

# 人工智能周报（26年第17周）

OpenAI 发布 GPT 5.5, DeepSeek V4 上线

优于大市

## 核心观点

**人工智能动态:** 1) **公司动态:** ①OpenAI 三线并进, GPT-5.5、Codex 大更新、GPT-image-2 相继发布, 全面强化“真实工作”能力; ②Anthropic 发布 Claude Opus 4.7 与 Claude Design, 覆盖编程与设计工具市场; ③DeepSeek 发布 V4, 极致低价冲击全球 AI 定价体系; ④阿里通义千问 Qwen3.6-27B 正式开源; ⑤腾讯双线出击, 混元 3D 世界模型 2.0 与 Hy3 Preview 相继亮相; ⑥字节 Seed 3D 2.0、月之暗面 Kimi K2.6、小米 MiMo-V2.5 密集发布, 国内大模型竞争进入多线并发阶段; ⑦Cursor 以 500 亿美元估值融资, SpaceX 寻求收购; ⑧Google 与 Amazon 双双加码 Anthropic; ⑨DeepSeek 启动首轮外部融资。

2) **底层技术:** ①DeepSeek V4 引入混合注意力与多 Token 预测, 算效大幅跃升; ②Kimi K2.6 大规模多智能体集群架构突破行业上限; ③腾讯混元 Hy3 Preview 重建预训练基础设施, MoE 架构全面升级; ④腾讯混元 3D 世界模型 2.0 实现 3DGS 与 Mesh 双表征融合生成; ⑤字节跳动 Seed 3D 2.0 以两阶段解耦策略攻克 3D 生成精度难题; ⑥阿里 Qwen3.6-27B 验证稠密架构在 Agent 编程场景的参数效率极限; ⑦小米 MiMo-V2.5 实现全模态统一推理, Agent 成本压缩 50%。

3) **行业政策:** ①联合国密集推进全球 AI 治理对话, 诺贝尔奖得主呼吁为 AI 装上“方向盘”; ②国常会部署科技创新重点工作, 强调加强原创性引领性科技攻关。

**投资建议:** 密切关注国产模型和国产芯片。Deepseek 的进一步在架构设计层面“降本增效”, 推动国产模型更普惠实现百万上下文长度, 密切关注国产模型厂商进展。寒武纪、华为昇腾的 Day 0 适配表明, 表明国产芯片在已达到商业可用的成熟度, 密切关注国产算力进展。

**风险提示:** 政策落地不及预期风险, 技术商业化瓶颈风险, 供应链波动风险等。

## 重点公司盈利预测及投资评级

公司代码	公司名称	投资评级	昨收盘 (港元/ 美元)	总市值 (百万港元/ 美元)	EPS		PE	
					2025E	2026E	2025E	2026E
0700.HK	腾讯控股	优于大市	493.4	4,502,570	30.9	35.0	16	14
3690.HK	美团-W	优于大市	82.5	509,079	-1.3	5.1	-66	16
9988.HK	阿里巴巴-SW	优于大市	131.8	2,529,557	3.9	5.4	34	25
9618.HK	京东集团-SW	优于大市	118.5	333,968	10.7	15.1	11	8
PDD.O	拼多多	优于大市	98.0	139,168	22.3	25.9	4	4

资料来源: Wind、国信证券经济研究所预测

## 行业研究 · 行业周报

### 互联网

#### 优于大市 · 维持

**证券分析师: 张伦可** 0755-81982651  
zhanglunke@guosen.com.cn  
S0980521120004

**证券分析师: 陈淑媛** 021-60375431  
chenshuyuan@guosen.com.cn  
S0980524030003

**证券分析师: 王颖婕** 0755-81983057  
wangyingjie1@guosen.com.cn  
S0980525020001

**证券分析师: 刘子谭** liuzitan@guosen.com.cn  
S0980525060001

**证券分析师: 张昊晨**  
zhanghaochen1@guosen.com.cn  
S0980525010001

### 市场走势



资料来源: Wind、国信证券经济研究所整理

### 相关研究报告

《DeepSeek V4 发布点评-百万上下文进入普惠时代, 国产算力成功适配需求爆发将至》——2026-04-26

《人工智能行业专题(16): Agent 驱动全球模型厂 MaaS 收入爆发, 国产模型各有优势》——2026-04-17

《人工智能周报(26年第15周)-腾讯 QClaw 正式开启公测, 小米推出三款面向 Agent 时代的大模型》——2026-04-13

《互联网行业 2026 年 4 月投资策略-美伊冲突加剧算力紧缺, “Tokens” 通胀链再强化》——2026-04-07

《人工智能周报(26年第12周)-腾讯 QClaw 正式开启公测, 小米推出三款面向 Agent 时代的大模型》——2026-03-24

## 内容目录

<b>AI 相关网站流量数据</b> .....	<b>4</b>
(一) 公司动态 .....	6
(二) 底层技术 .....	8
(三) 行业政策 .....	9
(四) 重点事件预告 .....	10
<b>投资建议</b> .....	<b>10</b>
<b>风险提示</b> .....	<b>10</b>

## 图表目录

图1: AI 相关网站周访问量 (百万) .....	4
图2: 第一梯队 AI 相关网站周访问量变动 (百万) .....	5
图3: 第二梯队 AI 相关网站周访问量变动 (百万) .....	5
图4: 第三梯队 AI 相关网站周访问量变动 (百万) .....	6
图5: 新增聊天、教育类 AI 相关网站周访问量变动 (百万) .....	6

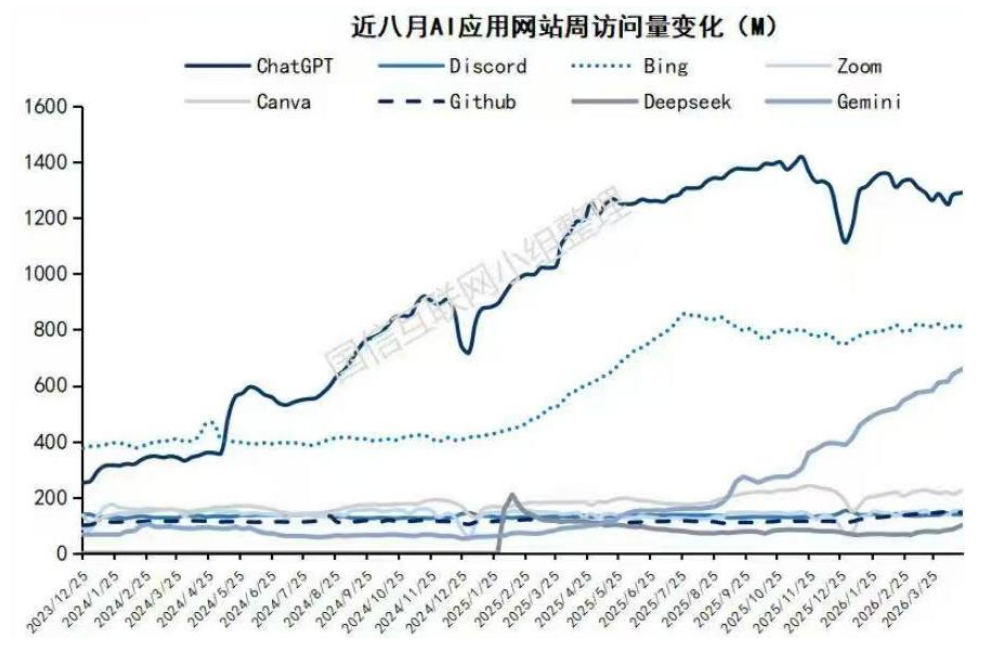
## AI 相关网站流量数据

图1: AI 相关网站周访问量 (百万)

AI相关网站流量数据汇总						
From Apr 16, 2026 to Apr 22, 2026 国信互联网小组						
应用名称	应用说明	3月独立访客 (M)	访客量环比增长	周平均访问 (M)	访问量环比增长	平均访问停留时间
ChatGPT	聊天机器人	476.00	↑ 3.41%	1289.00	↑ 0.55%	0:05:50
Bing	搜索	190.80	↑ 4.66%	809.80	↓ -0.44%	0:07:38
Gemini	聊天机器人	288.00	↑ 13.83%	659.30	↑ 2.90%	0:07:12
Canva	在线设计	223.90	↑ 8.69%	234.60	↑ 3.76%	0:05:53
Claude	聊天机器人	66.99	↑ 80.32%	200.20	↑ 11.35%	0:05:54
Zoom	在线会议	162.90	↑ 7.17%	153.10	↑ 0.99%	0:03:11
Github	代码托管	139.40	↑ 8.57%	145.60	↓ -1.82%	0:06:33
Discord	游戏社区	97.43	↑ 3.86%	138.50	↓ -0.36%	0:11:02
Deepseek	聊天机器人	49.29	↓ -0.14%	90.42	↓ -10.48%	0:05:00
Grok	聊天机器人	42.45	↓ -2.77%	66.89	↓ -3.98%	0:11:02
Character.AI	聊天机器人	8.17	↑ 5.30%	42.46	↓ -0.14%	0:16:13
NotionAI	文本/笔记	23.36	↑ 5.70%	40.51	↓ -1.60%	0:08:04
豆包	聊天机器人	17.09	↑ 8.30%	39.95	↑ 7.05%	0:04:17
Perplexity	聊天机器人	25.11	↑ 3.33%	36.18	↓ -1.26%	0:04:48
Slack	工作效率	13.66	↑ 2.63%	32.41	↑ 3.25%	0:16:07
Quizlet	学习支持	22.58	↓ -1.05%	19.20	↑ 6.25%	0:07:04
夸克	聊天机器人	19.57	↓ -13.86%	15.33	↓ -3.58%	0:04:08
Remove	抠图	23.85	↑ 4.70%	14.73	↓ -1.80%	0:03:05
Grammarly	英语写作批改	27.59	↑ 4.87%	13.33	↓ -2.84%	0:04:31
Duolingo	语言学习	12.42	↑ 0.65%	12.53	↑ 3.30%	0:07:54
Kimi	聊天机器人	8.37	↓ -2.86%	10.51	↑ 10.17%	0:08:26
道义千问	聊天机器人	6.25	↓ -20.51%	10.39	↑ 4.95%	0:03:17
Huggingface	模型社区	8.70	↑ 6.72%	7.02	↓ -0.28%	0:05:27
即梦	视频生成	2.98	↓ -20.53%	4.12	↓ -2.14%	0:06:09
元宝	聊天机器人	3.29	↑ 8.65%	3.96	↓ -5.94%	0:03:09
Kaggle	数据竞赛	4.28	↑ 15.27%	3.09	↑ 0.00%	0:06:30
Midjourney	图像生成	4.13	↑ 5.52%	2.76	↓ -2.82%	0:09:44
Replit	代码支持	4.05	↓ -2.90%	2.69	↓ -10.33%	0:07:15
Poe	聊天机器人	2.39	↓ -1.89%	2.07	↑ 2.48%	0:06:34
Tipranks	股票分析	3.04	↓ -15.13%	1.75	↑ 12.90%	0:01:55
Runway	视频生成	2.91	↑ 13.63%	1.53	↓ -1.92%	0:06:13
Chegg	学习支持	3.21	↓ -8.15%	1.38	↑ 2.22%	0:02:55
Talkie.ai	聊天机器人	1.33	↑ 13.10%	1.33	↑ 3.91%	0:13:54
可灵	视频生成	5.34	↓ -5.18%	0.76	↓ -25.49%	0:03:20
Langchain	AI应用	1.10	↑ 16.28%	0.70	↓ -14.63%	0:05:59
文心一言	聊天机器人	0.75	↓ -4.46%	0.57	↓ -1.72%	0:02:32
海螺	视频生成	0.99	↓ -14.88%	0.52	↓ -1.89%	0:07:52
tablediffusioi	图像生成	1.17	↑ 3.91%	0.42	↓ -4.55%	0:03:43
Pytorch	AI框架	0.78	↑ 7.59%	0.42	↓ -6.67%	0:02:25
Voice.AI	声音编辑	1.12	↑ 9.16%	0.35	↑ 6.06%	0:01:33
YouChat	聊天机器人	0.56	↓ -7.13%	0.33	↑ 10.00%	0:02:04
DreamBox	个性化学习	0.37	↑ 12.80%	0.31	↓ -11.43%	0:03:14
Jasper.AI	营销文案	0.38	↑ 5.56%	0.15	↓ -6.25%	0:01:50
Pi	聊天机器人	0.22	↓ -6.78%	0.13	↑ 8.33%	0:01:21
Stability	AI生成平台	0.39	↑ 7.44%	0.11	↓ -15.38%	0:01:05
Writesonic	文本生成	0.25	↓ -0.40%	0.09	↑ 12.50%	0:01:16
讯飞星火	聊天机器人	0.21	↑ 32.91%	0.07	↓ -12.50%	0:02:32
美图设计室	图像设计	0.11	↑ 22.22%	0.03	↓ -25.00%	0:01:08
Cici	AI聊天	0.05	↓ -13.79%	0.02	↑ 0.00%	0:00:26

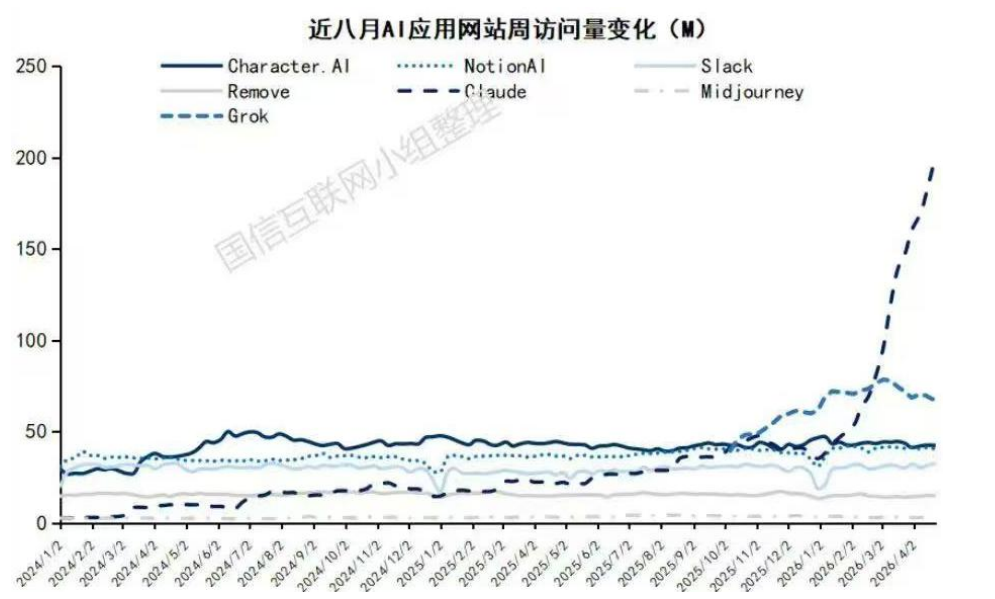
资料来源: Similarweb, 国信证券经济研究所整理

图2: 第一梯队 AI 相关网站周访问量变动 (百万)



资料来源: Similarweb, 国信证券经济研究所整理

图3: 第二梯队 AI 相关网站周访问量变动 (百万)



资料来源: Similarweb, 国信证券经济研究所整理

图4: 第三梯队 AI 相关网站周访问量变动 (百万)



资料来源: Similarweb, 国信证券经济研究所整理

图5: 新增聊天、教育类 AI 相关网站周访问量变动 (百万)



资料来源: Similarweb, 国信证券经济研究所整理

### (一) 公司动态

**OpenAI 发布 GPT-5.5, 定位“面向真实工作的全新智能层级”。**根据 OpenAI 官网, 4月23日, OpenAI 正式发布新一代旗舰模型 GPT-5.5, 在 Terminal Bench 2.0 基准测试中取得 82.7% 的成绩, 较 GPT-5.4 (75.1%) 大幅领先, 同时 Token 消耗更少、重试次数更少、输出质量更高。在测试 AI 独立操作真实电脑的 OSWorld-Verified 中, GPT-5.5 成功率达 78.7%, 超越人类基线; 在跨 44 种职业知识工作的 GDPval 中, 84.9% 的任务达到或超过行业专家水平。官方将其定位为“面向真实工作的全新智能层级”, 也是 OpenAI 迈向全新计算机工作方式的重要一步。

**OpenAI Codex 迎来重大更新, 一次性新增六大能力。**根据 OpenAI 官网, 4月16

日, OpenAI Codex 迎来重大更新, 新增 Computer Use (可在 Mac 后台操作其他应用, 支持测试 iOS 模拟器、复现图形界面 Bug 等场景, 首批限美国 macOS 用户)、内置浏览器 (目前支持本地 localhost 及公开页面预览, 未来将扩展至全网页与登录态操作)、图像生成 (接入 gpt-image-1.5 模型, 可生成风格一致的素材包并自动插入项目)、90+新插件 (覆盖 Atlassian、Slack、CircleCI、GitLab 等主流工具链)、记忆 (记住用户偏好与历史修改, 先向美国用户开放)、自动化 (支持跨天/周安排任务, 可用于 PR 监控、客户反馈追踪等场景) 六大能力, 推动编程智能体向“全流程自主工作台”演进。

**OpenAI 发布 GPT-image-2, Arena 排行榜反超谷歌登顶第一。**根据 OpenAI 官网, 4 月 21 日, GPT-image-2 正式推出, 凭借核心创新“思考”模式——在生成前先进行推理规划而非直接根据提示词出图——以领先 242 分的优势在 Arena 图像排行榜位列第一, 历时四个月完成对谷歌 Nano Banana 2 的反超。新模型同时具备强大的指令遵循与理解能力、卓越的多语言文字渲染能力以及灵活可变的输出尺寸与高分辨率, 标志着 OpenAI 在图像生成领域重新夺回技术制高点。

**Anthropic 发布 Claude Opus 4.7 与 Claude Design, 编程、视觉、设计能力全面升级。**根据 Anthropic 官网, 4 月 6 日, Anthropic 发布 Claude Opus 4.7, 编程能力大幅提升, 支持长程、自主编程, 可独立完成高难度任务, 减少无意义包装函数和兜底脚手架, 能自我发现问题并修正; 视觉能力同步翻倍升级, 图片长边支持 2576 像素 (约 3.75 兆像素), 是此前 Claude 模型的三倍多, 无需手动压缩, 模型层直接支持。新 tokenizer 使同文本 Token 数增至旧版的 1.0 至 1.35 倍。此外, 4 月 17 日, Anthropic 旗下孵化器 Anthropic Labs 同步发布 Claude Design, 基于 Opus 4.7 模型, 支持通过自然语言输入快速生成可交互原型、文档、演示文稿等, 核心特点包括一句话生成原型、智能调节与自定义、多场景覆盖, 进一步拓展 Claude 在创意与生产力工具领域的应用边界。

**DeepSeek 发布 V4, 以极低定价重塑全球 AI 价格格局。**根据 Fortune 报道, 4 月 24 日, DeepSeek 正式发布 V4-Pro 与 V4-Flash 两款旗舰模型, 在 Agent 能力、世界知识和推理性能 (数学、STEM、竞赛代码等) 上均达到或接近 GPT-5.4、Claude Opus 4.6 等世界顶级闭源模型水平。V4-Pro 每百万输出 Token 定价仅 3.48 美元, 全面低于 Gemini 3.1 Pro、GPT-5.5、Claude Opus 4.7 及 GPT-5.4 等竞品; V4-Flash 则以 0.28 美元的超低价格下探至轻量模型价位带。在训练层面, DeepSeek 此次与华为深度合作, 采用华为昇腾 950 芯片及“超节点”集群技术, 被业界视为中国 AI 摆脱对英伟达依赖的重要里程碑。

**阿里通义千问 Qwen3.6-27B 正式开源, “以小博大”震撼业界。**根据阿里云开发者社区及 IT 之家, 4 月 22 日, 阿里云通义千问团队宣布开源拥有 270 亿参数的稠密多模态模型 Qwen3.6-27B, 支持多模态思考与非思考双模式, 在智能体编程方面达到旗舰级表现, 全面超越了前代开源旗舰 Qwen3.5-397B-A17B——后者为总参数 3970 亿、激活参数 170 亿的 MoE 模型。开源权重已同步发布于 Hugging Face 和 ModelScope 平台, 开发者可无缝集成至 Claude Code、Qwen Code 等主流编程助手工具链。

**腾讯双线出击, 混元 3D 世界模型 2.0 与 Hy3 Preview 相继亮相。**根据腾讯云公众号及腾讯混元公众号, 4 月 16 日, 腾讯混元发布 3D 世界模型 2.0, 仅需输入文字描述或上传 2D 图片, 即可生成兼具 3D 高斯泼溅 (3DGS) 与 Mesh 表征的真实 3D 资产, 可直接导入 Unity、UE 等主流游戏引擎进行二次编辑, 用于快速生成游戏地图、关卡原型或数字孪生场景。4 月 23 日, 腾讯正式发布 Hy3 preview 大模型。该模型为混元团队经历重组、架构重构及由首席 AI 科学家姚顺雨牵头关键团队后推出的首份成果。总参数 295B、激活参数 21B, 支持 256K 上下文, 在复杂推理、代码与智能体、上下文学习等能力维度均实现显著提升, 是混元迄今最具智

能的模型。

**字节跳动 Seed 3D 2.0 发布，3D 内容生成迈向“生产可用”里程碑。**根据字节跳动 Seed 公众号，4月23日，字节跳动 Seed 团队发布新一代 3D 生成大模型 Seed 3D 2.0，旨在解决当前 3D 内容生成在几何精度、材质真实性方面难以达到生产级要求的问题。核心创新包括两阶段（Coarse-to-Fine）生成策略，将“整体结构”与“几何细节”解耦优化；以及将纹理生成从级联式模型简化为统一的 PBR 生成模型，支持任意光照下的一致视觉效果，技术报告已同步公开，API 已上线火山引擎。

**月之暗面 Kimi K2.6 发布并开源，长程 Agent 能力突破 13 小时连续运行。**根据月之暗面 Kimi 公众号，4月20日，Kimi K2.6 正式发布并开源，长程编码能力突出，可不间断运行 13 小时，编写或修改超过 4000 行代码，完成复杂系统开发。Agent 集群架构同步升级，最多支持 300 个子 Agent 并行完成 4000 个协作步骤，实现大规模任务并行处理。自主 Agent 与 OpenClaw、Hermes Agent 等框架协同良好，支持长达 5 天的持续自主运行，在监控、运维等后台任务中表现出色。

**小米 MiMo-V2.5 发布，原生全模态 Agent 模型加速开源生态布局。**根据小米技术公众号，4月23日，小米发布 MiMo-V2.5 系列，专为长程、高难度任务设计，在通用 Agent 能力和复杂软件工程方面宣称可与 Claude Opus 4.6、GPT-5.4 较量；MiMo-V2.5 定位原生全模态 Agent 模型，支持代码、文本、图像、视频、语音全模态，平均推理速度达 100 至 150 tokens/s，API 成本较 V2-Pro 降低约 50%，支持 1M 上下文，覆盖绝大多数通用 Agent 场景。

**Cursor 以 500 亿美元估值寻求融资 20 亿美元，同期 SpaceX 寻求收购。**根据 TechCrunch 报道，AI 编程工具 Cursor 正以 500 亿美元估值寻求融资 20 亿美元，现有投资者 Thrive 和 a16z（Andreessen Horowitz）预计将领投；预计 2026 年底 ARR 将超过 60 亿美元。Cursor 此前长期处于负毛利率状态，知情人士透露，去年 11 月推出的自研 Composer 模型以及引入 Kimi 等低调用成本模型，帮助公司实现了略微盈利，目前企业端已实现正毛利率，个人开发者账户仍处于亏损。与此同时，SpaceX 宣布将在今年晚些时候拥有以 600 亿美元收购 Cursor 的选择权，或支付 100 亿美元作为双方合作成果的对价，交易采用二选一机制，被视为 SpaceX 从航天硬件向 AI 生态延伸的关键一步。

**Google 与 Amazon 双双加码 Anthropic，总承诺投资额合计高达 650 亿美元。**根据 Anthropic 官方公告及彭博社报道，4月20日，Amazon 率先宣布向 Anthropic 追加投资最高 250 亿美元，其中 50 亿美元即时到位，余下 200 亿美元与商业里程碑挂钩；Anthropic 同步承诺未来 10 年在 AWS 上花费超 1000 亿美元，用于采购 Trainium 系列定制 AI 芯片。4月24日，Google 进一步宣布投资最高 400 亿美元，以 3500 亿美元估值先行注入 100 亿美元现金扩充算力，达成特定绩效目标后再追加 300 亿美元。两大科技巨头密集加码的背后，是 Anthropic 惊人的商业化速度——其年化营收（ARR）已在 4 月突破 300 亿美元，较 2025 年底的约 90 亿美元大幅增长，首次超越 OpenAI 的 250 亿美元，跃居全球 AI 企业营收第一。

**DeepSeek 启动创立以来首轮外部融资，腾讯阿里争相入股。**根据 IndexBox 报道，4月22日，腾讯和阿里巴巴正就参与 DeepSeek 首轮融资展开洽谈，腾讯提议收购最多 20% 股权，但 DeepSeek 不愿出让如此大比例的控制权；目标融资额为至少 3 亿美元，估值超 200 亿美元，谈判仍在进行中。此次开放融资的直接触发因素是人才流失压力：R1 模型成功后已有至少 5 名核心研发成员相继离职，包括 V3 模型主要贡献者罗福利（加入小米）、GRPO 算法核心作者郭达雅（加入字节跳动）、前核心多模态研究员阮冲（加入自动驾驶公司元戎启行）等。

## （二）底层技术

**DeepSeek V4 引入混合注意力与多 Token 预测，算效大幅跃升。**根据阿里云开发者社区，DeepSeek V4 在架构层面引入“压缩稀疏注意力”（GSA）与稀疏注意力（DSA）相结合的全新注意力机制，在 Token 维度进行压缩，实现了全球领先的 1M 长上下文处理能力，并大幅降低了对计算和内存的需求；同时启用深度为 1 的多 Token 预测策略。相比 V3，V4-Pro 单 Token 推理浮点运算仅为 V3 的 27%，KV 缓存大小仅为其 10%，算力效率实现跨代提升。官方宣布，1M 上下文将成为 DeepSeek 所有官方服务的标配。

**Kimi K2.6 大规模多智能体集群架构突破行业上限。**根据月之暗面 Kimi 公众号，Kimi K2.6 在 Agent 集群架构设计上实现关键突破，最多支持 300 个子 Agent 并行协作，可完成多达 4000 个协作步骤；长程任务方面，单次不间断运行可达 13 小时，编写或修改代码量超过 4000 行，支持长达 5 天的持续自主运行。该架构已与 OpenClaw、Hermes Agent 等主流框架完成深度适配，为后台监控、自动化运维等超长周期任务提供了工程化落地路径。

**腾讯混元 Hy3 Preview 重建预训练基础设施，MoE 架构全面升级。**根据腾讯官网，腾讯混元团队 Hy3 Preview 是在 2026 年 2 月重建预训练和强化学习基础设施后推出的首个成果，采用总参数 295B、激活参数 21B 的混合专家（MoE）架构，支持 256K 上下文窗口。相比前代，Hy3 在复杂推理、指令遵循、上下文学习、代码与 Agent、自然对话与用户理解等多个维度均实现显著提升，并已正式开源。

**腾讯混元 3D 世界模型 2.0 实现 3DGS 与 Mesh 双表征融合生成。**根据腾讯混元公众号，腾讯混元 3D 世界模型 2.0 在底层生成架构上实现关键创新：同时输出 3D 高斯泼溅（3DGS）与 Mesh 两种表征形式，前者适合实时渲染与视觉效果，后者便于游戏引擎导入与二次编辑。两种表征的统一生成，打通了从“AI 生成内容”到“工程可用资产”的全链路，使得 AI 生成的 3D 内容可直接进入 Unity、UE 等主流商业游戏引擎的生产流程。

**字节跳动 Seed 3D 2.0 以两阶段解耦策略攻克 3D 生成精度难题。**根据字节跳动 Seed 公众号，Seed 3D 2.0 在底层技术上引入两阶段 Coarse-to-Fine 生成策略，将“整体结构建模”与“几何细节优化”解耦处理，有效解决以往端到端方案中结构精度与细节质量难以兼顾的矛盾；纹理侧则将此前的级联式模型架构简化为统一 PBR 材质生成模型，直接联合建模完整的物理渲染贴图，实现任意光照条件下视觉效果的一致性，推动 3D 内容生成从“展示级”迈向“生产级”。

**阿里 Qwen3.6-27B 验证稠密架构在 Agent 编程场景的参数效率极限。**根据阿里云开发者社区，Qwen3.6-27B 以 270 亿参数的稠密架构，在 SWE-bench Verified 编程基准中取得 77.2 分，全面超越总参数量 3970 亿的前代 MoE 旗舰 Qwen3.5-397B-A17B，证明在充分优化的训练策略下，稠密模型无需 MoE 路由即可达到旗舰级编程辅助能力，大幅降低开发者部署门槛，单 GPU 即可完整运行。

**小米 MiMo-V2.5 实现全模态统一推理，Agent 成本压缩 50%。**根据小米技术公众号，MiMo-V2.5 在模型架构层面实现代码、文本、图像、视频、语音五种模态的原生统一推理，无需多模型切换；平均推理速度达 100 至 150 tokens/s，API 调用成本较上一代 MiMo-V2-Pro 降低约 50%，同时保持 1M 上下文窗口，在多项 Agent 能力评测中的表现超越上一代 MiMo-V2-Pro，视频与音频理解能力超越 MiMo-V2-Omni，为轻量化多模态 Agent 的规模化落地提供了更具性价比的技术路线。

### （三）行业政策

**联合国密集推进全球 AI 治理对话，诺贝尔奖得主呼吁为 AI 装上“方向盘”。**根据联合国新闻（UN News）报道，4 月 21 日，联合国在日内瓦举办 2026 数字世界

大会，聚焦 AI 治理、数字普惠与全球合作。AI“教父”、诺贝尔奖得主 Geoffrey Hinton 将当前 AI 发展比作“一辆没有方向盘、没有刹车的快车”，呼吁建立切实可行的监管机制。联合国独立国际 AI 科学小组于 4 月 23 日在马德里召开首次线下会议，由诺贝尔和平奖得主 Maria Ressa 与 Yoshua Bengio 共同主持，研究成果将为 2026 年 7 月联合国全球 AI 治理对话提供支撑。

**国常会部署科技创新，强调加强原创性引领性科技攻关。**据新华社报道，4 月 24 日，国务院总理李强主持召开国务院常务会议，研究部署科技创新重点工作。会议指出，要紧盯全球科技前沿动向，围绕国家战略需求与产业发展需要，加强原创性引领性科技攻关，提升基础研究水平，打造具有全球影响力的创新高地。会议强调，要完善科技创新体制机制，强化企业科技创新主体地位，推动科技与产业深度融合，为高质量发展提供有力支撑。

#### （四）重点事件预告

- 5 月 13 日-14 日：百度 Create 2026 AI 开发者大会
- 5 月 19 日-20 日：Fortune 职场创新峰会
- 5 月：Google I/O 2026 开发者大会
- 5 月：Microsoft Build 2026 开发者大会
- 6 月 10 日-11 日：SuperAI 2026
- 7 月 15 日：《人工智能拟人化互动服务管理暂行办法》正式施行

## 投资建议

**密切关注国产模型和国产芯片。**Deepseek 的进一步在架构设计层面“降本增效”，推动国产模型更普惠实现百万上下文长度，密切关注国产模型厂商进展。寒武纪、华为昇腾的 Day 0 适配表明，表明国产芯片在已达到商业可用的成熟度，密切关注国产算力进展。

## 风险提示

政策落地不及预期风险，技术商业化瓶颈风险，供应链波动风险等。

## 免责声明

### 分析师声明

作者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，力求独立、客观、公正，结论不受任何第三方的授意或影响；作者在过去、现在或未来未就其研究报告所提供的具体建议或所表述的意见直接或间接收取任何报酬，特此声明。

### 国信证券投资评级

投资评级标准	类别	级别	说明
报告中投资建议所涉及的评级（如有）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 6 到 12 个月内的相对市场表现，也即报告发布日后的 6 到 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作为基准。A 股市场以沪深 300 指数（000300.SH）作为基准；新三板市场以三板成指（899001.CSI）为基准；香港市场以恒生指数（HSI.HI）作为基准；美国市场以标普 500 指数（SPX.GI）或纳斯达克指数（IXIC.GI）为基准。	股票 投资评级	优于大市	股价表现优于市场代表性指数 10%以上
		中性	股价表现介于市场代表性指数 $\pm 10\%$ 之间
		弱于大市	股价表现弱于市场代表性指数 10%以上
		无评级	股价与市场代表性指数相比无明确观点
	行业 投资评级	优于大市	行业指数表现优于市场代表性指数 10%以上
		中性	行业指数表现介于市场代表性指数 $\pm 10\%$ 之间
		弱于大市	行业指数表现弱于市场代表性指数 10%以上

### 重要声明

本报告由国信证券股份有限公司（已具备中国证监会许可的证券投资咨询业务资格）制作；报告版权归国信证券股份有限公司

关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以我公司向客户发布的本报告完整版本为准。

本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。我公司不保证本报告所含信息及资料处于最新状态；我公司可能随时补充、更新和修订有关信息及资料，投资者应当自行关注相关更新和修订内容。我公司或关联机构可能会持有本报告中所提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或金融产品等相关服务。本公司的资产管理部门、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意见或建议不一致的投资决策。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，我公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

### 证券投资咨询业务的说明

本公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询，是指从事证券投资咨询业务的机构及其投资咨询人员以下列形式为证券投资人或者客户提供证券投资分析、预测或者建议等直接或者间接有偿咨询服务的活动：接受投资人或者客户委托，提供证券投资咨询服务；举办有关证券投资咨询的讲座、报告会、分析会等；在报刊上发表证券投资咨询的文章、评论、报告，以及通过电台、电视台等公众传播媒体提供证券投资咨询服务；通过电话、传真、电脑网络等电信设备系统，提供证券投资咨询服务；中国证监会认定的其他形式。

发布证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

## 国信证券经济研究所

### 深圳

深圳市福田区福华一路 125 号国信金融大厦 36 层  
邮编：518046 总机：0755-82130833

### 上海

上海浦东民生路 1199 弄证大五道口广场 1 号楼 12 层  
邮编：200135

### 北京

北京西城区金融大街兴盛街 6 号国信证券 9 层  
邮编：100032