindex.html

免责声明

华西证券股份有限公司（以下简称“本公司”）具备证券投资咨询业务资格。本报告仅供本公司签约客户使用。本公司不会因接收人收到或者经由其他渠道转发收到本报告而直接视其为本公司客户。

本报告基于本公司研究所及其研究人员认为的已经公开的资料或者研究人员的实地调研资料，但本公司对该等信息的准确性、完整性或可靠性不作任何保证。本报告所载资料、意见以及推测仅于本报告发布当日的判断，且这种判断受到研究方法、研究依据等多方面的制约。在不同时期，本公司可发出与本报告所载资料、意见及预测不一致的报告。本公司不保证本报告所含信息始终保持在最新状态。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，投资者需自行关注相应更新或修改。

在任何情况下，本报告仅提供给签约客户参考使用，任何信息或所表述的意见绝不构成对任何人的投资建议。市场有风险，投资需谨慎。投资者不应将本报告视为做出投资决策的惟一参考因素，亦不应认为本报告可以取代自己的判断。在任何情况下，本报告均未考虑到个别客户的特殊投资目标、财务状况或需求，不能作为客户进行客户买卖、认购证券或者其他金融工具的保证或邀请。在任何情况下，本公司、本公司员工或者其他关联方均不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告而导致的任何可能损失负有任何责任。投资者因使用本公司研究报告做出的任何投资决策均是独立行为，与本公司、本公司员工及其他关联方无关。

本公司建立起信息隔离墙制度、跨墙制度来规范管理跨部门、跨关联机构之间的信息流动。务请投资者注意，在法律许可的前提下，本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。在法律许可的前提下，本公司的董事、高级职员或员工可能担任本报告所提到的公司的董事。