



多模态AI：逐光前行，加速进击
传媒互联网行业2026年度中期投资策略

姓名 方光照（分析师）

证书编号：S0790520030004

邮箱：fangguangzhao@kysec.cn

2026年05月18日

1. 全球多模态技术持续迭代，国产模型从追赶到局部超越，整体商业化提速

自2021年Open AI DALL-E首次将大语言模型引入图像生成领域起，海外科技大厂、高校聚焦多模态技术持续迭代模型架构，多模态大模型生成内容质量、效率、成本不断优化。国内科技大厂快速追赶，并在视频生成、音频/音乐生成领域实现局部反超。技术成熟度提升及应用推广驱动AI原生应用ARR快速增长，Midjourney、快手可灵、ElevenLabs ARR达数亿美元，Gemini对Nano Banana的引入助其MAU三个月内增加2亿。全球人才、资本聚集或推动多模态技术继续高速发展，模型商业化迎来广阔市场并有望提速

2. 多模态大模型下游广泛，应用拓展或加速国产模型Token调用量/ARR增长

多模态大模型可深度赋能内容生产、营销、工业制造等，只要AI在工作流中创造价值高于Token成本，需求星辰大海。2026年头部模型Token调用量跃升，豆包3月实现两年千倍增长，智谱3月ARR同比增长60倍，年初以来增长6.4倍。对比Open AI ARR超250亿美元、Anthropic ARR超440亿美元（较2025年底提升389%），我们认为国产模型商业化空间依然广阔；参考主流多模态大模型，单秒音频/单张图像/单秒视频对比单个文字，输入Token消耗量高1-2/3/2-3个量级，输出Token消耗量高1/2-3/4个量级，我们测算仅国内视频消费场景每日所需Token即有望达350万亿，看好应用拓展带动各模型收入延续高增长。**重点推荐：腾讯控股、快手、受益标的：Minimax、阿里巴巴、智谱、昆仑万维。**

3. 多模态大模型深度赋能内容产业链，积极布局“AI+视频/游戏/营销”赛道

（1）**视频**：过去动漫/海外实拍等成本高，AI制片下单部短剧成本低至数万元。AI真人剧登顶红果热播总榜、入选戛纳展映，或反映AI内容已获观众认可；《三星堆：未来往事》获“龙标”标志AI电影上映跑通；“内容+成本”闭环或打开增量市场。**重点推荐：阅文集团，受益标的：中文在线、德才股份、博纳影业。**

（2）**游戏**：AI赋能研发全流程，技术门槛下降有助于分发平台/UGC生态游戏/开放世界游戏促成玩家向创作者的转化，打造“内容-用户”正循环。AI重塑交互体验，其中叙事类、社交竞技类游戏或率先受益。**重点推荐：完美世界、网易、巨人网络、恺英网络、心动公司、哔哩哔哩，受益标的：世纪华通。**

（3）**营销**：多模态大模型助广告系统实现深度个性化，以ChatGPT为代表的高CPM AI应用商业化加速，AI提效下内容激增将提升广告重要性，或共驱广告市场扩容。程序化广告技术领先及与头部AI应用公司合作紧密的营销公司或率先受益。**重点推荐：汇量科技、引力传媒，受益标的：蓝色光标、易点天下。**

4. 多模态大模型加快渗透或进一步扩大算力缺口，算力租赁/AIDC或充分受益

供需失衡推高算力价格，2026Q1国内头部云厂商价格上涨5%-34%。多模态输入提升单位Token算力占用，应用发展或进一步扩大算力缺口。字节跳动/阿里巴巴/腾讯2025年资本开支同比增长88%/71%/3%，2026年算力投资力度加大，看好算力产业链公司。**重点推荐：顺网科技，受益标的：华策影视、浙数文化。**

5. 风险提示：大模型迭代速度、AI应用商业化进展、算力投资力度不及预期等。

目 录

CONTENTS

- 1 全球多模态技术持续迭代，国产模型从追赶到局部超越
- 2 多模态AI商业化开启进击之路
- 3 投资建议及风险提示

1.1 全球多模态技术持续迭代，商业化逐步启动

■ 海外模型厂商引领多模态技术升级，生成内容质量、效率、成本持续优化。科技大厂模型变现主要靠API调用收费及在已有APP中拓展AI相关功能带动活跃用户及ARPU提升，原生AI应用商业化程度有限。多模态AI各领域独角兽涌现，Midjourney 2022年7月发布beta版模型，截至2023年2月实现千万社区成员，营收破亿美元；Runway持续迭代视频模型，推动ARR由2022年底的100万美元提升至2026年的9000万美元；ElevenLabs及Suno深耕音频/音乐，2026年ARR分别破5/3亿美元。

图1：海外大模型厂商起步早，持续推动多模态技术架构优化，商业化逐步启动

	2022年及以前	2023年	2024年	2025年至今
图像	2021年 Open AI发布DALL-E，首次将LLM用于图像生成，可生成256×256像素 2022年 4月，Open AI发布DALL-E2，生成架构由VAE转向扩散模型，引入自研CLIP，生成图像分辨率提升 7月，Midjourney Openbeta版本发布，Meta Make-A-Scene模型允许通过场景图控制构图	2023年 Midjourney产品搭载在Discord，截至2023年2月，拥有超1000万社区成员，年营收约1亿美元。据创始人披露，2023年全年营收“远超”2亿美元 2023-2024年 Google迭代Imagen，专注于增强照片级真实感和语义理解；利用扩散模型并整合LLM以改进提示词理解 Meta基于Emu架构扩展产品，强调速度和质量，探索扩散模型变体 Open AI迭代DALL-E，关注分辨率和图像生成一致性，进一步优化基于扩散架构 Anthropic探索与Claude模型结合的视觉生成		2025年 Open AI发布原生多模态模型GPT-image-1，从扩散模型架构转向自回归架构，显著提升图片可编辑性、文字渲染、物理规律遵循能力，可生成1024x1536像素图片。模型作为ChatGPT生图功能推出，一周内超1.3亿用户生成超7亿张图片，且被Adobe创意工具引入 8月，Google发布Nano Banana，拥有Gemini世界知识，集成实时谷歌搜索，带动Gemini月活从7月的4.5亿涨到10月的6.5亿。11月的pro版支持4K，整合Gemini 3 Pro的多模态理解能力 2026年 Nano Banana 2发布，对比pro版成本减半（1K）且效率更高。 GPT-image-2发布，内置原生推理架构，可联网获取实时信息，文字渲染能力、图片逼真性提升
视频	2022年 Meta Make-A-Video基于文生图模型、超分辨率网络实现文生视频	2023年 2023年初，Runway发布Gen-1、Gen-2，允许用文本对视频进行编辑，Gen-2生成时长延长到18秒 12月，Google Video Poet发布，对比主流的扩散架构，使用单一的LLM实现多项视频生成功能，且可给视频生成音频，允许交互式编辑	2024年 2月，基于DIT架构的Sora发布，证明视频生成领域Scaling Law，模型生成视频最长60秒，物理常识理解和复杂场景建模能力突出 5月，Google Veo发布，可生成1080p分辨率视频，时间可超一分钟；Veo集成到YouTube Shorts	2025年 5月，Veo 3音画同步生成，实现真实世界物理效果与精准的唇形同步 9月，Sora 2发布，音视频同步生成，更能遵循遵守物理定律；同步推出独立APP，上线不到五天下载量突破100万次 Runway ARR超9000万美元，2022年底ARR仅100万美元 Sora APP下线
音频	2022年 Open AI发布开源自动语音识别系统Whisper，模型幻觉较严重	2023年 1月，ElevenLabs首次推出语音设计和克隆产品，上线6个月后台累积超100万注册用户，创作总计时长超10年的音频内容 11月，Google Lyria发布，能生成包含乐器和人声的高品质音乐。与YouTube合作开发音乐创作工具Dream Track 12月，文生音乐模型Suno V1发布，专注于创作全新旋律和人声	2024年 Suno V3发布，提供丰富音乐风格和流派选项，用户可在几秒钟的时间内创作出2分钟的完整歌曲，其中包含前奏、主歌、副歌、歌词、结尾等必备元素 11月，ElevenLabs用户数超3300万	2025年Suno全年营收约1.5亿美元；截至2026年2月底，Suno付费订阅用户超200万名，ARR破3亿美元。2026年4月，Suno V5.5发布，支持用户上传人声生成音轨；可学习用户的创作风格、追踪收听习惯。 ElevenLabs 2025年初ARR达1亿美元，2026年初ARR达3.3亿美元，2026年5月ARR达5亿美元 Google Lyria 3发布，生成音乐时长提升至3分钟，模型能理解歌曲结构，用户可以通过提示词指定前奏、主歌、副歌、桥段等段落，对完整歌曲进行精细控制
世界		2023年 Open AI Shap-E可在几秒钟内生成用户语言描述的3D资产	2024年 3月，Google发布无监督训练的生成交互模型Gene 1，可生成可控制动作的2D视频游戏 12月，Gene 2发布，可生成3D场景	2026年1月，Gene 3发布，可生成交互式3D世界；可在交互过程中修改角色和环境；在移动时实时生成环境；整体世界能够保持约一分钟的一致性

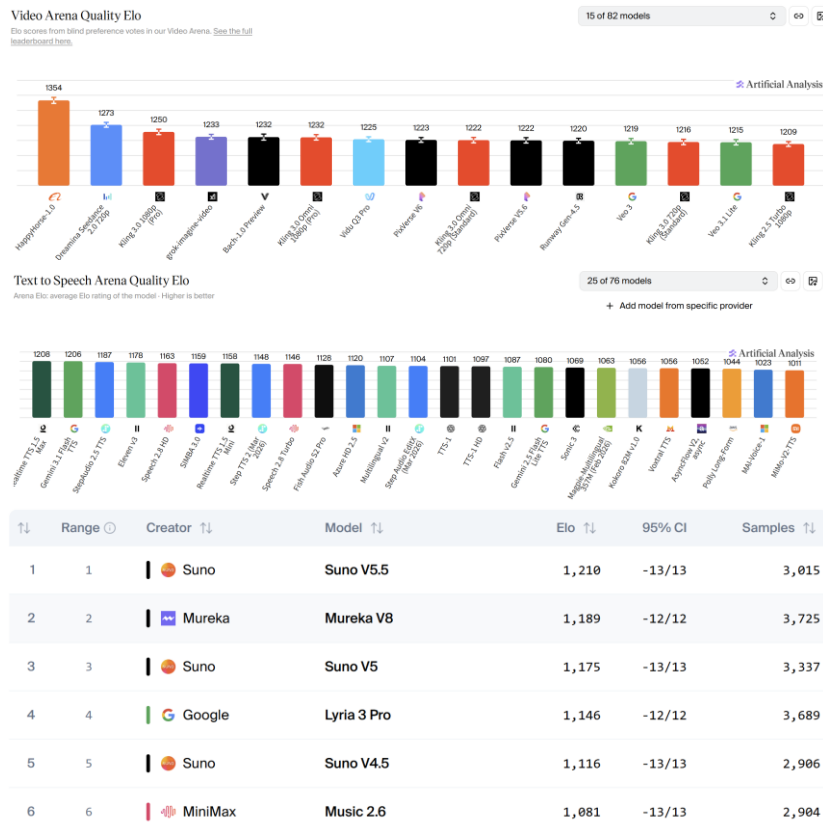
1.2 国产大模型厂商技术快速跟进，积极推进模型商业化

- 2026年，国产大模型在视频生成、音频/音乐生成领域居全球前列，截至5月14日，Artificial Analysis文生视频榜单前三模型分属于阿里巴巴、字节跳动、快手，而昆仑万维、MiniMax在音频领域全球领先。
- 商业化方面，快手、字节跳动结合旗下短视频平台探索视频生成模型变现，据晚点LatePost公众号，可灵2026年5月ARR已达5亿美元，比春节前翻倍。

图2：国产大模型厂商技术快速跟进，积极推进模型商业化

	2023年	2024年	2025年	2026年
图像	7月，阿里巴巴发布文生图模型通义万相，可基于参考图生成相似图及风格迁移 10月，腾讯混元文生图功能对外开放，同时建模中英文实现双语理解	7月，快手可图Kolors 1.0发布，采用 ChatGLM3 作为文本编码器，支持 256 字符的中英文提示词输入，远超过传统 CLIP 模型的 77 字符。模型应用在AI玩评、主站魔表、快影等多个场景	字节跳动4月发布Seedream 3.0，原生2K直出，3秒出图；9月发布4.0，推理速度较3.0提升超10倍，首次原生支持4K，首次将文生图、图像编辑、组图生成整合到单一模型架构 9月，混元图像3.0发布，全局准确率超业界领先的模型，原生多模态模型支持利用世界知识进行推理	Seedream 5.0首次支持检索生图；知识推理、编辑响应与一致性保持等能力提升显著，在真实落地场景中进步突出 Qwen-Image-2.0集生图和编辑于一体，支持1K tokens的超长文字输入和2K高分辨率
视频		字节跳动5月上线“即梦AI”网页版、AI作画和AI视频生成功能，9月发布豆包视频生成模型 6月，可灵上线，采用3D时空联合注意力机制，支持1080p分辨率、最长2分钟的视频生成；7月启动公测并推出付费会员，11月推出独立APP 9月，阿里巴巴通义万相上线首批文/图生视频功能 12月，腾讯HunyuanVideo可原生生成5-10秒时长的480p和720p视频，并可通过超分模型提升至1080p；可在14G显存的消费级显卡上流畅运行	快手4月发布可灵2.0，ARR由3月底的1亿美元提升至年底的2.4亿美元 字节跳动发布Seedance 1.5实现精准音画同步与多语言、方言支持 MiniMax发布Hailuo 2，原生支持1080p视频输出，价格约为同期发布的Google Veo3的1/9 通义万相Wan 2.2，采用高压压缩3D VAE架构，是上线时24帧每秒、720P像素级生成速度最快的基础模型；Wan 2.5生成时长提升至10秒，支持24帧每秒的1080P高清视频生成	1月，可灵3.0发布，支持3-15秒灵活时长生成；模型原生2K/4K像素级输出，最多10张参考图导入，ARR达3亿美元 字节跳动发布Seedance 2.0，在复杂交互场景下生成可用率达到业界 SOTA 水平，多模态能力显著强化，支持15秒多镜头音视频输出，上线供不应求 阿里巴巴发布原生多模态模型happyhorse 1.0，支持15秒多镜头叙事、多画幅适配及1080P超分输出
音频		昆仑万维Sky Music能生成80秒44100Hz采样率双声道立体声AI歌曲，并可根据用户输入的歌词风格生成对应歌曲风格。AI人声合成能够达到业内SOTA水平，中文效果显著好于国外产品 MiniMax发布abab-music-1、abab-speech-1	MiniMax持续迭代，2025年8月发布Speech 2.6，端到端延迟低于250毫秒，支持多种语言；2026年发布Music 2.5，开放全段落标签控制；扩充音色库至100+种乐器；speech 2.8支持40多种语言；语音模型接入智元机器人等硬件 昆仑万维Mureka V8靠自研MusicCoT技术在Artificial Analysis音乐模型榜单上同时拿下人声和乐器双料冠军，超过Suno V4.5和Udio；V9提升段落级歌词语义控制	
世界			7月，腾讯HY-World 1.0发布，业界首个开源可编辑的世界生成模型，将模型显存需求从原本的26GB优化至17GB以下 12月，HY-World 1.5发布，可按24FPS的速度生成720P的高清视频，实现实时交互，支持分钟级内容的几何一致性生成	多模态世界模型HY-World 2.0发布，对比主流世界模型仅能生成视频文件，HY-World 2.0支持多格式3D资产导出 阿里巴巴发布可实时构建和交互的开放式世界模型HappyOyster，可连续生成长达3分钟的720p实时视频

图3：据Artificial Analysis，国产大模型在视频生成、音频/音乐生成领域居全球前列



目录

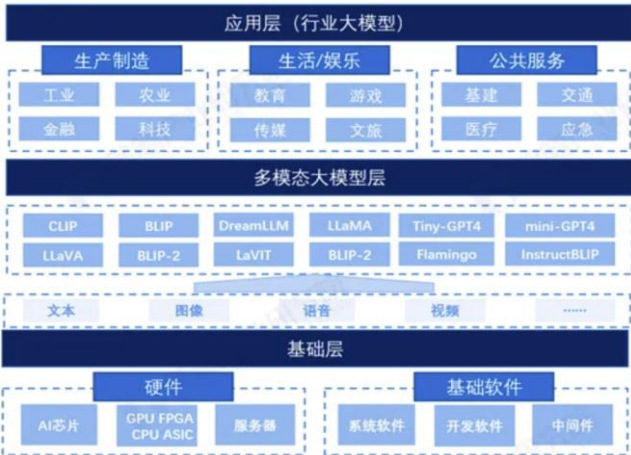
CONTENTS

- 1 全球多模态技术持续迭代，国产模型从追赶到局部超越
- 2 多模态AI商业化开启进击之路
- 3 投资建议及风险提示

2.1.1 AI大模型：多模态大模型下游应用广泛，市场高速增长

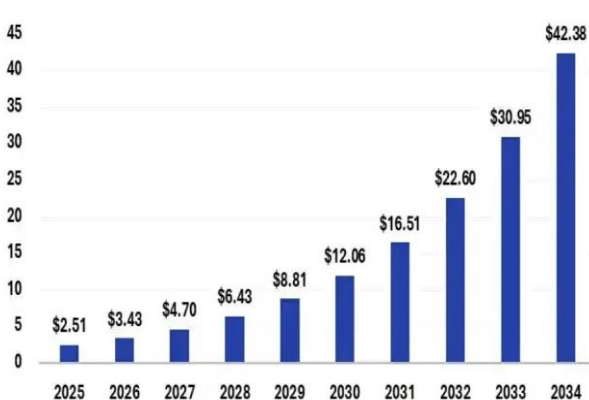
■ 图像生成领域，下游用户包括平面设计、影视、广告、电商等领域的专业创作者和个人爱好者，模型帮助其激发灵感并实质提升设计工作效率；视频生成领域，模型改变影视、短视频、短剧、广告、电商等行业工作流，允许一人工作室落地；音频生成领域，AI语音Agent在招聘、金融、医疗等场景的销售和客服中替代人力，在有声书教育、配音、游戏等场景提供贴近真人的语音能力。据前瞻产业研究院及Precedence Research，2025-2030年全球及中国多模态大模型市场规模CAGR或分别达37%/67%。

图4：多模态大模型下游应用覆盖生活娱乐、生产制造、公共服务等多领域



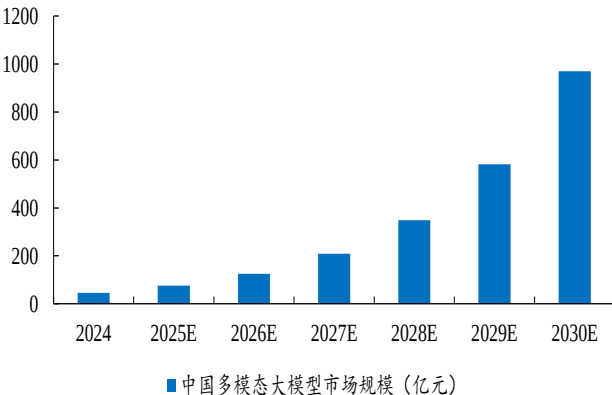
资料来源：前瞻产业研究院

图5：2025-2034年全球多模态大模型市场规模CAGR有望达37%（单位：十亿美元）



资料来源：Precedence Research

图6：2024-2030年，中国多模态大模型市场规模CAGR有望达67%



数据来源：前瞻产业研究院、开源证券研究所

2.1.2 AI大模型：快手、字节跳动、阿里巴巴居视频生成模型第一梯队

■ 快手可灵模型能力长期维持第一梯队，强调影视级质感（画质支持4K）；字节跳动Seedance 2.0、阿里巴巴HappyHorse 1.0快速崛起，其中Seedance 2.0的物理一致性、镜头调度能力广受好评，HappyHorse 1.0的生成效率、简单场景生成质感领先；昆仑万维推出全球首个同时支持多模态输入、联合音视频生成、统一生成/修复/编辑任务的视频基础模型。商业化上，可灵ARR已达5亿美元；Seedance 2.0上线后，即梦入口多次“涨价”；昆仑万维以自研模型赋能短剧业务。

表1：快手、字节跳动、昆仑万维等公司布局视频生成模型

公司	模型名称	模型发布时间	模型产品介绍	商业化
字节跳动	Seedance 2.0	2026/2	允许用户同时输入多达9张图片、3段视频、3段音频以及自然语言指令，在多主体交互和复杂运动场景中表现达SOTA水平，15秒多镜头音视频输出，双声道沉浸音效，最高支持1080P	据量子位实测，不含视频输入的标准视频生成（720p、24fps）对应1秒1元。面向部分企业客户推出千万保底协议（年消耗1000万元）
快手	可灵视频 3.0/3.0Omni	2026/2	AI自动调度景别与机位，图生视频支持额外添加多图/视频主体，支持多语种生成及国内多地方言，最长支持15秒，最高支持4K画质	据AI新榜实测，以尺寸16:9、分辨率1080p、长度15s为标准，每秒0.43元
Mini Max	Hailuo 2.3	2025/10	最高支持1080P、10s视频生成，刷新全球视频模型效果成本纪录，提供业内最优性价比	Hailuo系列累计生成超6亿个视频，单条1080P、6秒视频3.5元
智谱	CogVideoX-3	2025/7	支持生成最高4K、最长10s视频	1元/次
昆仑万维	SkyReels V4	2026/2	全球首个同时支持多模态输入、联合音视频生成、统一生成/修复/编辑任务的视频基础模型，实现1080p分辨率、32FPS 帧率、15秒时长的音视频同步生成，上线时Artificial Analysis 基准测试全球第二	每分钟8.4美元
阿里巴巴	HappyHorse 1.0、Wan 2.7	2026/4	HappyHorse 1.0支持15秒多镜头叙事、1080P超高清输出，上线时以1333Elo分登顶Artificial Analysis视频竞技场排行榜	HappyHorse 1.0：720P0.9元/秒，1080P为1.6元/秒
腾讯控股	HunyuanVideo 1.5	2025/11	可原生生成5-10秒时长的480p和720p高清视频，并可通过超分模型提升至1080p，对比主流20-30B体量，模型参数仅8.3B	开源模型，第三方平台模力方舟API每秒0.3元

资料来源：各公司公众号、各公司官网、智东西公众号、白鲸实验室公众号等、开源证券研究所

2.1.2 AI大模型：图像、音频生成模型能力各有侧重

■ 各公司主要基于自研模型与其他业务的协同效应选择发力方向，如阿里巴巴图像生成模型可编辑性强、文字输出准确、清晰度高，语音模型注重攻克客服场景痛点，可赋能淘系电商小B商家；MiniMax、昆仑万维模型生成内容“好听”，助力同公司旗下AI伴侣或音乐内容产品。

表2：字节跳动、阿里巴巴、快手等公司布局图像生成模型

公司	模型名称	模型发布时间	模型产品介绍	商业化
字节跳动	Seedream 5.0 Lite	2026/2	模型采用 多模态统一架构 ，提升跨模态理解和推理能力，内置丰富的世界知识，生成结果更符合物理规律， 首次引入实时检索增强能力 ，可通过联网检索更精准地回应具有时效性的创作需求， 支持4K图像生成	据火山引擎，单张0.22元
快手	可灵图片 3.0/3.0 Omni	2026/2	原生直出2K/4K，支持最多10张参考图	API调用下，4K单张原价0.4元，开通会员有折扣
MiniMax	Image-01	2025/4	支持文本描述生成多种尺寸的图片	每张图片0.025元
智谱	GLM-Image	2026/1	模型全程基于国产芯片完成训练，采用独创的“自回归+扩散解码器”混合架构，兼顾全局指令理解与局部细节刻画，支持生成多分辨率图片	0.1元/次
阿里巴巴	Wan2.7-Image	2026/4	支持 最高3K Token文字输入 ，文字输出能力强，支持12种语言、参考图色彩精准迁移、交互式编辑模块、 9张参考图，一次最多生成12张组图，pro版支持4K	0.20元/张
腾讯控股	HunyuanImage 3.0-Instruct	2026/1	首个工业级原生多模态生图模型，参数规模80B，可解析千字级别语义	开源模型

资料来源：各公司公众号、各公司官网、开源证券研究所



表3：MiniMax、昆仑万维等布局音频生成、音乐生成模型

公司	模型名称	模型发布时间	模型产品介绍	商业化
字节跳动	语音大模型 Seeduplex	2026/4	首个 全双工语音模型 ，抗干扰及动态判停能力提升，对比豆包App原半双工模型架构，误回复率和误打断率减少一半，判停延迟降低约250ms，AI抢话比例相对减少40%，打断响应延迟缩短约300ms	已在豆包App全量上线
MiniMax	音频生成模型 Speech 2.8	2026/1	支持40多种语言 ，每种语言配备不同的音色	Speech系列覆盖超200个国家和地区，累计生成超2亿小时语音内容，3.5元/万字符
MiniMax	音乐生成模型 Music-2.6	2026/4	模型更新引入翻唱功能，首包延迟压缩到20秒以内，乐器种类更丰富，中低频音质提升	1元/首
智谱	语音合成模型 GLM-TTS	2025/12	仅使用10万小时训练数据，远低于行业主流商用模型，单机4天可得到开源SOTA“发音准确度”与超高“音色还原度”，单机1天可完成精品音色LORA和强化学习训练， 以较低的价格获得行业领先的MOS分数	2元/万字符
昆仑万维	音乐生成模型 Mureka V9	2026/3	V8已在Artificial Analysis登顶人声、乐器双榜单，V9在段落内文本控制、生成效率、混音质量与整体听感等多个关键维度持续进化	基于订阅及用量付费，API调用下单首歌曲低至0.25元，截至2025年3月底，Mureka ARR约1200万美金
阿里巴巴	语音生成模型 Qwen3-TTS	2026/1	音色创造任务上，指令遵循能力和生成表现力整体大幅领先其余开源模型；音色控制任务上，一次性合成10分钟语音的中英词错率为2.36/2.81%；音色克隆任务上，中英文克隆的语音稳定性表现上均超越MiniMax和SeedTTS	qwen3-tts-vd-2026-01-26每万字符输入单价0.8元

资料来源：各公司公众号、各公司官网、艾瑞网、量子位公众号、开源证券研究所

2.1.2 AI大模型：腾讯、昆仑万维、阿里巴巴等布局世界模型

■ 3D生成/世界模型方面，头部厂商模型均可生成720P、分钟级稳定世界，为游戏、自动驾驶、具身智能等场景应用奠定基础，其中对比主流世界模型仅能生成视频文件，腾讯HY-World 2.0支持多格式3D资产导出，与现有的游戏工作流无缝对接。

表4：腾讯、昆仑万维、阿里巴巴等布局世界模型

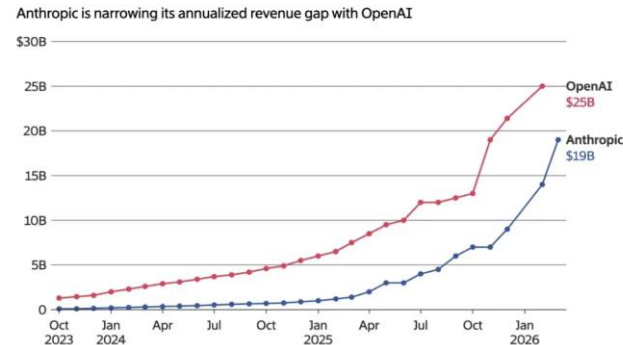
公司	模型类型及名称	模型发布时间	模型产品介绍	商业化
字节跳动	3D生成模型Seed3D 2.0	2026/4	几何生成精度、空间结构推理能力提升，在行业常用的Pairwise Comparison盲测中，相比所有参与对比的主流模型均呈现更高的偏好率	2.40元/次（3万Token/次*0.80元/万Token）
昆仑万维	世界模型Matrix-Game 3.0	2026/3	同时实现720P实时生成与分钟级长时序稳定演化	/
阿里巴巴	世界模型HappyOyster	2026/4	全球首个主动式实时交互世界模型，支持多模态输入与音视频联合生成，支持720P、480P，导演/漫游模式世界分别可维持3/1分钟	/
腾讯控股	世界模型HY-World 2.0	2026/4	多模态世界模型，支持多格式3D资产导出	/

资料来源：各公司公众号、各公司官网、机器之心公众号、量子位公众号、开源证券研究所

2.1.3 AI大模型：多模态大模型应用深化加速国产模型Token调用量/ARR增长

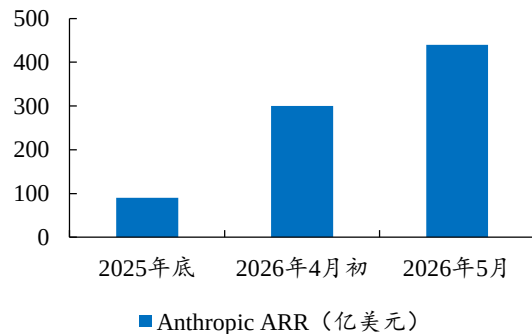
■ 据火山引擎，豆包大模型日均Token调用量由2024年5月的1200亿增长千倍至2026年3月的120万亿，2月发布的Seedance 2.0视频模型加速Token调用量提升。智谱披露，截至2026年3月，API ARR已突破2.5亿美元，同比增长60倍，年初以来增长6.4倍。MiniMax披露2026年2月公司ARR超1.5亿美元，面向企业客户和个人开发者的开放平台2026年2月新注册用户数达2025年12月的4倍以上。2023年以来，Open AI、Anthropic收入增速持续提升，2026年5月，Anthropic ARR超440亿美元，比2025年底的90亿美元提升389%。对比海外大厂，我们认为，国产模型厂商多模态AI业务空间广阔，技术追赶及算力扩充有望推动Token调用量/ARR增长进一步加速。

图8：2023年以来，Open AI、Anthropic收入增速逐步提升
Revenue Duel



资料来源：The Information

图9：2026年5月，Anthropic ARR超440亿美元，对比2025年底90亿美元提升389%



数据来源：Saasverse公众号、第一财经公众号、新财富公众号、开源证券研究所

图7：大模型厂商Token调用量/ARR高速增长



资料来源：各公司公众号、各公司官网、机器之心公众号、量子位公众号、开源证券研究所

2.1.4 AI大模型：多模态大模型应用拓展或显著拉升模型Token消耗量

■ 参考主流多模态大模型，单秒音频/单张图像/单秒视频对比单个文字，输入Token消耗量高1-2/3/2-3个量级，输出Token消耗量高1/2-3/4个量级，我们测算仅国内视频消费场景每日所需Token即有望达350万亿。

表5：多模态理解、生成模型Token消耗量显著高于文本模型

环节	类型	Token消耗量计算	注释
输入/输出	文本	1 Token=1-2个汉字	换算关系与模型分词方式相关
		1 Token=0.6-0.8英文单词	
输入（理解）	图像	1张1080P低细节理解模式图片=1312 Token 1张1080P高细节理解模式图片=2645 Token	换算关系与图片缩放方式等相关，参考doubao-1-5-vision-lite
输入（理解）	音频	1秒=32 Token	参考Gemini API
输入（理解）	视频	1秒=263 Token	视频可视作音频及从视频中截取的各帧图像的集合，参考Gemini API
输出（生成）	图像	1张1536*1024低质量图片=408 Token 1张1536*1024高质量图片=6208 Token	参考gpt-image-2
输出（生成）	音频	1秒=6.25 Token	参考VoxCPM 1.5
输出（生成）	视频	仅参考图片生成5秒480P视频=50220 Token 仅参考图片生成5秒720P视频=108000 Token 仅参考图片生成5秒1080P视频=243000 Token	换算关系与生成视频帧数、分辨率、时长等相关，参考doubao-seedance-2-0 API

资料来源：各公司公众号、各公司官网、机器之心公众号、量子位公众号、开源证券研究所

表6：国内AI视频或带来每日近350万亿Token消耗量

指标	数量	备注
国内中长视频日新增量（万条）	100	2025年B站月均视频投稿量近2400万条，考虑B站为国内主要中视频平台，假设中长视频新增量为B站日均增量的1.25倍
中长视频平均时长（分钟）	10	据新闻世界公众号，B站视频内容时长集中于5-15分钟
国内短视频日新增量（亿条）	1.3	据《中国短视频发展研究报告（2025）》，到2024年12月，我国短视频创作者账号规模已达16.2亿，每日全国上线短视频突破1.3亿条
短视频平均时长（分钟）	0.5	据东南传播公众号，抖音对一般用户的时长限制是15秒~60秒
国内视频日新增时长（亿秒）	45	
AI视频生成所需Token（万亿）	349.9 ^{40%} 2	考虑视频生成模型降低内容创作门槛，假设AI视频日新增时长为国内日新增视频时长，对应18亿秒；考虑创作成功率，假设每生成4条视频可获得1条成品；考虑用户画质要求，参考doubao-seedance-2-0生成5秒1080P视频需243000Token，对应单秒视频消耗4.86万Tokens

资料来源：新闻世界公众号、特看观察公众号、火山引擎、东南传播公众号、《中国短视频发展研究报告（2025）》、开源证券研究所

2.1.5 AI大模型：看好多模态模型发展带动各厂商Token调用量及收入高增

■ 重点推荐：腾讯控股、快手，受益标的：Minimax、阿里巴巴、智谱、昆仑万维。

表7：腾讯控股、快手等多模态大模型厂商有望受益

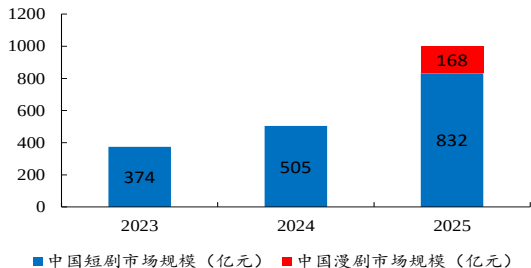
公司	业务布局
腾讯控股	多模态大模型布局广泛，对比行业主流世界模型，2026年4月发布并开源的HY-World 2.0不再只输出无法编辑的视频，而是直接生成可二次修改、可交互的3D资产文件，且文件可无缝对接主流游戏引擎，与游戏业务深度协同
快手	公司拥有图像生成及视频生成模型，是国内最早实现视频模型规模商业化的厂商，视频生成技术稳居第一梯队，服务超过6000万创作者，累计生成超过6亿条视频，并与超过3万家企业用户建立合作。可灵2025年12月/2026年1月ARR分别超2.4/3亿美元
MiniMax	公司构建了全模态的研发能力，语言、视频、语音、音乐等各主要模态均拥有具备全球竞争力的模型，广泛技术储备支撑公司开发全模态全能创作、一键成片的创作工具，亦赋能旗下AI陪伴产品《talkie》。2026年2月公司ARR超1.5亿美元，面向企业客户和个人开发者的开放平台产品2026年2月新注册用户数达2025年12月的4倍以上。
阿里巴巴	各主要模态模型水平居全球一线，其中视频生成模型HappyHorse 1.0上线时登顶Artificial Analysis视频技法排行榜。2026年3月19日，阿里百炼MaaS平台上公共模型服务市场的Token消耗规模三个月内提升了6倍
智谱	公司布局图像、视频、音频生成模型，其中图像生成模型GLM-Image完全基于国产芯片训练，语音合成模型GLM-TTS以较低的价格获得行业领先的MOS分数。截至2026年3月，API ARR已突破2.5亿美元，同比增长60倍，年初以来增长6.4倍。
昆仑万维	视频生成模型SkyReels V4上线时Artificial Analysis 基准测试全球第二，音乐生成模型Mureka V8登顶Artificial Analysis人声、乐器双榜单，世界模型Matrix-Game 3.0可实现实现720P实时生成与分钟级长时序稳定演化。自研模型与旗下AI社交产品《Linky》、AI音乐创作平台《Mureka》协同发展，其中《Mureka》2025年3月访问量达333万，ARR约1200万美元

资料来源：豚基Wisbase公众号、字母AI公众号、量子位公众号、涌金楼公众号、寰球新锐商业公众号等、开源证券研究所

2.2.1 AI视频：2025年国内短剧市场翻倍增长，海外短剧市场潜力初步释放

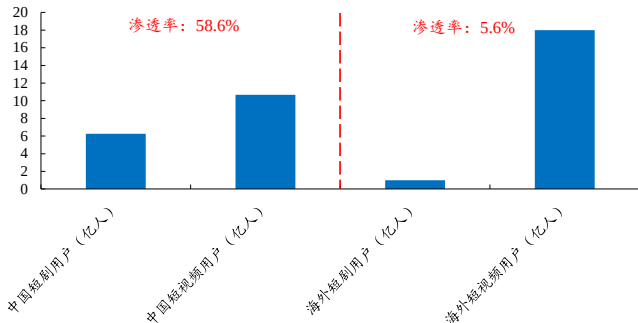
■ 2025年中国短剧市场同比+98%，主要得益于免费微短剧市场的持续扩容（约占真人微短剧整体市场的2/3）。2025年海外短剧收入23.8亿美元，同比+263%，其中AI短剧/漫剧市场规模仅1亿美元。

图10：2025年中国短剧市场规模达1000亿元，其中漫剧市场规模约168亿元



数据来源：DataEye短剧观察公众号、开源证券研究所

图12：参考国内短剧渗透率，海外短剧用户规模提升空间广阔



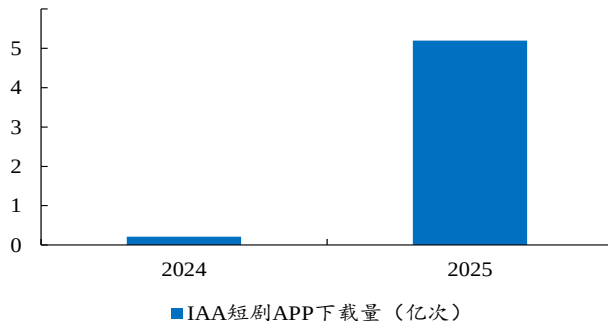
数据来源：澎湃新闻公众号、DataEye短剧出海公众号、开源证券研究所
(注：中国数据为截至2025年6月情况，海外数据为《DataEye 2025年海外微短剧数据报告》中预估)

图11：2025年海外短剧收入23.8亿美元，同比+263%，下载量突破12.1亿次，同比+135%



资料来源：短剧自习室公众号（收入为付费点播、会员订阅的净收入总额，纵坐标单位为收入/美元、下载/次）

图13：活跃IAA短剧应用下载量由2024年的0.2亿次增长至2025年的5.2亿次



数据来源：Appfigures、短剧自习室公众号、开源证券研究所

2.2.1 AI视频：AI真人剧越过奇点，有望重塑海内外真人剧市场

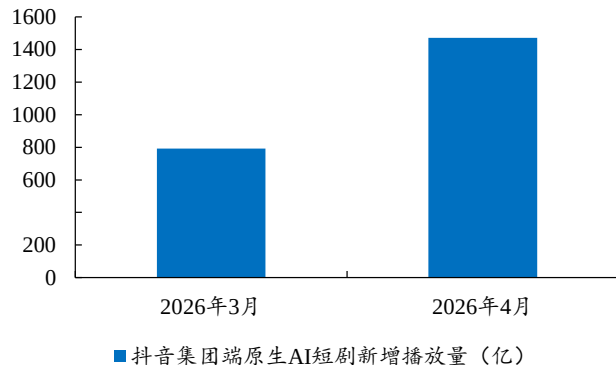
- 我们认为AI真人剧内容质量已获观众认可。2026年4月《菩提临世真人AI版》超越系列真人短剧，连续数日登顶红果热播总榜；两部由中国团队、全AI制作的竖屏短剧《摸金之天机入梦》、《饿塔》入选第79届戛纳国际电影节Fantastic Pavilion官方展映单元。
- AI大幅降本，据短剧自习室公众号，一部100分钟左右的短剧真人实拍成本普遍在50-100万元，头部精品项目成本可达200万元，而AI技术将单部短剧的制作成本下探至5-10万元。此外，AI降低塑造宏大世界观所需的复杂场景变换和特效设计成本，有效拓宽真人剧题材选择范围。针对海外市场本地化难题，AI翻译不仅贴合当地语言习惯与流行表达，成本也仅为人工翻译的10%，且AI制作视频成本仅为传统海外实景拍摄成本的20%。

图14：《菩提临世真人AI版》登顶红果热播总榜



资料来源：短剧自习室公众号

图15：2026年4月，仅抖音集团端原生AI短剧新增播放量突破1472亿，环比增长85.8%

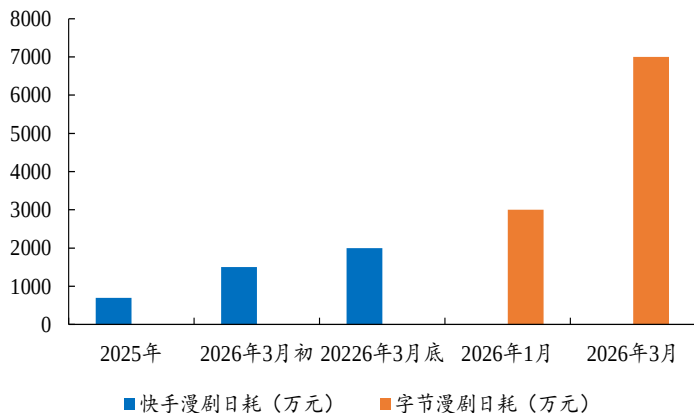


数据来源：短剧自习室公众号、开源证券研究所

2.2.1 AI视频：二次元文化受众广泛，AI大幅提升漫剧产能

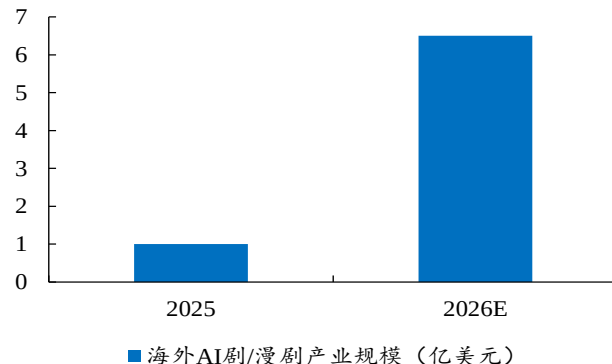
■ 据DataEye短剧出海公众号，全球动画市场规模约4500亿美元，狭义动漫市场约400亿美元。AI赋能下，动漫制作成本由传统方式下数万元/分钟降低至0-1000元/分钟，据吉星文化，应用AI工具后单人生产效率提升10倍。视频模型进步有望推动漫剧市场高增，2026年3月，快手及字节漫剧日耗峰值加总达9000万元，考虑整体投放回报应高于费用，国内漫剧单日流水或已近亿元；据DataEye研究院预估，2026年海外AI剧/漫剧市场规模预计达6.5亿美元，同比实现6倍增长。

图16：2026年3月，快手及字节漫剧日耗峰值加总达9000万元



数据来源：新腕儿公众号、快手磁力引擎商业洞察公众号、短剧自习室公众号、开源证券研究所

图17：据DataEye研究院预估，2026年海外AI剧/漫剧市场规模预计达6.5亿美元，同比实现6倍增长



数据来源：DataEye短剧出海公众号、开源证券研究所

2.2.1 AI视频： AI电影跑通上映路径，制作方式变革或创造电影行业新增长点

- 2026年4月，博纳影业耗时超2年出品制作的《三星堆：未来往事》正式获得国家电影局颁发的“龙标”，标志AI原生电影符合中国电影审查标准，可登陆院线。我们认为《三星堆：未来往事》上映仅是AI电影发展的起点，多模态大模型进步将持续推动AI电影成片效果、制作效率及成本优化，孕育行业机遇。

图18：电影《三星堆：未来往事》由博纳影业 AIGMS 制作中心集体创作，已获“龙标”



资料来源：博纳影业公众号

图19：博纳影业系列AI电影项目陆续推进

博纳影业 2026 年 AI 电影项目经营计划

序号	作品名称	类型	项目进度
1	《三星堆：未来往事》	科幻/剧情	已通过内容审查并取得龙标
2	《中华志怪之罗刹海市》	动画	已立项公示
3	《计划外：广寒行动》	军事/科幻	立项备案中
4	《东风必达》	军事/科幻	立项备案中

资料来源：博纳影业公告

2.2.1 AI视频：看好布局IP、AIGC工具、短剧分发平台及AI影视制作公司

■ 我们认为，AI技术进步将持续提升优质视频产能，优质内容上线或有助于短剧分发平台聚拢更多流量并提升ARPU；短剧景气或吸引更多制作团队入局，AIGC工具作为“卖铲人”，或持续受益于下游行业增长；成熟IP能为短剧提供完备世界观与流量基础，价值或进一步凸显。重点推荐：阅文集团，受益标的：中文在线、德才股份、博纳影业。

表8：阅文集团、中文在线、德才股份积极布局短剧行业

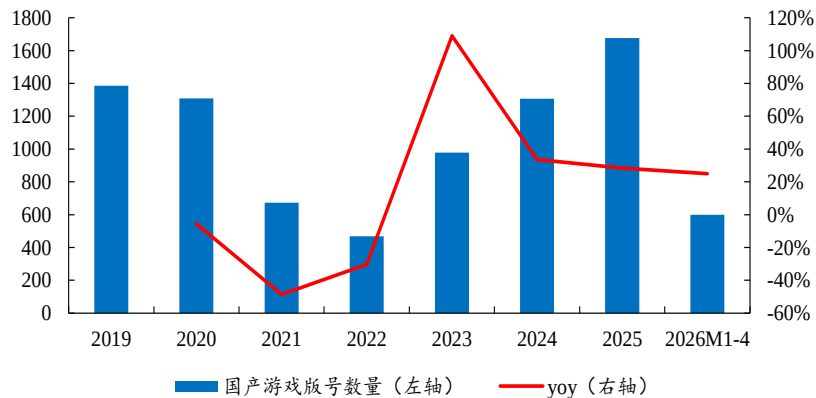
公司	业务布局
阅文集团	公司推出漫剧助手一站式全链路工具，内容供给层面，开放10万+优质版权IP，通过AI实现5分钟完成成长篇网文深度解读，助力编剧深度挖掘故事细节、精准设定钩子，快速产出贴合需求的基础剧本；技术提效层面，工具一站式整合剧本生成、风格定制、角色/场景定妆、分镜制作、视频生成等全环节，可实现单部作品效率提升300%、综合成本降低50%。公司亦通过投资酱油动漫等参与剧集制作，2025年短剧上线数超120部，其中标杆项目总流水超8000万元；AI漫剧H2推出后，已上线近千部作品，其中超百部播放量破千万，12部播放量破亿，收入超1亿元。
中文在线	据短剧自习室公众号，旗下海外短剧APP FlareFlow 2025年收入/下载量市占率均为1.84%，分列排行榜第10/17；参股公司CMS（持股超49%）旗下ReelShort 2025年收入/下载量市占率分别为23.14%/11.11%，分列排行榜第1/2。公司自有内容平台积累丰富IP资源及创作者，截至2025年末，平台数字内容创作者注册作家超400万名。2025年，公司出品的现象级短剧《一品布衣》总播放量突破30亿次，成为2025年红果平台上仅有的两部观看量突破该里程碑的男频短剧之一。截至2025年末，公司漫剧类产品产量达数百部，累计播放量达数十亿次。
德才股份	旗下奇想无限率先接入并深度应用Seedance2.0等国产顶级多模态模型，生产内容获得火山引擎官方的高度认可，与红果、阅文等头部平台深度合作。自研AIGC智能体平台奇想AI将传统动画11个环节压缩至5步，实现“1人1天完成1分钟精品内容”。
博纳影业	AI影视项目中，《三星堆：未来往事》已获得国家电影局颁发的“龙标”，改编自《聊斋志异》的动画《中华志怪之罗刹海市》已经立项公示；科幻电影《计划外：广寒行动》、《东风必达》在积极筹备和立项备案中。

资料来源：各公司公告、各公司公众号、短剧自习室公众号、DataEye短剧观察公众号、开源证券研究所

2.2.2 AI游戏：游戏行业供需两旺

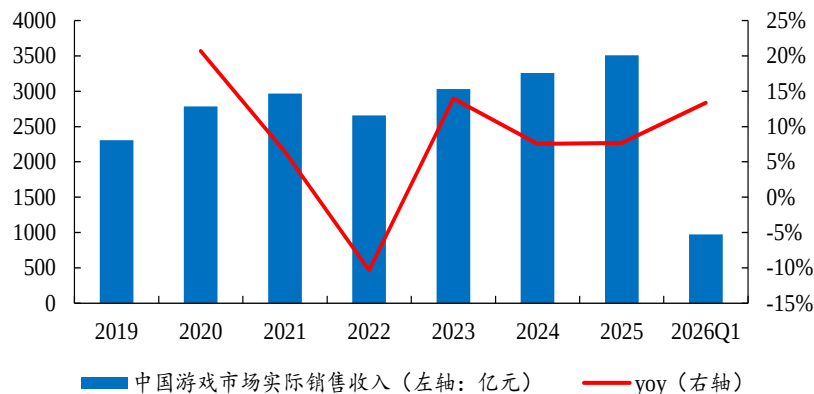
■ 供给侧，版本号发放数量自2022年触底以来稳步回升，2026年1-4月国产游戏发放版本号数为600个，同比增长25%。需求侧，游戏作为情绪消费的重要赛道，市场规模持续扩容，2026Q1中国游戏市场累计收入约972亿元，同比增长13%。

图20：2026年1-4月国产游戏发放版本号数为600个，同比增长25%



数据来源：国家新闻出版署、财经网科技公众号、游戏葡萄公众号、游戏智库公众号、蓝鲸新闻公众号等、开源证券研究所

图21：2026Q1中国游戏市场累计收入约972亿元，同比增长13%



数据来源：伽马数据、游戏产业报告公众号、开源证券研究所

2.2.2 AI游戏：AI逐步赋能游戏开发全流程

■ 从早期的美术资源生成、代码辅助编写等辅助性工作，逐步向玩法设计、剧情架构等核心创意领域渗透，有望持续提升游戏研发效率。据腾讯内部统计，AI赋能下，光子工作室的角色动画制作周期缩短了40%，部分项目组角色技能原型验证时间从2周压缩至3天。

表9：AI重构游戏开发流程，大幅降本增效

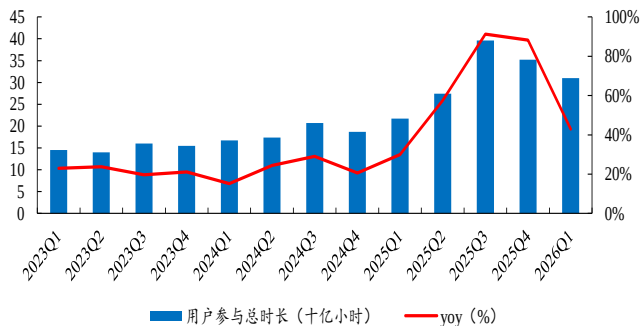
环节	原流程及工作量	AI赋能下流程及优化效果
动画制作-骨骼蒙皮	蒙皮是动画制作中骨骼绑定环节的重头戏，占该环节70%-80%的工作量，是人力重投入的环节	AI自动蒙皮工具将原本需要1.5天到3.5天才能完成的工作，缩短至1到3小时，节省70%-90%人工
动画制作-关键帧生成	手工K帧（逐帧调整）是行业主流，占单角色动画制作总工时（可长达数周）的60%-70%，是人力成本的主要消耗点	使用动画生成大模型，根据少量关键帧（5~10帧）自动生成完整的流畅动画（如60帧），将效率提升3~5倍。
3D场景与资产构建	游戏原型外景需用无人机和3D扫描设备扫描，过程依赖取景团队、现场指导、当地地陪等	次级资源可由3D生成工具直接生成，对于美术成本超亿元的3D类产品来说，其中20%-30%的次级资源可以用AI工具解决，由此达到千万元级别的降本效果

资料来源：第一财经公众号、开源证券研究所

2.2.2 AI游戏：分发平台/UGC生态/开放世界游戏或受益于AI带来创作技术门槛下降

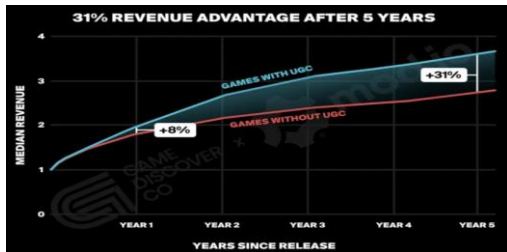
■ 我们认为世界模型及3D生成模型的发展或进一步提升UGC创作工具能力，进而强化分发平台及UGC生态游戏的内容正循环；此外，开放世界游戏具备容纳丰富玩法的潜力，亦有望受益于AI发展。

图22：2023Q1以来Roblox用户总时长同比持续增长



数据来源：海豚研究公众号、开源证券研究所

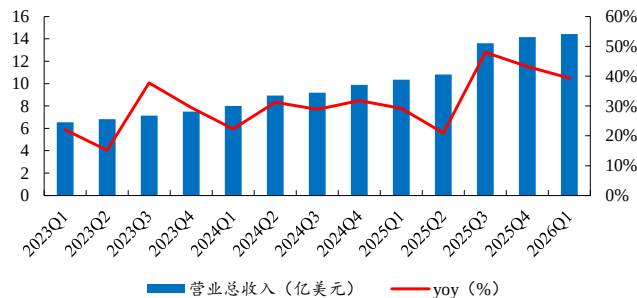
图24：据Mod.io，对比未支持UGC的产品，支持UGC或Mod的游戏五年累计收入提升约31%



资料来源：DataEye游戏观察公众号

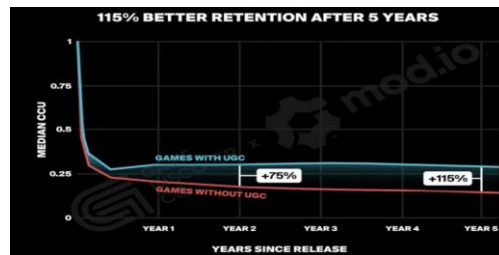
KYSEC

图23：2023Q1以来Roblox营收逐季攀升



数据来源：Wind、开源证券研究所

图25：据Mod.io，对比未支持UGC的产品，支持UGC或Mod的游戏五年用户留存提升超过115%

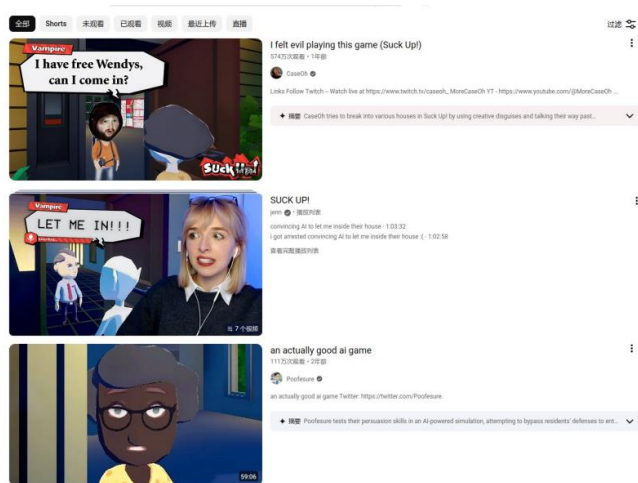


资料来源：DataEye游戏观察公众号

2.2.2 AI游戏：AI重塑游戏交互体验，其中叙事类、社交竞技类游戏率先受益

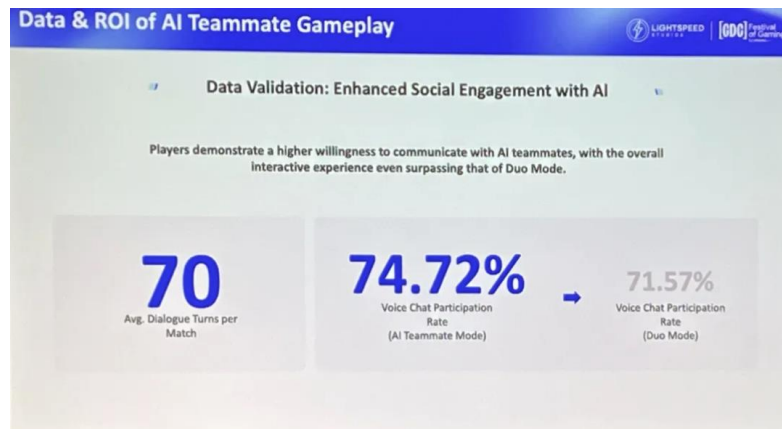
■ AI驱动的动态剧情生成系统能够根据玩家实时决策，自动生成后续剧情，打破线性叙事局限，提升游戏可玩性。《Suck Up!》在零营销预算的条件下，在YouTube上积累超1亿次播放量，反映AI赋能下叙事类玩法潜力。在大DAU竞技游戏中，社交压力影响部分用户游戏体验。《和平精英》通过引入“AI队友”显著提升玩家总对局数，截至2026年3月，所有AI NPC相关模式已吸引超1.1亿玩家体验，最高日活达1770万，“AI队友”模式下玩家开麦交流率超过真人双排模式，达到近75%的历史新高。

图26：YouTube中《Suck Up!》相关创作内容播放量累计破亿



资料来源：GameLook公众号

图27：《和平精英》“AI队友”系统显著提升社交压力大的玩家的总对局数



资料来源：游戏那点事Gamez公众号

2.2.2 AI游戏：看好AI对游戏产业整体赋能

■ 我们看好AI对游戏开发运营的整体降本增效，以及多模态大模型技术进步丰富原生AI玩法，扩大增量市场；游戏分发平台或凭门槛更低、能力更强的AI创作工具打造增长飞轮。重点推荐：完美世界、网易、巨人网络、恺英网络、心动公司、哔哩哔哩，受益标的：世纪华通。

表10：完美世界等游戏公司或受益于AI游戏发展

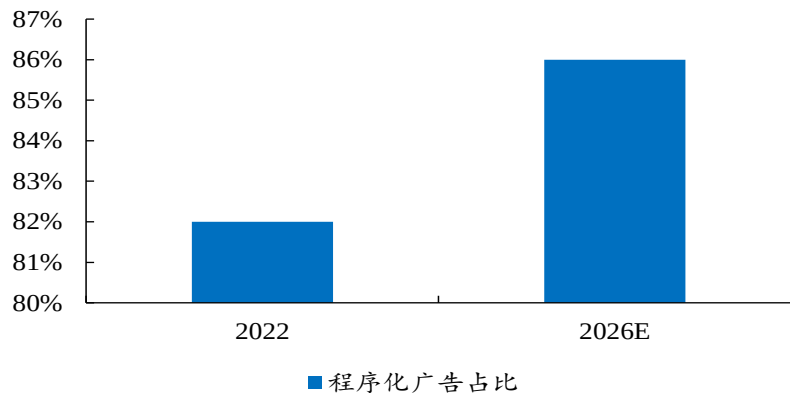
公司	业务布局
完美世界	旗下原创超自然都市开放世界RPG游戏《异环》全渠道预约量超3500万，全球首日流水超过1亿元人民币，玩家新增与留存等核心指标均显著优于公司此前旗舰产品《幻塔》的同期水平，同时《异环》的都市题材可容纳丰富副玩法，AI赋能或提升游戏成为大DAU长青游戏概率。此外，公司运营Steam本土化平台“蒸汽平台”。
网易	旗下拥有《蛋仔派对》《第五人格》等大DAU游戏，其中《蛋仔派对》UGC生态繁荣，2023年8月达成月活用户和UGC地图双破亿成就，2025年10月，《蛋仔派对》联合火山引擎，基于豆包图像生成模型及网易3DAI管线实现3分钟内输出可直接交互的3D模型。
巨人网络	旗下《超自然行动组》DAU超千万，依托阿里云、火山引擎、腾讯云等功推出国内首个千万DAU级游戏原生AI玩法“AI假人”。截至2026Q1末，该AI玩法参与对局数已突破4亿场，累计生成AI假人数量超29亿。
恺英网络	投资企业自然选择推出全球首款3DAI智能陪伴应用《EVE》。旗下极逸人工智能与中手游合作，推出由AI驱动的市场洞察、游戏开发到全球发行的一站式智能体GamePartner.AI，平台由极逸人工智能提供核心AI游戏开发技术支持。
心动公司	旗下TapTap为国内头部游戏社区，2025年MAU达4497万。2026年1月，公司推出“TapTap制造”，产品底层由“AI智能体+AI Native引擎+TDS（TapTap开发者服务）”三部分构成，覆盖开发技术、美术素材、音乐音效、TDS功能接入、一键上架等，创作者无需脱离聊天窗口即可完成从零到上线的全过程，真正实现游戏开发“0门槛”。截至2026年3月，“TapTap制造”每日上线新游数最多超20款。
哔哩哔哩	公司拥有国内头部游戏社区，且计划2026年加大与核心业务高度契合的AI应用相关投入。公司加大小游戏布局，2025年初，B站在不到2个月的内测时间里上线超100款活跃的小游戏。2026年，小游戏平台已有单款游戏月耗超1000万元。
世纪华通	旗下SLG、休闲游戏表现优异，《Whiteout Survival》/《Kingshot》/《Tasty Travels: Merge Game》分列2025年Sensor Tower中国手游境外收入榜第1/3/27，多模态大模型发展或加速公司游戏副玩法迭代并提升营销素材生产效率。

资料来源：各公司公告、各公司公众号、游戏新知公众号、游戏日报公众号、游戏葡萄公众号、GameLook公众号、竞核公众号、开源证券研究所

2.2.3 AI营销：程序化广告市场稳步增长

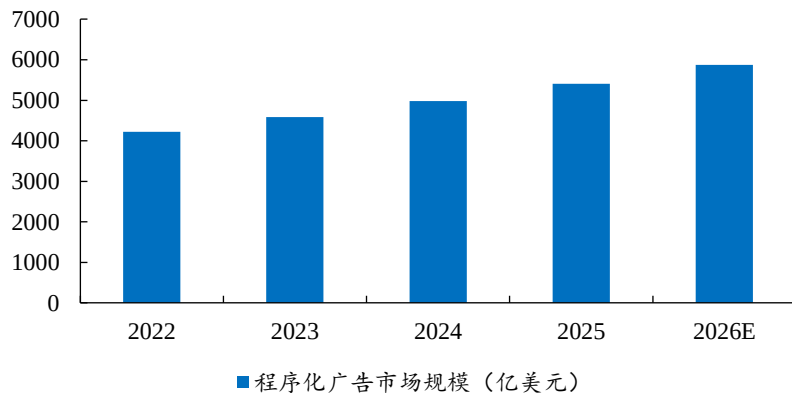
■ 传统的人工广告优化依赖经验和试错，而AI驱动的自动化系统能实时分析海量数据，做出更精准的决策。据Statista，2022-2026年，程序化广告收入有望从4225亿美元增长至5875亿美元，CAGR达8.6%，对应广告行业收入占比由82%提升至86%。

图28：程序化广告收入在广告行业的占比有望从2022年的82%提升至2026年的86%



数据来源：Statista、开源证券研究所

图29：2022-2026年，程序化广告收入有望从4225亿美元增长至5875亿美元，CAGR达8.6%



数据来源：Statista、开源证券研究所

2.2.3 AI营销：大模型有助实现深度个性化广告，AI应用商业化或驱动全球市场增长

- 快手生成式推荐广告系统GR4AD已全量部署于内部广告平台，服务超4亿用户，对比基准模型，该广告系统商业化广告收入高4.2%，中小广告主投放量提升17.5%，广告转化率提升10.17%，低活用户转化率提升7.28%。
- 据21世纪经济，2026年2月，Open AI启动ChatGPT广告测试，6周实现超1亿美元的年化收入，CPM最初定价60美元起，显著高于Instagram等海外主流媒体2025Q4 10美元左右的水平（数据源自EMARKETER），我们认为这或与ChatGPT可基于多模态上下文内容进行精准推荐相关。以ChatGPT为代表的拥有高单位流量价值的AI应用商业化推进或驱动全球广告市场增长。

表11：多模态大模型推动广告从“千人千面”迈向“一人千面”的深度个性化

新技术	传统方案局限性	新技术优势
生成式召回	传统召回方法下，系统推荐与广告边界被素材库的大小和多样性所严格限制，另外，对于新用户的冷启动问题或兴趣独特用户的长尾问题，检索式召回难以找到合适的匹配项，常常出现“词不达意、意图漏检”的局限性	生成式召回根据对用户需求和品牌调性的深度语义理解，即时生成全新的文案、视觉组合甚至互动叙事，从而突破素材库的限制
AI数据增强	大规模标注数据的获取通常伴随较高成本与较长周期，易使模型在有限样本下出现过拟合现象，进而降低模型在新数据上的泛化性能	数据增强技术借助对已有数据实施多类型变换处理，可人工生成全新训练样本，从而显著扩大训练数据集的规模
多模态理解	广告引擎无法理解图片与视频，对内容的理解仅停留在标签层	多模态理解技术通过融合异构数据，能够构建更丰富的用户画像和广告内容表征，超越了单一模态的局限

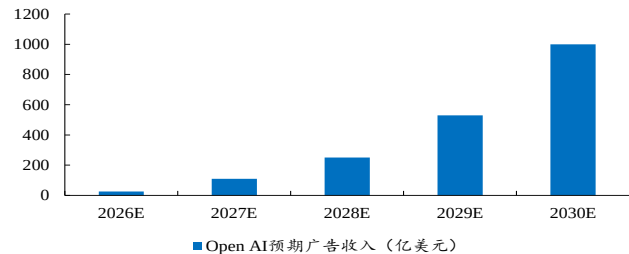
资料来源：现代广告杂志社公众号、开源证券研究所

图30：快手生成式推荐广告系统GR4AD在广告收入和推理性能上均较基准模型优秀

Table 1: Overall performance and ablation of GR4AD.		
Model Settings	ΔRevenue vs. Base	ΔQPS vs. GR-Base
<i>Baselines</i>		
DLRM (Base)	-	-
OneRec-V2 [40] (GR-Base)	+1.68%	-
<i>Tokenization Optimizations</i>		
+ UA-SID	+1.92%	0%
<i>Learning Optimizations</i>		
+ VSL	+2.80%	-25%
+ VSL + DPO [24]	+3.16%	-25%
+ VSL + GRPO [26]	+3.21%	-25%
+ VSL + RSPO	+3.86%	-25%
+ Unified VSL & RSPO (UVR)	+4.01%	-25%
<i>Serving Optimizations</i>		
+ UVR + DBS	+4.32%	+20%
+ UVR + DBS + DeepSeek-MTP [5]	+3.98%	+117%
GR4AD (+ UVR + DBS + LazyAR)	+4.28%	+117%

资料来源：机器之心公众号

图31：Open AI预计其广告收入将由2026年的25亿美元提升至2030年的1000亿美元，CAGR达151%



数据来源：科创板日报公众号、开源证券研究所

2.2.3 AI营销：AI显著提升移动应用/数字内容产量，广告曝光重要性或提升

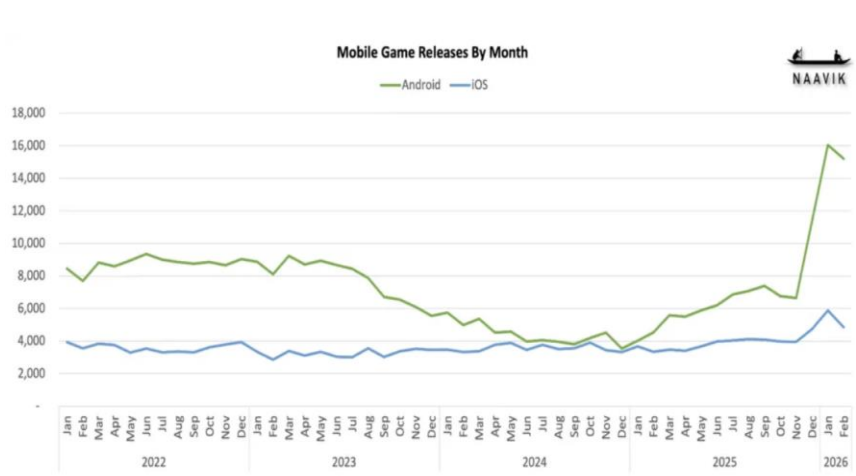
■ 据a16z，自AI编程工具普及以来，iOS新应用的上架数量明显增长，2025年12月iOS新应用数同比增幅达60%。游戏方面，2025年以来安卓端移动游戏月新增数震荡上升，整体趋势与移动应用一致。视频/音频方面，据《中国网络视听发展研究报告（2026）》，2025年由AI生成的视频/音频累计超20亿条，较2024年增长了14倍以上。据DataEye短剧观察公众号，仅2026年3月，抖音平台新增上传的AI生成微短剧数量便接近5万部，几乎追平2025年全年上传总量。我们认为不断膨胀的内容供给与有限的曝光机会间的矛盾或推动广告位价格提升。

图32：2025年2月AI编程工具普及后，iOS新应用的上架数量明显增长



资料来源：a16z

图33：2025年以来安卓端移动游戏月新增数震荡上升

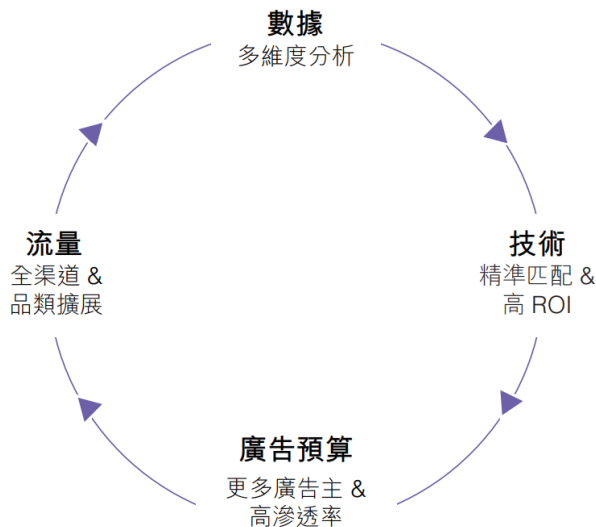


资料来源：NAAVIK

2.2.3 AI营销：看好头部程序化广告平台及与头部AI应用公司合作紧密的营销公司

■ 我们认为，多模态大模型可用于营销素材制作，推动全行业降本提效；程序化广告技术领先企业或持续受益于广告市场扩容。此外，Open AI已正式推出ChatGPT自助广告管理工具测试版，谷歌明确透露旗下AI产品Gemini未来将加入广告功能，Meta宣布旗下AI助手Meta AI将于2026Q3启动广告试点，微软已将广告整合到Copilot中，亚马逊的AI购物助手也已向品牌开放赞助回答，豆包推出付费订阅模式或标志国内头部AI应用商业化加速，我们认为与头部AI应用公司合作紧密的营销公司或率先受益于AI应用商业化带来的广告市场增长。重点推荐：汇量科技、引力传媒，受益标的：蓝色光标、易点天下。

图34：程序化广告行业的“飞轮效应”或助强者恒强



资料来源：汇量科技公告

表12：汇量科技、引力传媒等或受益于AI游戏发展

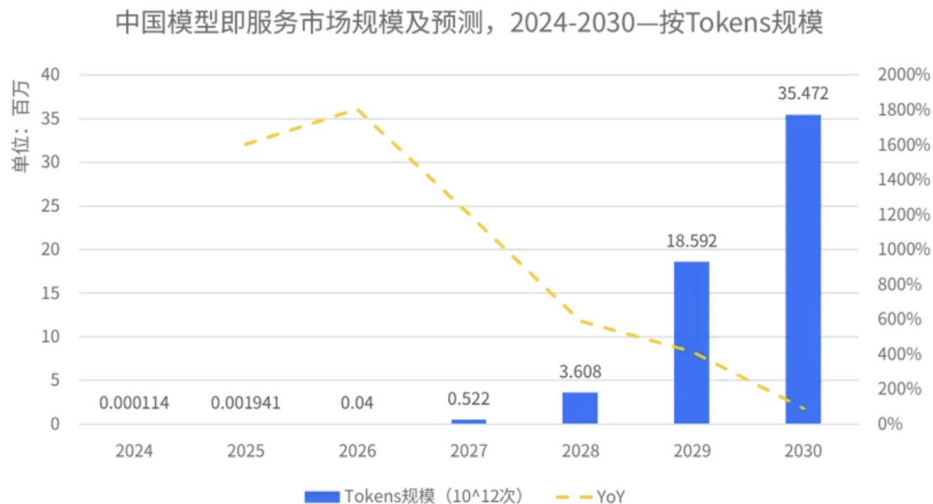
公司	业务布局
汇量科技	公司2015年开始孵化AI驱动的程序化广告平台Mintegral，据AppsFlyer《广告平台综合表现报告-2025版》，iOS端中，Mintegral排全球游戏类体量榜第3，排全球非游戏类体量榜第9；安卓端中，排全球游戏类体量榜第2，排全球非游戏类体量榜第5。公司依托大模型技术推出Multi-Agent架构的MaxAgent系统，覆盖研发、运营等核心业务场景，提升异常流量排查、监测测试等高频复杂场景的自动化水平。创意平台Playturbo已支持AI配音、AI翻译、AI生图、数字人视频等功能，将视频创意制作周期由数天压缩至数分钟。
引力传媒	在品牌营销端，公司是巨量引擎、腾讯广告、哔哩哔哩等主流平台的头部/核心代理商；在效果端，公司是巨量千川金牌照服务商和AD核心授权服务商，也是阿里妈妈全域六星合作伙伴。引力国际已重点围绕TikTok、Amazon生态着手团队搭建与扩张。公司基于积累的海量营销数据，自研“九合创意2.0”一站式AIGC创作平台，完成电商广告全链路视觉资产的流水线闭环生产，将原需数日的素材制作周期压缩至数小时。“AI用研平台”以营销实战落地为导向，帮助品牌在数天内完成以往需数周才能实现的用户认知分析。
蓝色光标	公司是国内各媒体排名前列的采购商，也是Meta、Google、TikTok for Business、X等平台国内官方授权2025年BlueX广告变现SDK完成全平台聚合兼容性战略部署，实现与Max、AdMob、TopOn、TradPlus等全球主流聚合平台的深度适配。BlueAI在出海与国内营销场景实现规模化落地，持续升级在社交媒体与达人营销、智能广告投放、创意生产与数据洞察深度融合的AI驱动的营销模式。
易点天下	公司是Google、Meta、TikTok、Microsoft Advertising等多个头部媒体平台的合作伙伴或一级代理商，同时是AppLovin、Amazon Ads及The Trade Desk等多家领先DSP平台的重要合作伙伴。程序化广告平台zMaticoo接入全球50+主流广告平台SDK，在Paxlate发布的亚太地区SSP市场份额报告中稳居前列。

资料来源：各公司公告、各公司公众号、开源证券研究所

2.3.1 算力租赁/AIDC：算力需求旺盛，供需缺口持续扩大

■ 中国信通院数据显示，2026Q1国内AI算力需求同比暴涨417%，供给增速仅128%。据21世纪经济，供需失衡推高算力价格，国内头部云厂商价格涨幅达5%-34%。多模态模型技术成熟、Agent类应用规模化落地有望推动国内算力需求持续高增，据IDC，2024-2030年，中国MaaS的Token消耗量年复合增长率或达1154.9%。

图35：IDC预计2024年至2030年中国MaaS的Token消耗量年复合增长率约为1154.9%

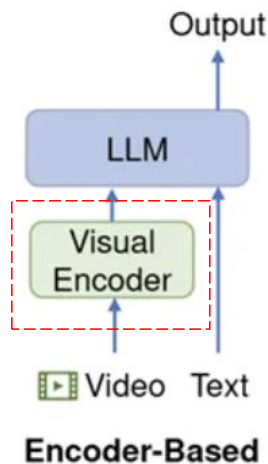


资料来源：IDC咨询公众号

2.3.2 算力租赁/AIDC：多模态提升单位Token算力占用，应用发展或扩大算力缺口

- 以图像/视频输入为例，对比文本输入，其在进入LLM前需经过视觉编码器提取特征，随视频帧数增加，特征提取阶段的内存和计算需求呈线性甚至指数级增长；当模型处理128帧或更长的视频时，编码器端的延迟可能会超过LLM本身。

图36：多模态大模型架构引入的视觉编码器显著提升计算开销

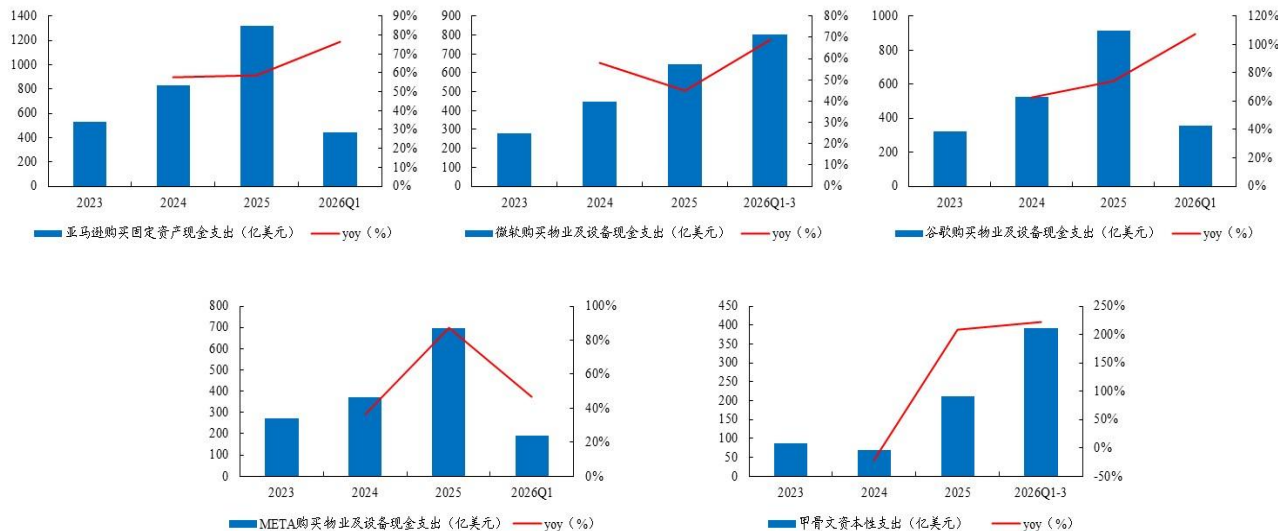


资料来源：《Breaking the Encoder Barrier for Seamless Video-Language Understanding》、开源证券研究所

2.3.3 算力租赁/AIDC：海外科技大厂资本开支逐年攀升，2026年预计维持高增

■ 2023-2025财年，亚马逊/微软/谷歌/META/甲骨文资本开支CAGR分别为58%/52%/68%/60%/56%。亚马逊/微软/谷歌/META 2026年资本开支总规模或超7000亿美元，同比增速中值分别为52%/61%/102%/94%。

图37：2023年以来，海外科技大厂资本开支持续攀升



数据来源：Wind、开源证券研究所（注：图中时间为美股财年时间，亚马逊、谷歌、Meta财年截止时间为12月31日，微软财年截止时间为6月30日，甲骨文财年截止时间为5月31日）

表13：亚马逊等海外四大科技企业2026年资本开支规模总值或超7000亿美元

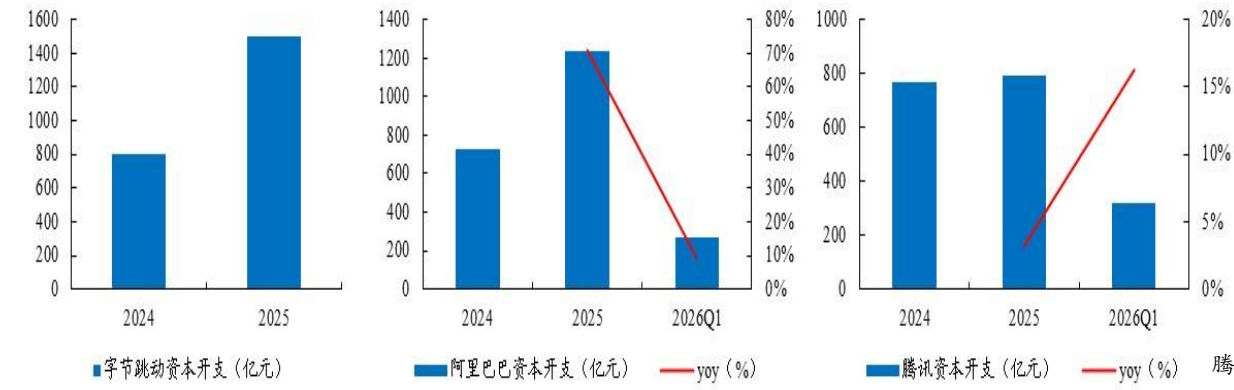
公司	资本开支规划
亚马逊	2026年2000亿美元资本支出，同比增长52%
微软	预计2026年全年资本支出将达到1900亿美元，同比增长61%
谷歌 (Alphabet)	2026财年资本支出预计为1800亿-1900亿美元，同比增长97%-108%
Meta	预计2026年资本支出将在1250亿-1450亿美元，同比增长79%-108%

资料来源：财联社、开源证券研究所

2.3.3 算力租赁/AIDC：国内科技大厂资本开支增速有望提升

■ 2025年，字节跳动（第三方估算）/阿里巴巴/腾讯资本开支规模分别为1500/1238/792亿元，同比增长88%/71%/3%。2026年，三家大厂均计划加大算力投资力度，我们认为随多模态模型及应用商业化加速推进、算力硬件限制解除，模型大厂资本开支及算力服务采购或进一步加码，有望拉动算力租赁及AIDC基建需求及订单高增。

图37：2023年以来，海外科技大厂资本开支持续攀升



数据来源：Wind、花椒财经公众号、另镜公众号、开源证券研究所（注：字节跳动资本开支为第三方估计值）

表14：字节跳动、阿里巴巴、腾讯控股均计划加大资本开支

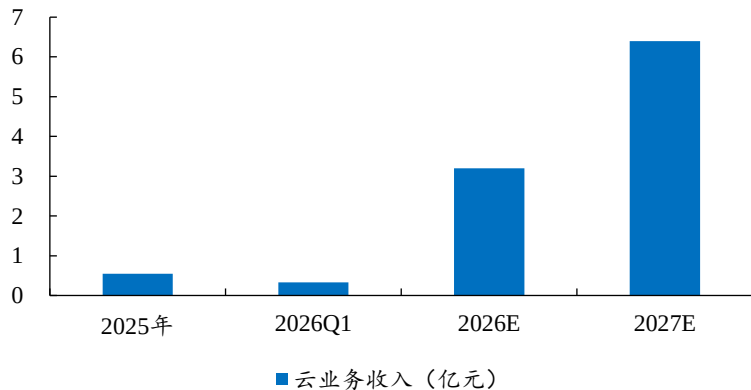
公司	资本开支规划
字节跳动	据《南华早报》，已将2026年AI资本开支计划上调至逾2000亿元人民币，较2025年底初步讨论的1600亿元方案增幅至少25%。
阿里巴巴	2025-2027年资本支出或将超过2025年2月宣布的3800亿元，据36氪，2026年资本开支将上涨至1500-1700亿元
腾讯控股	2026年资本支出相比2025年会有所增加，特别是在下半年

资料来源：财联社、开源证券研究所

2.3.4 算力租赁/AIDC：看好有先发/卡位/规模优势，开始兑现高增长的算力相关公司

■ 顺网科技：云增长陡峭，股权激励彰显信心，产品矩阵及下游行业拓展或驱动高增。公司为电竞场域客户提供长期稳定的存储及算力上云服务；同时，针对电竞场域算力“忙闲时段”特性，将日常经营的冗余算力调度至电竞场域外，供各类场景用户使用。2025年公司可运营算力规模5万路，服务近80万终端，云业务收入5489万元，2026Q1末公司可运营算力5.3万路，服务近82万终端，云业务收入3306万元，云业务快速扩张。2026年2月，公司发布股权激励计划，云业务2026/2027年目标收入分别为3.2/6.4亿元，反映团队信心。公司推出针对不同场景的云服务产品，适配酒店、商场、小区等各类数字消费场景，并向校园、机器人等行业延伸，我们看好算力规模及覆盖范围扩大、产品及客户拓展驱动云业务持续高增。

图39：2025-2027年，顺网科技云业务收入CAGR或达241%

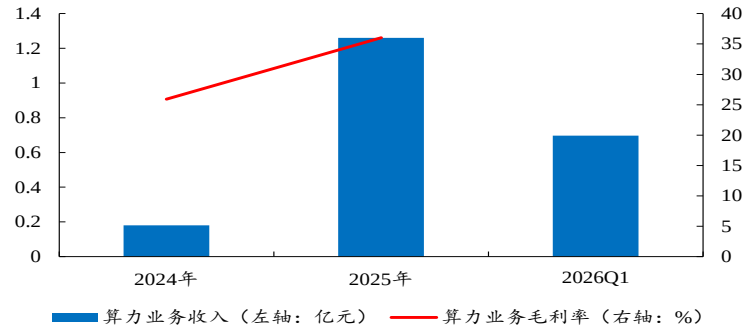


数据来源：顺网科技公告、开源证券研究所（注2026、2027年云业务收入预测值为2026年2月股权激励方案中的云业务目标收入值）

2.3.4 算力租赁/AIDC：看好有先发/卡位/规模优势，开始兑现高增长的算力相关公司

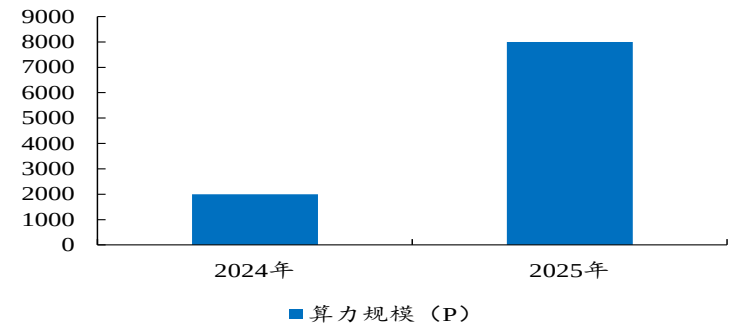
- **华策影视：算力业务高增，大额服务器采购合同或再为增长充能。**公司以“轻资产算力综合解决方案与运营平台”为核心定位，布局算力贸易、云算力服务业务。2025年，算力业务收入1.26亿元，同比增长6倍，2026Q1算力业务收入约0.7亿元，达2025全年的55%。公司加速算力储备，截至2025年末，自有智算规模达8000P（同比增长3倍），已全部面向头部科技赛道客户签订长期协议。2026年5月，公司披露拟向多家供应商采购服务器，采购合同总金额预计不超过人民币33亿元，主要用于为客户提供云算力服务。
- **浙数文化：联手阿里云，持续推进国产智算中心建设。**旗下杭州富春云科技加速推动算力基础设施改造，杭州富阳数据中心基本建成浙江省域文化算力中心，北京四季青数据中心与多家国内领先的科技、算力企业共同打造国内算力重点项目，加快建设新一代国产智能算力中心基地。2025年9月，富春云与阿里云达成深度合作，共同推进国产化算力生态建设，在政企客户联合服务领域，合作提供从咨询、规划、建设到运营的全链条服务。2025年11月，公司间接投资数字化产品一站式技术服务平台云尖信息，其在ICT基础设施领域具有深厚技术积累。

图40：2025年华策影视算力业务收入同比增长6倍，2026Q1算力业务收入达2025全年的55%



数据来源：华策影视公告、开源证券研究所

图41：截至2025年末，华策影视已有智算规模8000P，同比增长3倍



数据来源：华策影视公告、开源证券研究所

目录

CONTENTS

- 1 全球多模态技术持续迭代，国产模型从追赶到局部超越
- 2 多模态AI商业化开启进击之路
- 3 投资建议及风险提示

■ 我们认为，全球多模态AI将加速进击，国产模型优势突出，将持续赋能视频、游戏、营销等下游应用领域，助力多模态模型厂商Token调用量及收入高增，并扩大算力供需缺口，拉动算力租赁、AIDC等算力产业链需求及订单高增，建议积极布局大模型、核心多模态AI应用（视频、游戏、营销）、算力租赁/AIDC方向。

- AI大模型：重点推荐：腾讯控股、快手，受益标的：Minimax、阿里巴巴、智谱、昆仑万维；
- AI视频：重点推荐：阅文集团，受益标的：中文在线、德才股份、博纳影业；
- AI游戏：重点推荐：完美世界、网易、哔哩哔哩、巨人网络、恺英网络、心动公司，受益标的：世纪华通；
- AI营销：重点推荐：汇量科技、引力传媒，受益标的：蓝色光标、易点天下；
- 算力租赁/AIDC：重点推荐：顺网科技，受益标的：华策影视、浙数文化。

表15：重点推荐及受益标的涉及大模型、游戏、视频、营销等行业

股票代码	公司名称	评级	收盘价（元）	市值（亿）	EPS（2025A，元/股）	EPS（2026E，元/股）	EPS（2027E，元/股）	PE（2025A）	PE（2026E）	PE（2027E）
1024.HK	快手-W*	买入	49.34	2145	4.28	3.32	3.86	10.0	12.9	11.1
0772.HK	阅文集团*	买入	25.40	259	-0.76	1.07	1.22	-29.0	20.6	18.0
002624.SZ	完美世界*	买入	16.00	310	0.39	0.87	1.15	41.0	18.4	14.0
0700.HK	腾讯控股*	买入	456.40	41615	24.75	26.70	28.94	16.0	14.8	13.7
9999.HK	网易-S*	买入	179.00	5714	10.66	11.75	13.23	14.6	13.2	11.7
9626.HK	哔哩哔哩-W*	买入	153.50	652	2.85	3.74	5.91	46.7	35.6	22.5
002558.SZ	巨人网络*	买入	30.12	572	0.94	2.30	2.58	32.0	13.1	11.7
002517.SZ	恺英网络*	买入	17.65	377	0.89	1.30	1.54	19.8	13.5	11.5
1860.HK	汇量科技*	买入	14.63	230	0.27	0.59	1.02	46.3	21.5	12.4
603598.SH	引力传媒*	买入	25.33	68	0.06	0.30	0.41	422.2	85.1	62.4
300113.SZ	顺网科技*	买入	27.27	184	0.62	0.73	0.93	44.0	37.5	29.3
0100.HK	Minimax-W	未有评级	794.00	2490	-116.93	-10.79	-12.83	-5.9	-63.7	-53.6
9988.HK	阿里巴巴-W	未有评级	132.30	25392	6.89	5.74	7.65	16.6	20.0	15.0
2513.HK	智谱	未有评级	1040.00	4637	-12.03	-10.42	-9.80	-74.9	-86.4	-91.9
300418.SZ	昆仑万维	未有评级	49.21	618	-1.28	-0.32	0.10	-38.4	-153.0	474.1
300364.SZ	中文在线	未有评级	27.90	203	-0.92	0.03	0.15	-30.3	882.9	188.4
605287.SH	德才股份	未有评级	48.16	67	0.52	-	-	92.6	-	-
002602.SZ	世纪华通	未有评级	44.00	1098	0.78	1.21	1.46	19.1	12.3	10.2
300058.SZ	蓝色光标	未有评级	17.66	634	0.06	0.08	0.10	280.3	211.0	184.9
301171.SZ	易点天下	未有评级	43.70	268	0.34	0.41	0.56	128.5	106.7	77.4
300133.SZ	华策影视	未有评级	9.63	182	0.10	0.19	0.22	96.3	51.8	44.4
600633.SH	浙数文化	未有评级	13.02	165	0.45	0.55	0.60	28.9	23.7	21.7

数据来源：Wind、开源证券研究所（“*”公司盈利预测取自开源证券研究所，其余公司盈利预测取自Wind一致预测，股价及市值截至2026年5月15日收盘，港股公司所示市值及股价为港币，计算采用1港币= 0.8663人民币、1美元=6.7862人民币进行换算）

（1）大模型迭代速度不及预期：大模型更新迭代受技术路径、算力成本、人员投入及监管等影响，其进展存在不确定性。

（2）AI应用商业化进展不及预期：受大模型能力、实际应用效果、消费者接受度及收入预期等因素影响，AI应用商业化进展存在不及预期的风险。

（3）算力投资力度不及预期：算力投资受算力硬件产能、AI应用变现预期、AI应用厂商融资能力等因素影响，投资计划落地进展存在不及预期的风险。

分析师声明

负责准备本报告以及撰写本报告的所有研究分析师或工作人员在此保证，本研究报告中关于任何发行商或证券所发表的观点均如实反映分析人员的个人观点。负责准备本报告的分析师获取报酬的评判因素包括研究的质量和准确性、客户的反馈、竞争性因素以及开源证券股份有限公司的整体收益。所有研究分析师或工作人员保证他们报酬的任何一部分不曾与，不与，也将不会与本报告中具体的推荐意见或观点有直接或间接的联系。

特别声明

《证券期货投资者适当性管理办法》、《证券经营机构投资者适当性管理实施指引（试行）》已于2017年7月1日起正式实施。根据上述规定，开源证券评定此研报的风险等级为R4（中高风险），因此通过公共平台推送的研报其适用的投资者类别仅限定为专业投资者及风险承受能力为C4、C5的普通投资者。若您并非专业投资者及风险承受能力为C4、C5的普通投资者，请取消阅读，请勿收藏、接收或使用本研报中的任何信息。因此受限于访问权限的设置，若给您造成不便，烦请见谅！感谢您给予的理解与配合。

股票投资评级说明

	评级	说明	备注： 评级标准为以报告日后的6~12个月内，证券相对于市场基准指数的涨跌幅表现，其中A股基准指数为沪深300指数、港股基准指数为恒生指数、新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）、美股基准指数为标普500或纳斯达克综合指数。我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议；投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者应阅读整篇报告，以获取比较完整的观点与信息，不应仅仅依靠投资评级来推断结论。
证券评级	买入（buy）	预计相对强于市场表现20%以上；	
	增持（outperform）	预计相对强于市场表现5%~20%；	
	中性（Neutral）	预计相对市场表现在-5%~+5%之间波动；	
	减持（underperform）	预计相对弱于市场表现5%以下。	
行业评级	看好（overweight）	预计行业超越整体市场表现；	
	中性（Neutral）	预计行业与整体市场表现基本持平；	
	看淡（underperform）	预计行业弱于整体市场表现。	

分析、估值方法的局限性说明

本报告所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。本报告采用的各种估值方法及模型均有其局限性，估值结果不保证所涉及证券能够在该价格交易。

法律声明

开源证券股份有限公司是经中国证监会批准设立的证券经营机构，具备证券投资咨询业务资格。

本报告仅供开源证券股份有限公司（以下简称“本公司”）的机构或个人客户（以下简称“客户”）使用。本公司不会因接收人收到本报告而视其为客户。本报告是发送给开源证券客户的，属于商业秘密材料，只有开源证券客户才能参考或使用，如接收人并非开源证券客户，请及时退回并删除。

本报告是基于本公司认为可靠的已公开信息，但本公司不保证该等信息的准确性或完整性。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他金融工具的邀请或向人做出邀请。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、价值及投资收入可能会波动。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。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突，不应视本报告为做出投资决策的唯一因素。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。本公司未确保本报告充分考虑到个别客户特殊的投资目标、财务状况或需要。本公司建议客户应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。若本报告的接收人非本公司的客户，应在基于本报告做出任何投资决定或就本报告要求任何解释前咨询独立投资顾问。

本报告可能附带其它网站的地址或超级链接，对于可能涉及的开源证券网站以外的地址或超级链接，开源证券不对其内容负责。本报告提供这些地址或超级链接的目的纯粹是为了客户使用方便，链接网站的内容不构成本报告的任何部分，客户需自行承担浏览这些网站的费用或风险。

开源证券在法律允许的情况下可参与、投资或持有本报告涉及

的证券或进行证券交易，或向本报告涉及的公司提供或争取提供包括投资银行业务在内的服务或业务支持。开源证券可能与本报告涉及的公司之间存在业务关系，并无需事先或在获得业务关系后通知客户。

本报告的版权归本公司所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

开源证券研究所

上海：上海市浦东新区世纪大道1788号陆家嘴金控广场1号楼3层

邮箱：research@kysec.cn

北京：北京市西城区西直门外大街18号金贸大厦C2座9层

邮箱：research@kysec.cn

深圳：深圳市福田区金田路2030号卓越世纪中心1号楼45层

邮箱：research@kysec.cn

西安：西安市高新区锦业路1号都市之门B座5层

邮箱：research@kysec.cn

THANKS

感 谢 聆 听



开源证券