下载日志已记录，仅供内部参考

西南证券
SOUTHWEST SECURITIES

传媒行业2026年Q4投资策略

Agent迭代成熟，看好AI影视和AI营销

西南证券研究院
传媒研究团队
2026年9月

下载日志已记录，仅供内部参考

核心观点

- **AI影视：1）AI短剧，**2026年中国微短剧及AI剧漫剧市场规模预计突破1210亿元（其中AI剧漫剧超400亿元），用户规模达8.51亿，海外上半年双端内购约12.7亿美元，头部平台DramaWave 7月新剧集超90%由AI生成、下载环比增长39.82%。红果免费短剧MAU达3.57亿，超越多个长视频平台。 2026年9月，微信短剧内容中心推出免费短剧小程序“绿泡泡短剧”，通过免费观看、广告变现的模式，聚合微信生态内的短剧内容。 **2）中长剧：**芒果超媒制作国内首部上星黄金档全 AI 长剧《后西游记》8 月 31 日开播，首播当晚实时收视峰值0.3384%，市占率 1.7862%，拿下省级卫视同时段收视第一。AI制作成本和周期大幅缩短：制作周期4个月，成本大幅缩减，虚拟拍摄、搭景、数字人、特效等环节均大幅节省成本。**3）AI电影：**首部AI超写实院线电影《三星堆：未来往事》正式定档2026年10月23日，《计划外：广寒行动》，已于2026年4月通过国家电影局立项备案，目前已开机，进入制作阶段。**4）互动剧：**9月29日广电总局公示首部互动剧《恐龙快打·华夏秘境》备案信息，代表互动剧正式进入规范化发展阶段。

- **AI营销：**AI大模型已深度渗透营销领域，AIGC实现文案、图像、视频的秒级批量生产，营销智能体自主完成调研、策略、投放及客服互动。2026年上半年，AI营销Agent赛道进入规模化商业兑现阶段，头部厂商产品架构从“单点工具”升级为“多Agent协同系统”。迈富时推出AI-Agentforce 3.0智能体中台，集成500+智能体应用，覆盖营销、销售、服务全链路；趣致集团发布“全感物理AI智能体”体系，将线下终端进化为具备环境感知与内容生成能力的端侧智能体。汇量科技关注后续催化，AI infra进入冲刺阶段，预计10月上线。新架构GPU的使用比例显著提升，且用于模型训练。原先基于CPU的训练架构数据底无法训练复杂用户行为数据，新架构加入时序特征利于训练复杂模型，且能大幅缩短IAP模型数据积累时间。

- **办公AI Agent：**2026年国内桌面端 AI 原生办公智能体行业步入爆发式规模扩张阶段，头部大厂凭借流量和大模型优势初步跑出，**1）腾讯系：**易观发布的2026Q2报告显示，6 月 WorkBuddy月访问量达2097万次，位居行业第一；**2）阿里系：**2026年7月27日，阿里将QoderWork、MuleRun、悟空三款产品全面整合，推出全新Agent办公产品“千问办公”，千问办公上线满一个月，用户数突破3000万，企业用户占比超过一半。**3）字节系：**整合 AI Agent 工具“扣子”和 AI Coding 工具 “Trae”的产品团队并入了“豆包”体系，企业办公协作工具“飞书”也并入了“豆包”团队。

- **游戏：**版号发放延续增长态势，1-7月累计发放国产网络游戏版号1110个、进口版号36个，合计1146个；其中7月单月国产版号193个，创2024年以来单月新高。优质内容和IP强驱动，赋能游戏产业稳健发展。2026年上半年，中国游戏市场实际销售收入1,884.50亿元，同比增长12.17%，用户规模6.84亿，同比增长0.82%。2026年上半年自研游戏海外销售收入123.72亿美元，同比增长30.22%。关注新游戏流水释放与新游戏上线预期：网易关注《遗忘之海》、《无限大》（2027.1.15）、《诡影藏锋》；腾讯关注《洛克王国：世界》、《王者万象棋》；完美世界《异环》上半年流水达14亿，关注后续业绩释放。

- **投资策略：**2026年我们主要看好三个方向： 1）AI影视：随着seedance2.5，minimaxH3，可灵3.0模型发布，模型性能在时长和画质取得重大突破，AI漫剧和AI短剧逐步扩展至AI中长剧和AI电影。相关标的：AI中长剧芒果超媒、AI电影博纳影业。2）AI营销：关注汇量科技AI Infra上线催化；3）游戏游戏方面重点推荐恺英网络、巨人网络，其他标的：网易，公司后续自研新游上线有望贡献增量。

- **风险提示：**行业政策趋严、多模态大模型相关研发不及预期、游戏新品表现不及预期、票房表现不及预期、市场竞争加剧等风险。

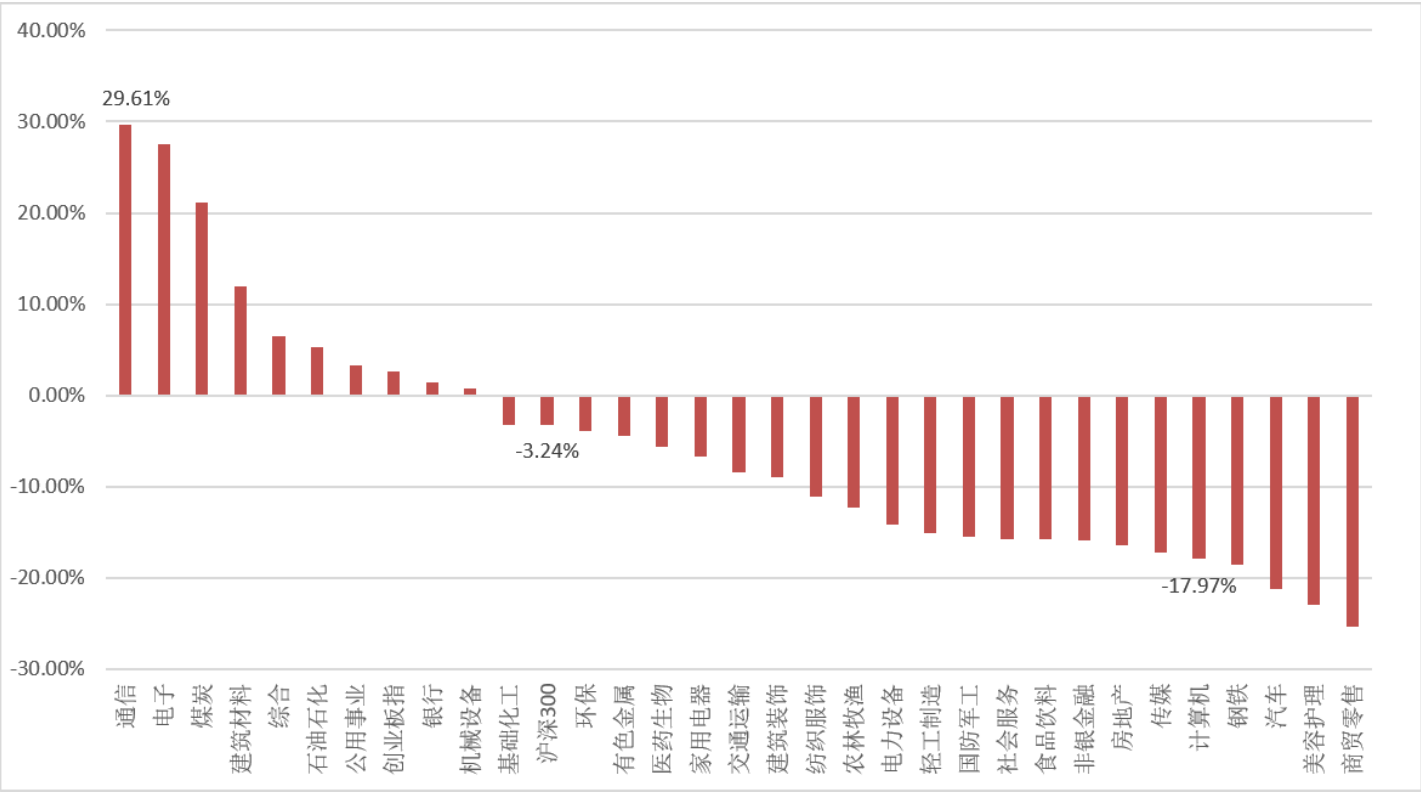
下载日志已记录，仅供内部参考

传媒行业2026年市场表现回顾

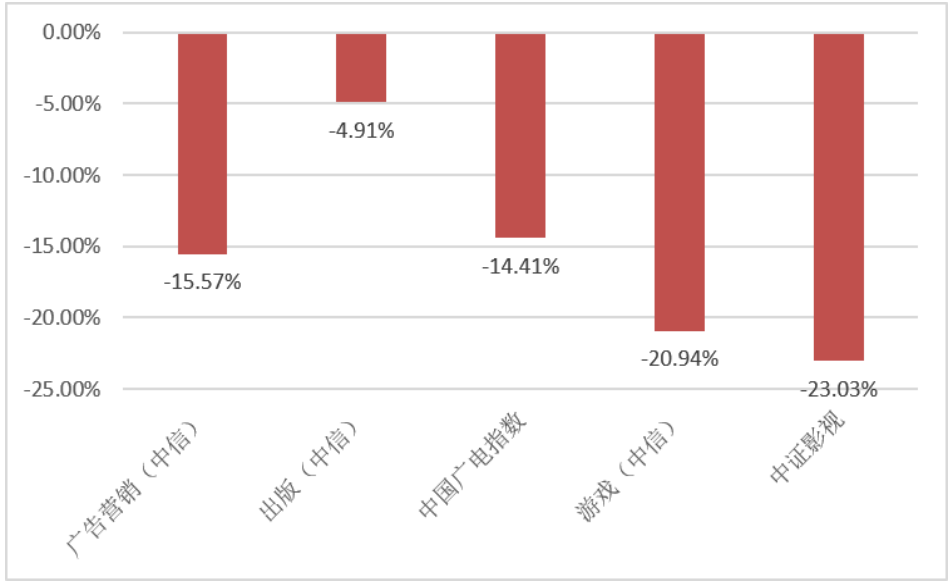
2026年传媒行业复盘-涨跌幅方面：

- 截至2026年9月15日，传媒行业指数下跌17.97%，跑输沪深300指数14.73pp。年初以来，传媒行业跌幅最大的子行业是影视，跌幅约23.03%；出版指数下跌4.91%。

2026年初以来全行业行情



年初以来传媒行业主要子行业行情



数据来源：ifind，西南证券整理

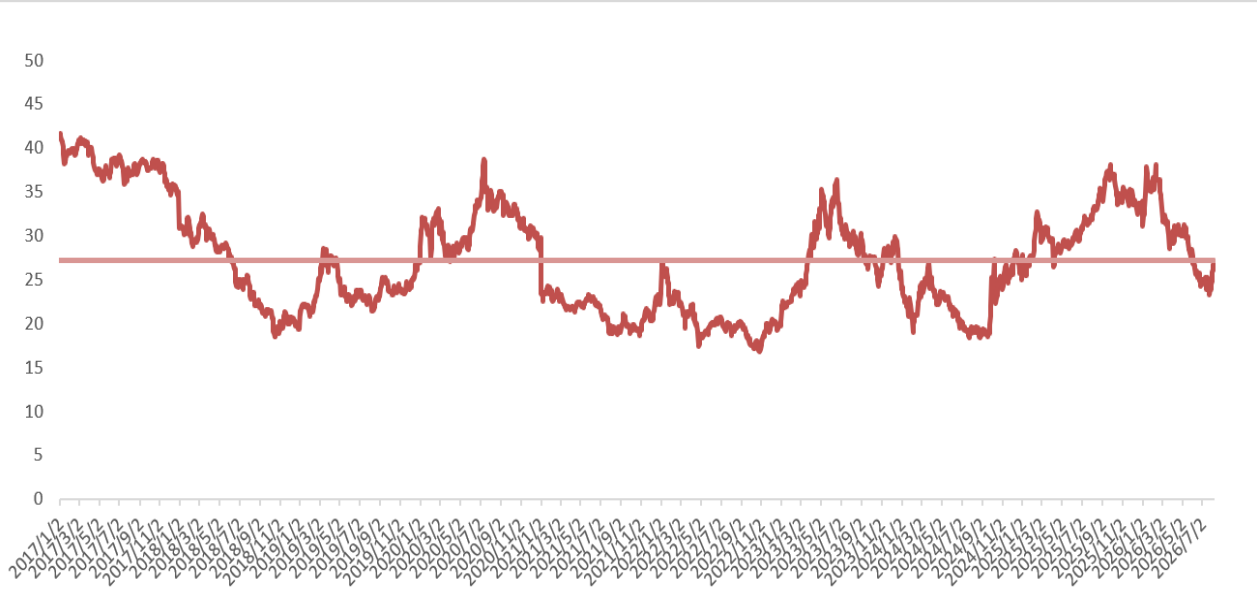
下载日志已记录，仅供内部参考

传媒行业2026年市场表现回顾

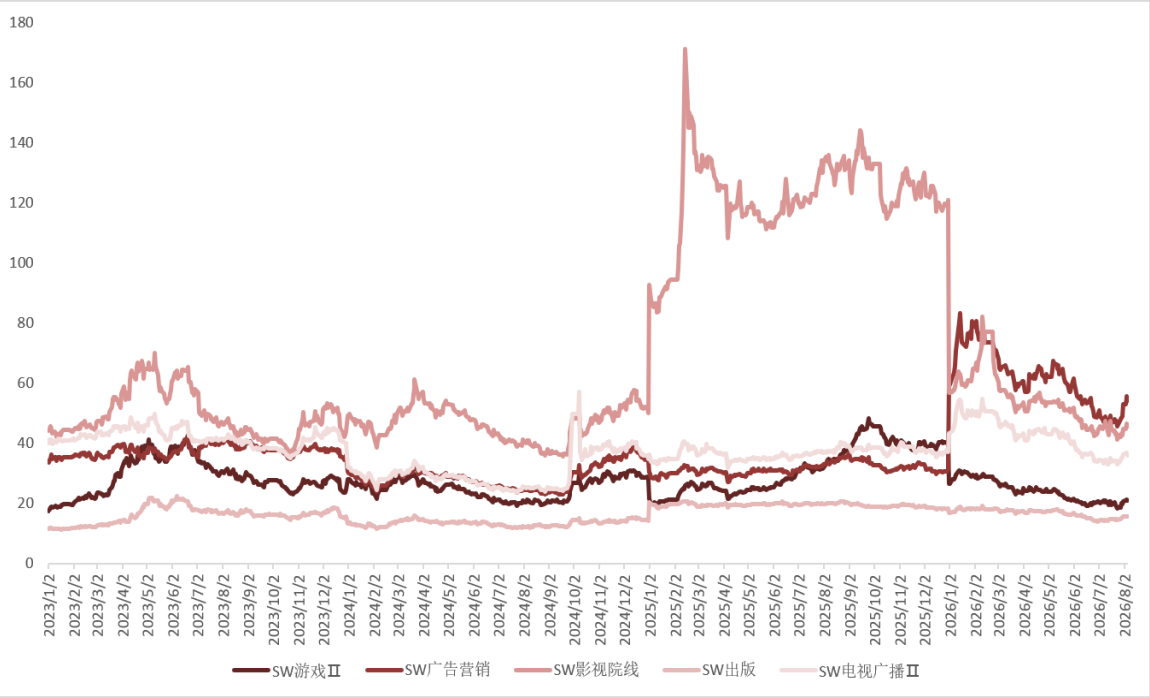
2026年传媒行业复盘-估值方面：

- 横向看，2026年传媒板块各子行业上半年市盈率整体呈现先降后升趋势；其中，出版行业市盈率（16倍）相对较低，广告营销行业市盈率（54倍）相对较高。
- 纵向看，当前传媒行业PE（TTM）为27倍左右，与历史估值中枢相当。

2017年至今传媒行业估值水平PE（TTM）



传媒行业部分子行业市盈率（TTM）



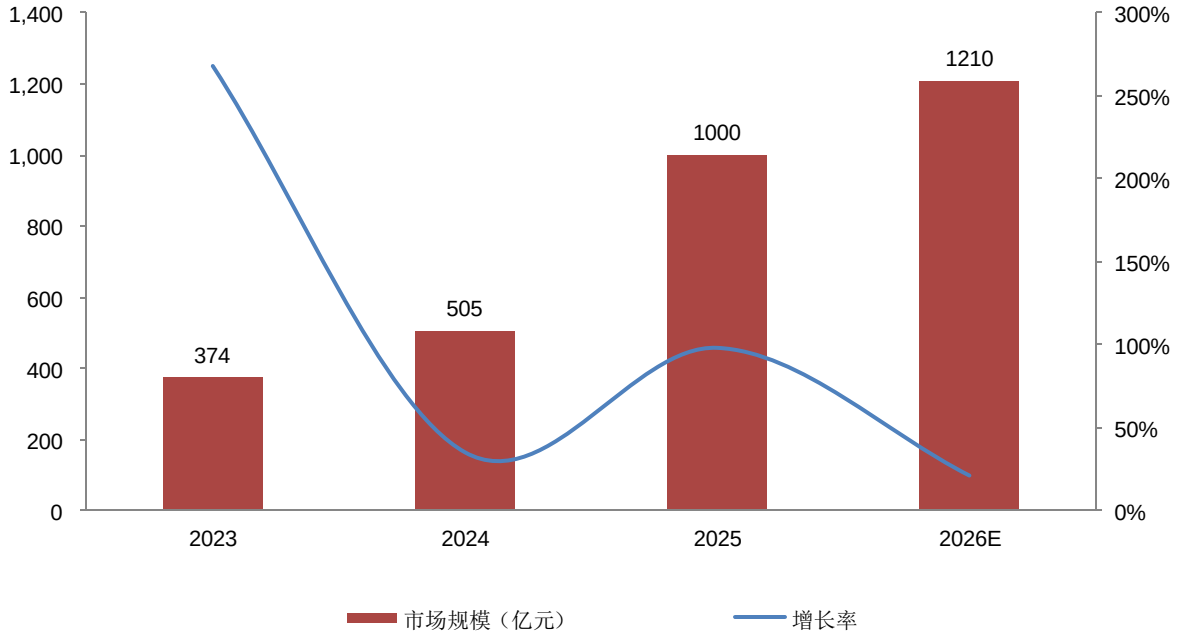
数据来源：wind，西南证券整理

下载日志已记录，仅供内部参考

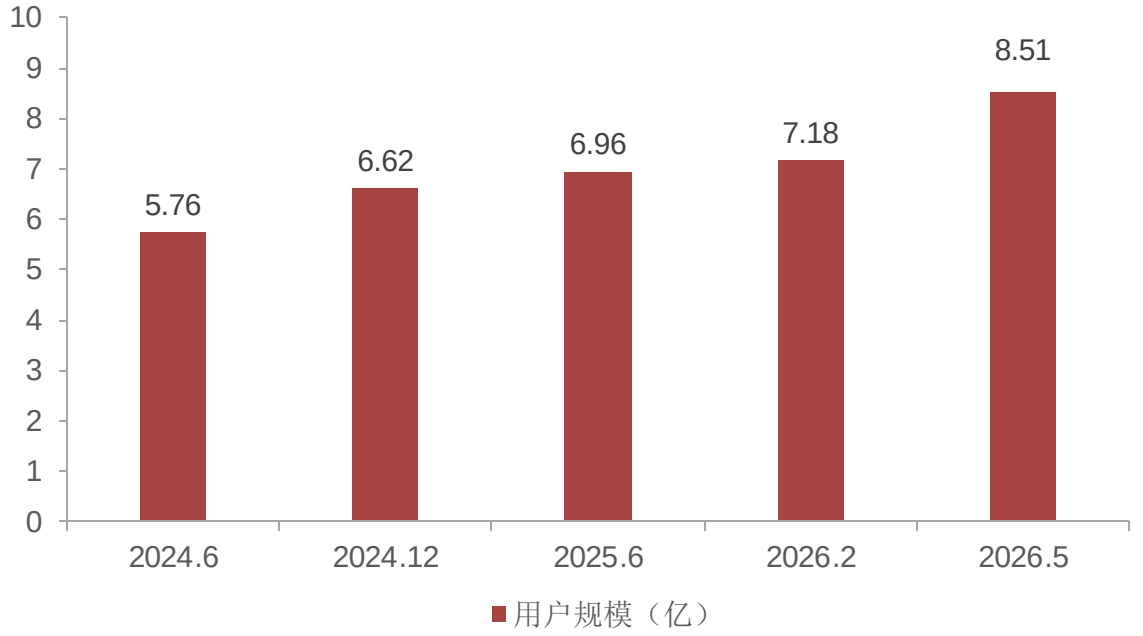
上半年短剧市场规模与数量变化：AI短剧成增长核心引擎

- **国内：微短剧市场规模持续扩张，**根据dataeye数据，2026年中国微短剧、AI剧漫剧市场规模预估合计将突破1210亿元，其中AI剧漫剧超400亿元。用户端呈现爆发态势，截至2026年5月，微短剧用户规模已达8.51亿，占网民总数66.7%；近4个月新增用户约1.33亿，涨幅接近20%，增速有所放缓，创近两年新高，主要受AI剧/漫剧内容供给增加驱动。
- 平台层面：红果免费短剧MAU达3.57亿，超越多个长视频平台。 2026年9月，微信短剧内容中心推出免费短剧小程序“绿泡泡短剧”，通过免费观看、广告变现的模式，聚合微信生态内的短剧内容。

中国微短剧（含漫剧）市场规模



微短剧（含漫剧）用户规模（亿人）



数据来源：DataEye研究院，西南证券整理

下载日志已记录，仅供内部参考

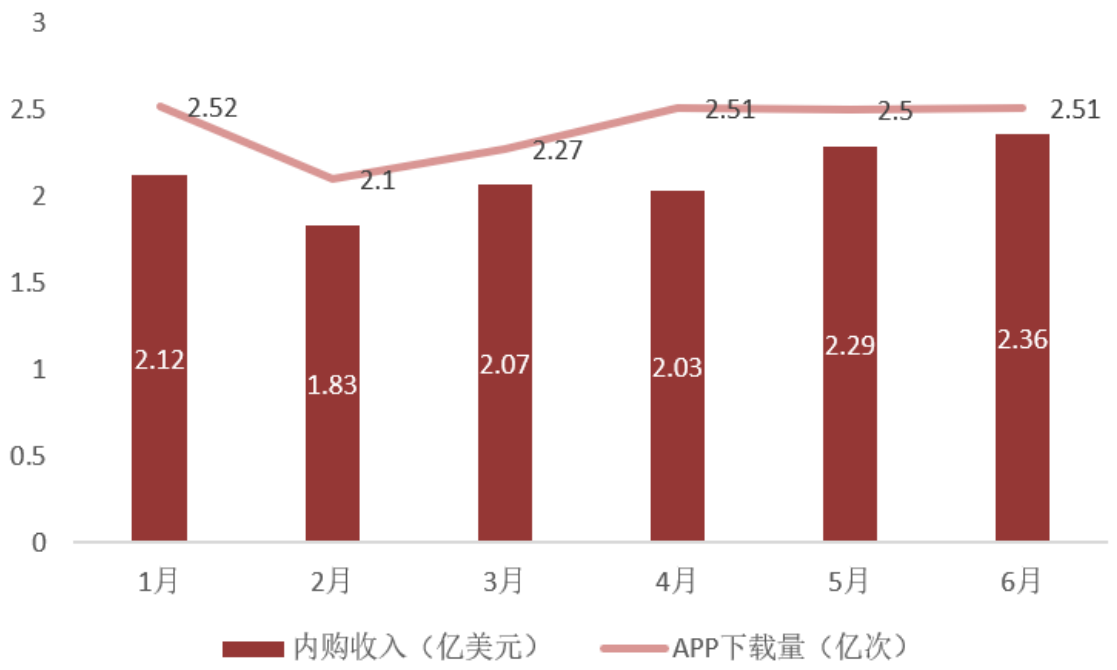
上半年短剧市场规模与数量变化：AI短剧成增长核心引擎

- 2026年上半年，海外微短剧市场双端内购收入累计约12.7亿美元，应用下载量近15亿次。月度走势呈"量稳价升"态势：下载量自3月起稳定在2.5亿次以上，内购收入逐月走高，6月达2.36亿美元，市场进入成熟增长期。AI内容成为核心增长引擎。 DramaWave作为AI短剧平台，7月新剧集超90%由AI生成，下载环比增速39.82%居首；VibeShort 2月上线后快速起量，7月收入环比激增84.66%。AI技术正推动行业从"人工精品化"向"AI规模化"跃迁。
- 中国厂商持续主导海外格局：NetShort以Q2海外收入超9100万美元登顶综合实力榜；字节跳动通过PineDrama、Melolo双平台以"免费+补贴"策略半年冲入下载榜TOP4；昆仑万维形成付费与广告变现的差异化矩阵，上半年下载量达2.18亿次全球第一。市场已从"流量红利期"迈入"内容效率竞争期"，AI生成内容的技术突破正在重塑行业成本结构与增长逻辑。

海外短剧平台排名（下载）

排名	平台名称	所属公司	主要市场	商业模式	核心特色
1	NetShort	麦芽传媒（新加坡）	北美、日本、韩国、欧洲	IAP金币+IAA广告	Q2海外收入超$9100万环比+31%，下载登顶，定价$29.99/周(高于竞品)
2	ReelShort	Crazy Maple Studio	北美(69%)、全球200+国家	IAP金币+订阅	累计收入$1.2B+，MAU 70M+，美国市场下载/收入/MAU三榜统一，好莱坞合作
3	DramaWave	昆仑万维	东南亚、拉美、全球	IAP金币+订阅	AI短剧平台，7月新剧集超90%为AI生成，下载环比+39.82%增速头部第一
4	PineDrama	字节跳动	美国、巴西	免费无广告(字节补贴)	字节跳动补贴免费平台，2025年末上线，半年进入下载榜TOP4，Q2下载环比+460%
5	DramaBox	StoryMatrix Pte. Ltd.	美国(57%)、东南亚、欧洲	IAP金币(70%)+订阅+广告	唯一盈利的主要平台，2026H1收入$1.95亿，7月重回收入第1
6	DramaReels	昆仑万维	东南亚、拉美、印度、菲律宾	IAA广告(完全免费)	昆仑万维免费平台，2026H1下载2.18亿次全球第一，IAA广告变现
7	Melolo	字节跳动	印尼、东南亚	免费+IAA广告(字节补贴)	字节跳动，印尼市场转化闭环强，7月下载第7
8	Rigi TV	Rigi	印度、东南亚	IAP金币+IAA广告	7月下载第8，环比+1.22%
9	StoryReel	Equinox	美国、墨西哥、哥伦比亚	IAP金币+IAA广告	7月下载环比+218.35%，收入上升6位，AdMob渠道占比99.99%
10	VibeShort	VibeShort	全球(AI短剧)	IAP金币+订阅(AI内容)	AI短剧黑马，2月上线，7月收入环比+84.66%(增速第一)，下载第10

2026年1-6月海外微短剧市场双端内购收入与应用下载量



数据来源：DataEye研究院，Insightrackr，西南证券整理

下载日志已记录，仅供内部参考

AI短剧政策变化：扶持政策发布，精品化导向明确

- 2026年6月，国家广播电视总局发布《微短剧发展管理办法（征求意见稿）》，7月《微短剧发展管理办法》正式发布，微短剧正式进入法治化监管时代。此外，中国网络视听协会发布《微短剧创作指引（2026年Q2）》，为行业提供创作规范指引。广电总局发布《"微短剧精品创作传播计划"实施方案》，提出6家重点平台将投入至少60亿元资金支持优秀真人微短剧创作传播，聚焦内容创新与分账透明化。 红果短剧同步推出配套举措，投入5亿元专项资金，聚焦真人实拍内容创新与现实题材深耕。

广电总局公布《微短剧发展管理办法》

《微短剧发展管理办法》9月1日起施行

来源：新华网 | 2026年08月01日 04:09:33

原标题：《微短剧发展管理办法》9月1日起施行

新华社北京7月31日电（记者白瀛）记者7月31日从国家广播电视总局了解到，《微短剧发展管理办法》自2026年9月1日起施行。

办法明确，本办法所称微短剧，是指单集时长少于二十分钟，主题主线明确、故事情节连续完整、人物角色突出的剧集。在中华人民共和国境内，通过互联网站、互联网应用程序和分发平台、广播电视频道和专区、互联网电视、交互式网络电视（IPTV）、有线电视等播出渠道，以电视机、手机、平板、电脑、公共电子屏、智能（可穿戴）和车载设备等为接收终端，面向公众播出的微短剧，适用本办法。

根据办法，微短剧按照投资额度、题材等分为一类微短剧、二类微短剧和三类微短剧，分类实行备案公示和发行许可制度。分类标准由国务院广播电视主管部门制定，并结合行业发展实际情况适时调整。

办法要求微短剧不得载有危害国家安全、歪曲民族历史、诋毁中华优秀传统文化、侵害未成年人合法权益等11项内容，同时规定使用人工智能技术生成、制作的微短剧，在每集明显位置添加提示标识。

广电总局公布《"微短剧精品创作传播计划"实施方案》

国家广播电视总局召开“微短剧精品创作传播计划”工作部署推进会

2026-05-14 20:55 来源：网络视听节目管理司 视力保护色： 【字号： 大 中 小 】 分享到：

5月14日，国家广播电视总局召开“微短剧精品创作传播计划”工作部署推进会，国家广播电视总局党组成员、副局长何飚出席会议并讲话。

会议通报了《“微短剧精品创作传播计划”实施方案》，介绍了广电总局、11个省级广电主管部门、6家网络视听平台支持优秀微短剧创作传播的各项措施，以及2026年指导推出千部优秀微短剧的目标任务。

《方案》提出，健全完善“总局、省、平台”三位一体的协同推进机制，更好发挥有效市场和有为政府作用，拓宽微短剧题材类型维度，畅通创作传播链路，激发优秀微短剧创作动力和创新活力，培育微短剧精品持续迭出的良性生态。

北京、辽宁、上海、江苏、浙江、福建、江西、河南、湖南、广西、陕西等11个省份，安排专门扶持资金，同时出台剧本创作、拍摄制作、上线播出、宣传推广、推优评奖、国际传播等配套政策措施，扶持激励微短剧精品创作。

6家重点平台，将投入至少60亿元资金支持优秀真人微短剧创作传播。抖音（红果短剧）加大优秀真人微短剧扶持，为创新题材作品提供启动资金，为优质、创新、积极正向的现实题材作品提供保底支持，并提升分账比例。腾讯视频坚持对真人微短剧赛道持续性战略性投入，支持优秀真人微短剧创作。芒果TV、咪咕、快手、点众等将进一步加大优秀真人微短剧扶持力度。

数据来源：国家广播电视总局，中国网络视听协会，西南证券整理

下载日志已记录，仅供内部参考

AI影视：从AI短剧拓展至中长剧、电影、互动剧

- **AI中长剧：**芒果超媒制作国内首部上星黄金档全 AI 长剧《后西游记》 8 月 31 日开播，首播当晚实时收视峰值0.3384%，市占率 1.7862%，拿下省级卫视同时段收视第一；上线首轮 5 集实现酷云、欢网同时段省卫收视五连冠；后续战天宫篇持续多次拿下同时段省级卫视实时收视第一名。从“花果山”到“战天宫”，据美兰德数据显示，《后西游记》全网传播已超40亿次，全网热搜上榜311次。制作成本和周期大幅缩短：制作周期4个月，成本大幅缩减，虚拟拍摄、搭景、数字人、特效等环节均大幅节省成本。
- **AI电影：**首部AI超写实院线电影《三星堆：未来往事》正式定档2026年10月23日。博纳影业智能原生科幻电影项目《计划外：广寒行动》，已于2026年4月通过国家电影局立项备案，目前已开机，进入制作阶段。影片以“广寒”这一东方文化意象为起点，将中国人对月亮的想象融入近未来科幻故事，并以“智能原生”方式推进创作。
- **AI互动剧：**9月29日广电总局公示首部互动剧《恐龙快打·华夏秘境》备案信息。《恐龙快打·华夏秘境》，当代科幻题材，共30集，《恐龙快打·华夏秘境》纳入备案，代表互动剧正式进入规范化发展阶段。备案制度之下，互动剧和常规网络剧一样，要经过备案公示与成片审查两道关，坚守价值导向，约束猎奇、低俗的分支内容。值得一提的是，针对微短剧形态的互动内容，规则已经先行一步：总局网络视听司2026年8月发布管理提示，明确具有游戏特征、以分支叙事形成多个情节供用户选择的互动微短剧，按真人或AI制作方式分别执行分类分层标准。长剧进备案公示、微短剧按分类分层管理——互动内容的规范化，已经是长短两轨的完整拼图。

《后西游记》剧照



数据来源：芒果TV官方公众号，西南证券整理

博纳影业AI电影作品



数据来源：博纳影业官方公众号，西南证券整理

下载日志已记录，仅供内部参考

多模态大模型迭代升级：视频生成与推理能力跨越式突破

- 2026年7月31日，Seedance 2.5与MiniMax H3同日发布，标志着国产AI视频模型在"长叙事能力"与"开源生态"两条路线上同时实现关键突破。当前国产文生视频模型在部分权威评测中表现领先，但海外厂商在开发者生态、专业创作套件等领域仍具竞争力。

Seedance 2.5（字节跳动）

核心性能突破：单次生成延长至30秒并支持多轮续写至数分钟，实现"编剧级叙事"，支持最多50个全模态参考（含30张图片、10段视频、10段音频）及动态记忆锚点，原生4K与主体迁移技术赋能专业创作。

MiniMax H3（MiniMax）

核心性能突破：登顶Artificial Analysis视频编辑能力榜单，以2K原生音视频统一生成、全模态架构及仅为闭源竞品约三分之一价格的极致性价比，提供开源视听底座，两者共同推动内容生产降本增效。

2026年短剧模型五维对比

维度	开源	单次时长	核心优势	API计费方式	适用场景
MiniMax H3	开源（社区许可）	最长15秒	视频编辑榜第一、原生立体声、低价	按生成秒数计费	短剧批量生产、本地部署
Seedance 2.5	闭源	最长30秒（可多轮续写）	长叙事、多模态参考、导演式运镜	按可计费Token计费	专业短剧、广告片、影视特效
可灵 Kling 3.0	闭源	较长（视档位）	国产画质标杆、物理一致性强	按积分/秒计费	广告片、宣传片、海外发行
Sora	闭源	Sora 2: 4/8/12秒；Pro: 10/15/25秒	物理一致性、时间稳定性	按秒计费	概念验证、物理模拟
Runway	闭源	视版本而定	专业创作套件、风格一致性	订阅制 + API积分制	专业影视后期、风格化创作

数据来源：字节Seed团队官方博客，MiniMax官方公告，Artificial Analysis Video Arena榜单，OpenAI官方，西南证券整理

下载日志已记录，仅供内部参考

多模态大模型迭代升级：视频生成与推理能力跨越式突破

- **2025-2026年全球视频生成模型竞争已从“海外定义前沿”切换为“中美双极并跑、部分维度国产反超”。**行业评价标准全面转向“单位成本下的可用成片率、参考控制能力、长时一致性、音画一体化与工作流嵌入能力”；**国产模型实现多维同步补齐：**与过去单点追平不同，以Seedance 2.5、Kling 3.0、MiniMax H3为代表的国产模型，在“能力接近甚至领先 + 成本显著更低 + 商业生态更适配”三条线上同时形成冲击；**商业范式逆转：**海外模型率先证明技术可行性，国产模型则快速将前沿能力压缩为更低成本、行业领先可用性的标准化生产力供给，商业闭环验证速度加快。

- **排名来看，文生视频（Text-to-Video）方面，**根据 Artificial Analysis 最新榜单，Google（Gemini Omni / Veo系列）、Black Forest Labs（FLUX.3 Video）、字节跳动（Seedance 2.0/2.5）、Meta（Muse-Video）以及 MiniMax（MiniMax H3）等模型位居前列，在复杂指令理解、镜头运动平滑度及画质细腻度上建立起明显优势；**图生视频（Image-to-Video）方面，**字节跳动（Seedance 2.0）、MiniMax（H3）、xAI（Grok Imagine Video 1.5）、Google（Gemini Omni / Veo 3.1系列）以及阿里巴巴（HappyHorse 1.0 / Wan 2.7）占据榜单头部，展现出极强的图像角色一致性、物理规律模拟与首尾帧掌控力；**竞争演进趋势方面，**行业竞争维度已由单一的“生成画质”转向“原生音视频同步、多镜头叙事、长时序一致性（30s+）与高精准控制力”的综合工程能力比拼；同时，开源基座（如 Wan 2.2、Hunyuan Video）与闭源商业化模型双轨并行，进一步加速了视频生成在影视制作、广告营销与社交内容领域的落地进程。

文生视频（Text-to-Video）排名

排名	厂商	模型名称	发布时间	开源/闭源	Elo	API 价格（美元 / 分钟）
1	Google	Gemini Omni Flash	2026.05	闭源	1242	6
2	MiniMax	MiniMax H3	2026.07	开源	1237	7.8
3	字节跳动	Dreamina Seedance 2.0 720p	2026.03	闭源	1223	9.07
4	阿里巴巴	Wan2.7-260612	2026.06	开源	1160	9
5	阿里巴巴ATH	HappyHorse-1.1	2026.06	开源	1147	9.9
6	阿里巴巴ATH	HappyHorse-1.0	2026.04	开源	1127	13.2
7	阿里巴巴	Wan 2.7	2026.04	开源	1110	9
8	快手	Kling 3.0 1080p (Pro)	2026.02	闭源	1109	20.16
9	昆仑万维	SkyReels V4	2026.03	闭源	1107	21
10	快手	Kling 3.0 720p (Standard)	2026.02	闭源	1101	15.12

数据来源：Artificial Analysis、西南证券整理；注：排名为榜单名义排序；第 1-2 名、第 7-9 名、第9-10名在盲测中 95% 置信区间重叠，主观偏好无统计学显著差异，属同一梯队。

图生视频（Image-to-Video）排名

排名	厂商	模型名称	发布时间	开源/闭源	Elo	API 价格（美元 / 分钟）
1	字节跳动	Dreamina Seedance 2.0 720p	2026.03	闭源	1199	9.07
2	MiniMax	MiniMax H3	2026.07	开源	1193	7.8
3	Google	Gemini Omni Flash	2026.05	闭源	1191	8
4	xAI	grok-imagine-video-1.5	2026.05	闭源	1115	8.4
5	阿里巴巴 ATH	HappyHorse-1.1	2026.06	开源	1112	9.9
6	Sand.ai	MAGI-2 Preview	2026.08	开源	1108	即将上线
7	阿里巴巴	Wan 2.7	2026.04	开源	1093	9
8	阿里巴巴 (ATH)	HappyHorse-1.0	2026.04	开源	1090	13.2
9	昆仑万维	SkyReels V4	2026.03	闭源	1087	21
10	Google	Veo 3.1	2026.01	闭源	1086	24

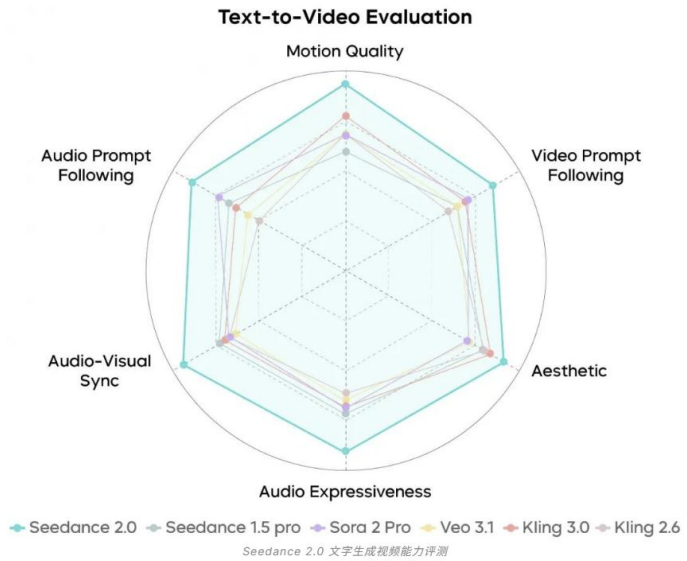
数据来源：Artificial Analysis、西南证券整理；注：排名为榜单名义排序；第 1-3 名、第 4-6 名、第 7-10 名在盲测中 95% 置信区间重叠，主观偏好无统计学显著差异，属同一梯队。

下载日志已记录，仅供内部参考

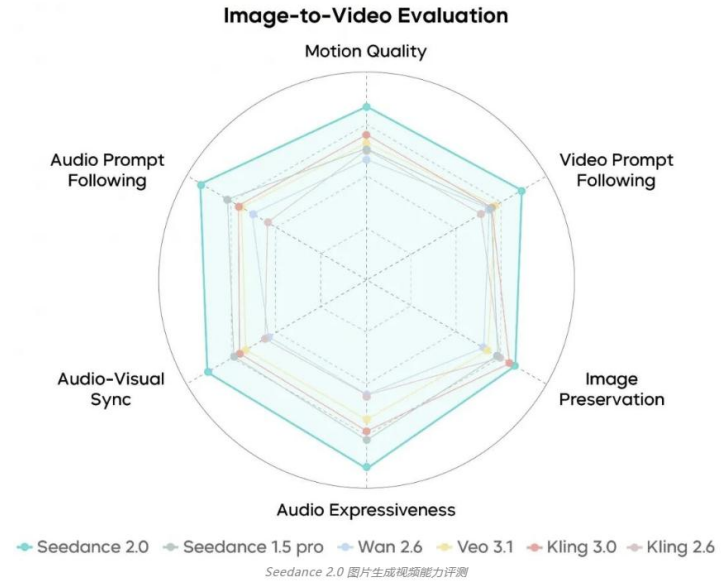
AI视频：字节Seedance迭代迈入2.0/2.5时代，国产AI视频领跑工业级创作

- **技术迭代与商业化闭环同步推进，字节生成式视频能力树立行业标杆。** 2026年2月，字节跳动推出现象级音视频大模型 Seedance 2.0，基于统一的多模态音视频联合生成架构，实现文本、图片、音频及视频四模态无缝输入。相较于前代产品，2.0版本在物理规律还原、复杂动作交互及控制精度上取得实质性突破，全面切入工业级视听内容生产场景。截至2026年中，最新升级的 Seedance 2.5 进一步在单段原生输出长度（提升至30秒）、全模态参考上限（最高50组多模态素材融合）及精准局部视频编辑三大维度实现技术跨越。
- **双轨商业化落地迅速，Token定价具备极高性价比。**商业模式上，Seedance 采用“C端订阅+B端API”双轮驱动。自2026年2月起，模型已接入豆包App、即梦AI及剪映等产品生态；4月火山引擎及海外BytePlus正式上线API服务。在价格端，Seedance 2.0包含视频输入是28元/百万tokens，不含视频输入的价格是46元/百万tokens；Seedance 2.5包含视频输入是42元/百万tokens，不含视频输入的价格是70元/百万tokens。经财联社测算，Seedance2.0 生成单条15秒标准视频约消耗30.88万Tokens，纯生成模式下单条生产成本约15元（折合每秒1元）。

Seedance 2.0文生视频能力评测



Seedance 2.0图生视频能力评测



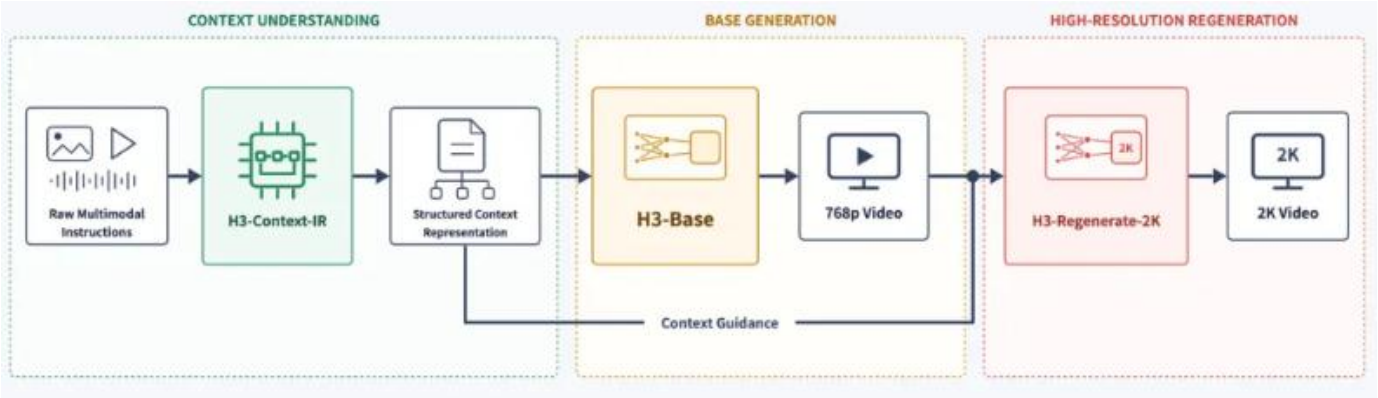
数据来源：公司官网、西南证券整理

下载日志已记录，仅供内部参考

AI视频：MiniMax开源H3，技术突破音画原生生成

- **技术迭代与模型性能同步跃升，MiniMax H3树立原生音视频联合生成与高性价比行业标杆。**根据 MiniMax 官方技术博客，自研视频大模型 H3 摒弃了传统的“视频生成+后期配音”分拆架构，率先实现 Contextual Omni Representation（全模态统一表征），将文本、图像、视频及音频统一在单个 Context 中建模，原生支持 32kHz 立体声（Native Stereo Audio）与画面精细对齐。在底层架构上，H3 搭载自研 H3-VisualVAE（时空因果自编码器，有效序列长度提升 4 倍）与理解生成解耦训练架构，训练吞吐量提升近 30%；结合 In-Context Regeneration（上下文再生成）机制，可输出原生 2K / 24FPS 视频。在控制力上，模型突破性支持 Omni Reference（全能多模态参考），单次最高可融合 9 张图片、3 段视频与 3 段音频（共 12 个参考素材），精准锁定角色一致性、镜头调度与声音特征，并具备对话式局部精细编辑与高精度文字/品牌 Logo 渲染能力，全面切入影视后期、广告营销及短视频工业级内容生产场景。
- **双轨商业化落地迅速，Token/按秒计费定价具备高性价比。**商业模式上，MiniMax 采用“C端海螺 AI 应用生态 + B端开放平台 API”双轮驱动。C端通过海螺 AI 网页及 App 积累高黏性创作者生态，B端通过开放平台输出高并发 API 服务。在价格端，Minimax API 视频生成采取按秒计费，2K分辨率的视频输出价格为0.8元/秒，768P分辨率的视频输出价格为0.5元/秒，具备价格优势。性价比优势极大地降低了企业级 API 调用与工业级内容量产的门槛，加速推动 AI 视频在泛娱乐、电商广告及自动化工作流中的大规模商业落地。

MiniMax H3 系统概览



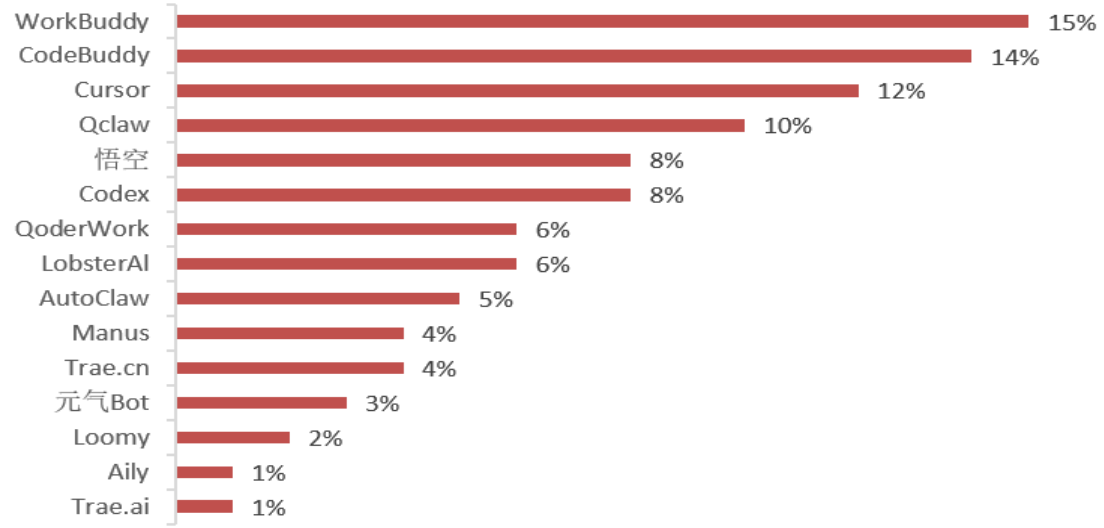
数据来源：公司官网、公司公众号，西南证券整理

下载日志已记录，仅供内部参考

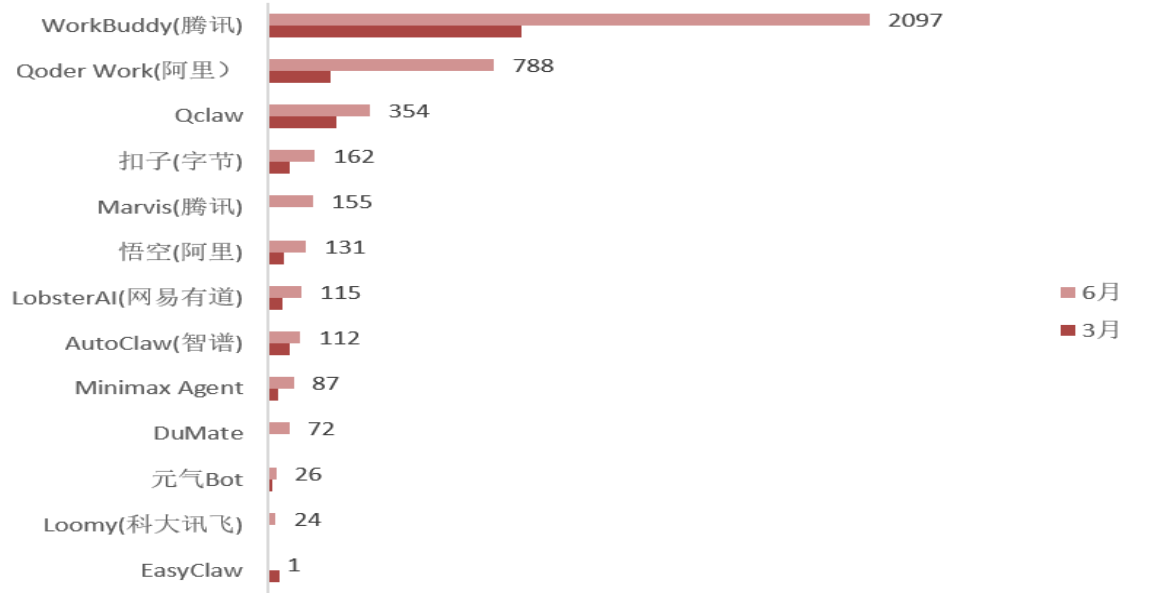
桌面端AI原生办公智能体应用迎来爆发式增长

- 2026年国内桌面端 AI 原生办公智能体行业步入爆发式规模扩张阶段，头部大厂凭借资源优势形成集聚效应，市场分化出大厂生态整合、独立厂商深耕垂直场景两条差异化竞争路线。在Agent时代的应用演进中，PC 端依托算力、系统环境、办公场景适配、企业生态对接四大核心优势，自身价值从单纯操作界面升级为 AI 智能体执行任务的核心承载环境。
- 流量数据维度：2026年3月数据腾讯WorkBuddy以885万月访问量断层领跑全行业；字节Trae.cn、腾讯 QClaw、阿里 Qoder Work 月访问量均突破 200 万，构成行业第一梯队；CodeBuddy、Auto Claw 等其余产品体量依次递减，头部四家大厂产品合计月访问量突破 2000 万。2026 年 6 月数据行业整体规模大幅走高，WorkBuddy 月访问量增至 2097 万稳居第一，大幅领先第二名字节 TRA E（1279 万）；腾讯系 CodeBuddy、QClaw，阿里系 QoderWork 依旧稳居上游梯队，中小厂商产品访问量体量普遍偏低，头部优势进一步拉大。
- **用户渗透维度：**2026年5月用户日常使用渗透率上，腾讯系两款产品占据前两位，WorkBuddy 渗透率 15%、CodeBuddy14%；Cursor、腾讯 QClaw 紧随其后，渗透率分别为 12%、10%；阿里悟空、Codex 均为 8%，前六款产品合计覆盖超六成用户首选需求。此外，不同产品定位分化明显CodeBuddy、Cursor、Codex 侧重代码开发场景，WorkBuddy、QClaw 主打通用办公，其余产品分别聚焦细分垂直办公场景。

用户调研最常用的办公智能体渗透率(5月)



主要办公智能体平台月访问量(万)(3月，6月)



数据来源：易观，西南证券整理

下载日志已记录，仅供内部参考

Work Buddy：全场景 AI 办公工作台

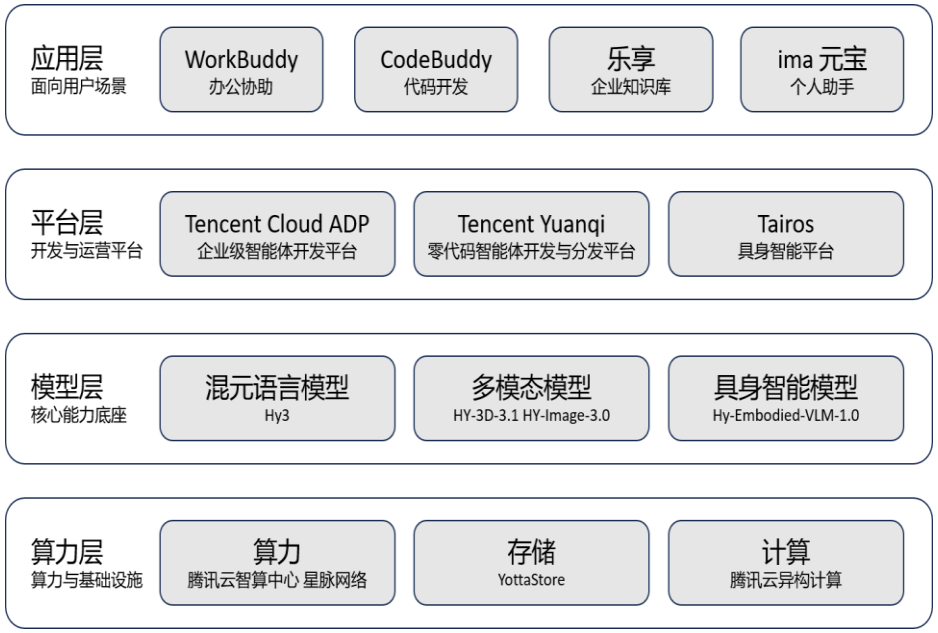
- **WorkBuddy 是腾讯出品的全场景 AI 办公工作台。该产品并非传统对话式AI助手，而是具备自主规划、任务拆解与执行交付能力的智能体。**
- **WorkBuddy核心能力体现在专家、Skill、连接器三个板块：**

1）专家用于为 WorkBuddy 配置专业角色，在特定领域中能以更明确的方法和视角完成任务。专家板块包含专家中心和我的专家。在专家中心的专家和专家团支持按行业分类浏览，召唤他们为你服务；在我的专家创建属于你的专家，分享专业知识。

2）配置技能能够扩展 WorkBuddy 的能力范围，赋予 WorkBuddy 更强大的能力。技能板块分为技能市场、已安装两个内容。在技能市场查看推荐技能，可按需一键安装；已安装板块中已安装的本地技能，可直接在对话中调用。

3）连接器用于将 WorkBuddy 与外部服务进行对接，实现数据互通与能力扩展。当前已支持 QQ 邮箱、腾讯文档、腾讯乐享、腾讯会议、TAPD 等连接器，并支持自定义连接器；新增连接器将同步更新隐私协议相关条款。

腾讯AI布局



数据来源：公司公告，西南证券整理

Workbuddy界面



数据来源：WorkBuddy，西南证券整理

下载日志已记录，仅供内部参考

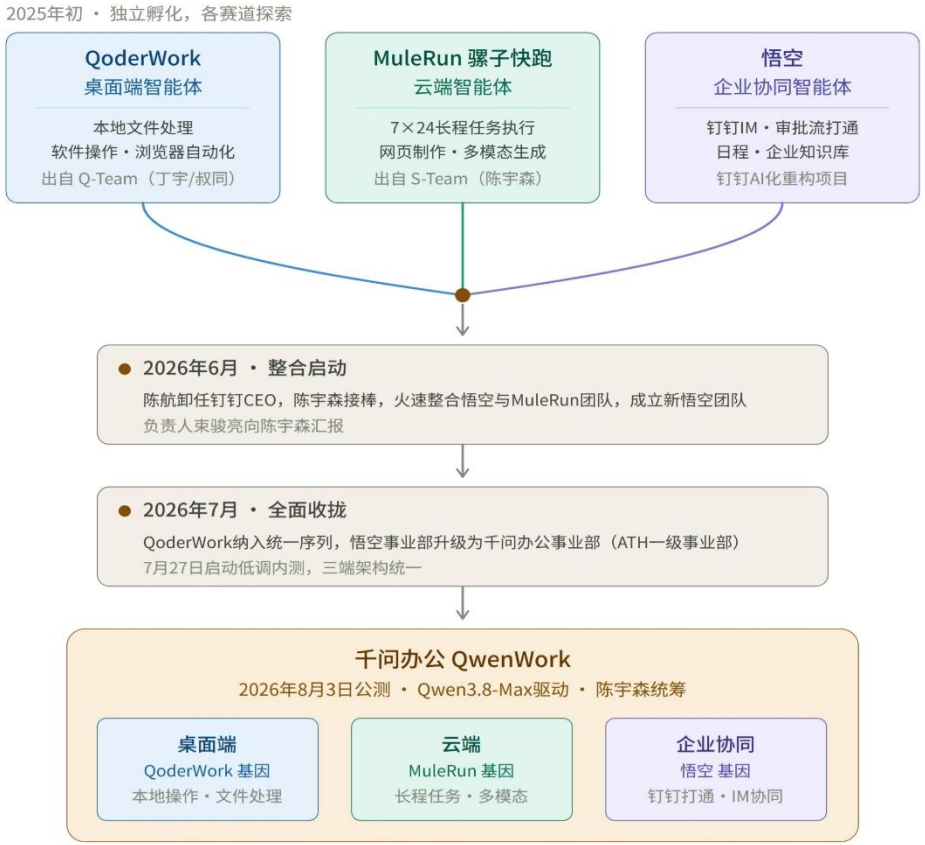
阿里悟空：根植钉钉组织生态的企业级一体化AI工作平台

在2025年初，阿里AI办公智能体QoderWork、MuleRun、悟空三款办公智能体开始独立孵化，各自运营。最终通过企业内部赛马，在2026年7月27日，阿里将QoderWork、MuleRun、悟空三款产品全面整合，推出全新Agent办公产品“千问办公”。

阿里AI布局

应用层 面向用户场景	Qoder 智能体编程平台	悟空 企业级工作平台	千问办公 企业级办公平台	千问 个人AI助手
平台层 开发与运营平台	阿里云 PAI 一站式机器学习平台	阿里云百炼 一站式大模型服务平台		
模型层 核心能力底座	通义千问 Qwen3.8 Qwen3.7	通义万相 Wan2.7-Image-Pro	通义听悟 Qwen-Audio-3.0	
算力层 算力与基础设施	算力 飞天智算平台 平头哥半导体	存储 Tablestore	计算 阿里云弹性计算	

阿里AI办公智能体内部架构



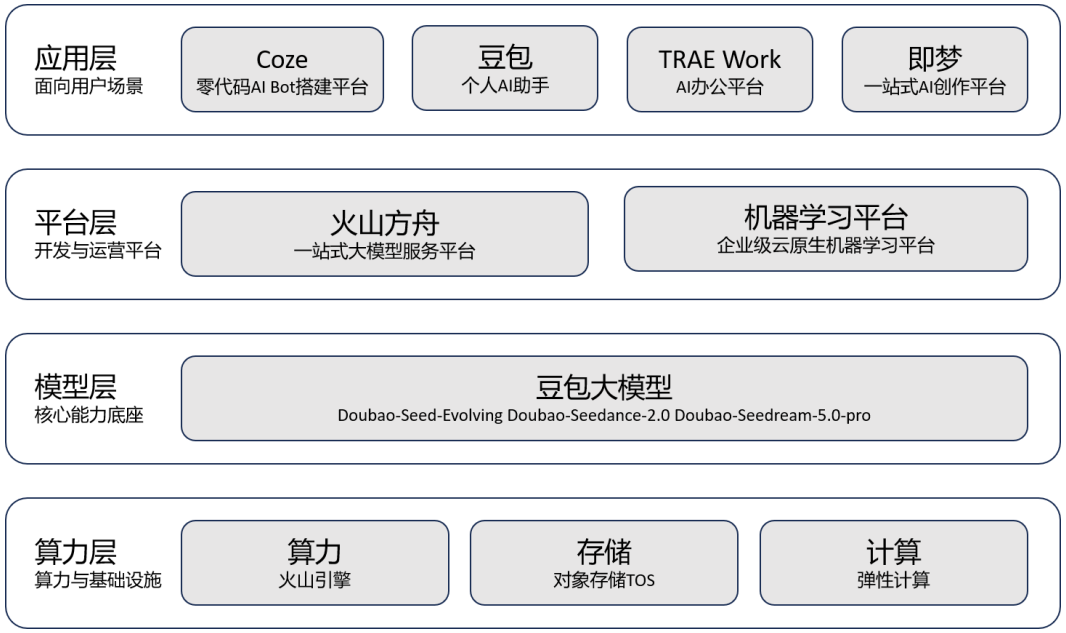
数据来源：公司公告，西南证券整理

下载日志已记录，仅供内部参考

字节系 AI Agent 办公产品整合

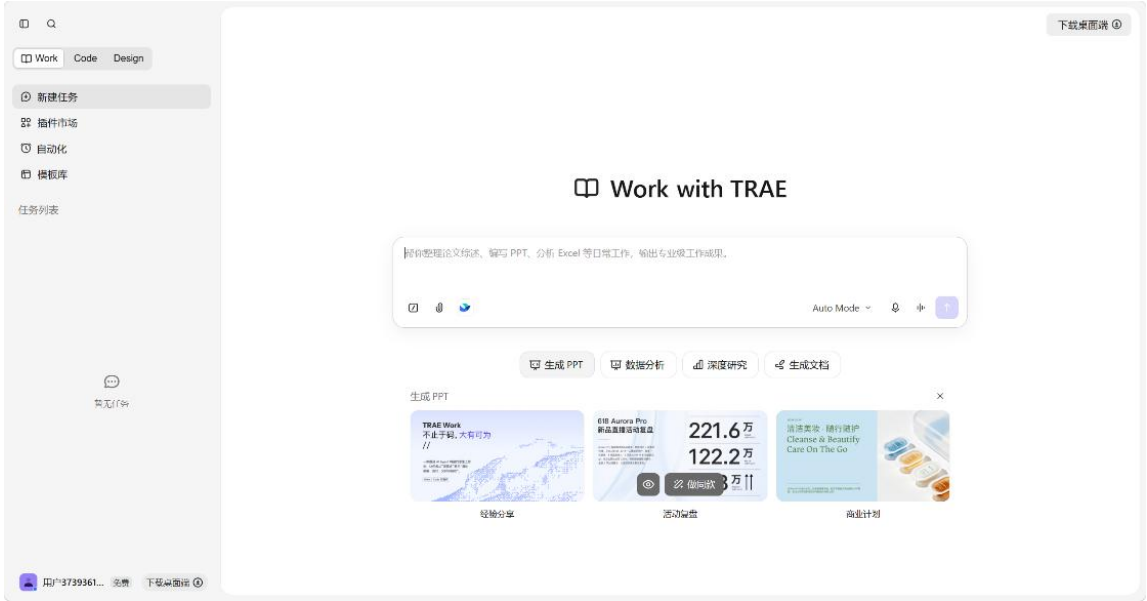
- TRAE Work 是一款AI原生工作台，提供网页版、桌面版和移动版三种形态，并设有为不同用户群体设计的Work、Code与Design模式：1）Work办公模式核心功能包括：多格式混合文档全域处理、智能数据分析、行业研究与信息检索、会议与内容自动化。2）Code开发模式核心功能包括： Builder端到端开发智能体、全链路AI编码辅助、 Git与研发流程自动化、云端隔离代码沙箱、全代码库全局理解。3） Design模式核心功能包括： Figma文件解析，提取组件、配色、布局规范，蒸馏可复用企业设计系统Skill；输入风格需求自动生成整套页面UI、运营活动页面；内置多套行业设计模板，可一键复用生成完整视觉素材。
- 字节为实现战略聚焦、解决内部多产品赛马与销售内耗问题，顺应生产力入口从 IM 文档向 AI Agent 转移的趋势，同时依托豆包 2 亿 DAU 的流量基础，借助飞书企业场景补齐 ToB 变现能力，启动旗下 AI Agent 办公产品整合；2019 至 2025 年字节陆续落地飞书商业化、豆包内测、扣子 Coze、飞书 Aily、TRAE 等产品，在 2026 年 7-8 月用 26 天完成分步收口：7 月 30 日飞书产品团队并入豆包，8 月中旬飞书 Aily 更名为豆包工作伙伴，8 月 24 日 TRAE 与扣子的办公能力并入豆包工作，8 月 25 日正式对外发布统一品牌「豆包工作」，整合后形成分层能力架构，豆包作为大模型与流量主干，飞书承担 ToB 场景上下文底座，TRAE 和扣子负责智能体与工具生态，叠加云端执行能力，打造集生成编辑、电脑操作、多智能体协作于一体的 AI 办公产品。

字节AI布局



数据来源：公司公告，西南证券整理

TRAE Work界面



数据来源：TRAE Work，西南证券整理

下载日志已记录，仅供内部参考

个人AI Agent：加速C端应用商业化进程

- 个人AI Agent本质是面向用户个体生活场景的个人数字替身，目标是替个人处理生活、办公的跨应用长周期事务，追求帮助用户完成任务；和其他Agent的核心差异集中在**运行环境、外部集成、主动性能力、身份与记忆、权限体系与安全模型、定义与容错边界上：1）运行环境：**专属持久云端沙箱VM是标志性特征，普通对话助手依附聊天会话，会话关闭任务即终止；而个人Agent为每一个普通用户分配独立隔离的云端虚拟机，App关闭，任务依然可以后台持续执行；**2）外部集成：**实现用户自主授权即可连接消费类 App、个人SaaS 服务，无需管理员介入，能够操作无开放API 的网页场景；**3）主动性能力：**从被动应答升级为主动式双向交互，支持后台主动研究、主动汇报进度、请求用户决策，不只是等待用户输入；**4）身份与记忆：**绑定个体，跨多端统一身份与长期记忆，不是一次性会话上下文，而是持续沉淀该用户的个人偏好。**5）权限与安全模型：**个人自主全权管控，由用户自主决定访问、执行、审批，普遍采用模型原生对齐 + 外部独立安全审查双层防护；**6）定位与容错边界：**个人Agent面向个人生活、私人办公事务，出错由个人修正；企业 Agent 面向业务流程，容错极低，优先合规可控。
- 海外以 Meta Muse、OpenAI Dots、为代表，分别依托社交生态、模型能力、知识工作场景与云办公套件构建差异化；国内以微信小微为代表，以微信生态原生调用为核心。 9 月 8 日META推出个人助手Muse正式上线后，仅仅 10 天，Muse 就冲上美国 App Store 免费榜第一，第二天又登顶 Google Play。Sensor Tower 估算，其上线前两周获得了约 280 万次下载，日均增速达到 55%。

主流个人AI Agent 对比

产品	厂商 / 上线时间	核心定位	核心能力	底层模型	生态依托
Muse	Meta / 2026.9.8	主动式个人 Agent，全球首款为每个人打造的个人 AI 助手	主动式目标感知与任务推进、日常任务代办、服务商连接器生态对接、沙箱虚拟机浏览器操作、端到端任务闭环交付	Muse Spark 1.3	谷歌全家桶、OpenTable、Expedia 等第三方 Connector API
Dots	OpenAI / 2026.9.29	全天候在线智能体，拥有独立云端电脑的个人 AI 助理	独立云端电脑与浏览器、连接超 4,000 个应用、主动研究、全天候后台推进任务、多通道交互（ChatGPT / Slack / Teams / 短信）	GPT-6 Astra	OpenAI 插件生态（4,000+ 应用），Codex 工具集成
微信小微	微信团队 / 2026.6 灰度	微信生态原生 AI 助手，生态内服务调度器	微信原生功能操作、小程序服务调用、群聊总结与文件理解、创建小工具、内容场景延申	WeLM-80B（MoE）；部分调用 DeepSeek	微信生态（小程序、公众号、视频号、支付、社交）

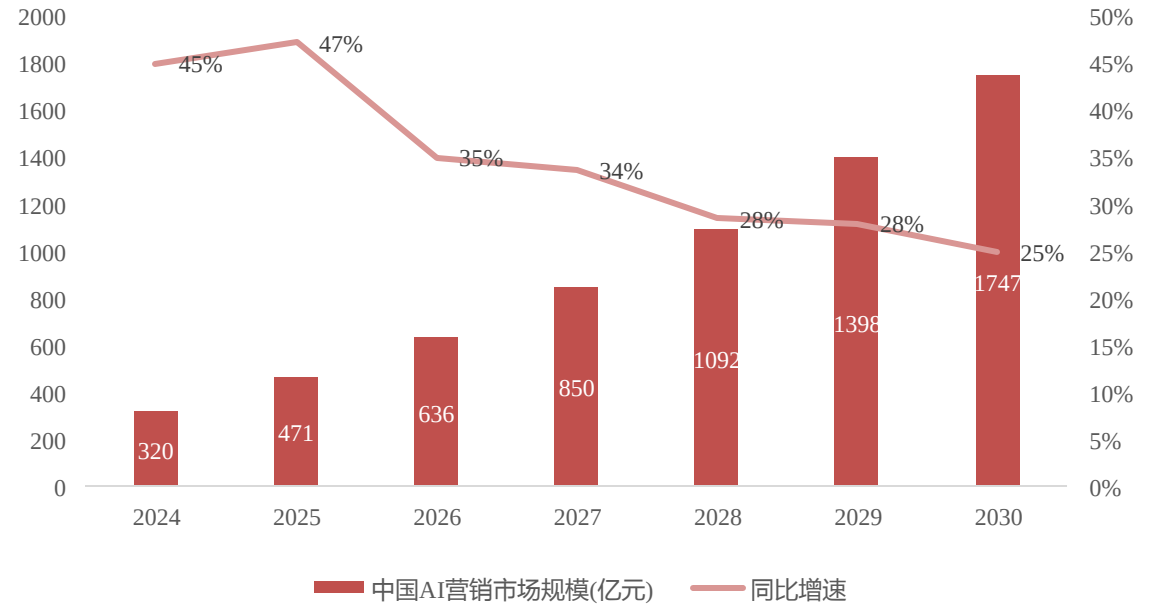
数据来源：各公司官网、官方 API 文档，西南证券整理

下载日志已记录，仅供内部参考

“AI新科技+营销”板块-AI营销

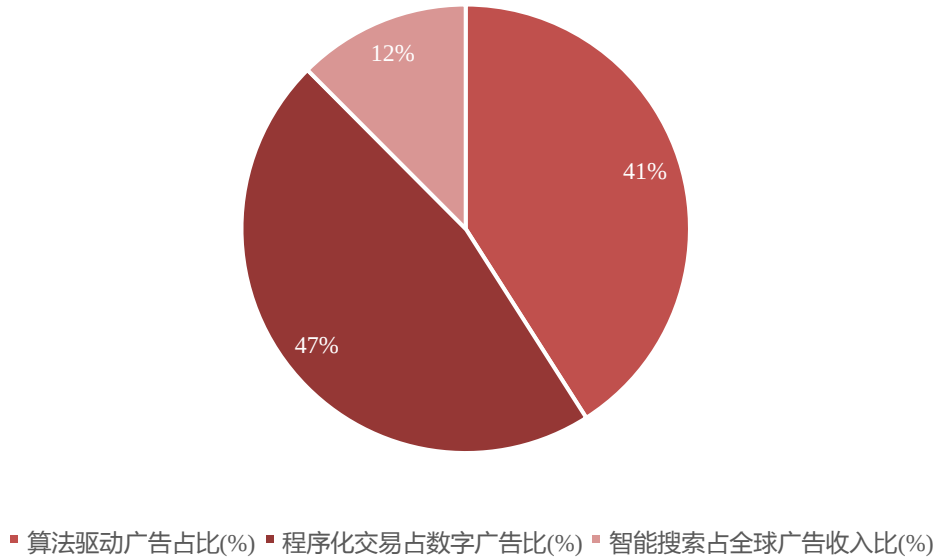
- 2026年，AI大模型已深度渗透营销领域，AIGC实现文案、图像、视频的秒级批量生产，营销智能体自主完成调研、策略、投放及客服互动。随着搜索行为向AI问答迁移，GEO（生成式引擎优化）取代传统SEO成为新技术高地，企业投入已从试验预算升级为核心营销战略。
- 根据预测，2026年全球广告收入中，算法驱动广告占比达71.6%，程序化交易占数字广告81.4%；同时WPP首次将传统搜索与生成式搜索合并为"智能搜索"统计，该类别占全球广告收入21.8%，预计2030年成为史上最快突破千亿美元规模的广告渠道。
- 中国市场同步高速增长，预计2026年AI营销市场规模达636亿元，同比增长35%，未来三年保持30%左右增速，2030年突破1700亿元；2026年第一季度国内互联网广告发布收入2642.8亿元，同比增长26.7%。

中国AI营销市场规模预测



数据来源：易观分析，西南证券整理

全球广告收入AI相关占比



数据来源：Dentsu，WPP Media，西南证券整理

下载日志已记录，仅供内部参考

AI营销Agent：产品矩阵全面成型，从工具辅助迈向智能体协同

- 2026年上半年，AI营销Agent赛道进入规模化商业兑现阶段，头部厂商产品架构从“单点工具”升级为“多Agent协同系统”。迈富时推出AI-Agentforce 3.0智能体中台，集成500+智能体应用，覆盖营销、销售、服务全链路；趣致集团发布“全感物理AI智能体”体系，将线下终端进化为具备环境感知与内容生成能力的端侧智能体；微盟“星系列”AI原生矩阵成型，以“Agent+Skills”架构实现对话即经营；有赞以“有赞龙虾”AI经营智能体为核心，串联获客、转化、复购、经营四大环节；易点天下构建数百个AI Agent日常运行的Token服务体系，覆盖出海营销全链条。

主流厂商AI营销Agent产品布局对比

厂商	核心Agent产品/平台	技术架构	覆盖营销环节	关键进展
迈富时	AI-Agentforce 3.0智能体中台、T-GEO、EVA、Nora、KnowForce AI	Tforce千亿级垂直营销大模型+多Agent协同	市场洞察、内容生成、线索获取、客户培育、经营分析、GEO优化、出海获客	500+智能体应用上线，多Agent协同效率提升300%，自然语言构建智能体，5分钟完成场景诊断
趣致集团	全感物理AI智能体体系：Q-Agent、Q-Sense、Q-Gen	AI-Omni多模态垂直模型+端侧边缘智能	新品营销认知建立、体验激发、反馈收集、转化承接、物理空间互动	WAIC 2026正式发布，线下终端升级为具备环境感知、用户理解、资源调度、内容生成能力的物理AI智能体
微盟	微盟星元、星启、星枢、星拓、星创（WIMO）	Agent+Skills架构，LUI语言交互驱动	对话式店铺搭建、导购线索自动化跟进、ChatBI归因分析、AI创意自动化生产	星元托管任务6月商户数为5月的6.8倍；经营分析场景商户数季度环比+102.7%；AI素材渗透率70%
有赞	有赞龙虾、AI客销、加我推荐官、AI数据分析师	AI原生智能体+FDE前线部署工程师模式	商品管理、营销活动策划与自动执行、私域会员精细化运营、连锁门店统一管控、店铺诊断	使用AI产品商家超2万家；AI客销半年承接87.7万顾客、完成613万次对话
易点天下	Cycor多云平台、Kreado AI、AdsGo.ai、CortexAI	三层Token服务体系+多Agent路由+MCP工具链	广告创意生产、投放策略调优、程序化广告、数据分析、异常预警	数百个AI Agent日常运行，沉淀1500+可复用Skills及130+ MCP工具；AI生成素材占内部创意产出50%+

数据来源：各公司官网、各公司2026年半年报、西南证券整理

下载日志已记录，仅供内部参考

AI营销Agent：全链路渗透营销环节，效率提升从“降本”走向“重构”

- AI营销Agent正从单点效率工具升级为系统性生产力重构，覆盖“获客—内容—转化—运营—决策”全链路。在获客环节，迈富时GEO智能体实现一次优化、全域生效，内容在主流AI搜索平台首屏引用率从18.5%提升至62.3%；趣致集团物理AI智能体打通线下“最后一公里”，实现从认知建立到转化承接的完整物理闭环。在内容生产环节，微盟星创WIMO实现“创意-脚本-分镜-成片”一站式自动生成，用户工作效能提升3倍；易点天下AI生成素材占内部创意产出50%以上。在转化与运营环节，迈富时AI销售助手推动某零售企业销售转化率提升35%，AI KYC助手将人工审核时间从2小时缩短至10分钟；有赞AI客销将客服、销售、线索挖掘融为一体，某商家销售人效直接提升3倍，AI数据分析师几分钟完成传统两三天工作量。

AI营销Agent分环节效率提升实证

营销环节	代表厂商	Agent功能	效率提升/效果数据
智能获客/GEO优化	迈富时	GEO全域优化智能体，覆盖豆包、文心、通义等8大AI搜索平台	首屏引用率从18.5%提升至62.3%；多Agent协同效率提升300%
内容生产/创意生成	微盟、易点天下	微盟星创WIMO一站式视频生成；易点天下Kreado AI数字营销创作	工作效能提升3倍；AI生成素材占创意产出50%+；AdsGo.ai累计发布广告1276条
销售转化/线索跟进	迈富时、有赞	迈富时AI销售助手（线索评级+话术推荐）；有赞AI客销（自动识别高意向客户+跟进清单）	销售转化率提升35%；导购培训周期缩短50%；某商家销售人效提升3倍；AI客销带来5570万+交易额
私域运营/会员管理	有赞、趣致集团	有赞龙虾自动分层触达、复购提醒；趣致集团物理AI互动	会员运营自动化率大幅提升；AI互动营销客户数同比+96.2%
经营决策/数据分析	迈富时、有赞、易点天下	迈富时AI-BI助手跨系统关联；有赞AI数据分析师自然语言查询；易点天下CortexAI异常预警	AI KYC审核从2小时缩短至10分钟；AI数据分析几分钟替代2-3天人工；累计节省近200小时高价值人工耗时
客服接待/售后	光云科技、有赞	快麦小智客服机器人；有赞龙虾7×24小时自动回复	首次响应时间缩短至0.5秒；问题识别准确率99%；咨询转化率提升8%

数据来源：公司官网、各公司2026年半年报、西南证券整理

下载日志已记录，仅供内部参考

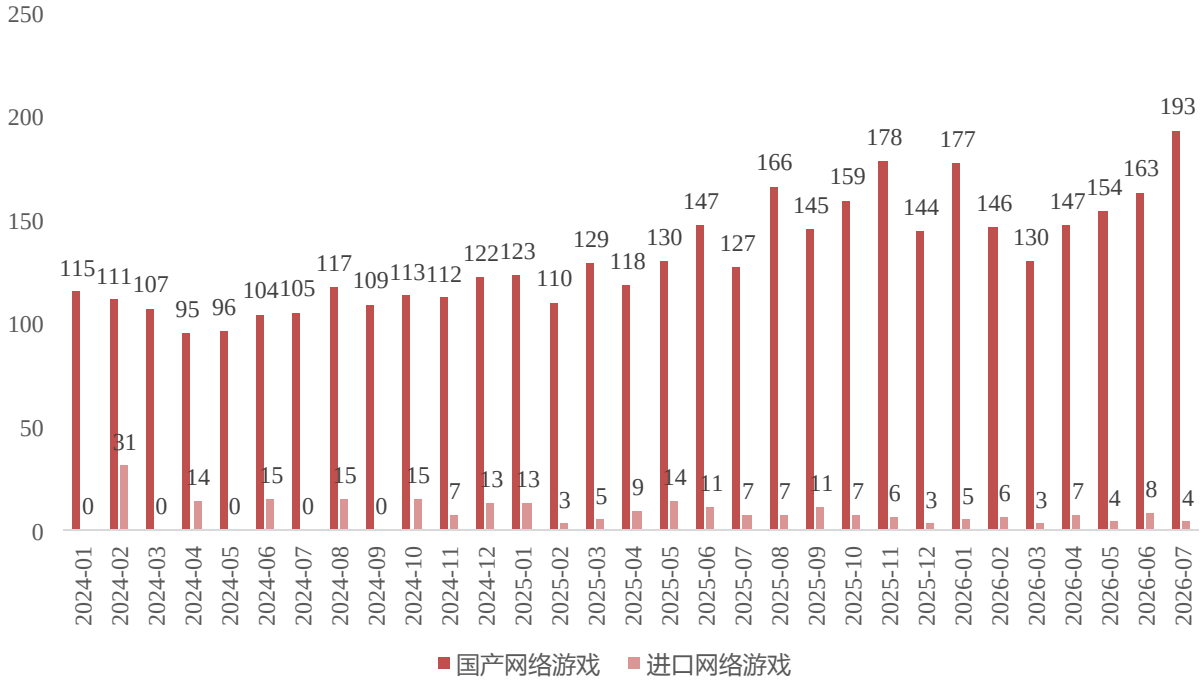
供给侧：7月版号供给数量创新高

- 进入2026年，版号发放延续增长态势，1-7月累计发放国产网络游戏版号1110个、进口版号36个，合计1146个；其中7月单月国产版号193个，创2024年以来单月新高。优质内容和IP强驱动，赋能游戏产业稳健发展。2025年以来国内游戏市场增速显著改善，2026年市场规模延续扩张态势，在新品供给充足以及技术赋能背景下，游戏行业量利双升趋势持续。 2026年上半年，中国游戏市场实际销售收入1,884.50亿元，同比增长12.17%，用户规模6.84亿，同比增长0.82%。2026年上半年自研游戏海外销售收入123.72亿美元，同比增长30.22%；策略类占出海前100移动游戏46.91%，合成类占16.22%、增长9.74%。

中国游戏市场规模（月度;亿元）



2024.01-2026.07 游戏版号发放情况（个）



数据来源：游戏产业网，伽马数据，西南证券整理

数据来源：国家新闻出版署，西南证券整理

下载日志已记录，仅供内部参考

产品侧：多端协同发行成为标配，小游戏"重核化"提速

- 多端发行与全域运营成为2026年小游戏增长的核心趋势。头部小游戏产品已普遍采用多平台、多端并行的发行策略：小游戏平台承担快速获客、轻量体验和前端转化功能，APP端则更多承担稳定留存、深度付费和用户生命周期管理。随着小游戏承载能力提升及产品中重度化趋势增强，微信小程序畅销榜TOP100中已有77%的产品采用MMO、SLG等类App重度玩法。例如冰河工作室的《三国：冰河时代》和点点互动的《奔奔王国》。

《三国：冰河时代》



数据来源：《三国：冰河时代》官网，西南证券整理

《奔奔王国》



数据来源：《奔奔王国》官网、西南证券整理

下载日志已记录，仅供内部参考

游戏：AI深度融入研运全链条，关注新游业绩

产品侧：AI深度融入研运全链条，驱动产品体验重构

- AI技术向游戏产业链全面渗透，在降本增效基础上重塑研发与玩法体验。据游戏工委专项调研，国内游戏研发端AI整体普及率达86.36%，其中美术设计环节渗透率超84%。Unity《2026游戏开发报告》显示，73%的开发者表示AI显著提高了团队效率，62%主要将其用于编写与优化代码。在产品层面，AI不仅加速美术资产生成与界面设计，更开始驱动玩法创新——如AI游戏创作平台SOON已实现输入描述自动生成可玩版本，传奇小游戏《烈焰觉醒》借助AI推出“AI女友”功能实现情感互动，《历史模拟器：崇祯》引入AI实时对话模拟王朝兴衰。

AI游戏创作平台SOON



数据来源：极逸SOON官网，西南证券整理

《烈焰觉醒》与《历史模拟器：崇祯》的AI模拟功能



数据来源：TapTap，Steam、西南证券整理

下载日志已记录，仅供内部参考

游戏：AI深度融入研运全链条，关注新游业绩

- **游戏公司上半年业绩表现亮眼，估值处在洼地。1）巨人网络：**2026年上半年归母净利润为21.4亿元，同比增长175.96%，扣非归母净利润22.4亿元，同比增长177.23%。**2）吉比特：**2026年上半年归母净利润为10.9亿元，同比增长69.31%，扣非归母净利润为10.9亿元，同比增长67.77%。**3）恺英网络：**2026年上半年归母净利润为14.3亿元，同比增长50.34%，扣非归母净利润为11.7亿元，同比增长24.31%；**4）世纪华通：**2026年上半年归母净利润为45.0亿元，同比增长69.51%，扣非归母净利润为43.0亿元，同比增长66.40%。腾讯：2Q26本土游戏收入473亿元，同比+17%，国际游戏收入186亿元，同比-1%。

主要游戏公司盈利预测及PE

公司	收入（亿元）				归母净利润（亿元）				PE	
	2026E	YoY	2027E	YoY	2026E	YoY	2027E	YoY	2026E PE(倍)	2027E PE(倍)
腾讯控股	8,307.21	10.50%	11,057.56	33.11%	2,346.76	4.37%	2,563.49	9.24%	14.98	13.71
网易	1,206.19	7.10%	1,317.39	9.22%	385.04	14.05%	430.52	11.81%	14.05	12.56
世纪华通	478.35	26.22%	520.60	8.83%	89.71	60.05%	108.19	20.60%	11.56	9.59
完美世界	91.66	37.64%	102.26	11.56%	16.89	131.04%	21.44	26.98%	13.13	10.34
巨人网络	93.45	85.14%	105.85	13.28%	44.08	151.23%	50.16	13.79%	11.72	10.30
恺英网络	76.85	44.31%	85.04	10.66%	26.85	41.03%	30.38	13.17%	13.97	12.34
吉比特	67.17	8.25%	72.52	7.98%	19.37	8.02%	21.21	9.45%	14.50	13.25

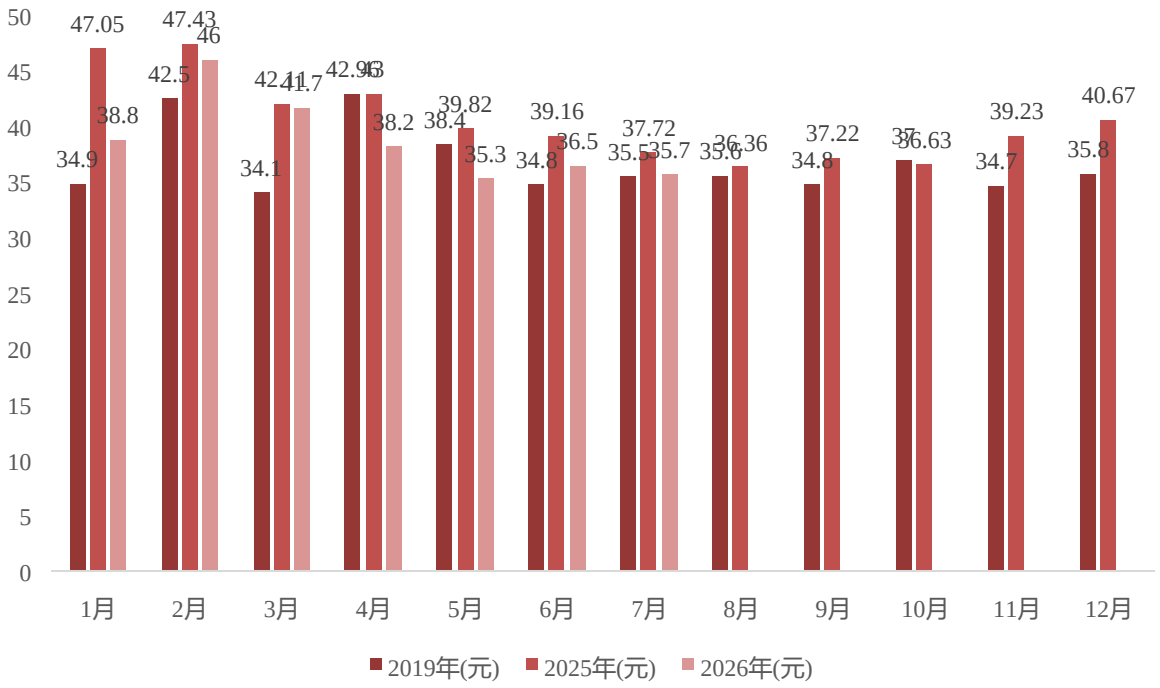
数据来源：wind一致预期、西南证券整理

下载日志已记录，仅供内部参考

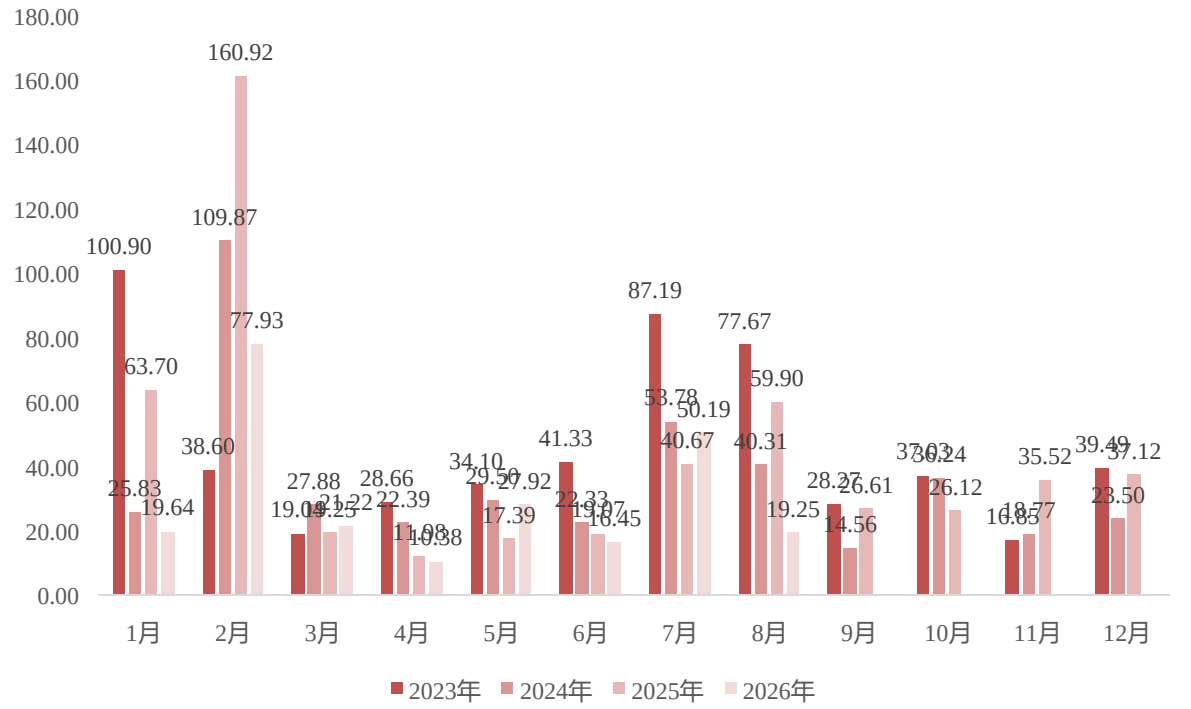
暑期票房小幅增长，“以价换量” 趋势明显

- **平均票价回落，观影人次上涨。**从月度对比来看，2026年1-8月平均票价均低于2025年同期水平。叠加国家启动“电影经济促进年”专项行动，全年计划投放总额不少于12亿元惠民观影补贴拉动观影人次提升。 2026年截至8月10日，累计观影人次6.13亿。截至8月16日，2026年暑期档总票房（含预售）已突破100亿元，同比增长约3.2%，累计观影人次约2.74亿（+5.7%），平均票价36.5元（-2.4%），“以价换量” 趋势明显。

2019-2026年(8月)年单月全国平均票价



2023-2026(8月)年各月份票房对比情况（亿元）



数据来源：猫眼电影，西南证券整理

下载日志已记录，仅供内部参考

2026年暑期档市场表现

- 档期票房TOP5为《功夫女足》《八仙！》《蜘蛛侠：崭新之日》《欢迎来龙餐馆》《给阿嬤的情书》。根据最新猫眼预测值，《欢迎来龙餐馆》预测约30亿元，《八仙！》预测约21.37亿元，《功夫女足》预测约24.72亿元，《蜘蛛侠：崭新之日》预测约16亿元。新片《欢迎来龙餐馆》后续潜力较大，或成为8月中下旬关键变量。

2026年暑期档日前票房Top10

排名	影片	上映时间	档期票房(亿元)	猫眼预测票房(亿元)	类型
1	功夫女足	7月11日	22.90	24.72	喜剧/运动/动作
2	八仙!	7月18日	16.00	21.37	国风动画/奇幻/合家欢
3	蜘蛛侠:崭新之日	7月29日	14.20	16.00	超级英雄/科幻/动作
4	欢迎来龙餐馆	8月11日	8.70	30.00	现实/反战/剧情
5	给阿嬤的情书	4月30日	5.60	20.13	剧情/亲情/潮汕方言
6	奥德赛	8月14日	2.50	7.88	史诗/冒险
7	想你了	7月17日	1.70	2.00	友情/剧情
8	痴迷	7月24日	1.55	2.00	惊悚/爱情
9	年会不能停!2	8月1日	1.51	5.07	职场喜剧
10	去你的岛	8月7日	1.02	1.30	奇幻/爱情动画

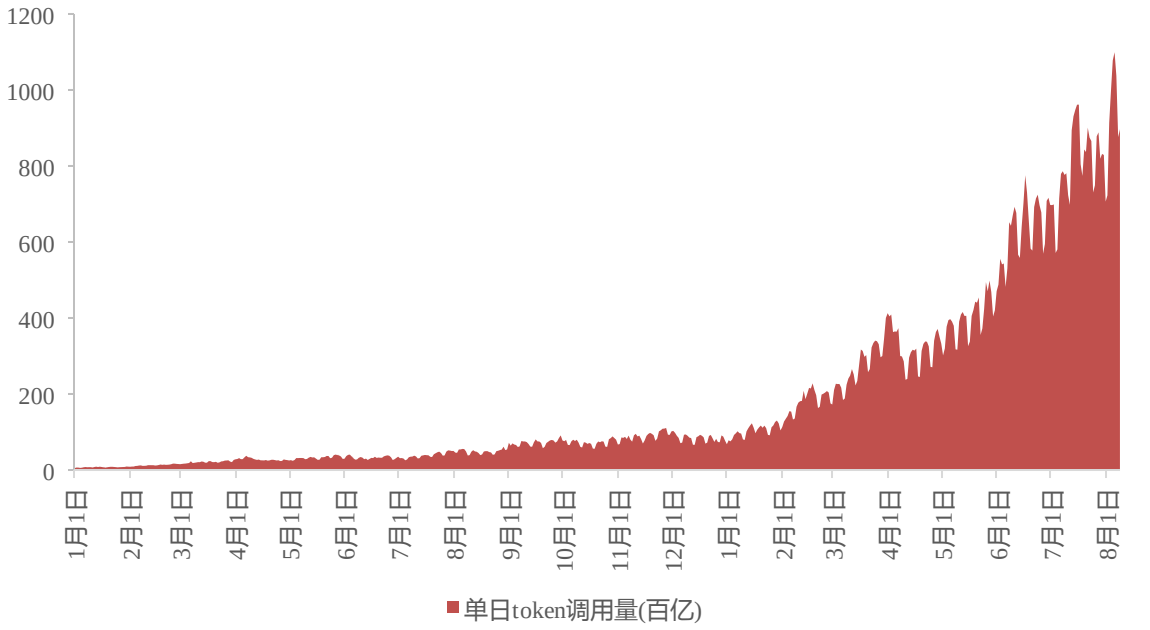
数据来源：猫眼电影，西南证券整理
注：数据截至2026.8.10

下载日志已记录，仅供内部参考

Token调用量：中国AI大模型调用量连续十五周领跑

- OpenRouter 最新统计数据显示，8/3-8/9当周，平台内 AI 大模型总调用量达 69 万亿 Token，环比+21.48%。分区域来看，上榜中国大模型周调用量 34.25 万亿 Token，环比 + 21.76%；同期美国大模型周调用量 9.17 万亿 Token，环比 + 109.36%，**中国大模型合计调用量已连续十五周高于美国，在该平台保持领先。**当周调用量全球前四名均为国产模型：DeepSeek-V4-Flash-0731（正式版）以 8.83 万亿 Token 居首，环比大增 570%；腾讯 Hy3 位列第二，调用量 8.05 万亿 Token，环比 + 67%；DeepSeek-V4-Flash-0423（预览版）排第三，调用量 5.88 万亿 Token，环比回落 19%。

2026年全球AI大模型总调用量



2026/8/3-2026/8/9 全球调用量排名

1.	DeepSeek V4 Flash 0731 by deepseek	8.83T tokens ↑570%	11.	MiniMax M3 by minimax	1.7T tokens ↓13%
2.	Hy3 by tencent	8.05T tokens ↑67%	12.	Kimi K3 by moonshotai	1.35T tokens ↓5%
3.	DeepSeek V4 Flash 0423 by deepseek	5.88T tokens ↓19%	13.	Claude Opus 5 by anthropic	1.32T tokens ↑35%
4.	MiMo-V2.5 by xiaomi	5.39T tokens ↓14%	14.	Step 3.7 Flash by stepfun	1.27T tokens ↓23%
5.	GPT-5.6 Luna by openai	4.43T tokens ↑128%	15.	Claude Sonnet 5 by anthropic	1.05T tokens ↑4%
6.	GLM 5.2 by z-ai	3.49T tokens ↑21%	16.	Gemini 3 Flash Preview by google	899B tokens ↓9%
7.	DeepSeek V4 Pro by deepseek	2.61T tokens ↓20%	17.	Claude Sonnet 4.6 by anthropic	841B tokens ↓16%
8.	Nemotron 3 Ultra (free) by nvidia	2.41T tokens ↓1%	18.	GPT-5.6 Terra by openai	802B tokens ↑63%
9.	Gemini 3.6 Flash by google	2.33T tokens ↑446%	19.	Ling-3.0-flash by inclusionai	627B tokens ↓55%
10.	Laguna S 2.1 (free) by poolside	1.87T tokens ↑101%	20.	Gemini 2.5 Flash Lite by google	619B tokens ↓5%

数据来源：OpenRouter，西南证券整理

下载日志已记录，仅供内部参考

大模型价格：国内模型性价比占优，海外旗舰定价更高

主流大模型 API 定价对比

地区	厂商	模型	是否开源	发布时间	输入价格($/百万Tokens)	输出价格($/百万Tokens)	输入价格(¥/百万Tokens)	输出价格(¥/百万Tokens)	上下文/备注
国内	DeepSeek	DeepSeek-V4 Pro	开源	2026/04/24（Preview）；正式版 2026/08/12 上线	$0.44	$0.88	¥3.00	¥6.00	1M上下文，最大输出384K；缓存命中 $0.0037/MTok（¥0.025）；DeepSeek已预告将整体上调API定价并引入峰谷机制（高峰 9:00-12:00 / 14:00-18:00 翻倍），当前价格未变
国内	DeepSeek	DeepSeek-V4 Flash	开源	2026/7/31	$0.15	$0.29	¥1.00	¥2.00	1M上下文；缓存命中 $0.0029/MTok（¥0.02）；高峰时段价格翻倍（计划中）
国内	字节/火山引擎	Doubao Seed 2.1 pro	闭源	2026/6/23	$0.88	$4.42	¥6.00	¥30.00	256K上下文；缓存命中 $0.18/MTok（¥1.2）
国内	MiniMax	MiniMax M3	开源	2026/6/1	$0.31	$1.24	¥2.10	¥8.40	1M上下文；缓存命中 $0.062/MTok（¥0.42）；≤512K档（官方标价 ¥4.2/¥16.8，当前永久五折后 ¥2.1/¥8.4）；>512K 价格翻倍
国内	智谱AI	GLM 5.2	开源	2026/6/16	$1.18	$4.12	¥8.00	¥28.00	1M上下文；缓存命中 $0.29/MTok（¥2）
国内	月之暗面	Kimi K2.7	开源	2026/6/12	$0.96	$3.98	¥6.50	¥27.00	256K上下文；缓存命中 $0.19/MTok（¥1.3）
国内	阿里/通义	Qwen3.7 Max	闭源	2026/5/21	$0.88	$2.65	¥6.00	¥18.00	1M上下文；缓存命中 $0.09/MTok（¥0.6）；当前限时5折（官网原价 ¥12/¥36）
国内	腾讯	Hunyuan Hy3（正式版）	开源	2026/7/6	$0.15	$0.59	¥1.00	¥4.00	256K上下文；缓存命中 $0.037/MTok（¥0.25）；原 Hy3 preview 将于 2026/08/31 下线，preview 定价 ¥2/¥8
国内	阶跃星辰	Step 3.7 Flash	开源	2026/5/29	$0.20	$1.19	¥1.35	¥8.10	256K上下文；缓存命中 $0.040/MTok（¥0.27）
国内	小米	MiMo-V2.5	开源	2026/4/23	$0.15	$0.29	¥1.00	¥2.00	1M上下文；缓存命中 $0.0029/MTok（¥0.02）
国内	小米	MiMo-V2.5 Pro	开源	2026/4/23	$0.44	$0.88	¥3.00	¥6.00	1M上下文；缓存命中 $0.0037/MTok（¥0.025）
海外	Google	Gemini 3.5 Flash	闭源	2026/5/19	$1.50	$9.00	—	—	1M上下文，最大输出66K；缓存命中 $0.15/MTok
海外	Anthropic	Claude Sonnet 5	闭源	2026/2/17	$2.00	$10.00	—	—	1M上下文；缓存命中 $0.20/MTok；限时 introductory 价至 2026/08/31，9月1日起恢复标准价 $3/$15
海外	Anthropic	Claude Opus 5	闭源	2026/5/24	$5.00	$25.00	—	—	1M上下文；缓存命中 $0.50/MTok
海外	Anthropic	Claude Fable 5	闭源	2026/6/9	$10.00	$50.00	—	—	1M上下文；缓存命中 $1.00/MTok（原表误作 Fable 8，当前官方型号为 Fable 5）
海外	OpenAI	GPT-5.6-sol	闭源	2026/7/9	$5.00	$30.00	—	—	1M上下文；缓存命中 $0.50/MTok；长上下文（>272K）输入×2、输出×1.5（约 $10/$45）
海外	OpenAI	GPT-5.6-terra	闭源	2026/7/9	$2.00	$12.00	—	—	1M上下文；缓存命中 $0.20/MTok；长上下文输入×2、输出×1.5
海外	OpenAI	GPT-5.6-luna	闭源	2026/7/9	$0.20	$1.20	—	—	1M上下文；缓存命中 $0.02/MTok；长上下文输入×2、输出×1.5；7/30 起降价 80%

数据来源：各公司官网、官方 API 文档、官方云平台定价页，西南证券整理；注：1）国内模型人民币价格按 2026/8/12 中间价 6.7882 折算为美元；2）Claude Sonnet 5 当前为限时 introductory 价，2026/09/01 起恢复 $3/$15；3）Hunyuan Hy3 preview 将于 2026/08/31 下线。

下载日志已记录，仅供内部参考

风险提示

- **行业政策趋严的风险。**AI短剧备案与标识要求趋严，对AI影视行业可能有政策不确定风险；
- **多模态大模型相关研发不及预期的风险。**AI应用依赖于大模型能力，多模态大模型不断迭代升级，随着AI短剧逐步发展到AI中长剧，需要的模型能力不断提升，模型迭代升级进度不及预期。
- **游戏新品和票房表现不及预期的风险。**新游戏上线可能表现不及预期。
- **市场竞争加剧风险。**AI短剧制作门槛降低，产能快速扩张；影视优质内容供给不足，核心档期资源稀缺，市场竞争加剧。

下载日志已记录，仅供内部参考

西南证券
SOUTHWEST SECURITIES

分析师：苟宇睿
执业证号：S1250524030001
邮箱：gyr@swsc.com.cn

分析师：阮雯
执业证号：S1250522100004
邮箱：rw@swsc.com.cn

分析师：潘妍洁
执业证号： S1250525100002
邮箱： S1250525100002

分析师：刘言
执业证号：S1250515070002
邮箱：liuyan@swsc.com.cn

下载日志已记录，仅供内部参考

西南证券投资评级说明

报告中投资建议所涉及的评级分为公司评级和行业评级（另有说明的除外）。评级标准为报告发布日后6个月内的相对市场表现，即：以报告发布日后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。其中：A股市场以沪深300指数为基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以恒生指数为基准；美国市场以纳斯达克综合指数或标普500指数为基准。

公司评级	买入：未来6个月内，个股相对同期相关证券市场代表性指数涨幅在20%以上
	持有：未来6个月内，个股相对同期相关证券市场代表性指数涨幅介于10%与20%之间
	中性：未来6个月内，个股相对同期相关证券市场代表性指数涨幅介于-10%与10%之间
	回避：未来6个月内，个股相对同期相关证券市场代表性指数涨幅介于-20%与-10%之间
	卖出：未来6个月内，个股相对同期相关证券市场代表性指数涨幅在-20%以下
行业评级	强于大市：未来6个月内，行业整体回报高于同期相关证券市场代表性指数5%以上
	跟随大市：未来6个月内，行业整体回报介于同期相关证券市场代表性指数-5%与5%之间
	弱于大市：未来6个月内，行业整体回报低于同期相关证券市场代表性指数-5%以下

分析师承诺

报告署名分析师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，报告所采用的数据均来自合法合规渠道，分析逻辑基于分析师的职业理解，通过合理判断得出结论，独立、客观地出具本报告。分析师承诺不曾因，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获取任何形式的补偿。

重要声明

西南证券股份有限公司（以下简称“本公司”）具有中国证券监督管理委员会核准的证券投资咨询业务资格。

本公司与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。

《证券期货投资者适当性管理办法》于2017年7月1日起正式实施，本报告仅供本公司签约客户使用，若您并非本公司签约客户，为控制投资风险，请取消接收、订阅或使用本报告中的任何信息。本公司也不会因接收人收到、阅读或关注自媒体推送本报告中的内容而视其为客户。本公司或关联机构可能会持有报告中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或财务顾问服务。

本报告中的信息均来源于公开资料，本公司对这些信息的准确性、完整性或可靠性不作任何保证。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可升可跌，过往表现不应作为日后的表现依据。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告，本公司不保证本报告所含信息保持在最新状态。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，投资者应当自行关注相应的更新或修改。

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。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，本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。

本报告及附录版权为西南证券所有，，任何机构和个人不得以任何形式翻版、复制和发布。如引用须注明出处为“西南证券”，且不得对本报告及附录进行有悖原意的引用、删节和修改。未经授权刊载或者转发本报告及附录的，本公司将保留向其追究法律责任的权利。

下载日志已记录，仅供内部参考

西南证券 SOUTHWEST SECURITIES

西南证券研究院

西南证券研究院

上海

地址：上海市浦东新区陆家嘴21世纪大厦10楼

邮编：200120

北京

地址：北京市西城区金融大街35号国际企业大厦A座8楼

邮编：100033

深圳

地址：深圳市福田区益田路6001号太平金融大厦22楼

邮编：518038

重庆

地址：重庆市江北区金沙门路32号西南证券总部大楼21楼

邮编：400025

西南证券机构销售团队

区域	姓名	职务	手机	邮箱	姓名	职务	手机	邮箱
上海	崔露文	上海公募销售主管	15642960315	clw@swsc.com.cn	毛玮琳	销售岗	18721786793	mwl@swsc.com.cn
	张方毅	私募销售主管	15821376156	zfyi@swsc.com.cn	张大炜	销售岗	13163027178	zhangdaw@swsc.com.cn
	戴剑箫	销售岗	13524484975	daijx@swsc.com.cn	张怀文	销售岗	18800275611	zhanghuaiwen@swsc.com.cn
	叶佳缘	销售岗	15800609605	yejy@swsc.com.cn	苏湖	销售岗	18913127448	suhu1997@swsc.com.cn
	贾文婷	销售岗	13621609568	jiawent@swsc.com.cn				
北京	李杨	北京公募销售主管	18601139362	yfly@swsc.com.cn	杨薇	销售岗	15652285702	yangwei@swsc.com.cn
	姚航	北京保险销售主管	15652026677	yhang@swsc.com.cn	王宇飞	销售岗	18500981866	wangyuf@swsc.com.cn
	张岚	销售岗	18601241803	zhanglan@swsc.com.cn	刘艳	销售岗	18456565475	liuyanyj@swsc.com.cn
广深	高欣	广深销售主管	13923418464	gaoxin@swsc.com.cn	黄诗洁	销售岗	18817316880	hsj@swsc.com.cn
	龚之涵	销售岗	15808001926	gongzh@swsc.com.cn	周玥	销售岗	18580531950	zhy1997@swsc.com.cn
	文柳茜	销售岗	13750028702	wlq@swsc.com.cn	彭志伟	销售岗	15521604635	pzhw@swsc.com.cn
	林哲睿	销售岗	15602268757	lzr@swsc.com.cn				