

证券研究报告·行业动态报告

DeepSeek激活创新竞争，AI应用迎来“安卓时刻”

分析师：杨艾莉

yangaili@csc.com.cn

010-85156448

SAC 编号：S1440519060002

SFC 编号：BQI330

分析师：杨晓玮

yangxiaowei@csc.com.cn

SAC 编号：S1440523110001

发布日期：2025年2月4日

➤ 1、Deepseek激发创新竞争之后，AI应用怎么看？

➤ 应用开发迎来“安卓时刻”

回顾安卓与iOS应用的发展，我们率先提出不应只关注大模型本身的用户数及活跃度，更应该关注开发者，尤其是中小开发者的数量。据GitHub，在Llama比DeepSeek开源时间早1年半的情况下，目前DeepSeek R1在GitHub上的开发者点赞数量已经达到约5.7万，接近Llama。根据GitHub、Hugging Face社区上的开发者实测，经过R1微调的80亿参数小模型可以在个人笔记本中运行，本地化部署门槛显著下降，应用的开发将迎来百花齐放。

➤ 有用户有产品能力的公司，仍将“赢在起跑线”

虽然春节期间Deepseek的关注度赶超字节豆包，但我们认为以字节跳动为代表的中国头部互联网公司，手握高粘性+大DAU产品，叠加强产品能力。在第二阶段的应用、场景等领域，用户数+产品力+商业变现能力，仍然将帮助他们在接下来的竞争中抢占先机。

目前豆包全球累计下载量（约9000万）仍然明显高于Deepseek（约2000万），而其他拥有用户基础和产品能力的公司，也有机会迎头赶上。

➤ 哪些公司有望迅速对齐这一趋势？

我们重申，有用户、有场景的公司将受益于此轮模型能力爆发，近期Meta、苹果、小米等公司明显的相对涨幅也印证这一点。

- ✓ 字节跳动：C端AI应用第一梯队，有抖音\TIKTOK等产品的用户红利。
- ✓ 腾讯：微信MAU超过14亿，腾讯元宝/优图/混元3D不断推出
- ✓ 阿里：推出通义千问系列效果最好的模型Qwen Max，各项性能看齐Deepseek
- ✓ 其他：美团（超6亿活跃交易用户）、哔哩哔哩（超3亿MAU）、拼多多（超8亿活跃交易用户）、小米（手机+智家等端侧AI）

2、相关公司

➤ 利好有用户、有场景的公司

- ✓ 互联网公司：腾讯、阿里巴巴、小米、美团、拼多多、哔哩哔哩
- ✓ 垂类及细分应用：汇量科技、第四范式、美图、值得买、焦点科技、易点天下、万兴科技、恺英网络、巨人网络

第一章

春节回顾：DeepSeek全球出圈，激发竞争活力

第二章

展望：DeepSeek利好哪些AI应用？

1、DeepSeek：低成本+高性能模型，引发全球AI产业的竞争活力

- 12月底上线V3模型，多项评测成绩接近GPT-4o。V3为开源MoE模型，当前版本暂不支持多模态输入输出。其多项评测成绩超越了 Qwen2.5-72B 和 Llama-3.1-405B 等其他开源模型，知识类、长文本、代码测试表现超过GPT-4o。但每百万token的输入和输出价格分别为0.27和1.10美元，显著低于GPT-4o的2.50和10美元。
- MLA（多头潜在注意力）创新技术降低模型推理成本。MLA与标准注意力机制相比，每个查询所需的 KV 缓存量（存储上下文对话）减少约93.3%，从而减少计算量和算力需求。我们认为这是DeepSeek实现超低API价格的关键一环。

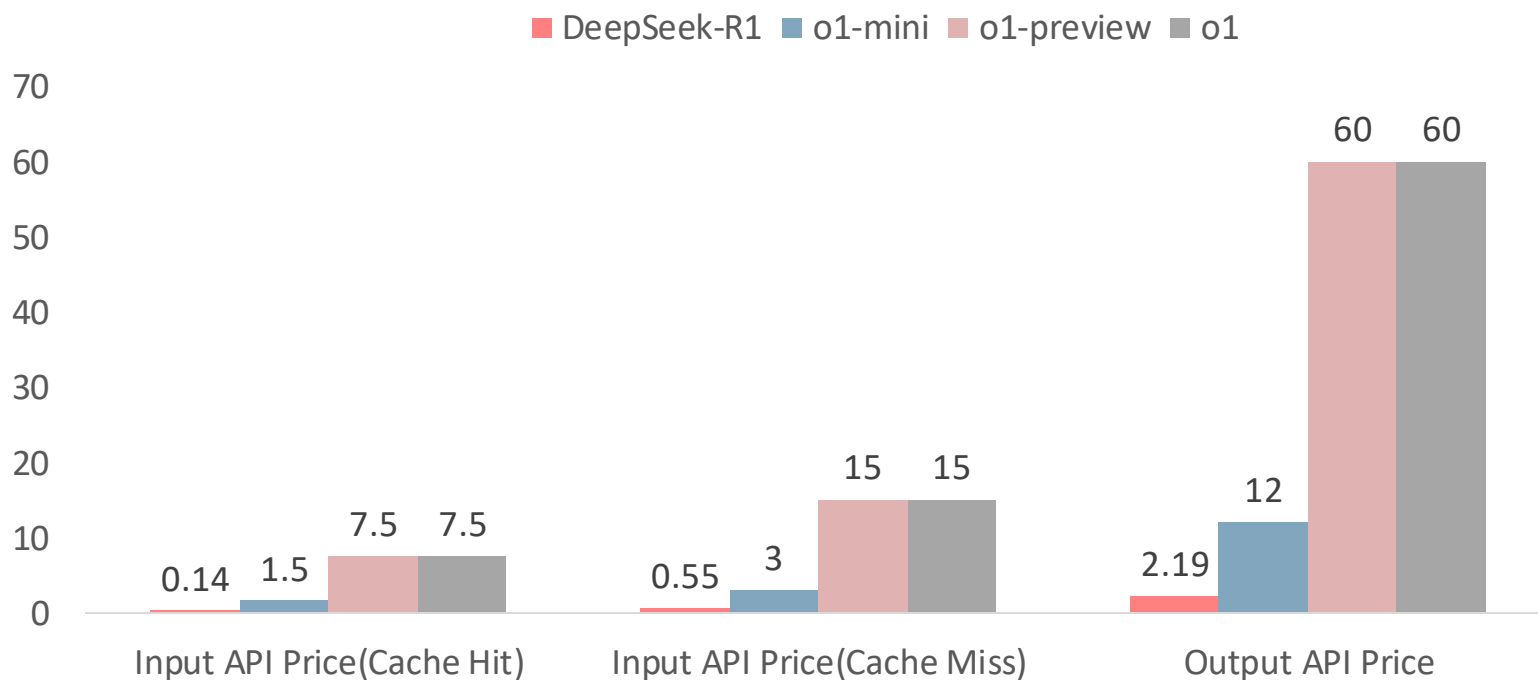
图表：各模型测试结果与API价格对比

模型	输入价格 (美元/百万 token)	输出价格 (美元/百万 token)	MMLU (Pass@1)	SWE Verified (Resolved)	AIME2024	MATH-500
Llama 3.1 405B	3.50	3.50	88.6	24.5	23.3	73.8
Claude 3.5 Sonnet	3.00	15.00	88.3	50.8	16.0	78.3
GPT-4o	2.50	10.00	87.2	38.8	9.3	74.6
Gemini 1.5 Pro	1.25	5.00	86.0		20.0	88.0
Llama 3.2 70B	0.59	0.73	86.0		20.0	64.0
DeepSeek V3	0.27	1.10	88.5	42.0	39.2	90.2
GPT-4o mini	0.15	0.60	82.0	33.2	6.7	79.0

1、DeepSeek：R1模型低成本、高效率，性能赶超o1模型

- Deepseek发布开源模型DeepSeek-R1。数学/编程/推理能力接近OpenAI的o1模型，但API调用成本仅为其2%。DeepSeek-R1系列包括DeepSeek-R1-Zero和DeepSeek-R1两大模型，核心创新点在于：
 - ✓ 传统大模型需要大量人工标注的监督数据进行训练，但DeepSeek-R1-Zero仅通过大规模强化学习训练，没有监督微调，便自发涌现出“反思”“多步验证”等复杂推理行为，训练成本较低。
 - ✓ 推理能力可以向小模型迁移。使用DeepSeek-R1作为教师模型生成800K数据对小模型微调，提升逻辑能力，通义千问15亿参数开源模型经过微调后，数学能力超过GPT-4o。

图表：各模型API价格对比（美元/百万token）



1、DeepSeek——C端视角：全球破圈，成为用户规模增速最快的移动端AI应用

- Deepseek AI助手成为今年春节爆款应用。
- ✓ 看国内：2021-2024年春节中国区iOS应用榜Top10，字节系普遍占5个以上，今年DeepSeek稳居榜首，成为首个在春节期间持续霸榜的AI应用。
- ✓ 看全球：春节期间Deepseek登顶中国、美国、德国等全球超100个国家和地区的iOS应用总榜第1，下载量超越ChatGPT。用户从早期的开发者+编成IT群体，拓宽到了目前的泛C端用户。

图表：DeepSeek成为首个春节持续登顶iOS榜的AI应用

排名	2021	2022	2023	2024	2025
1	快手极速版	抖音极速版	央视频	今日头条	DeepSeek
2	番茄小说	番茄畅听	微信	央视频	红果短剧
3	今日头条极速版	微信	支付宝	微信	豆包
4	抖音	今日头条极速版	快手	抖音极速版	DreamDesk AI
5	西瓜视频	抖音	抖音	支付宝	纳米AI搜索
6	快手	快手极速版	剪映	抖音	云闪付
7	抖音极速版	支付宝	WiFi万能钥匙	剪映	蛋仔派对
8	剪映	今日头条	快手极速版	番茄小说	天启圣源
9	腾讯微视	快手	QQ	汽水音乐	夸克
10	番茄畅听	番茄小说	拼多多	快手	铁路12306

图表：DeepSeek登顶超100个国家和地区的iOS应用榜首

国家和地区	#1	#2	#3	#4	#5
中国	DeepSeek	红果短剧	豆包	DreamDesk中文...	纳米AI搜索
中国香港	DeepSeek	拼多多	豆包	高合用户口	高合玩三合一...
中国台湾	台灣彩票	ChatGPT	DeepSeek	DramaBox	Google Gemini
美国	DeepSeek	ChatGPT	Threads	T	TurboTax
日本	DeepSeek	Google Gemini	ChatGPT	ENEOS公式ア...	Threads
韩国	DeepSeek	ChatGPT	Temu	대한민국 모바일	스마트 워터스
印度	DeepSeek	Blinkit	Zepto	Hotstar	ChatGPT
俄罗斯	DeepSeek	Remini	*** Proxy Fast	***	BEAR***
英国	DeepSeek	Ticketmaster	ChatGPT	GOV.UK ID Check	HMRC
加拿大	DeepSeek	ChatGPT	Threads	Temu	***
澳大利亚	DeepSeek	Klarna	ChatGPT	Ticketek AU	Temu

注：21-24年数据为除夕的排名；25年数据为2月1日的排名；
资料来源：七麦数据、中信建投

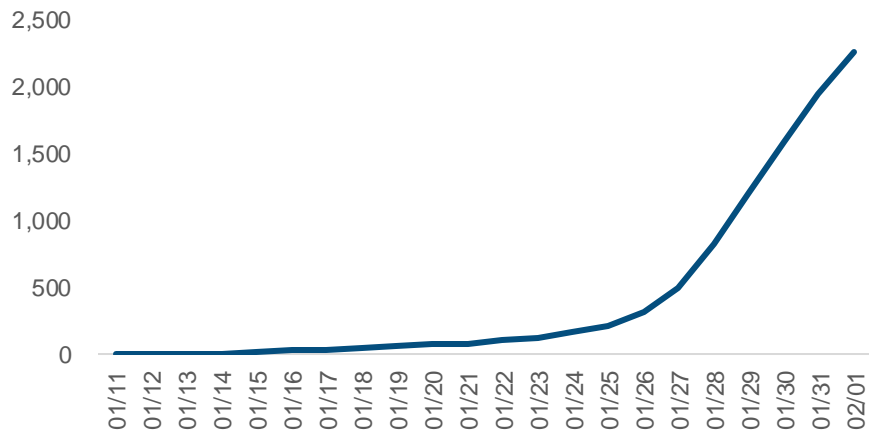
1、DeepSeek——C端视角：全球破圈，成为用户规模增速最快的移动端AI应用

➤ Deepseek成为全球用户规模增速最快的移动端AI应用。

✓ 日活（全球）：据AI产品榜，DeepSeek移动端上线第20天（1月30日）全球日活2161万，为同期ChatGPT移动端日活 14.8 倍。若对比海外移动端用户数（SensorTower口径），1月31日DeepSeek为1203万，环比增加4%。

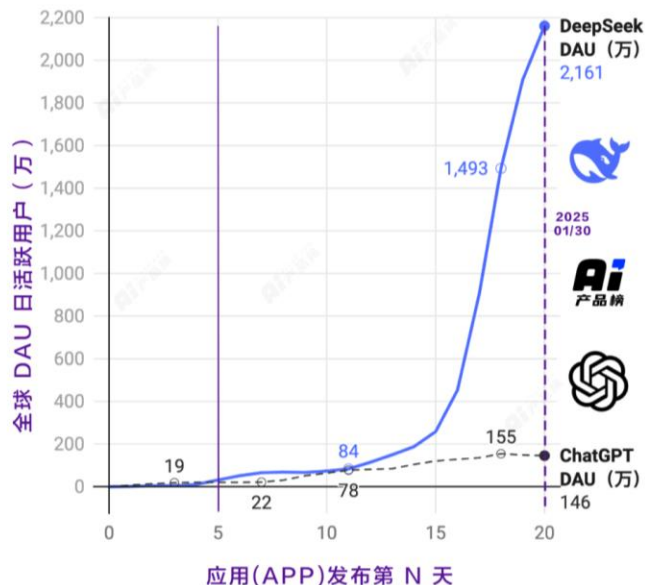
✓ 下载量（全球）：据点点数据，截至2月1日DeepSeek App全球累计下载量2261万创新高，环比前一天增加16%。

图表：DeepSeek全球累计下载量持续创新高（万）

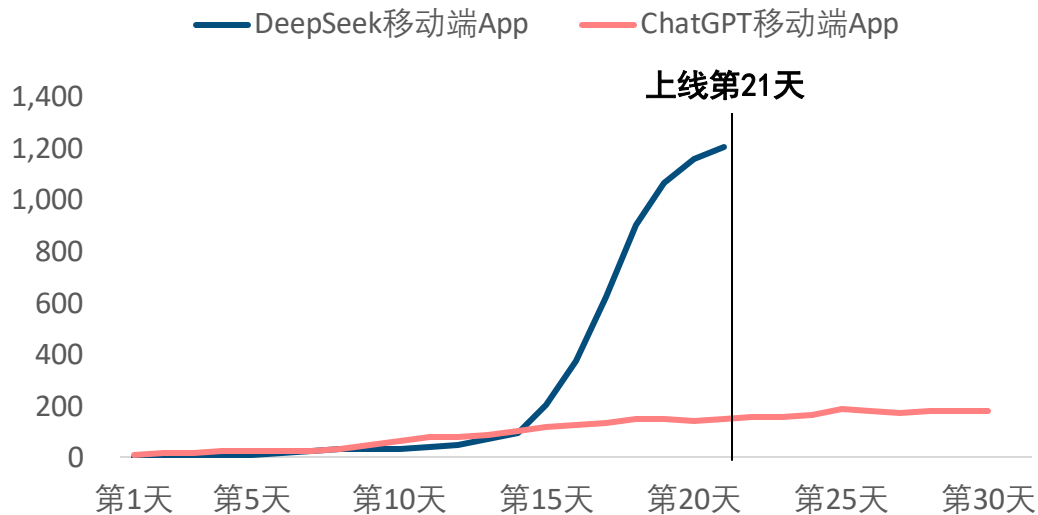


图表：上线同样天数DeepSeek移动端全球DAU与ChatGPT移动端对比

全球DAU对比



海外（不含中国大陆）DAU对比（万）



注：ChatGPT上线第1天为23年5月18日，DeepSeek上线第1天为25年1月11日
资料来源：AI产品榜、点点数据、Sensor Tower、中信建投

1、DeepSeek——大厂视角：积极采纳，模型已接入多个大模型聚合平台和AI应用

➤ 全球主流的公有云公司均在旗下的AI应用开发平台中接入DeepSeek模型，包括微软、亚马逊、腾讯、华为、案例、百度等。应用端，秘塔AI搜索成为首个接入DeepSeek模型的国产应用之一，弥补了当前DeepSeek AI助手未开放联网搜索功能的问题。

图表：R1模型已接入多个大模型聚合平台和AI应用

公司	DeepSeek上线的平台/产品名称	备注
微软	Azure AI Foundry和GitHub	Azure AI Foundry是微软在2024年11月发布的一站式AI开发平台，支持不同大语言模型之间的切换，目前超过1800种模型
亚马逊	Amazon Bedrock 和 Amazon SageMaker AI	均为亚马逊旗下AI应用开发平台，Amazon Bedrock 适合希望通过API快速集成预训练模型的团队；Amazon SageMaker AI 则适合自主进行高级定制、训练和部署、调用基础设施的团队
英伟达	—	发布采用DeepSeek R1 671b的NIM微服务预览版，以API形式供开发者调用
腾讯	HAI	HAI是一款面向AI、科学计算的GPU应用服务产品，开发者可以在线一键调用DeepSeek R1
华为	—	硅基流动与华为云团队联合首发并上线基于华为云昇腾云服务的DeepSeekR1/V3推理服务
阿里	PAI Model Gallery	支持云上一键部署DeepSeek V3、DeepSeek R1
百度	百度智能云千帆平台	V3和R1的调用价格仅为官方定价的3折和5折
秘塔AI搜索		联网搜索功能接入DeepSeek R1

1、DeepSeek——大厂视角：看好DeepSeek提升AI应用渗透率

➤ 多个海外AI应用公司高管认为DeepSeek提升大模型效率、降低推理成本。当前大模型性能已满足绝大部分应用场景，DeepSeek提升大模型训练和推理效率、降低API成本，有望提升AI应用在各个场景的渗透率。苹果、微软和SAP的CEO均对DeepSeek在大模型效率和成本方面的贡献给予积极评价，Meta作为海外头部开源模型研发厂商，CEO表示将吸收DeepSeek的模型开发经验。

图表：多个海外公司高管在业绩会中表达对DeepSeek的观点

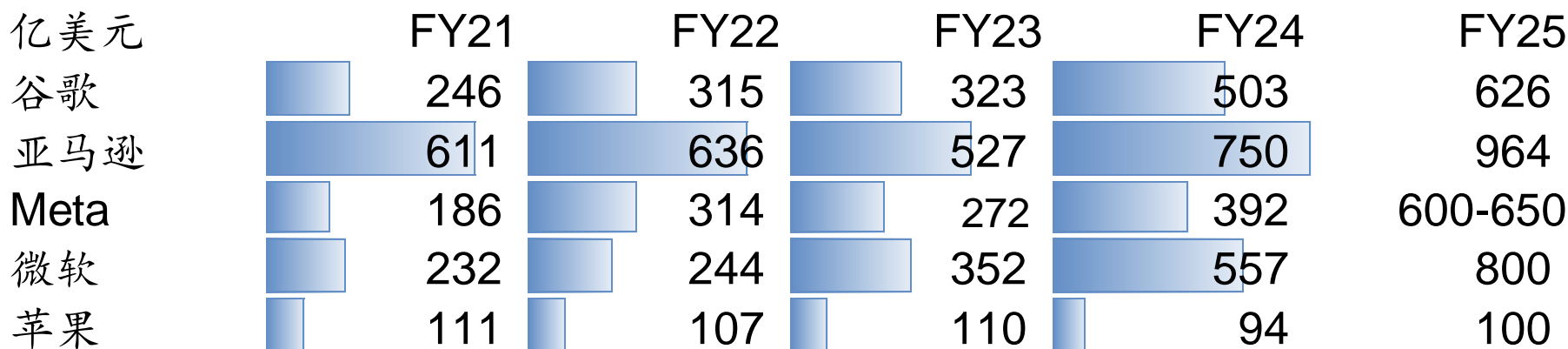
公司名称	高管名称	职务	相关观点
苹果	Tim Cook	CEO	推动效率创新是一件好事，我们将坚持芯片与软件高度集成的路线，以及云端、端侧模型混合调用的方法
微软	Satya Nadella	CEO	DeepSeek展示出“real innovations”，AI成本下降是趋势。在推理方面，我们通常看到每一代硬件的性价比提高2倍以上，每一代模型的性价比提高10倍以上
Meta	Zuckerberg	CEO	DeepSeek在基础设施优化方面所取得的突破性进展，对Meta在AI领域的雄心壮志具有长远的积极影响。这不会对公司的Capex计划产生实质性影响。我们学习他们的经验，并将其应用到我们的工作中
SAP	Christian Klein	CEO	DeepSeek这类高效且平价的大模型有望提升AI普及度，推动SAP AI需求

注：Zuckerberg观点综合自业绩会和Meta全体股东大会
资料来源：各公司官网、中信建投

1、DeepSeek——大厂视角：微软、Meta未下调资本开支预期，对推理需求保持乐观

- 微软和Meta并未下调25财年资本开支计划，我们认为这是对推理侧需求的积极预判。
- ✓ 微软：四季度（FY2Q25）资本开支226亿美元，公司指引未来两个季度与四季度相近，则25财年合计约800亿美元，与1月初的指引相近。CEO纳德拉提到DeepSeek反映了创新周期中的成本下降趋势，有望带来更多AI需求，利好微软这种AI服务提供商；微软未来资本开支会根据需求决定。
- ✓ Meta：公司重申25财年资本支出600-650亿美元，CEO扎克伯格提到重金投入AI基础设施是长期战略。

图表：海外大厂各年度资本开支（亿美元）



注：微软和Meta的25财年数据为公司指引，其余为彭博一致预期；谷歌和亚马逊的24财年数据为彭博一致预期
资料来源：各公司公告、彭博、中信建投

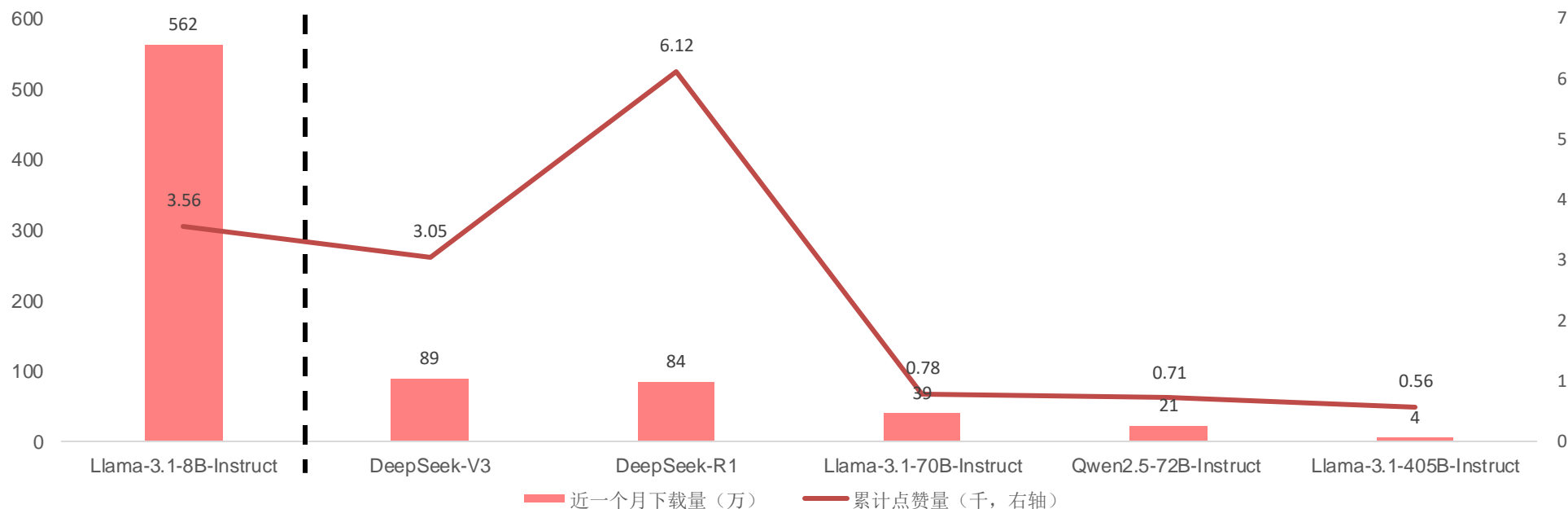
1、DeepSeek——开发者视角：适合本地化部署的大参数高性能模型

- Hugging Face数据反映DeepSeek V3和R1均为适合本地化部署的大参数高性能模型。Hugging Face是全球主流的开源模型社区之一，该社区的模型下载量虽然不能直接反映基于各模型开发的AI应用数量，但也侧面反映了开发者对不同模型的偏好。
- ✓ 小参数模型：性能较弱，但推理算力和存储需求较小，适合本地化微调和部署，因此Llama 3.1 8B等小参数模型下载量断层式领先。据Synaptic的统计，24年6月下载量前10的模型，参数量普遍低于100亿；
- ✓ 中大参数模型：性能较强，但对推理算力、部署设备均有较高要求。V3和R1性能处于第一梯队，同时推理算力需求相对较低，适合本地化部署和蒸馏，近一个月下载量显著高于Llama和通义千问2.5。

图：DeepSeek模型近一个月下载量显著超过Llama3的中大参数版本（根据Hugging Face统计）

小参数模型适合本地化部署
下载量断层式领先

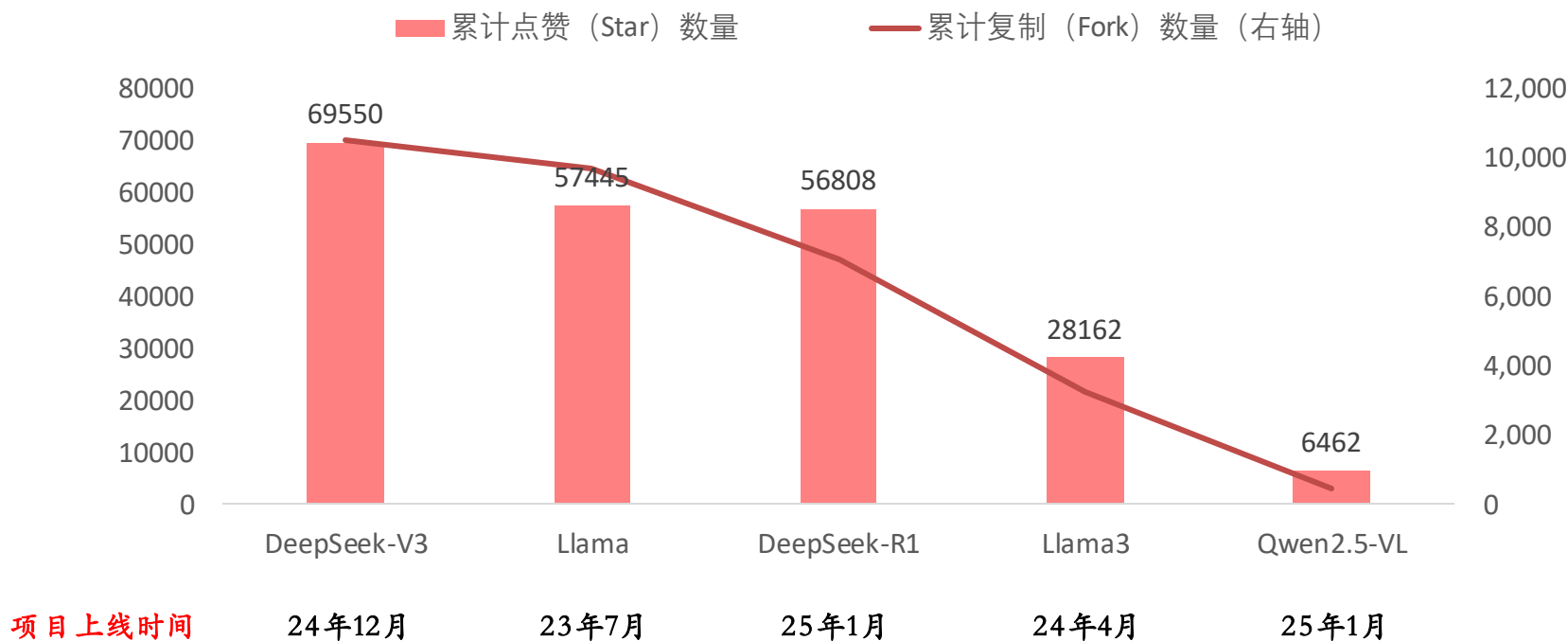
中大参数模型中，DeepSeek近一个月
下载量显著高过Llama和通义千问2.5



1、DeepSeek——开发者视角：适合本地化部署的大参数高性能模型

- GitHub的Stars是项目在社区中受欢迎程度的直接指标，Fork则表示项目累计被用户拷贝的数量，两个指标均代表项目上线至今的关注度和用户喜爱度。DeepSeek V3和R1两个项目上线至今均不足2个月，但它们的累计Star和Fork均与上线时间更早的Llama接近，显著高于24年4月发布的Llama3，直接反映了开发者对DeepSeek开源模型的高认可度。

图：DeepSeek累计关注度高于更早发布的Llama（根据GitHub统计）



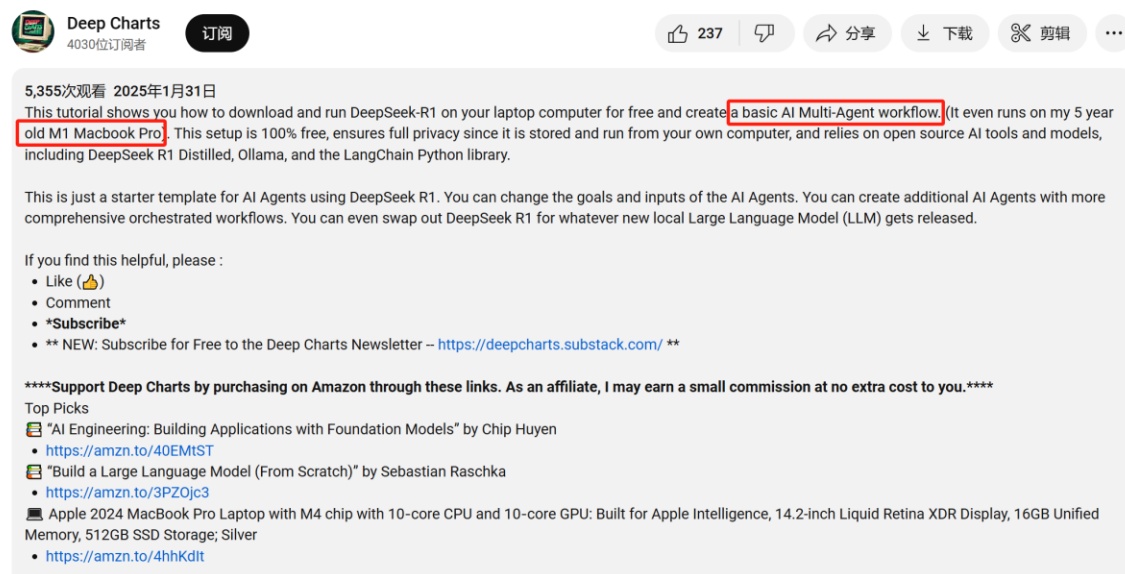
1、DeepSeek——开发者视角：模型运行成本低，降低AI应用开发门槛

- 开发者展示了R1模型进行本地化部署的案例。V3和R1模型均采用稀疏MoE架构，推理过程中激活的参数较少，算力和设备要求也相应较低。我们通过GitHub、Hugging Face两大开发者社区，以及Youtube相关内容梳理本地化部署和调用R1模型的实测体验，Youtube博主Deep Charts和Digital Spaceport的实测结果显示，671B大模型可以在较低成本的服务器中部署（满足特定配置要求），经过微调的80亿参数小模型则可以在普通的M1 Macbook Pro中赋能多Agent workflows。作为对比，Llama 3.1 405B需要在两台H100服务器中运行，年租金合计30万美元以上（SemiAnalysis测算）。

图：DeepSeek开发者实测案例

案例一：经过R1微调的80亿参数小模型可以在个人笔记本中运行，基于它构建多Agent workflows

Easily Create DeepSeek AI Agents on a Laptop for Free (Ollama + Langchain)



5,355次观看 2025年1月31日

This tutorial shows you how to download and run DeepSeek-R1 on your laptop computer for free and create a basic AI Multi-Agent workflow. (It even runs on my 5 year old M1 Macbook Pro). This setup is 100% free, ensures full privacy since it is stored and run from your own computer, and relies on open source AI tools and models, including DeepSeek R1 Distilled, Ollama, and the LangChain Python library.

This is just a starter template for AI Agents using DeepSeek R1. You can change the goals and inputs of the AI Agents. You can create additional AI Agents with more comprehensive orchestrated workflows. You can even swap out DeepSeek R1 for whatever new local Large Language Model (LLM) gets released.

If you find this helpful, please :

- Like (👍)
- Comment
- *Subscribe*
- ** NEW: Subscribe for Free to the Deep Charts Newsletter – <https://deepcharts.substack.com/> **

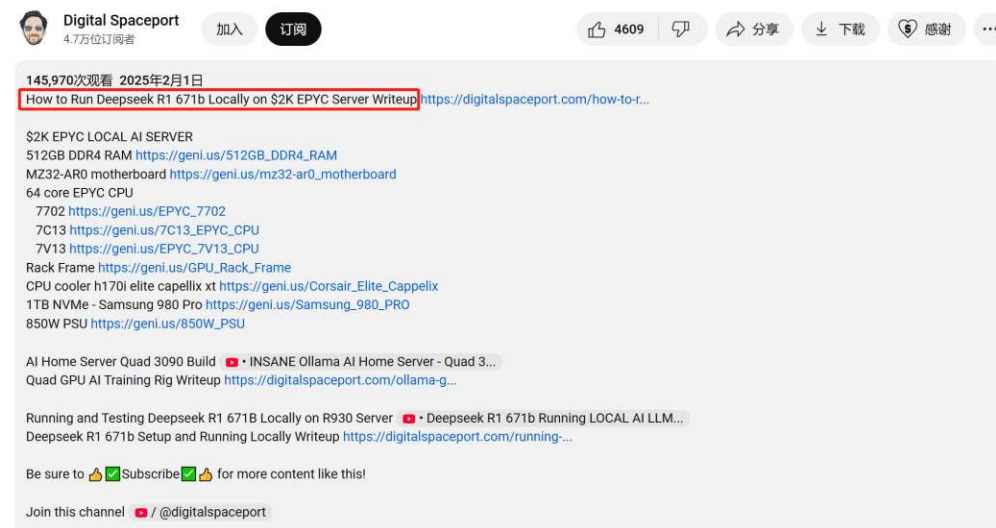
Support Deep Charts by purchasing on Amazon through these links. As an affiliate, I may earn a small commission at no extra cost to you.

Top Picks

- "AI Engineering: Building Applications with Foundation Models" by Chip Huyen
<https://amzn.to/40EMtST>
- "Build a Large Language Model (From Scratch)" by Sebastian Raschka
<https://amzn.to/3PZ0jc3>
- Apple 2024 MacBook Pro Laptop with M4 chip with 10-core CPU and 10-core GPU: Built for Apple Intelligence, 14.2-inch Liquid Retina XDR Display, 16GB Unified Memory, 512GB SSD Storage; Silver
<https://amzn.to/4hhKdlt>

案例二：R1 671B大模型可以在较低成本的服务器中实现本地化运行

Deepseek R1 671b Running and Testing on a \$2000 Local AI Server



145,970次观看 2025年2月1日

How to Run Deepseek R1 671b Locally on \$2K EPYC Server Writeup <https://digitalspaceport.com/how-to-r-...>

\$2K EPYC LOCAL AI SERVER

- 512GB DDR4 RAM https://geni.us/512GB_DDR4_RAM
- MZ32-AR0 motherboard https://geni.us/mz32-ar0_motherboard
- 64 core EPYC CPU
- 7702 https://geni.us/EPYC_7702
- 7C13 https://geni.us/7C13_EPYC_CPU
- 7V13 https://geni.us/EPYC_7V13_CPU
- Rack Frame https://geni.us/GPU_Rack_Frame
- CPU cooler h170i elite capellix xt https://geni.us/Corsair_Elite_Cappellix
- 1TB NVMe - Samsung 980 Pro https://geni.us/Samsung_980_PRO
- 850W PSU https://geni.us/850W_PSU

AI Home Server Quad 3090 Build • INSANE Ollama AI Home Server - Quad 3...
Quad GPU AI Training Rig Writeup <https://digitalspaceport.com/ollama-g-...>

Running and Testing Deepseek R1 671b Locally on R930 Server • Deepseek R1 671b Running LOCAL AI LLM...
Deepseek R1 671b Setup and Running Locally Writeup <https://digitalspaceport.com/running-...>

Be sure to 🍌👍Subscribe👍🍌 for more content like this!

Join this channel <https://t.me/digitalspaceport>

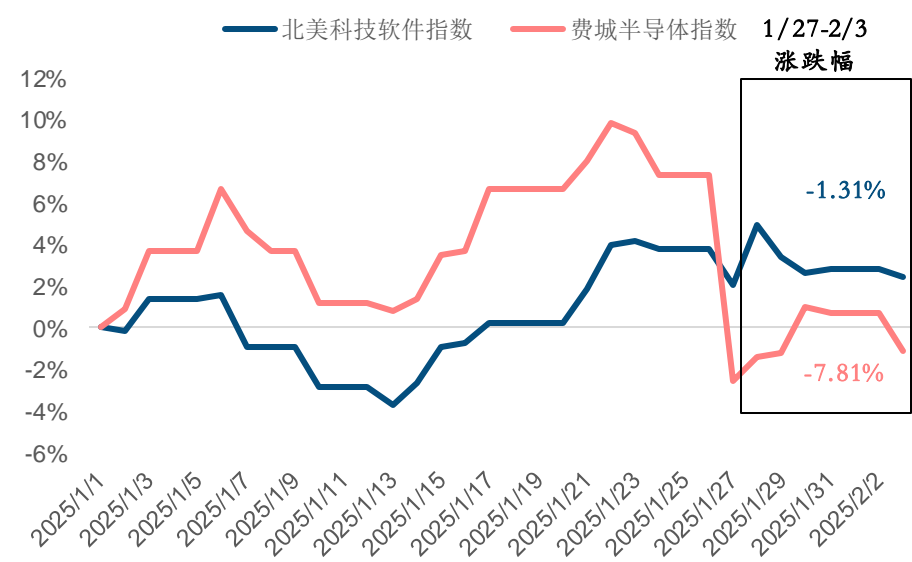
注：上述内容仅展示海外开发者的实测案例，一切以实际运行为准

资料来源：GitHub、Hugging Face、Youtube博主Deep Charts与Digital Spaceport的实测结果、中信建投

1、DeepSeek——行情视角：Meta、苹果等应用与端侧相关公司涨幅靠前

- DeepSeek全球出圈至今，相较半导体等硬件指数，北美科技软件指数仅下跌1.3%，走势相对平稳。DeepSeek驱动API成本下降并提升大模型推理效率，有望加速AI应用商业化落地。
- AI应用公司中，Duolingo、Reddit等垂类公司和Meta大厂涨幅靠前。综合大厂具备丰富的用户和场景积累，高额资本开支构筑护城河。

图表：DeepSeek出圈以来AI硬件和软件行业指数走势分化



注：数据截至2月3日
资料来源：彭博、中信建投

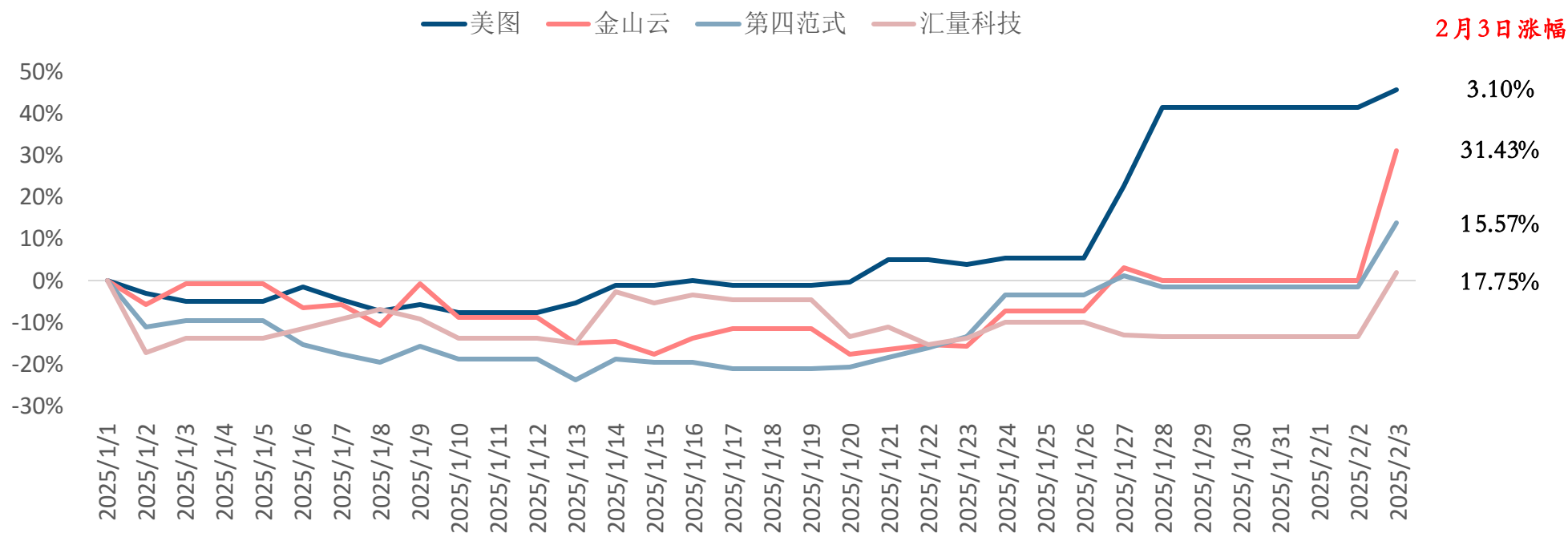
图表：DeepSeek出圈至今美股AI公司涨跌幅

分类	股票代码	公司名称	业务	市值 (亿美元)	24年涨跌幅	25年至今涨跌幅	DeepSeek出圈 (1月27日)至今涨跌幅
综合大厂	META.O	Meta	社交软件	17,671	63%	19%	8%
	AAPL.O	苹果	消费电子	34,252	26%	-9%	2%
	GOOGL.O	谷歌	搜索	24,711	32%	6%	0%
	AMZN.O	亚马逊	电商、公有云	24,965	47%	8%	1%
	MSFT.O	微软	办公、Azure	30,548	12%	-3%	-7%
电商	SHOP.N	Shopify	电商独立站建站工具	1,478	37%	8%	6%
广告	APP.O	Applovin	程序化广告投放	1,226	712%	13%	1%
	TTD.O	TradeDesk	程序化广告投放	582	64%	0%	-2%
教育	DUOL.O	Duolingo	口语学习软件	160	50%	12%	11%
SaaS/服务	NET.N	Cloudflare	网络性能与安全服务软件	470	32%	27%	9%
	ESTC.N	Elastic	B端搜索、数据可视化	116	-9%	13%	7%
	HUBS.N	Hubspot	客户管理SaaS	395	23%	10%	5%
	ZG.O	Zillow	房产估价与房源搜索软件	190	27%	11%	3%
	GDDY.N	GoDaddy	网站设计和建站工具	299	85%	8%	3%
	BILL.N	Bill Holdings	金融自动化软件	99	1%	13%	2%
	INTA.O	Intapp	面向金融、法律的流程自动化软件	55	87%	10%	2%
	CRM.N	Salesforce	CRM软件	3,247	26%	1%	2%
	DDOG.O	Datadog	数据分析和云监控公司	486	22%	0%	2%
	FRSH.O	Freshworks	CRM软件	56	-29%	14%	1%
	AI.N	C3.AI	定制化AI应用	40	29%	-9%	-5%
	NOW.N	ServiceNow	IT、人力等B端软件	2,103	51%	-4%	-9%
社交娱乐	RDDT.N	Reddit	国外版的贴吧	355	292%	24%	11%
	RBLX.N	Roblox	UGC游戏平台	464	33%	22%	5%
	DOCS.N	Doximity	面向医生的社交平台	110	91%	11%	4%
	U.N	Unity	AI游戏引擎+程序化广告	86	-42%	-5%	-4%
数据	PLTR.O	Palantir	数据分析与决策支持	1,908	361%	11%	6%
	SNOW.N	Snowflake	云数据存储和分析	603	-22%	18%	5%
	MDB.O	MongoDB	数据分析	200	-41%	15%	3%
	CFLT.O	Confluent	实时数据分析公司	95	27%	4%	1%
	ZETA.N	Zeta Global Holdings	营销数据分析	43	126%	1%	-2%
	INOD.O	Innodata	数据标注与分析	11	390%	-7%	-9%
通信语音	TWLO.N	Twilio	基于云的通信平台即服务公司	225	21%	36%	8%
	RNG.N	RingCentral	通讯服务公司	31	-1%	-2%	-1%
	SOUN.O	Soundhound	语音生成与识别	53	1301%	-28%	-10%
硬件	AMD.O	AMD	算力	1,854	-18%	-5%	-7%
	AVGO.O	博通	算力	10,206	108%	-6%	-11%
	NVDA.O	英伟达	算力	28,570	169%	-13%	-18%

1、DeepSeek——行情视角：港股AI公司涨幅显著，国产应用有望加速落地

- 国产模型全面赶超海外，市场看好国产应用长期受益。DeepSeek V3和R1模型是国产模型快速赶超海外的典型代表。国产模型追平海外的时间不断缩短，4o发布半年之后，国产豆包、Minimax就发布对标模型；Sora的demo4个月后，国产AI视频可用度与性价比更高；o系列模型发布4个月后，R1等国产o系列模型涌现。国产应用有望受益于国产模型能力跃升，春节后首个交易日港股AI公司涨幅明显，恰恰反映之一市场预期。

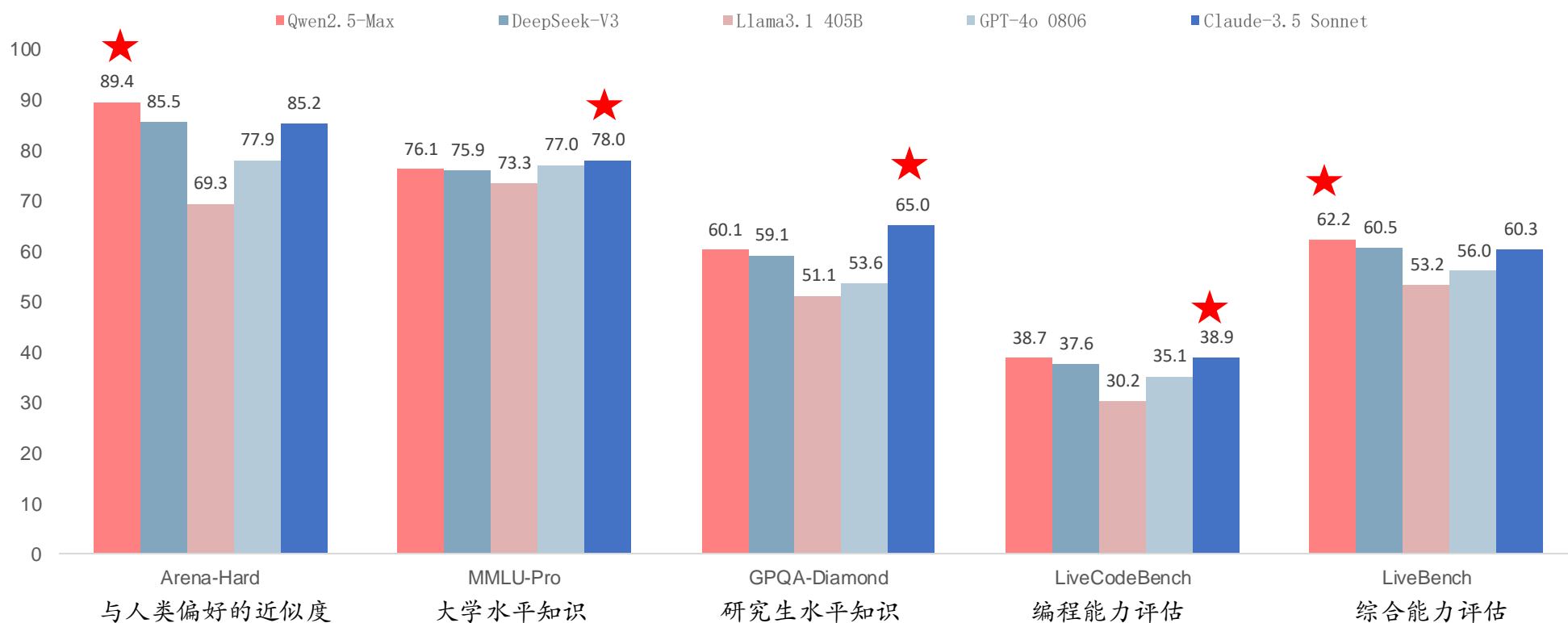
图表：春节后首个交易日港股AI公司涨幅明显



2、阿里：Qwen2.5-Max综合性能对标DeepSeek V3，国产开源模型再添重磅新品

- Qwen Max是通义千问系列效果最好的模型，春节期间更新了2.5版本。新版本同样为超大规模MoE模型，在大学水平知识测试MMLU-Pro、编程能力测试LiveCodeBench，综合能力测试LiveBench中，新版本结果均略高于DeepSeek V3，与GPT-4o的水平接近。当前已推出AI助手和API服务。新版本使用了超过20万亿token的预训练数据，显著超过Llama3的15万亿，是超大规模模型训练的典型案例。

图表：Qwen2.5-Max综合性能对标DeepSeek V3



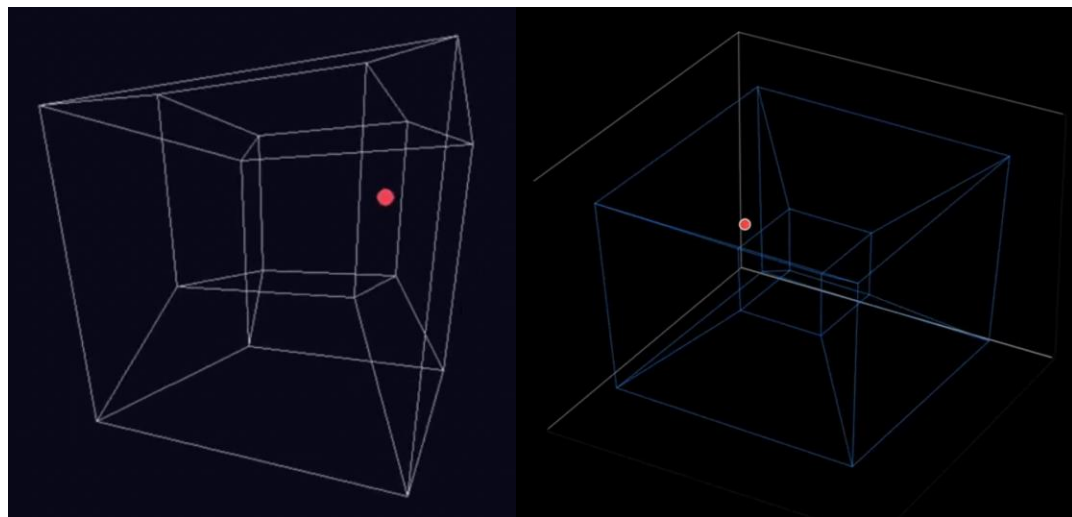
3、OpenAI: o3-mini和Deep Research相继发布，将发力端侧AI硬件

- o3-mini模型上线ChatGPT，物理模拟能力较强。目前Pro会员用户可以无限制使用，其他用户限量体验，API服务也已经同步推出，价格约为o1的1/15、R1的两倍。完整版o3模型将在4-6周内发布，将支持文件附件功能、更长的上下文窗口和记忆功能。
- ✓ 从基准测试看：在综合能力测试LiveBench中，o3-mini整体水平为73.94，介于R1的71.38和o1的75.67之间。o3-mini幻觉率从GPT-4o的52.4%大幅降至14.8%。
- ✓ 从实际体验看：根据University of Washington、Every公司的测试，o3-mini的代码能力和物理模拟效果优于R1，但在部分体验中，视觉艺术效果不及Sonnet 3.5。

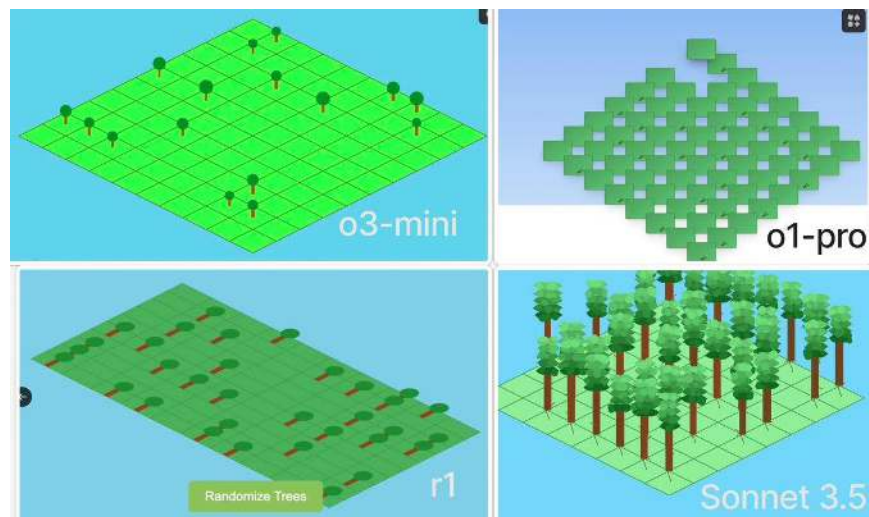
图表：o3-mini的物理模拟效果优于R1
(此处模拟在超立方体内弹跳的球)

o3-mini

DeepSeek R1



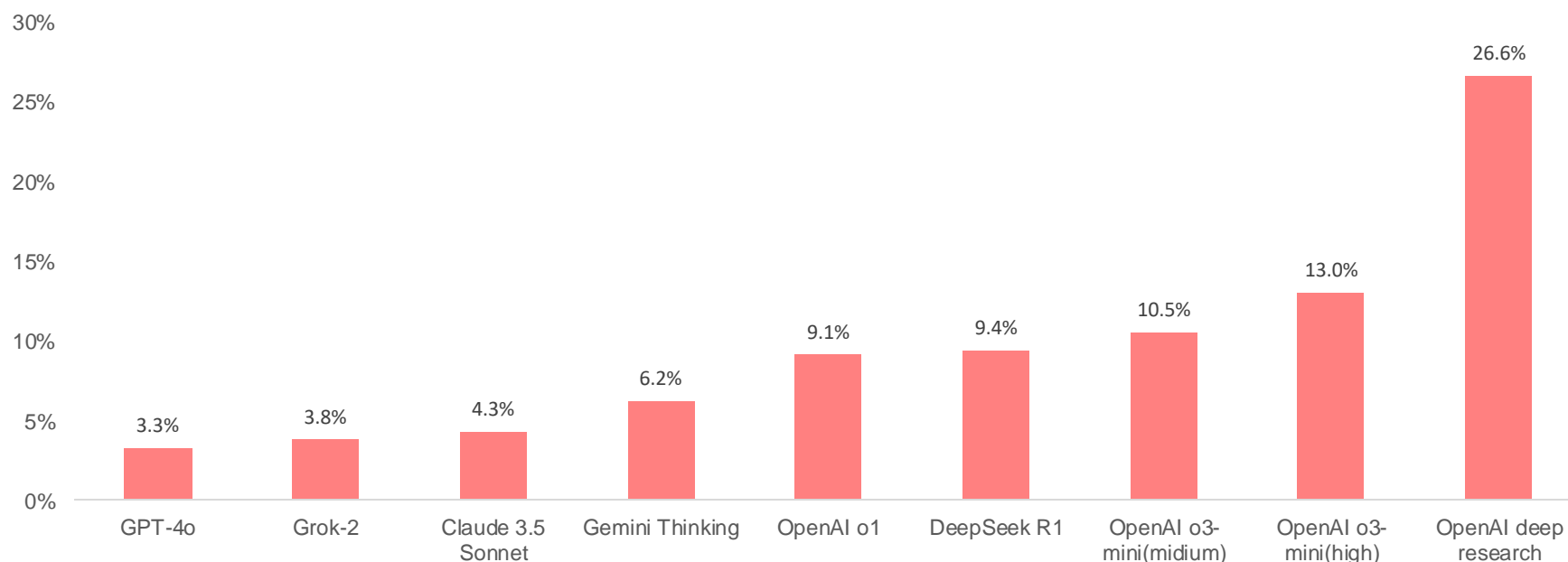
图表：在生成等距森林中，o3-mini优于o1 pro和r1，但视觉效果不及Sonnet 3.5



3、OpenAI: o3-mini和Deep Research相继发布，将发力端侧AI硬件

- Deep Research在短时间内完成深度研究任务，o3模型+联网搜索构成专业研究Agent。该功能在5-30分钟内围绕特定的主题完成深度研究，并输出研究报告。它可以运用推理能力，在互联网上搜索、解读和分析大量文本、图像和PDF文件，转为金融、科学、工程领域从事知识工作、需要深度研究的用户设计。Humanity's Last Exam (HLE) 由3000多道语言学、火箭科学等超100个学科的题目组成，用于评估模型的学术能力，Deep Research准确率为26.6%创新高，显著高于o1和R1的9.1%、9.4%。
- OpenAI近期发布会透露后续计划：1) 将开始开发专用于生成式AI的设备，以取代智能手机，引发自2007年iPhone推出以来的下一轮数字设备革命；2) Plus订阅目前不会涨价，希望在未来可以降价；3) 正在考虑更积极的开源策略，但这并非眼下的首要任务。

图表：Deep Research在HLE测试中的准确率达26.6%创新高



第一章

春节回顾：DeepSeek全球出圈，多方积极响应

第二章

展望：DeepSeek利好哪些AI应用？

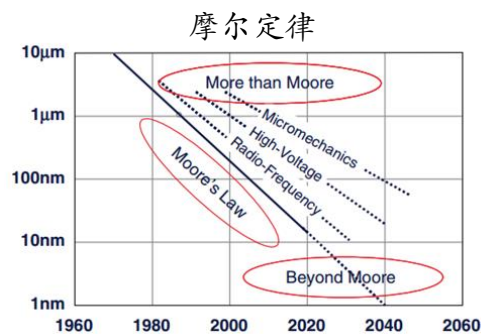
1、技术创新周期总会伴随成本下降趋势，DeepSeek引领AI成本下行

图表：技术创新周期总会伴随成本下降趋势

集成电路

集成电路成本下降，
创造全球超6000亿美元
规模的集成电路产业

摩尔定律：集成电路
上可容纳的晶体管数
目每隔约两年便会增
加一倍，驱动集成电
路成本下降

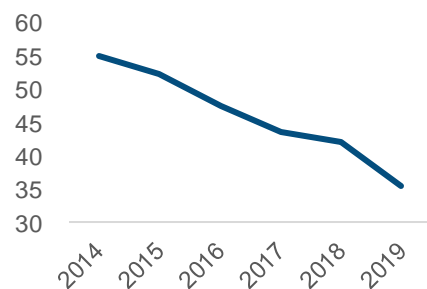


PC互联网

家庭宽带成本下降，
触发PC生态系统

我国固定宽带月户均
支出从14年的55.1元下
降至2019年的35.6元，
催生PC生态系统

我国固定宽带月户均支出（元）

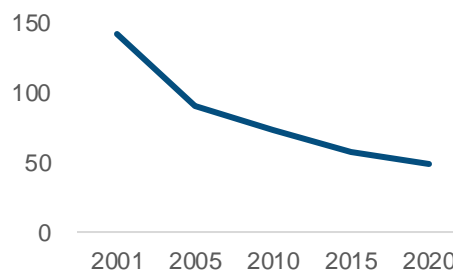


移动互联网

移动流量降费增速，
推动移动互联网繁荣
发展

中国移动ARPU由2001
年的141元/户/月降至
2020年的47元，侧面反
映移动资费下降趋势

中国移动ARPU（元/户/月）

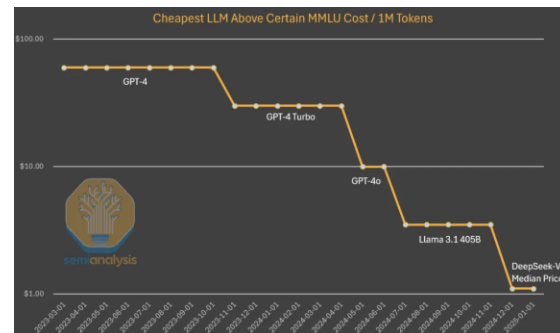


AI

大模型API成本下降，
有望推动应用百花齐
放

从GPT-4到DeepSeek
V3，23年以来头部大
模型API成本已下降超
90%

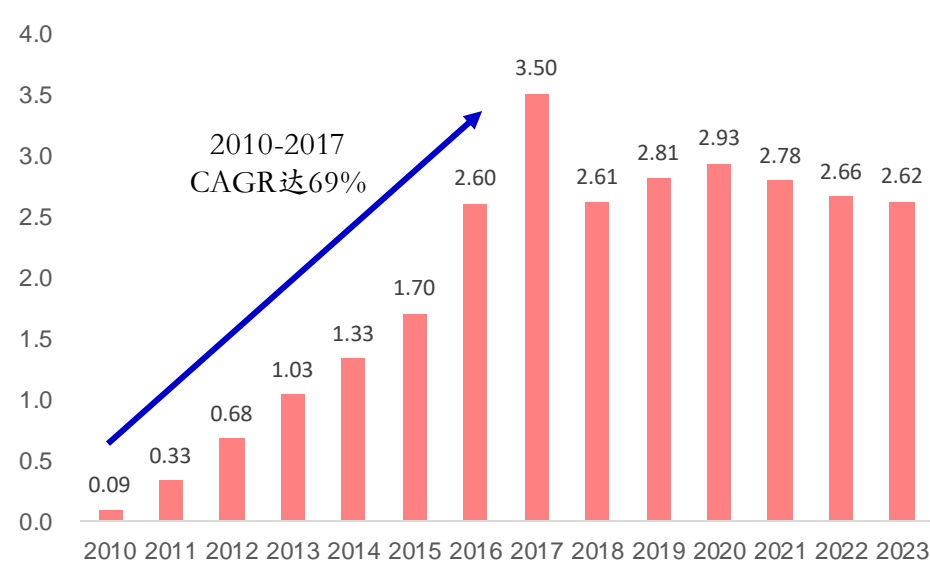
头部模型API价格（美元/百万token）



2、对照移动互联网，AI应用有望迎来“安卓时刻”

- 安卓系统开源成为移动互联网应用爆发的关键一环。
2007年11月Google与84家硬件制造商、软件开发商及电信营运商组建开放手机联盟共同研发改良Android系统，随后发布源代码。安卓系统使用简单易上手的Java语言，驱动基于该系统的应用和手机数量快速增长，Google Play上的应用数量从2010年的9万个快速增值2017年的350万个。
- DeepSeek同样免费开源，便于本地化部署和模型能力迁移，实质性降低AI应用开发成本，意义类似于移动互联网的安卓时刻。

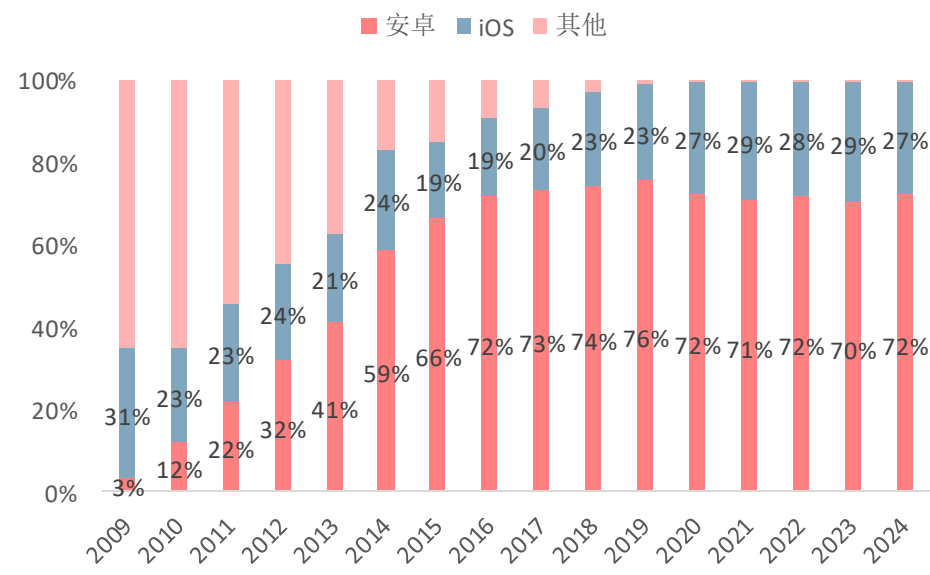
图表：安卓系统开源后Google Play应用数量快速增加（百万）



图表：DeepSeek与安卓具有三大共同点

分类	DeepSeek	安卓
免费开源	V3、R1 开源	2007年谷歌组建开放手机联盟并开源安卓
方便使用	推理算力要求较低，便于本地化部署	基于Java语言，简单易懂
深度定制	可以将R1推理能力迁移至小参数模型	自由度高，各家厂商基于安卓自主开发操作系统

图表：安卓快速成为市占率第一的移动设备操作系统

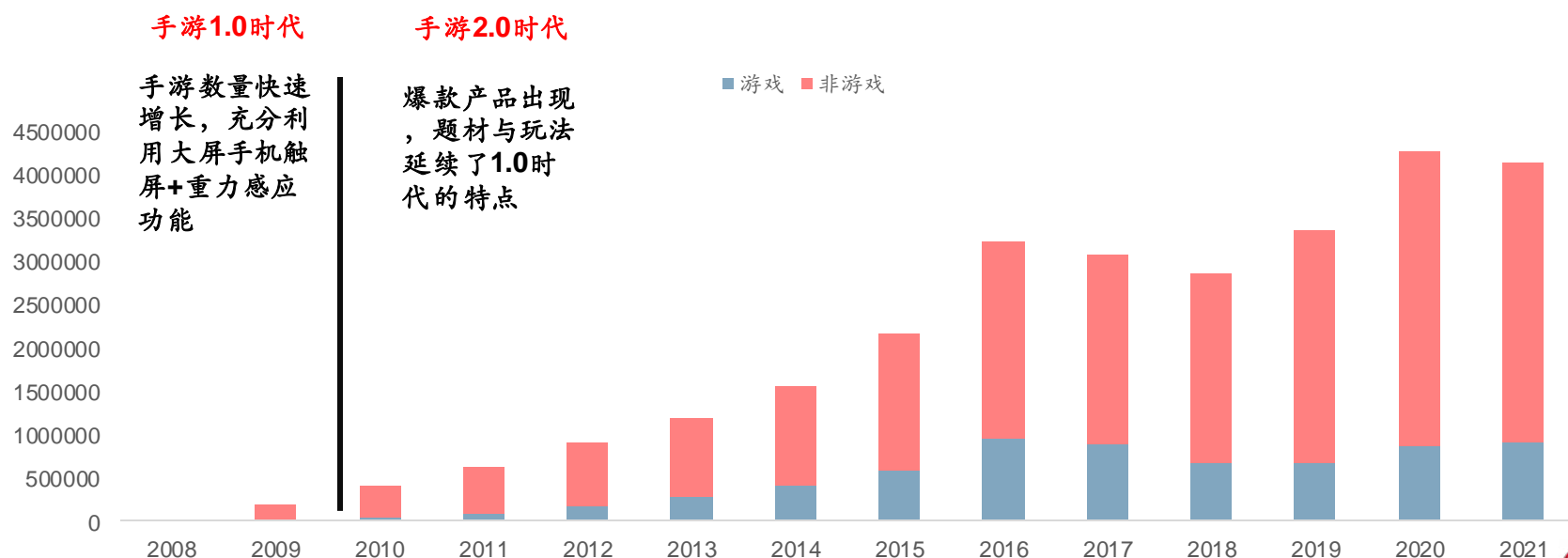


资料来源：AppBrain、Prioridata、Statistia、中信建投

3、对标手游发展初期，看新一代应用爆发的节点

- **手游1.0：2009年手游数量快速增长，但未见破圈产品。** App Store于2008年7月发布，发布首日App数量仅约500个，此后出现爆发性增长，游戏是IOS App的主要类型之一（数量占比约25%），手游时代开启。我们认为2008-2009年是手游1.0时代，这一时期手游数量快速增长，根据AppStore，2008年底App Store上线手游5136个，2009年底该数字变为2.94万个；2010年以后是手游2.0时代，《神庙逃亡》《水果忍者》等爆款产品出现，题材与玩法延续了1.0时代的特点。
- **手游2.0破圈：玩家创新+免费助推破圈。** 以《神庙逃亡》为例，1) 收费模式转变：公司将游戏改为免费下载，增加IAP弥补收入，在App Store免费榜上的排名也逐步提高，于2011年12月28日登上IOS免费榜第一，2012年1月登上IOS畅销榜第一。2) 玩法创新、简便：充分利用触屏和重力感应的功能，创新性地采取玩家控制镜头转动，成为3D跑酷游戏的开创者。

图表：2008-2020全球IOS App数量（单位：个）



4、四大类应用有望受益——端侧AI

- DeepSeek小参数版本端侧表现较好，R1推理能力可以迁移至其他小模型。根据技术报告，DeepSeek-Lite（2.1B）在端侧延迟、内存占用、常见任务精准度表现优于Llama-3（8B-Instruct）和Phi-3-mini（3.8B）。使用R1生成的800K数据，对Qwen和Llama系列的多个小模型(1.5B、7B、8B、14B、32B、70B)进行了微调。经过R1蒸馏的小模型，推理能力显著提升，部分模型的数学测试结果优于o1-mini。
- 看好DeepSeek提升端侧AI能力，除了AI手机、AI电脑以外，兼具实用性与时尚性的AI眼镜也有望进一步完善实时翻译与导航功能。

图表：经过R1蒸馏的小模型，推理能力显著提升

Model	AIME 2024		MATH-500	GPQA Diamond	LiveCode Bench	CodeForces
	pass@1	cons@64				
GPT-4o-0513	9.3	13.4	74.6	49.9	32.9	759
Claude-3.5-Sonnet-1022	16.0	26.7	78.3	65.0	38.9	717
OpenAI-o1-mini	63.6	80.0	90.0	60.0	53.8	1820
QwQ-32B-Preview	50.0	60.0	90.6	54.5	41.9	1316
DeepSeek-R1-Distill-Qwen-1.5B	28.9	52.7	83.9	33.8	16.9	954
DeepSeek-R1-Distill-Qwen-7B	55.5	83.3	92.8	49.1	37.6	1189
DeepSeek-R1-Distill-Qwen-14B	69.7	80.0	93.9	59.1	53.1	1481
DeepSeek-R1-Distill-Qwen-32B	72.6	83.3	94.3	62.1	57.2	1691
DeepSeek-R1-Distill-Llama-8B	50.4	80.0	89.1	49.0	39.6	1205
DeepSeek-R1-Distill-Llama-70B	70.0	86.7	94.5	65.2	57.5	1633

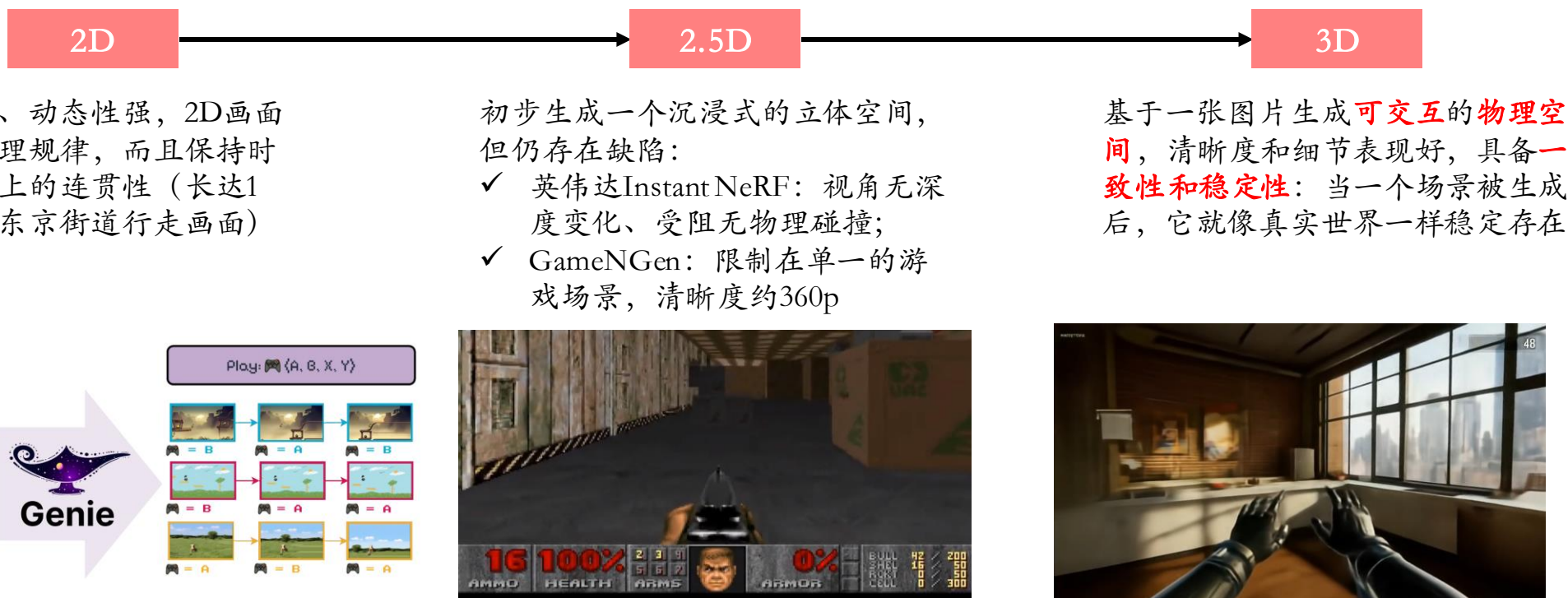
图表：安卓系统开源后Google Play应用数量快速增加（百万）



4、四大类应用有望受益——多模态视频/游戏

- DeepSeek擅长在算法通过大量创新优化降低训练和推理侧的算力需求，从而降低模型使用门槛。例如R1在算法层面引入专家混合模型、多头隐式注意力、多token预测，实现最优效率配置。多模态视频和游戏对算力需求较高，有望吸取DeepSeek的算法创新思路提升效率，DeepSeek也有望从图像向视频模态拓展，革新视频模型、世界模型的算法思路：
- ✓ **AI游戏**：动态开放世界的开发成为可能，支持每帧刷新更大面积的地图变化、剧情多条因果链，但推理效率仍有较大提升空间。
- ✓ **AI视频**：需要低成本、低延迟实现AI短剧的类型化批量生产，动态调整场景/光照自适应等。

图表：世界模型已完成从2D到3D的迭代，算力需求显著提升



24年2月—Deepmind Genie1（可交互）

24年11月—Deepmind AI游戏引擎
GameNGen生成的《毁灭公爵》

24年12月—Deepmind Genie2

4、四大类应用有望受益——情感陪伴/AI玩具

- 目前AI陪伴机器人以调用云端模型为主，DeepSeek小模型在端侧延迟、内存占用等方面表现较好，本地化部署时对算力和内存需求较低，与AI陪伴机器人外形小巧、硬件部署空间有限的情况相匹配。如果AI陪伴机器人本地化部署DeepSeek，语音互动时延性有望下降，语音识别、面部捕捉、人格微调体验有望优化。

图表：两类情感陪伴/AI玩具如果集成DeepSeek，产品体验均有望提升

分类一 儿童向 益智、教育

大模型提供知识问答、口语锻炼等

AI挂件BubblePal



售价：409元
销量：7月上市至今，销售近10万台

字节AI显眼包



不公开发售，
二手平台价格
最高500元

如果集成DeepSeek

口语互动的时延性下降
AI教育拓展至数理学科

除了语音交互，还有
眼神交流、触感反馈
，外形也追求拟真

分类二 成人向 情感陪伴

AI毛绒宠物Ropet



售价：约200美元
销量：已预售近千台

AI高端毛绒宠物LOVOT



售价：3万元以上
销量：已超一万台

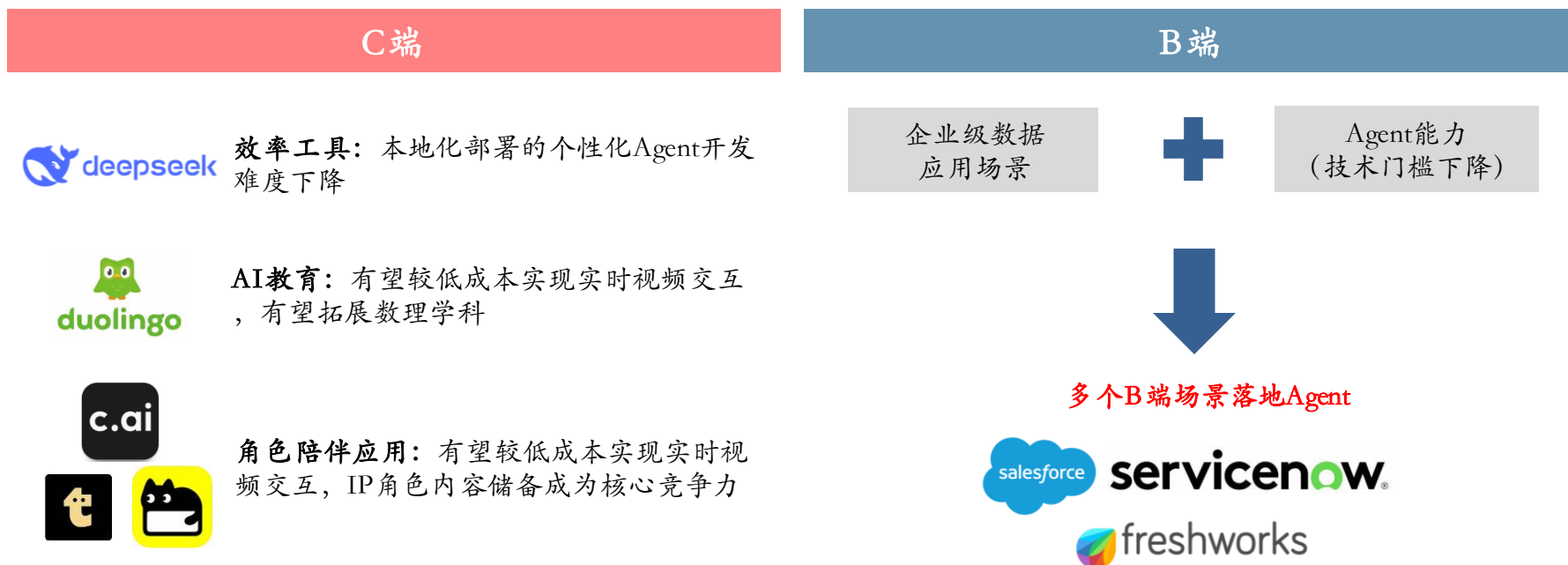
如果集成DeepSeek

AI能力进一步提升，语音识别、面部捕捉、人格微调体验优化

4、四大类应用有望受益——拥有稀缺行业数据/垂类场景

- DeepSeek降低AI应用的技术门槛，数据与场景对AI应用的重要性有望提升。无论是云端调用还是本地化部署，DeepSeek的免费开源、低API成本、高推理效率均能显著降低AI应用的开发成本，推动AI应用在各个场景加速落地。随着AI应用的技术门槛下降，稀缺行业数据和需求刚性的垂类场景，有望成为AI应用的核心竞争要素。
- ✓ C端：本地化部署的个性化Agent开发难度下降，AI教育、角色陪伴应用有望较低成本实现实时视频交互；
- ✓ B端：ERP、OA、RPA等B端SaaS集成大量企业数据， workflow也与企业的实际场景深度融合。将Agent能力与这类SaaS软件结合，能自动化执行复杂任务，提升各个企业场景的工作效率。

图表：AI应用有望在各个场景加速落地



风险提示

- 宏观经济风险，版权保护力度不及预期，知识产权未划分明确的风险，与IP或明星合作中断的风险，大众审美取向发生转变的风险，竞争加剧的风险，用户付费意愿低的风险，消费习惯难以改变的风险，关联公司公司治理风险，内容上线表现不及预期的风险，生成式AI技术发展不及预期的风险，产品研发难度大的风险，产品上线延期的风险，营销买量成本上升风险，人才流失的风险，人力成本上升的风险，政策监管的风险，商业化能力不及预期的风险。

分析师介绍

杨艾莉： 传媒互联网行业首席分析师，中国人民大学传播学硕士，曾任职于百度、新浪，担任商业分析师、战略分析师。2015年起，分别任职于中银国际证券、广发证券，担任传媒与互联网分析师、资深分析师。2019年4月加入中信建投证券研究发展部担任传媒互联网首席分析师。曾荣获2019年wind资讯传播与文化行业金牌分析师第一名；2020年wind资讯传播与文化行业金牌分析师第二名；2020年新浪金麒麟评选传媒行业新锐分析师第二名。

杨晓玮： 传媒互联网行业分析师，上海交通大学金融硕士。

评级说明

投资评级标准		评级	说明
报告中投资建议涉及的评级标准为报告发布日后6个月内的相对市场表现，也即报告发布日后的6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作为基准。A股市场以沪深300指数作为基准；新三板市场以三板成指为基准；香港市场以恒生指数作为基准；美国市场以标普500指数为基准。	股票评级	买入	相对涨幅15%以上
		增持	相对涨幅5%—15%
		中性	相对涨幅-5%—5%之间
		减持	相对跌幅5%—15%
		卖出	相对跌幅15%以上
	行业评级	强于大市	相对涨幅10%以上
		中性	相对涨幅-10-10%之间
		弱于大市	相对跌幅10%以上

分析师声明

本报告署名分析师在此声明：(i) 以勤勉的职业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，结论不受任何第三方的授意或影响。(ii) 本人不曾因，不因，也将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。

法律主体说明

本报告由中信建投证券股份有限公司及/或其附属机构（以下合称“中信建投”）制作，由中信建投证券股份有限公司在中华人民共和国（仅为本报告目的，不包括香港、澳门、台湾）提供。中信建投证券股份有限公司具有中国证监会许可的投资咨询业务资格，本报告署名分析师所持中国证券业协会授予的证券投资咨询执业资格证书编号已披露在报告首页。

在遵守适用的法律法规情况下，本报告亦可能由中信建投（国际）证券有限公司在香港提供。本报告作者所持香港证监会牌照的中央编号已披露在报告首页。

一般性声明

本报告由中信建投制作。发送本报告不构成任何合同或承诺的基础，不因接收者收到本报告而视其为中信建投客户。

本报告的信息均来源于中信建投认为可靠的公开资料，但中信建投对这些信息的准确性及完整性不作任何保证。本报告所载观点、评估和预测仅反映本报告出具日该分析师的判断，该等观点、评估和预测可能在不发出通知的情况下有所变更，亦有可能因使用不同假设和标准或者采用不同分析方法而与中信建投其他部门、人员口头或书面表达的意见不同或相反。本报告所引证券或其他金融工具的过往业绩不代表其未来表现。报告中所含任何具有预测性质的内容皆基于相应的假设条件，而任何假设条件都可能随时发生变化并影响实际投资收益。中信建投不承诺、不保证本报告所含具有预测性质的内容必然得以实现。

本报告内容的全部或部分均不构成投资建议。本报告所包含的观点、建议并未考虑报告接收人在财务状况、投资目的、风险偏好等方面的具体情况，报告接收者应当独立评估本报告所含信息，基于自身投资目标、需求、市场机会、风险及其他因素自主做出决策并自行承担投资风险。中信建投建议所有投资者应就任何潜在投资向其税务、会计或法律顾问咨询。不论报告接收者是否根据本报告做出投资决策，中信建投都不对该等投资决策提供任何形式的担保，亦不以任何形式分享投资收益或者分担投资损失。中信建投不对使用本报告所产生的任何直接或间接损失承担责任。

在法律法规及监管规定允许的范围内，中信建投可能持有并交易本报告中提公司的股份或其他财产权益，也可能在过去12个月、目前或者将来为本报告中所提公司提供或者争取为其提供投资银行、做市交易、财务顾问或其他金融服务。本报告内容真实、准确、完整地反映了署名分析师的观点，分析师的薪酬无论过去、现在或未来都不会直接或间接与其所撰写报告中的具体观点相联系，分析师亦不会因撰写本报告而获取不当利益。

本报告为中信建投所有。未经中信建投事先书面许可，任何机构和/或个人不得以任何形式转发、翻版、复制、发布或引用本报告全部或部分内容，亦不得从未经中信建投书面授权的任何机构、个人或其运营的媒体平台接收、翻版、复制或引用本报告全部或部分内容。版权所有，违者必究。

中信建投证券研究发展部

北京
朝阳区景辉街16号院1号楼18层
电话：(8610) 56135088
联系人：李祉瑶
邮箱：lizhiyao@csc.com.cn

上海
上海浦东新区浦东南路528号南塔
2103室
电话：(8621) 6882-1600
联系人：翁起帆
邮箱：wengqifan@csc.com.cn

深圳
福田区福中三路与鹏程一路交汇处
广电金融中心35楼
电话：(86755) 8252-1369
联系人：曹莹
邮箱：caoying@csc.com.cn

中信建投（国际）

香港
中环交易广场2期18楼
电话：(852) 3465-5600
联系人：刘泓麟
邮箱：charleneliu@csci.hk