

2025，国产AI机会正启——字节生态篇

证券分析师：张良卫
执业证书编号：S0600516070001
联系邮箱：zhanglw@dwzq.com.cn

证券分析师：鲍娴颖
执业证书编号：S0600521080008
联系邮箱：baoxy@dwzq.com.cn

证券分析师：马天翼
执业证书编号：S0600522090001
联系邮箱：maty@dwzq.com.cn

证券分析师：郭若娜
执业证书编号：S0600524080004
联系邮箱：guorn@dwzq.com.cn

2025年2月6日

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

■ **国内AI产业正进入加速时刻。**（1）**模型**：2024年是大模型继续探索能力边界的一年，也是大模型间技术差距逐步缩小的一年，体现在：其他大模型玩家与OPENAI的差距缩小、开源模型与闭源模型差距缩小、中国大模型与海外顶尖大模型差距缩小。大模型一家独大的局面并不可持续。进入2025年，1月DeepSeek开源推理模型DeepSeek-R1，以更低的成本实现接近OpenAI推理模型O1的性能，快速成为全球上线后增速最快的 AI 原生应用。我们看好在DeepSeek这类玩家贡献下，国内大模型整体能力与海外顶尖模型的差距进一步缩小，并以更低的推理成本推动AI技术更广泛更深入的落地。值得关注的是，在其他细分方向，如技术所处阶段更早的视频生成模型，国内玩家正逐渐追平海外顶尖水平，2025年我们期待国内以快手为代表的头部玩家在视频生成关键技术上进一步取得突破，推动AI视频在创意/娱乐行业渗透率加速提升。（2）**应用**：2024年国内AI原生APP用户规模及粘性正快速提升，MAU在网民中的渗透率从24年1月的3%提升到12月的11%。12月全球MAU TOP15的AI APP 中，中国公司产品占据4个席位，字节的豆包排名第二。国内AI技术正持续渗透多个应用场景，我们看好在技术进步和成本下降的背景下2025年加速发展。

■ 为什么说字节链是最值得重视的AI产业机会之一？

- **字节已成为目前大厂中在AI领域布局最激进的玩家。**TOP级资源（资金、流量）+创新者的姿态，相比百度、阿里等大厂，字节在大模型领域起步较晚，但内部极为重视，创始人张一鸣亲自下场，从其不断加码的资本开支、高薪聘请AI顶尖人才、大模型持续追赶、应用领域全方位推进可看出字节重视程度。
- **字节产业链充满机会：**（1）**豆包大模型虽非顶尖，但正在快速追赶。**根据MMLU等国际权威基准测试以及国内SUPERCLUE测试结果，豆包大模型与国际顶尖大模型、国内头部大模型的差距正逐渐缩小。2024年，豆包大模型支撑超过50个应用场景，其中豆包成为国内最受欢迎的AI产品，通过火山引擎豆包大模型服务30多个行业，日均tokens调用量超4万亿，较5月增长33倍。（2）**字节积极推进降低模型成本，推动应用加速。**大厂中字节大模型率先大幅降价，国内大模型价格进入“厘时代”。大模型作为AI产业重要基建，其价格降低有望通过更低的试错成本提高企业大模型使用意愿，助力更多开发者开发AI应用，推动国内AI产业加速。（3）**字节正在打造新的AI生态，且愿意让利合作伙伴，产业链上下游玩家迎来新红利。**字节目前在AI领域采用全面进攻姿态，不管是面向多个垂类行业寻找大模型落地场景，还是在硬件终端和C端应用展开更多关于AI AGENT形态的探索，这一过程都需要合作伙伴一起建设生态。字节跳动一向也偏好先让利和扶持合作伙伴，共同将蛋糕做大，并最终与优质的合作伙伴实现共赢。因此我们看好在字节当前大力投入打造AI大生态的过程中，多个产业链迎来新的发展红利。

■ 字节AI生态及产业机会框架初探：

- **云：火山引擎有望弯道超车，助字节卡位AI时代重要基建。** 字节向外推出云服务的时间较晚，在客户基础、传统云基础产品等方面存在不足。但随着大模型训练推理需求不断增多，云服务客户还关心大模型相关的AI infra等能力，这方面所有云厂商都在同一起跑线，而火山引擎的迭代速度几乎是厂商里面最快的，有望实现弯道超车。2024年上半年火山引擎在国内GENAI IaaS服务厂商中排名第2，豆包大模型日均token使用量从2024年5月的1200亿增长至12月的4万亿，增长32倍。展望2025年，随着字节豆包大模型能力增强和成本降低，火山引擎以更高性价比和更专业的GEN AI 云服务拓展更多行业客户，其份额有望进一步提升，助力字节卡位AI时代重要的基建环节。随着云调用量提升，特别是推理需求增加，字节有望加大算力相关投入，服务器、AI芯片、PCB、散热等产业链环节的合作公司有望受益于订单增长。
 - **模型服务/开发：火山方舟+扣子+HiAgent，完善模型服务生态。** HiAgent和扣子coze降低了企业/开发者开发agent的成本，更好加速AI应用迭代，建议关注字节降低模型开发门槛后对垂类AI应用加速发展带来的潜在影响。
 - **应用：（1）AI助手豆包。** 豆包提供了优秀的用户体验，叠加大规模买量等原因，豆包APP一跃成为全球用户数第2大的AI应用；期待其PC端在学习办公场景的渗透率提升。我们认为豆包有望成长为新的多端流量入口，与字节达成合作的营销、内容安全、多模态技术等服务商有望受益。（2）**AI娱乐猫箱：**猫箱作为一款类character.ai应用，已成为字节旗下用户规模第二大的C端AI原生应用。我们认为猫箱有望成为国内AI时代重要的娱乐内容形态代表，成为继抖音等短视频之后年轻人新的娱乐应用，这个过程中手握优质IP资源的公司有望拓展IP变现方式。（3）**AI教育豆包爱学：**建议关注大模型推理性能增强对AI教育行业的影响、产品重大迭代及用户增长趋势；（4）**AI创作即梦：**我们估计即梦视频生成能力在国内处在第二梯队，但基于模型亮点、生成效率和高效产品迭代等原因，我们看好其在创作者端占据一定心智及长期发展潜力，后续建议关注开源技术进步、团队技术人才变动及模型重大迭代。（5）**AI端侧：**随着字节豆包、DeepSeek、阿里通义等国产大模型生态共荣，我们看好端侧赛道的投资机会。举例来看，在AI玩具、AI宠物、AI眼镜、AI耳机、AI音箱等智能硬件中，豆包实时语音大模型能够为其赋予更强大的语音交互功能，使这些硬件不再仅仅是简单的播放设备或执行单一任务的工具，而是可以与用户进行深度互动的智能伙伴，拓展了智能硬件的应用场景和功能边界。
- **投资建议：**我们认为字节的投入加速了国内AI大模型及应用的进展，未来有望给算力、数据/IP提供商及下游细分链条上的合作伙伴（营销、内容安全、多模态技术、AI玩具、AI眼镜、AI耳机等）带来业务增量和投资机会；国内AI应用亦有望迎来更多落地。具体标的上：1）数据：建议关注掌阅科技、海天瑞声、视觉中国、丝路视觉等。2）AI玩具及情感陪伴方向：推荐昆仑万维、建议关注实丰文化、乐鑫科技、盛天网络、汤姆猫、奥飞娱乐等。3）AI硬件：端侧芯片及模组建议关注恒玄科技、全志科技、移远通信、乐鑫科技、润欣科技、中科蓝讯、星辰科技、兆易创新、瑞芯微、广和通、移远通信、美格智能、翱捷科技等；端侧零部件及组装环节建议关注歌尔股份、亿道信息、天键股份、龙旗科技、康耐特光学、华勤科技、瑞声科技等。4）AI工具：推荐美图公司、万兴科技、创业黑马等。5）AI游戏：推荐巨人网络、恺英网络等。6）AI电商：推荐焦点科技、值得买、华凯易佰、吉宏股份、建议关注返利科技、光云科技等。
- **风险提示：**模型能力提升不及预期，字节AI业务进展不及预期，竞争格局变化。



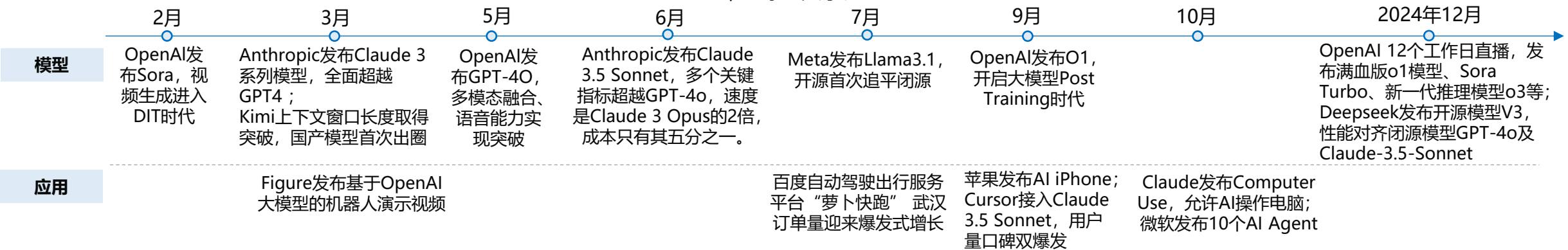
- 1、国内AI产业正进入加速时刻
- 2、为什么说字节链是最值得重视的AI产业机会之一
- 3、字节AI生态及产业机会框架初探
- 4、投资建议
- 5、风险提示

1、国内AI产业正进入加速时刻

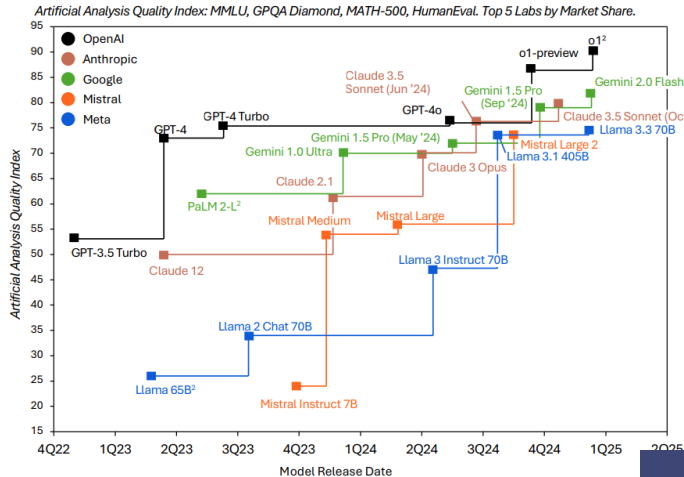
1.1 2024是重塑认知的一年：大模型间技术和份额差距逐步缩小

- 2024年是大模型继续探索能力边界的一年，也是大模型间技术差距缩小的一年，体现在：其他大模型玩家与OPENAI的差距缩小（代表玩家Anthropic）、开源模型与闭源模型差距缩小（代表玩家META、DEEPSEEK）；中国大模型与海外顶尖大模型差距缩小（代表玩家DEEPSEEK）。
- 差距缩小还体现在份额指标上，大模型一家独大的局面并不可持续：根据企业支付管理 SaaS 公司 Ramp的数据，23年11月底OpenAI token消耗占总量近90%，24年9月降到近80%；根据Artificial Analysis调研数据，大多数开发者会调用至少2个以上的大模型（N=265）。

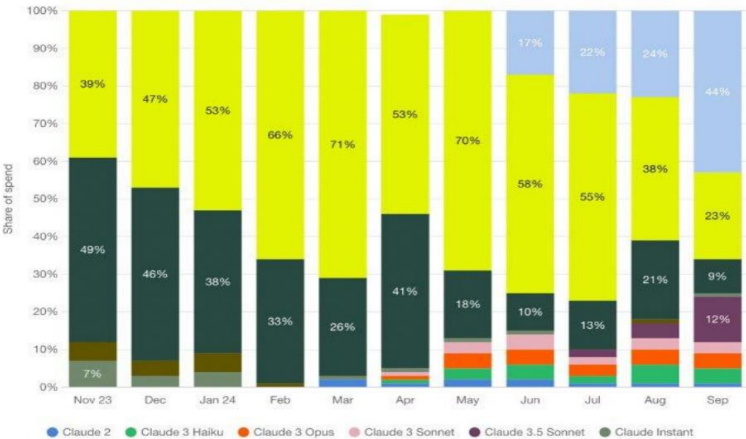
2024年全球AI大事记



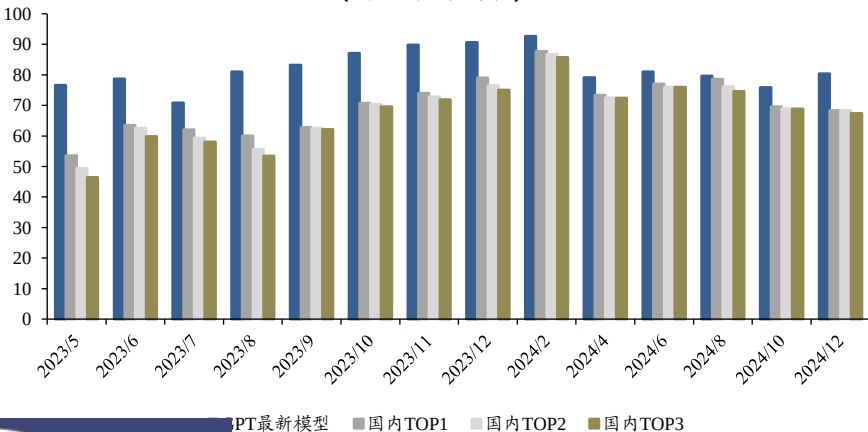
2024年其他大模型玩家与OPENAI的差距缩小



海外大模型token消耗量份额变化



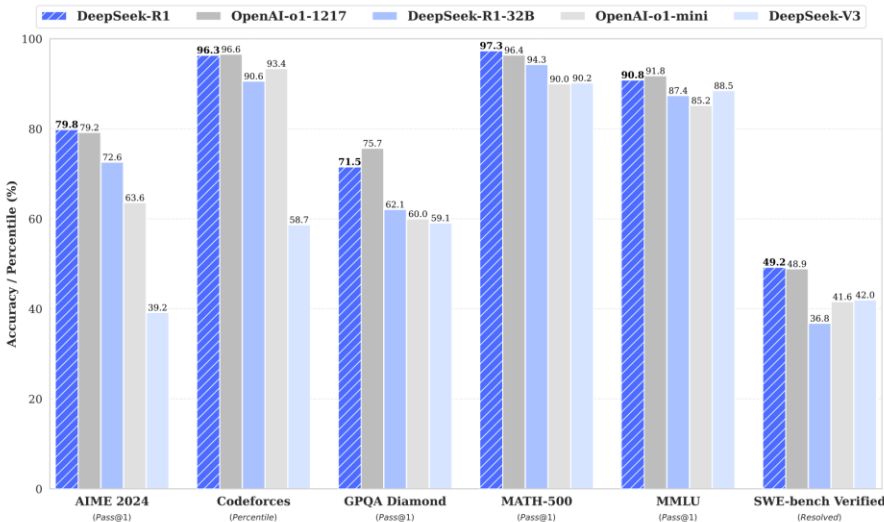
SUPERCLUE基准：国内TOP大模型与GPT最新模型的性能差距缩小（纵轴为分数）



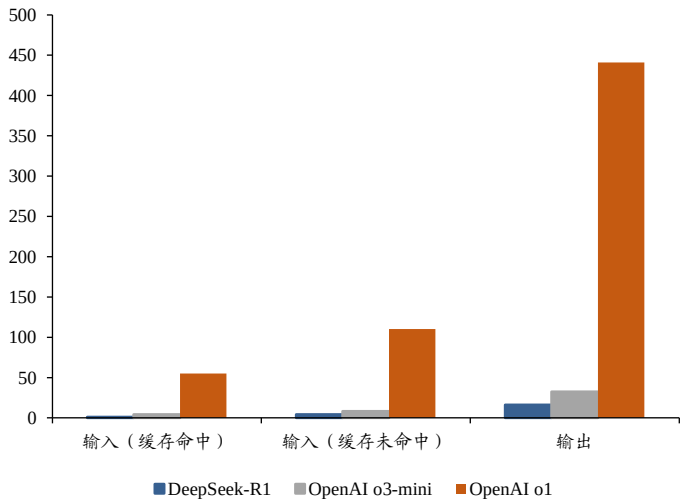
1.2 算力受限下找更优解，2025将是国产AI进击的一年

- **2025年1月，DeepSeek发布推理模型DeepSeek-R1，同步开源模型权重，以更低的成本实现接近OPENAI推理模型O1的性能。**DeepSeek首次验证了可通过纯强化学习（RL）激励大语言模型推理能力，无需监督微调，并提出包含冷启动数据和多阶段训练的 DeepSeek-R1 训练管道，提升模型性能与实用性，在推理任务上的性能与OpenAI-o1-1217相当，API定价上不到o3-mini的一半，仅为o1的4%。模型更新后DeepSeek热度急剧攀升，App登顶中国大陆、美国等75个地区的iOS免费榜总榜，成为全球上线后增速最快的 AI 原生应用，网页版访问量持续攀升。
- **DeepSeek在国内高端算力受限的情况下持续寻找更高效的模型训练方式和提高模型表现的更多路径，并将研究成果进行开源，将大幅推动国内AI产业前进。**根据暗涌对DeepSeek创始人梁文锋的采访， DeepSeek押注了三个方向，数学和代码、多模态以及自然语言。**我们看好在DeepSeek这类创新型玩家的贡献下，2025年国内大模型整体能力与海外顶尖模型的差距进一步缩小，并以更低的推理成本推动AI技术更广泛更深入的落地。**

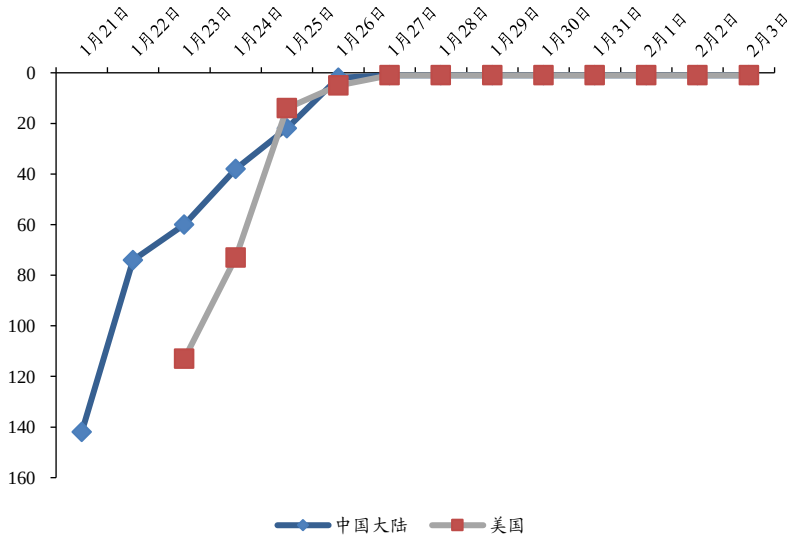
DeepSeek-R1基准测试结果对比openai模型



DeepSeek-R1 API定价对比openai推理模型（元人民币/百万tokens，汇率取2025/2/3数据1美元=7.35元人民币）



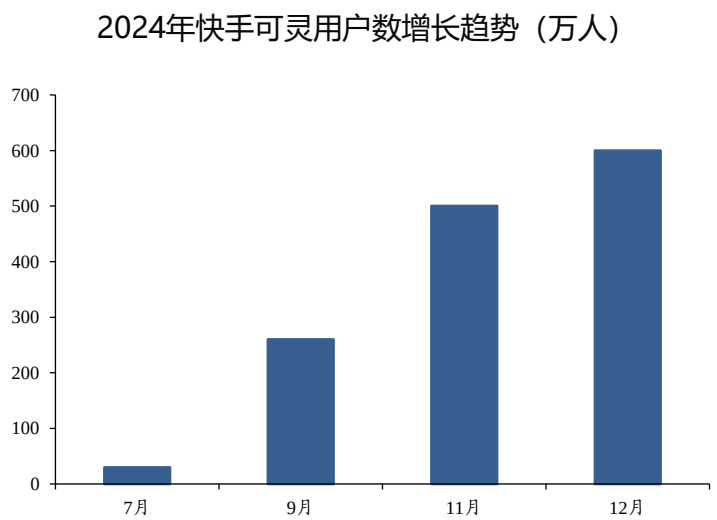
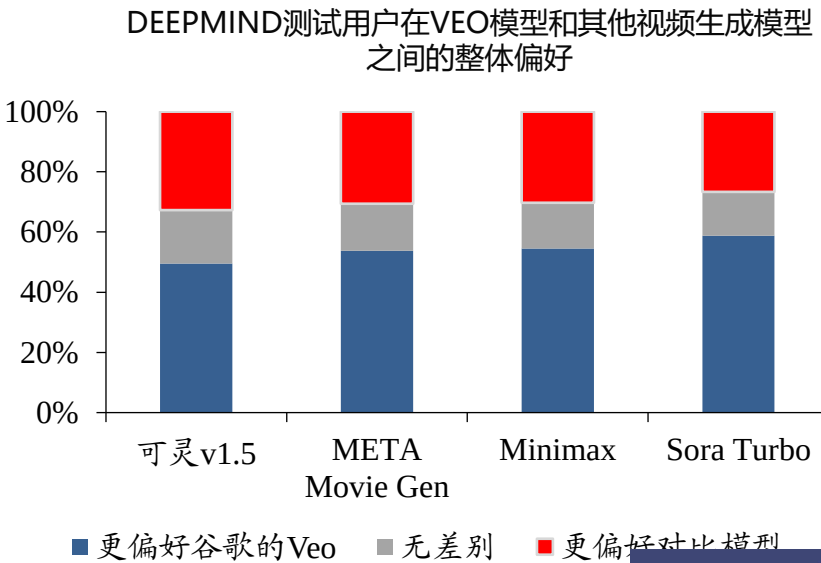
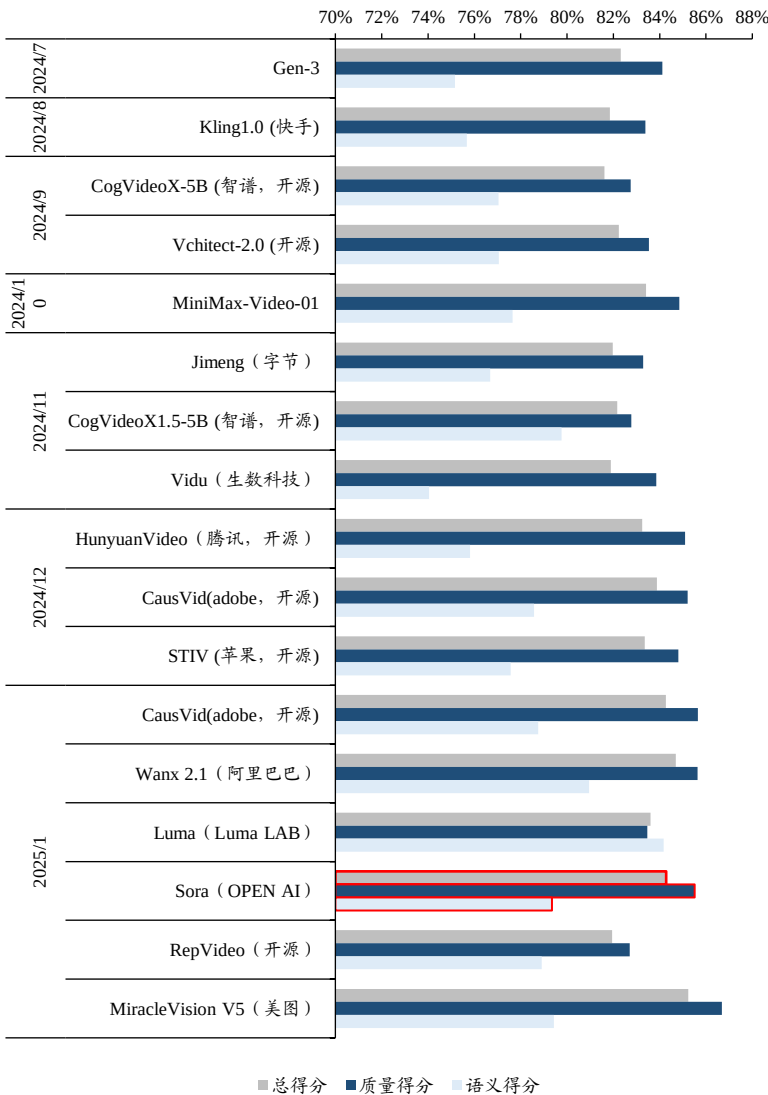
DeepSeek APP在中国大陆和美国iphone免费榜总榜排名



1.3 视频生成模型国内逐渐追平，期待2025年国内进展

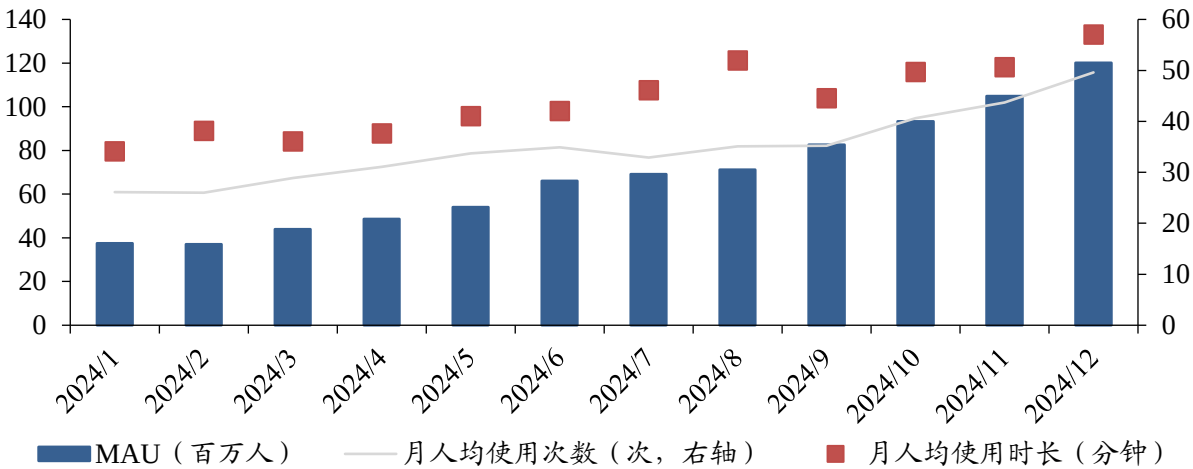
- 相比自然语言大模型，视频生成模型的技术所处阶段更早。2024年随着SORA的发布以及海内外多个玩家的努力，视频生成模型在生成单帧质量、语义与视频内容的一致性以及视频内容时序一致性等方面有所突破，主流视频生成模型基本可以在5分钟以内基于文字/图片生成5秒钟720p的视频；但区别于自然语言大模型，视频模型存在连续性弱、输出随机的局限性，用户难以一次性生成符合预期的视频，其中的技术难点仍需要玩家攻克。
- 国内玩家逐渐追平海外顶尖水平。海外Google的VEO2模型在生成视频的时长、清晰度和质量（更好的保真度、准确度和相机控制）上都强于SORA。而根据DEEPMIND测试结果，在VEO和其他模型的倾向对比上，偏好快手的视频生成模型可灵和MINIMAX视频生成模型的用户分别占33%和30%，高于SORA的27%，其中在与可灵对比中会偏好谷歌VEO的用户只占一半，我们认为国内头部玩家已逐渐追平海外顶尖水平。可灵2024年6月份发布，是全球首个用户可用的视频生成大模型，在模型能力上持续打磨、快速迭代，实现用户数环比快速增长，根据快手2024Q3业绩电话会，可灵AI商业化单月流水超千万人民币。**进入2025年，我们期待国内以快手为代表的头部玩家在视频生成关键技术上进一步取得突破，推动AI视频在创意/娱乐行业渗透率加速提升。**

视频生成模型VBENCH得分对比

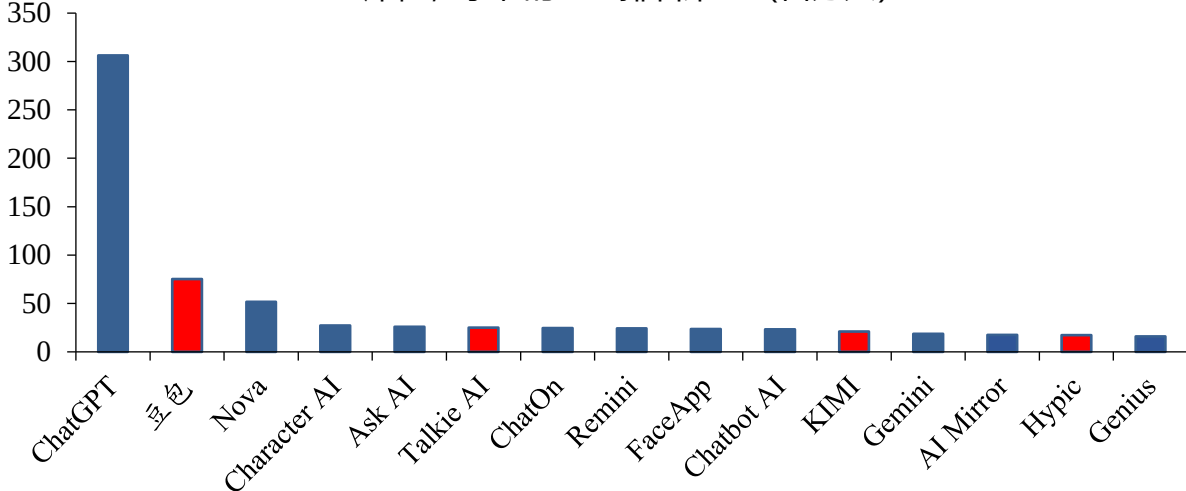


1.4 国内AI原生应用在网民中渗透率已超过10%，看好2025年加速

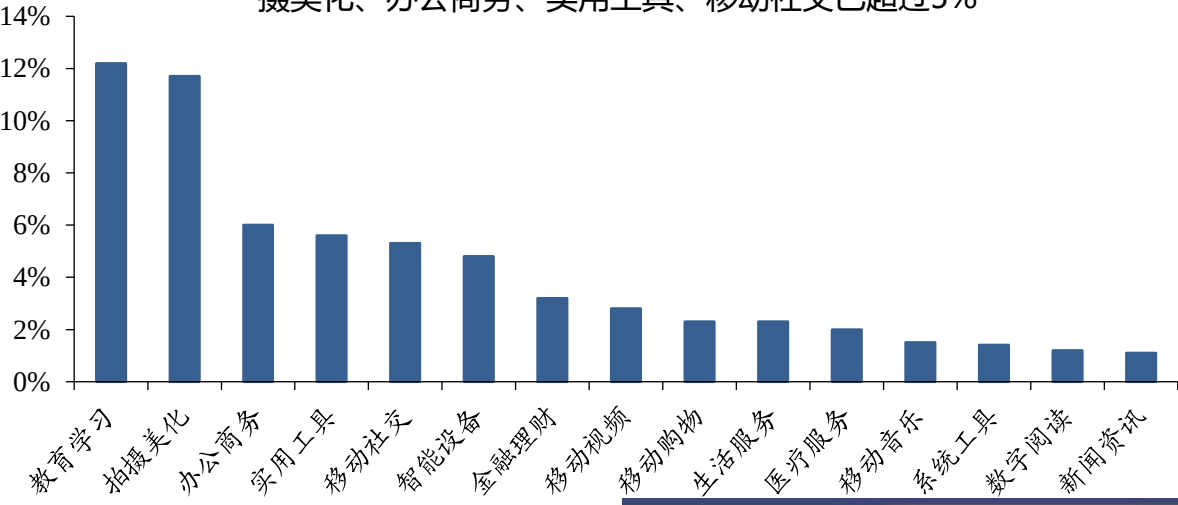
国内AI原生APP用户规模及粘性正快速提升，MAU在网民中的渗透率从24年1月的3%提升到12月的11%



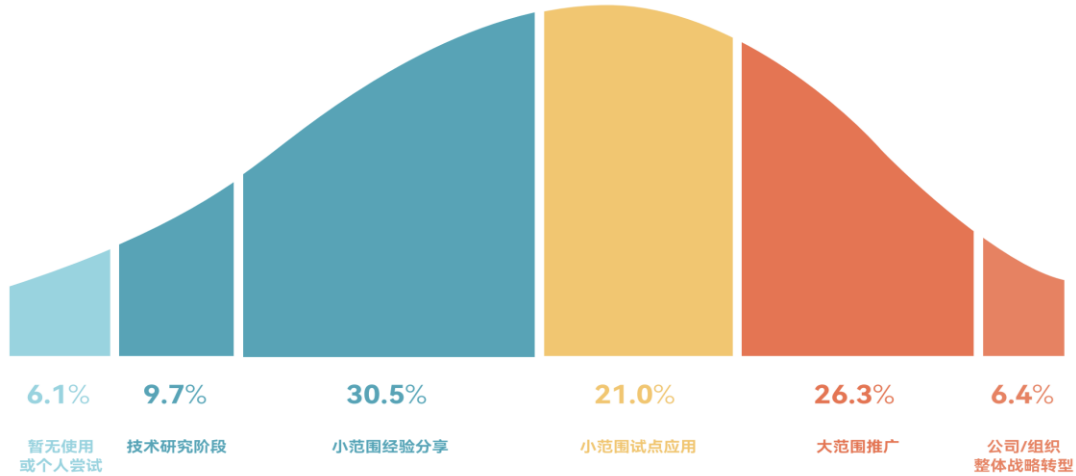
2024年12月全球MAU TOP15的AI APP 中，中国公司产品占据4个席位，字节的豆包排名第二（百万人）



2024年12月分行业算法备案中AI转型APP数量占比，教育学习、拍摄美化、办公商务、实用工具、移动社交已超过5%



国内企业评估或使用生成式AI的进度：企业中高层已开始广泛了解和应用生成式AI（2024调研,N=590）



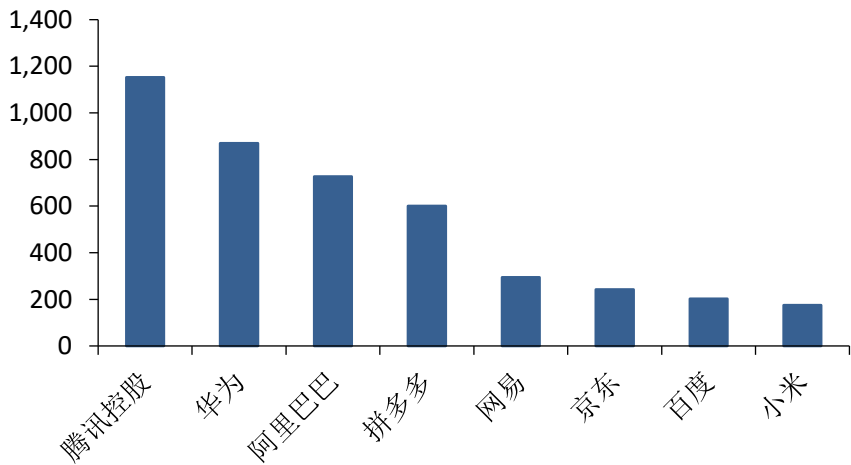
2、为什么说字节链是最值得重视的AI产业机会之一

2.1 TOP级资源+创新者的姿态，字节已成AI领域最激进的玩家



- 字节已成为目前大厂中在AI领域布局最激进的玩家，给产业链上下游合作伙伴带来新发展机遇。
- TOP级资源：1) 资金。作为国内头部互联网企业之一，字节强大的资金实力为其AI资本开支、人才引进、多赛道布局等提供支撑；2) 流量。根据QuestMobile，字节旗下抖音APP月活近8亿人、单用户每日使用时长超过2小时，此外还有数十款APP，包括今日头条、西瓜视频、懂车帝、番茄小说等等涵盖多个细分领域，流量优势明显，可为其AI应用导流（2024年11月豆包接近90%的用户与抖音APP重合）。
- 创新者的姿态：相比百度、阿里等大厂，字节在大模型领域起步较晚，2023年初才成立首个专注于大型模型研发的团队Seed，同年11月成立新部门Flow专注于AI技术应用。但内部极为重视，已退任CEO的创始人张一鸣亲自下场，定期复盘内部AI进展，并亲自监督AI核心人才招聘。字节在AI领域的激进表现在：不断加码的资本开支、高薪聘请AI顶尖人才（如阿里通义技术负责人周畅，零一万物前算法副总裁黄文灏，面壁智能原核心成员、序智科技创始人秦禹嘉等）、大模型持续追赶、应用领域全方位推进。

2023财年国内头部科技企业利润表现（亿元）



字节跳动AI业务部分核心人员背景

核心人员	职位	职业背景
朱文佳	大模型团队负责人	曾在百度担任搜索部主任架构师，在字节担任今日头条CEO、TikTok产品技术负责人。
朱骏	Flow部门负责人	Musical.ly联合创始人
杨震原	技术部门负责人	曾在百度工作9年，担任大搜索副总监，在加入字节跳动后主导推荐算法。
洪定坤	技术部门负责人	曾担任百度贴吧技术经理，后短暂加入小米一年后于2014年正式加入字节跳动。
项亮	AML和豆包大模型 Foundation团队负责人	其所在的AML团队为火山引擎的智能推荐技术服务提供支持。曾任职于Hulu和宜信公司，从事视频推荐系统的研究和开发工作；曾任瓜子二手车首席科学家。
李航	研究部门负责人、人工智能实验室总监	ACL Fellow、IEEE Fellow、ACM 杰出科学家。曾在NEC 公司中央研究所任研究员、微软亚洲研究院任高级研究员与主任研究员、华为技术有限公司诺亚方舟实验室任首席科学家
陆游	FLOW移动产品负责人	原抖音社交负责人，曾孵化虚拟社交功能“抖音仔仔”，密友社交“抖音时刻”等社交项目。
齐俊元	FLOW PC产品负责人	原飞书产品副总裁，曾创立协作平台teambition，后担任阿里云盘负责人

2.2 头部互联网公司在AI领域布局对比



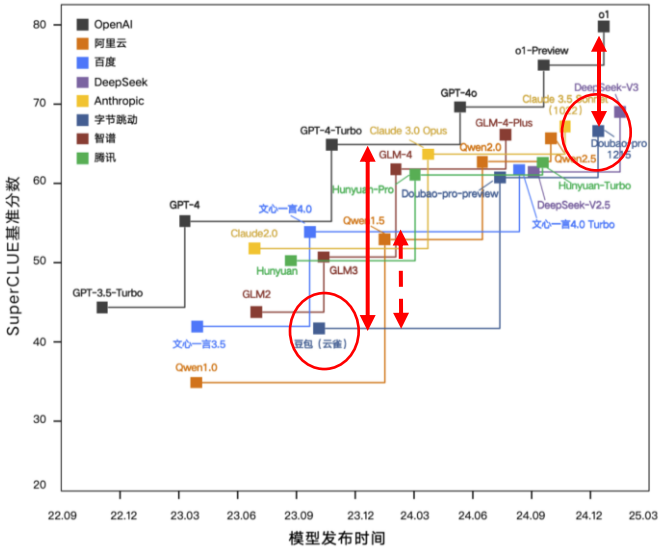
2.3 字节产业链充满机会：豆包大模型虽非顶尖，但快速追赶

- 字节的豆包大模型在总体能力上并非国内顶尖，但布局全面，且正在快速追赶。
- 豆包大模型家族2024年快速迭代，且已形成较全面的模型矩阵：7月推出豆包语音识别模型Seed-ASR和语音生成基座模型Seed-TTS，9月推出音乐生成模型框架Seed-Music、视频生成模型PixelDance与Seaweed，12月推出豆包代码大模型Doubao-coder和视觉理解模型，同时豆包大模型上下文窗口提升至300万字的业界极限；25年1月发布豆包大模型1.5Pro版本，综合能力显著增强，低训练/推理成本，全面提升多模态能力、推理能力，多项公开评测基准上全球领先。根据MMLU等国际权威基准测试以及国内SUPERCLUE测试结果，豆包大模型与国际顶尖大模型、国内头部大模型的差距正逐渐缩小。
- 2024年，豆包大模型支撑超过50个应用场景，其中豆包成为国内最受欢迎的AI产品，通过火山引擎豆包大模型服务30多个行业，日均tokens调用量超4万亿，较5月增长33倍。

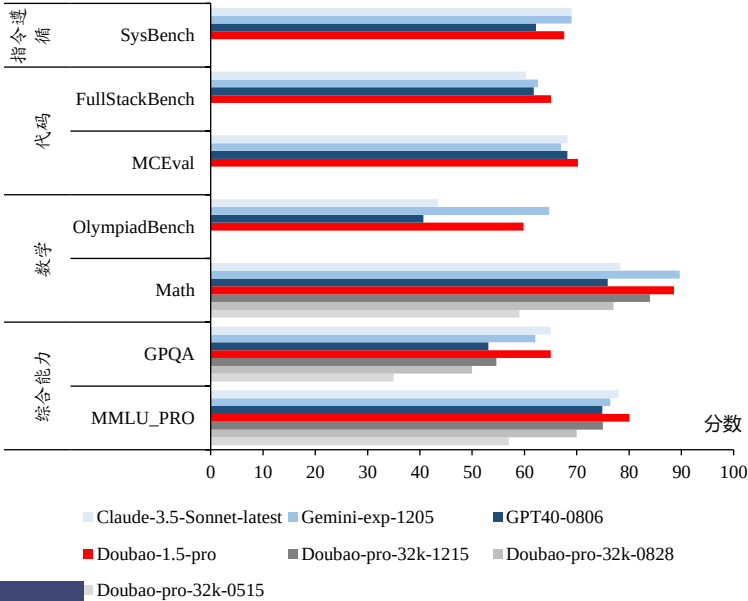
字节大模型布局全面



字节豆包大模型起步晚但快速追赶



注：红色实现表示豆包模型与同时段发布的国际领先模型间的能力差距，虚线表示



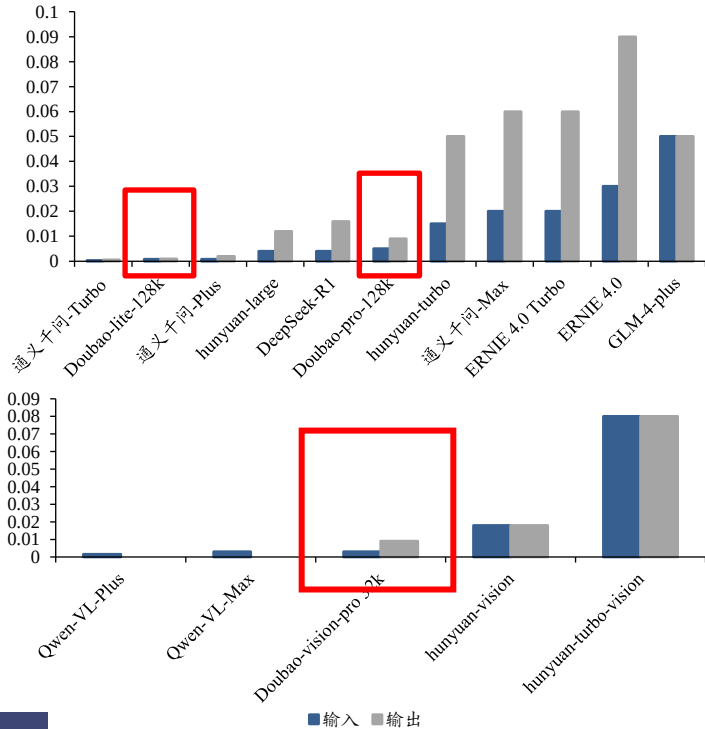
2.4 字节产业链充满机会：降低模型成本，推动应用加速

- **大厂中字节大模型率先大幅降价，国内大模型价格进入“厘时代”。**（1）2024年5月，在Deep Seek和智谱相继降低大模型价格后，字节的豆包大模型也跟进降价，旗舰模型价格降至每 0.0008 元 / 1K input tokens 和 0.005 元 / 1K output tokens，较行业均价降低99.3%，大模型价格进入“厘时代”。阿里云、百度云、腾讯云随之跟进，阿里大模型降价幅度接近97%，9月再次降价；百度ERNIE Speed、ERNIELite宣布免费，ERNIE 4.0 Turbo定价降70%；腾讯云混元多款大模型推理token降价，降价幅度50%-87.5%。（2）12月豆包视觉理解模型售价定为0.003元/ 1K input tokens，比行业均价便宜85%，视觉理解大模型价格进入“厘时代”；阿里也随之跟进。
- 大模型作为AI产业重要基建，其价格降低有望通过更低的试错成本提高企业大模型使用意愿，助力更多开发者开发AI应用，推动国内AI产业加速。

国内主要大模型2024年以来降价节点

	24年5-6月	24年7-9月	24年12月-25年1月
字节跳动	豆包大模型0.0008元/1K input tokens和0.005 元/1K output tokens，较行业平均价格降低99.3%		豆包视觉理解模型定价0.003元/ 1K input tokens，比行业平均价格便宜85%
阿里巴巴	Qwen-Long模型API输入价格降幅接近97%	Qwen-Turbo 输入价格降幅85%，Qwen-Plus与Qwen-Max降幅分别为80%和50%。	图像理解模型Qwen-VL-Plus 0.0015元/1K input tokens，降幅81%，Qwen-VL-Max 0.003元/1K input tokens，降幅85%。
百度	ERNIE Speed和ERNIE Lite全面免费	ERNIE 4.0 Turbo 定价相较文心一言4.0通用版降幅70%；ERNIE 3.5-8k、ERNIE 3.5-128k 输入价格以厘计。	
腾讯	混元-lite全面免费，混元-standard API 输入和输出价格分别降55%和50%，混元-standard-256k 模型API输入和输出价分别降87.5%和50%，混元 - pro 模型输入价格降幅70%。		
deepseek	DeepSeek-V2输入、输出价格分别降为1元 /100万tokens、2元/100万tokens，仅为GPT4-Turbo的1%		DeepSeek-V3 API价格每百万输入tokens 0.5元（缓存命中）/2元（缓存未命中），每百万输出tokens 8元。
智谱	个人版GLM-3Turbo调用价格降低80%。GLM-4 Air价格降至原来的1%；GLM-4-flash API 相较GLM-3 Turbo 降价10 倍		

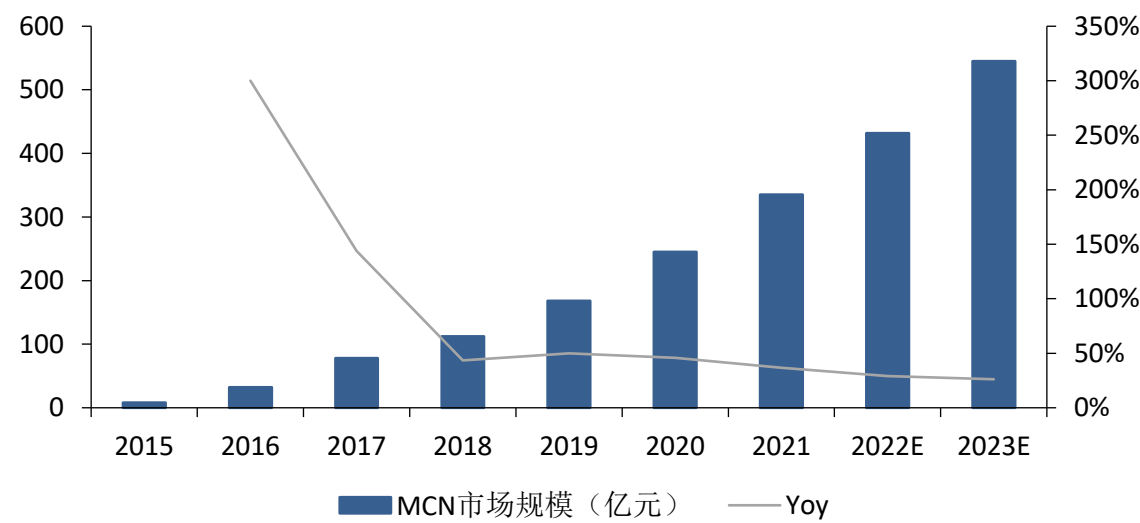
国内主要通用大模型（上图）和视觉理解大模型（下图）API调用价格对比（元/千tokens，截至2025年1月）



2.5 字节产业链充满机会：打造AI大生态，多个产业链迎红利

- **字节正在打造新的AI生态，且愿意让利合作伙伴，产业链上下游玩家迎来新红利。**
- 字节目前在AI领域采用全面进攻姿态，不管是面向多个垂类行业寻找大模型落地场景，还是在硬件终端和C端应用展开更多关于AI AGENT形态的探索，这一过程都需要合作伙伴一起建设生态。字节跳动一向也偏好先让利和扶持合作伙伴，共同将蛋糕做大，并最终与优质的合作伙伴实现共赢。因此我们看好在字节当前大力投入打造AI大生态的过程中，多个产业链迎来新的发展红利。

受益于抖音等短视频平台兴起，国内MCN产业实现高速发展



字节跳动在AI应用多个领域积极布局

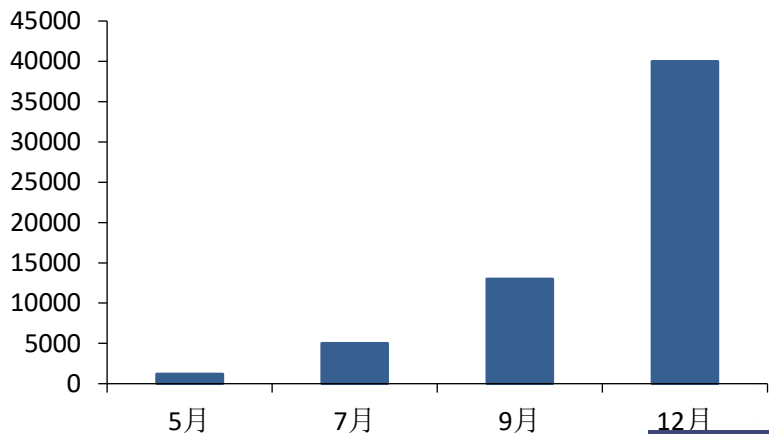
类型		产品
软件	智能助手	豆包、CiCi(海外)
	工具集	小悟空
	社交	猫箱、AnyDoor(海外)
	图像	星绘、PicPic(海外)
	图片/视频生成	即梦AI、Dreamina(海外)
	教育	豆包爱学、识典古籍、Gauth(海外)
	音乐生成	海绵音乐
	智能体开发平台	扣子、Coze(海外)
	抖音电商内容生成	即创
	编程助手	豆包MarsCode、MarsCode(海外)
	模型分享社区	炉米Lumi
硬件	智能体耳机	Ola Friend
	机器人	地瓜机器人
	智能台灯	大力智能学习灯
	智能玩偶	显眼包
已有应用添加AI功能	视频剪辑工具	剪映、Capcut(海外)
	修图工具	醒图、Hypic(海外)

3、字节AI生态及产业机会框架初探

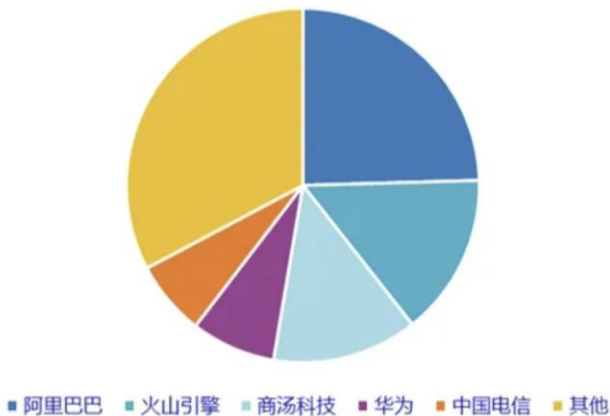
3.1 云：火山引擎有望弯道超车，助字节卡位AI时代重要基建

- **大模型对算力的需求给字节云服务一次弯道超车的机会。**
 - 相比阿里巴巴（2011年）、腾讯（2013年）、百度（2015年）、华为（2017年）等大厂，字节向外推出云服务的时间较晚（2021年火山引擎正式发布全系云产品），在客户基础、传统云基础产品等方面存在不足。
 - 但随着大模型训练推理需求不断增多，云服务客户还关心大模型相关的AI infra等能力，这方面所有云厂商都在同一起跑线，而火山引擎的迭代速度几乎是厂商里面最快的。火山引擎针对AI infra进行了多项优化（以GPU为中心的AI infra分布式业务体系、使用GPU直接远程内存访问技术等等），推动大模型成本降低，吸引更多细分行业的需求，反过来推动火山引擎寻找更高效云服务方案，实现飞轮效应。此外豆包作为国内日活最高的AI应用，也为火山引擎云服务在AI领域的能力提供标杆案例。
- **火山引擎在2024年进展飞速：**2024年上半年火山引擎在国内GENAI IaaS服务厂商中排名第2，好于其在整个云服务市场中的排名；根据智能超参数，2024年全年在公开市场中标24单项目，在大模型公司中排名第4，项目较多集中在四季度。火山引擎在大模型领域的增长也带动其传统云计算服务，一定比例的CPU 算力、数据库等需求来自大模型客户。从豆包大模型日均token使用量变化也可以看出火山引擎增长态势：从2024年5月的1200亿增长至12月的4万亿，增长32倍。**展望2025年，随着豆包大模型能力增强和成本降低，火山引擎以更高性价比和更专业的GEN AI 云服务拓展更多行业客户，其份额有望进一步提升，助力字节卡位AI时代重要的基建环节。**
- **产业机会：**随着云调用量提升，特别是推理需求增加，字节有望加大算力相关投入，服务器、AI芯片、PCB、散热等产业链环节的合作公司有望受益于订单增长。

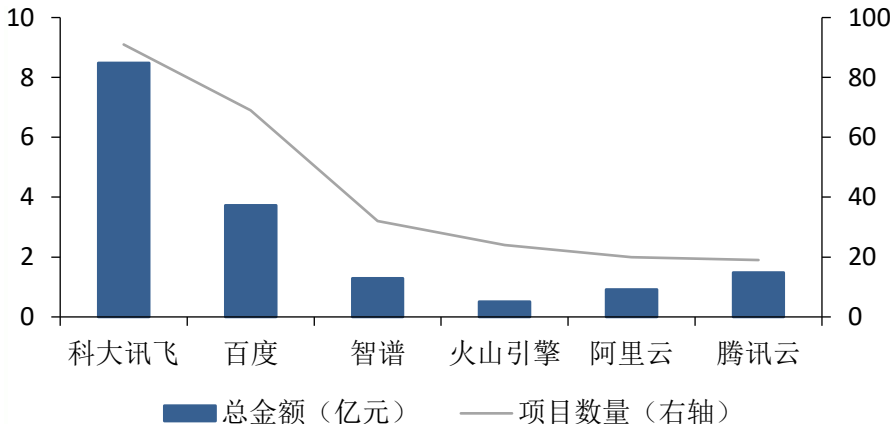
豆包大模型日均token使用量增长情况（亿）



2024H1国内TOP5 GENAI IaaS服务厂商市场份额



2024国内大模型厂商中标情况



3.2 模型服务/开发：火山方舟+扣子+HiAgent，完善模型服务生态



- 字节已针对模型服务及应用开发推出三个平台，火山方舟、HiAgent和扣子coze，企业使用门槛从高到低。尤其是HiAgent和扣子coze降低了企业/开发者开发agent的成本，更好加速AI应用迭代，当前企业大模型应用的真实需求并不十分明确，所以敏捷开发、迅速试错的能力是关键。
- 高代码工具 “火山方舟”：2023年上线的大模型服务平台，为企业客户提供大模型精调、推理、评测、部署等服务，较为依赖企业的开发和工程经验。
- 低代码工具 “HiAgent”：2024年5月上线的企业专属AI应用创新平台。企业在开发AI应用过程中面临AI专业人才不足、担心数据泄露、专属应用开发难、集成到业务系统中麻烦和效果难以评估等问题，HiAgent把企业开发 AI 应用的工具封装成 “聊天机器人” “插件” “工作流” “知识库” 等模块，方便调用，同时高度适配企业个性化需求，让业务人员可以轻松构建智能体。
- 无代码工具 “扣子coze专业版”：2024年2月上线的AI应用开发平台，各种大模型封装成API 接口，开发者可快速调用大模型，零代码开发、调试和优化 AI 应用。
- 产业机会：建议关注字节降低模型开发门槛后对垂类AI应用加速发展带来的潜在影响。

火山引擎模型和应用开发服务



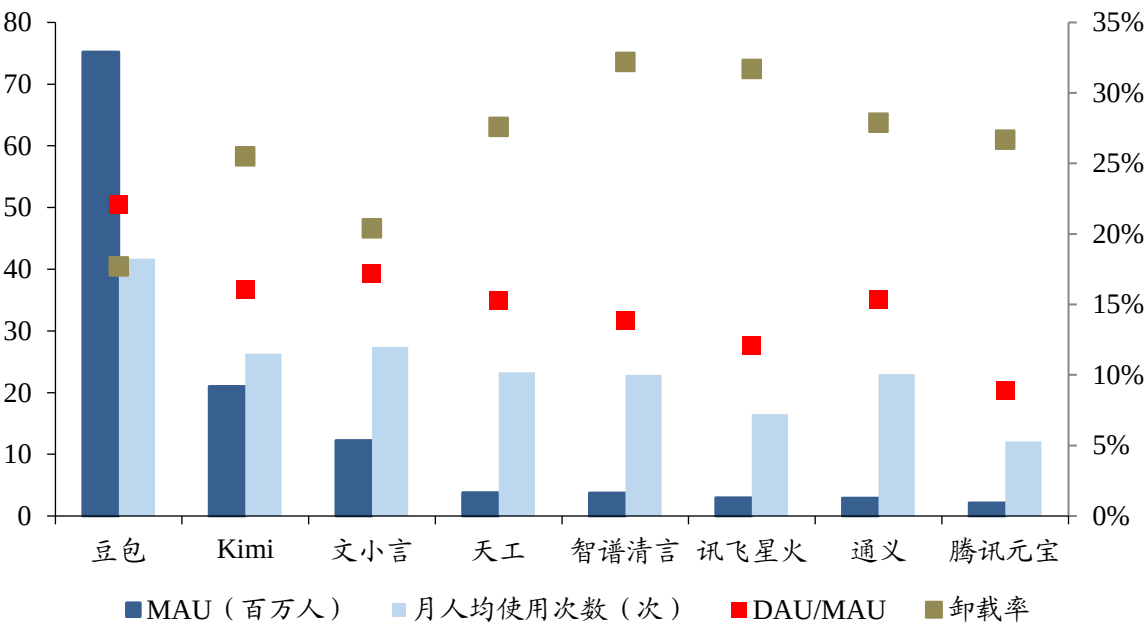
字节扣子对比腾讯元器

		字节扣子	腾讯元器
Agent编排能力	对话创建bot	是	否
	prompt	提供自动优化提示词能力	提供提示词模版
	插件	累计插件数量共768个，其中官方插件仅50+左右	累计117个插件，官方插件仅3个
	知识库	支持文档、URL、表格、图片等	仅支持文档格式上传
	工作流	支持自定义工作流，并提供工作流模版，同时还支持图像流功能	是
	大模型	支持豆包、通义、moonshot、minimax、百川等多个商业大模型	仅支持混元大模型
记忆		支持数据库、变量、长期记忆、文件盒子多种形式的记忆能力	否
C端开发者能力限制		有限的bot使用次数	无限bot使用次数
企业端商业化能力建设	团队空间	是	否
	API/SDK	是	仅支持API调用
	企业版	是	否

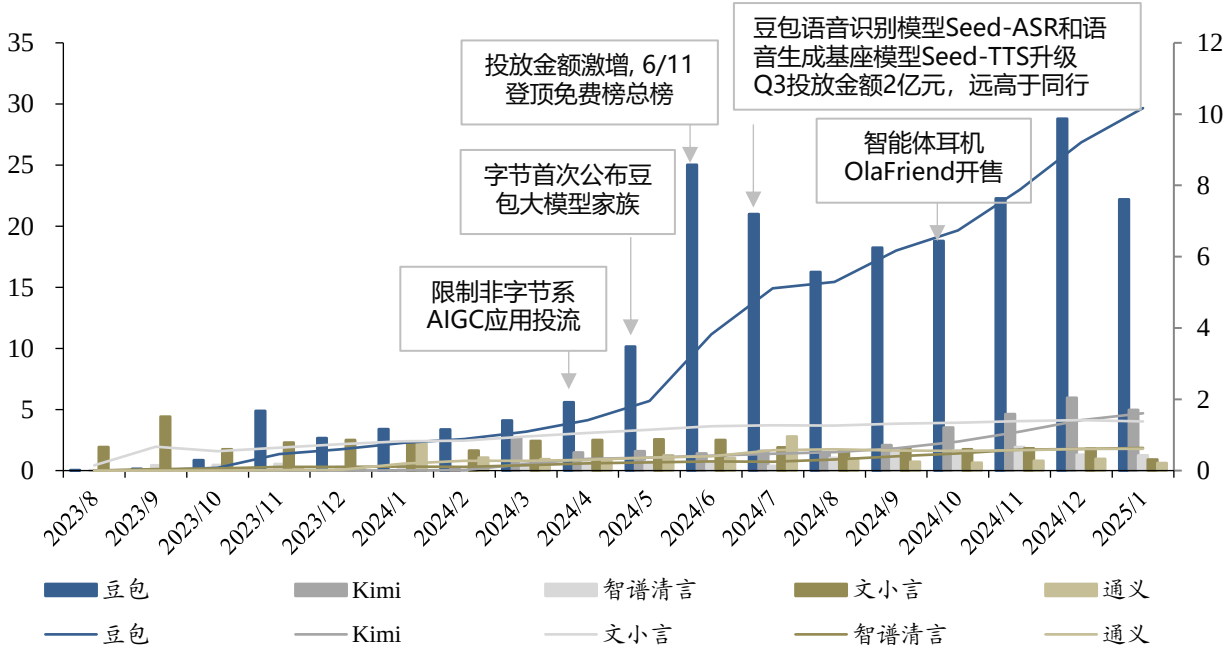
3.3 AI助手豆包：高举高打+优秀用户体验，用户规模与质量双高

- 豆包是字节基于豆包大模型推出的类chatgpt产品（海外版为cici），相比其他聊天机器人应用，**豆包**在具备较好模型性能和较全面功能的基础上，**提供了优秀的用户体验，实现了更好的用户活跃和留存。叠加大规模买量等原因，豆包应用用户数在2024年6月份之后迎来高速增长，一跃成为全球用户数第2大的AI应用。**
- **豆包移动端提升用户体验的核心能力之一：领先同行的实时语音聊天功能，在情感理解和表达能力等方面表现优秀。**2024年7月，基于全新的豆包语音识别模型Seed-ASR和语音生成基座模型Seed-TTS，豆包大模型能听懂20种以上方言夹杂的对话，学会在说话中表达情感，可随时被打断，能在交互中“插话”，还能保留吞音、口音等人类习惯。2025年1月更新推出实时语音通话功能，根据豆包团队测试结果，豆包实时语音大模型在测试用户整体满意度的评分为4.36，GPT-4o 为 3.18。其中50% 的测试者对豆包实时语音大模型表现打出满分；在“一听就是 AI 与否”评测中，反馈GPT-4o和豆包实时语音大模型“过于 AI ”的比例分别为超过 30%和2%以内。模型在情感理解、情感承接以及情感表达等方面取得显著进展，能较为准确地捕捉、回应人类情感信息。

国产综合类AI原生APP 2024年12月用户规模及用户质量对比



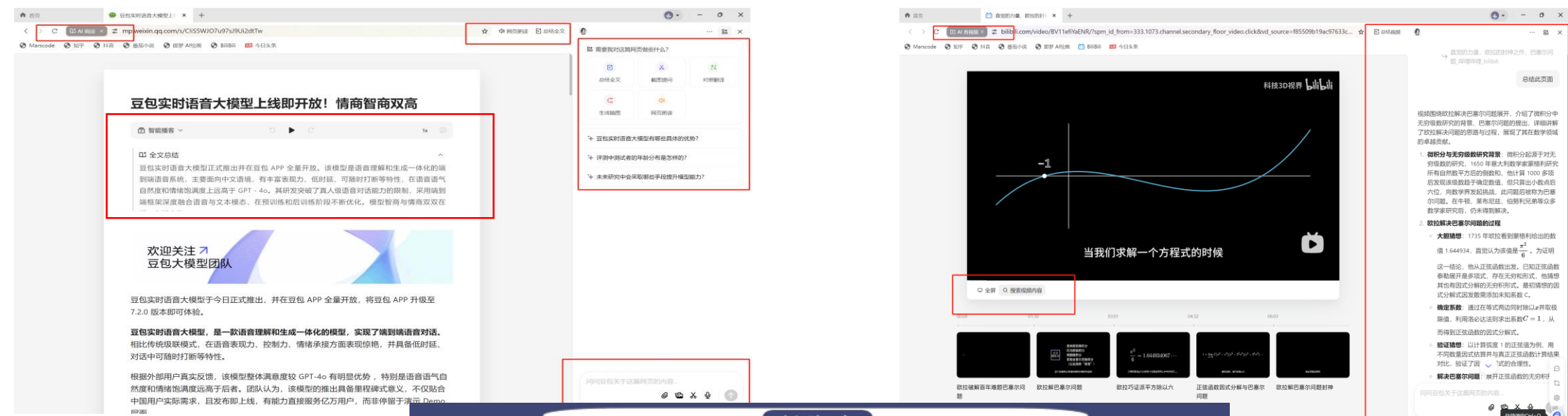
国内AI APP iOS端MAU（折线图，百万人）及下载量（柱状图，百万次，右轴）变化趋势



3.3 AI助手豆包：期待PC端在学习办公场景的渗透率提升

- 豆包电脑版在定位和功能全面性上领先同行，可兼做聊天机器人、浏览器和类微软copilot产品，且免费，我们预计有望迅速占据AI核心用户群体在学习办公场景下的心智。
 - 信息处理类（相关功能可在其他应用电脑端上划词使用）：AI 搜索、划词搜索、AI 总结、AI 伴读。
 - 写作辅助类：AI写作、学术搜索。
 - 多模态类：截图提问、图像生成。
 - 办公协作类：一键录音与实时转写、云端存储、任务管理。
 - 其他功能：个性化定制、跨平台同步等
- 产业机会：我们认为豆包有望成长为新的多端流量入口，与字节达成合作的营销、内容安全、多模态技术等服务商有望受益。

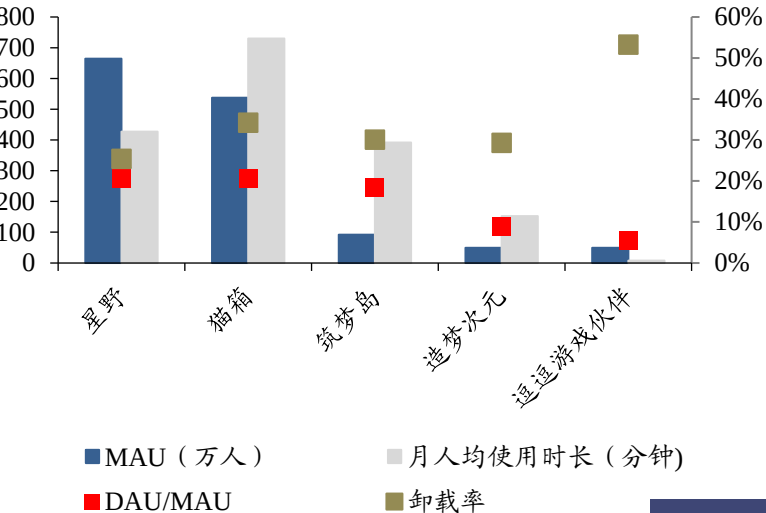
豆包电脑端对文章和视频进行AI总结



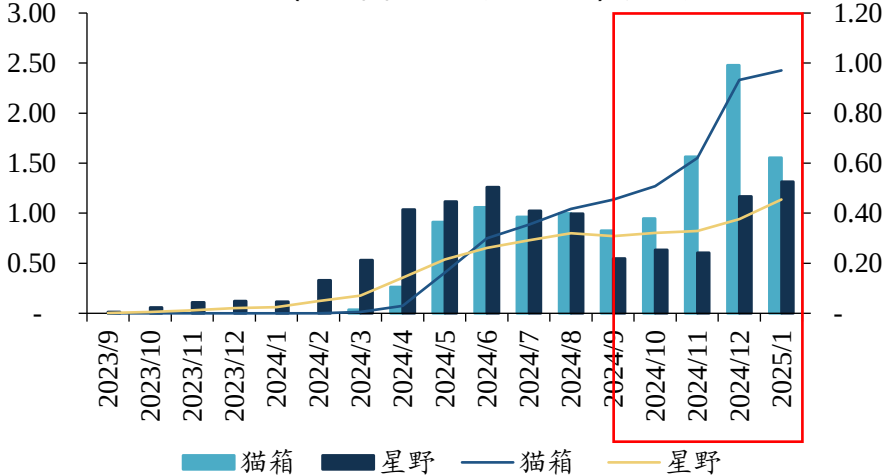
3.4 AI娱乐猫箱：2C重要产品，有望成为年轻人新娱乐应用

- 猫箱作为一款类character.ai应用（海外版为Anydoor），已成为字节旗下用户规模第二大的C端AI原生应用。
- 猫箱2024年10月开始增加素材投放，根据SENSOR TOWER，猫箱iOS端下载量和MAU Q4快速爬坡，迅速超过原来在娱乐类AI原生应用中排名第一的星野（minimax旗下应用）；根据QUEST MOBILE数据，2024年12月猫箱APP MAU 537万人，相比6月增长8%，增长表现好于星野（12月MAU相比6月下滑5%）。
- **我们认为猫箱“娱乐内容”属性更重，星野“乙女游戏”属性更重。**猫箱在AI角色智能程度以及声音等方面表现更好（前文所提的豆包实时语音大模型有望进一步提升用户体验），内容题材更多元（我们抽样测试中，趣味、工具类推荐占比高于星野），用户时长好于星野；但在用户与AI角色的情感深度、互动玩法多样性、角色记忆能力等方面比星野逊色，卸载率高于星野（我们认为这是因为星野用户对AI角色的粘性更强）。这也使得猫箱有着更均衡的男女比例，更年轻化的用户群体。
- **产业机会：猫箱有望成为国内AI时代重要的娱乐内容形态代表，成为继抖音等短视频之后年轻人新的娱乐应用，这个过程中手握优质IP资源的公司有望拓展IP变现方式。**

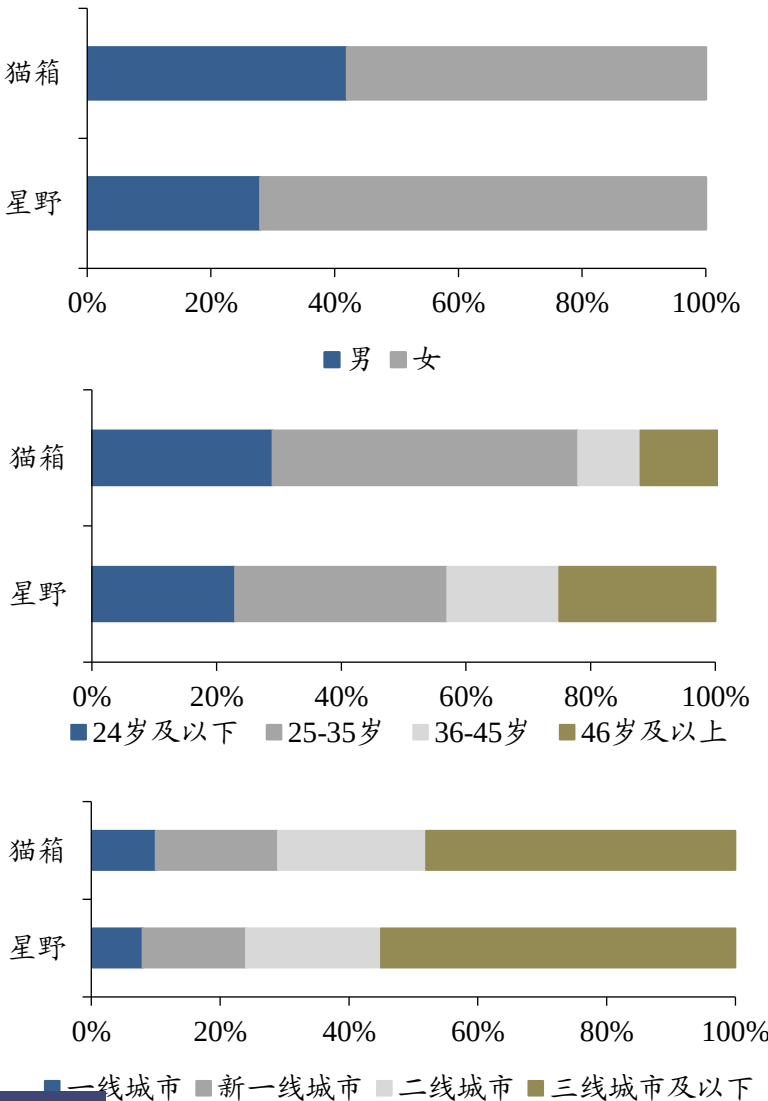
国产社交/娱乐类AI原生APP 2024年12月用户规模及质量对比



猫箱和星野APP iOS端MAU（折线图，百万人）及下载量（柱状图，百万次，右轴）变化趋势



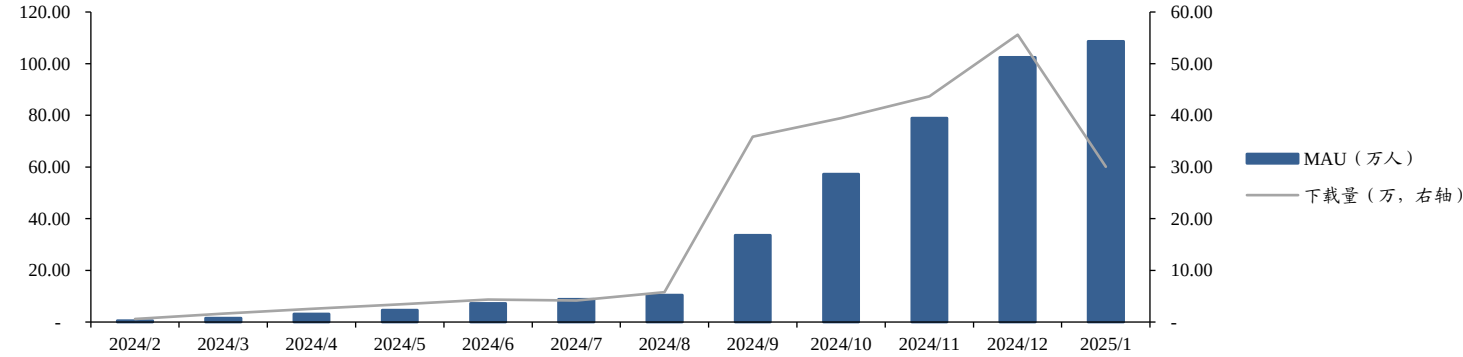
星野及猫箱用户画像对比（2024年12月）



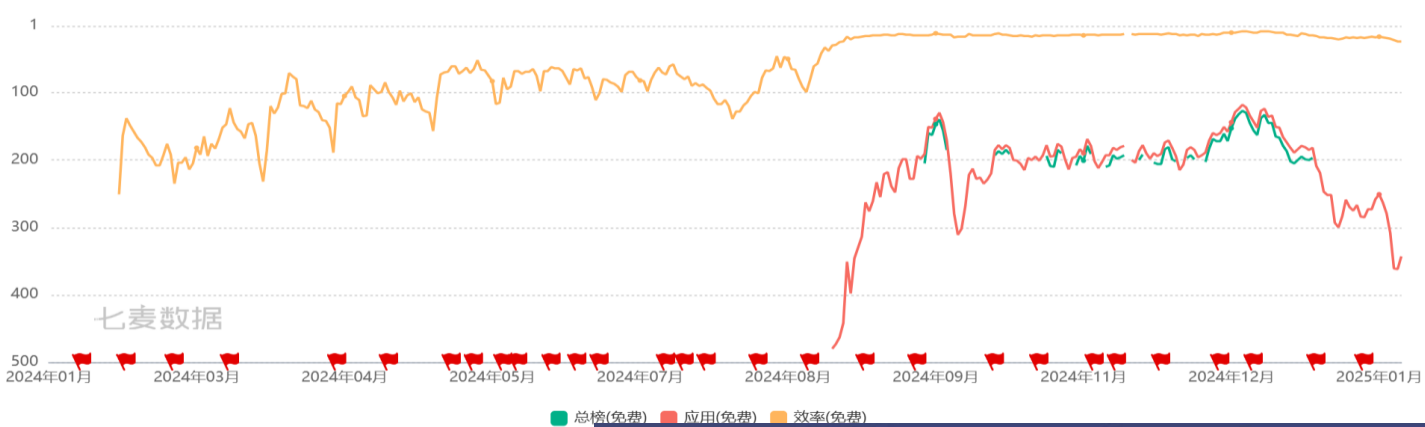
3.5 AI教育豆包爱学：建议关注产品重大迭代和用户增长趋势

- 豆包爱学是一款AI教育应用，原字节大力教育旗下的河马爱学，后更名为“豆包爱学”。2024年9月以来，豆包爱学APP在iPhone效率APP免费榜排名稳居前20名，下载量和MAU提升迅速。
- 目前国内尚未跑出特别有代表性的AI教育应用，主要系模型幻觉问题难以彻底解决以及教育场景容错率低等原因，但考虑到字节大模型在与学习场景相关的数学推理、空间与计数理解等多项视觉理解基准上性能领先，我们看好豆包爱学后续继续以优于同行的产品体验实现用户快速增长。
- 建议关注大模型推理性能增强对AI教育行业的影响、豆包爱学的重大迭代以及用户增长趋势

豆包爱学APP iOS端MAU 及下载量变化趋势



豆包爱学APP在iPhone APP榜单排名情况



豆包PRO 1.5模型在数学推理、空间与计数理解等多项视觉理解基准上性能领先

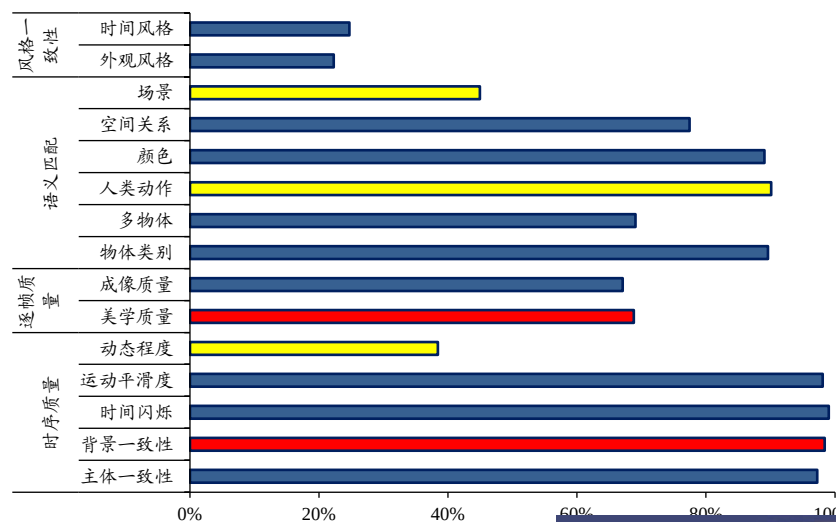
Benchmark		Doubao-1.5-pro	GPT4o-1120	Claude3.5-Sonnet	Gemini-2-flash	Qwen2-VL-72B	InternVL-2.5-78B
College-level Problems	MMU(val)	73.8	70.7	70.4	70.7	64.5	70.1
	MMU-Pro	59.3	54.5	54.7	57.0	46.2	48.6
Mathematical Reasoning	MathVision	48.6	30.4	38.3	41.3	25.9	32.2
	OlympiadBench	48.5	25.9	27.8	43.6	11.2	25.1
	MathVista	78.8	63.8	65.4	73.1	70.5	76.6
Document and Diagrams Reading	TextVQA(val)	84.7	81.4	76.5	75.6	85.5	83.4
	ChartQA(test avg)	88.0	86.7	90.8	85.2	88.3	88.3
	InfoVQA(test)	88.0	80.7	74.3	77.8	84.5	84.1
	DocVQA(test)	96.7	91.1	95.2	92.1	96.5	95.1
	Charxiv(RQ/DQ)	54.4 / 84.3	52.0 / 86.5	60.2 / 84.3	55.2 / 81.8	43.0 / 81.3	42.4 / 82.3
General Visual Question Answering	RealWorldQA	78.9	75.4	66.6	74.5	77.8	78.7
	MMStar	71.9	63.9	65.1	69.4	68.6	73.1
	MMBench-en	87.5	83.5	81.7	83.0	85.9	88.3
Spatial and Counting Understanding	MMBench-cn	86.0	82.1	83.4	82.9	83.4	88.5
	Blink	68.4	68.0	59.6	62.6	61.1	63.8
Video Understanding	CountBench	89.6	85.1	86.8	88.2	88.6	84.1
	Video-MME	74.1	73.4	61.7	78.2	71.2	72.1
	EgoSchema-subset	75.4	74.8	64.4	71.8	80.6	78.2

3.6 AI创作即梦：处在追赶阶段，看好长期潜力，建议关注技术迭代

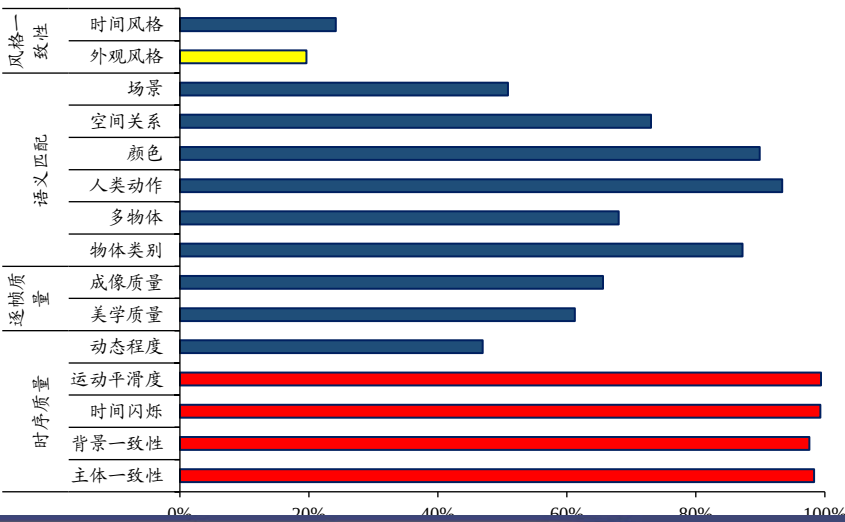


- 即梦是字节旗下AI创作平台，核心功能包括图片生成、智能画布、视频生成、故事创作及音乐生成，底层的模型包括图片生成模型豆包文生图模型、视频生成模型Seaweed和PixelDance、音乐大模型豆包音乐生成模型等。考虑到图片生成技术相对成熟，音乐生成商业化空间相对有限，这里我们更多讨论字节在视频生成领域的现状和潜力。
- 我们估计即梦视频生成能力总体在国内处在第二梯队，现阶段对创作者的吸引力弱于快手的可灵、minimax的海螺等产品。基于VBENCH的测试结果，同时考虑竞对模型版本的迭代，我们估计即梦视频生成模型整体能力在国内视频生成模型中处在第二梯队。前文提到时序一致性及语义匹配是目前视频生成模型仍在攻克的难点，在一致性上表现优秀的可灵背后是快手对其更多的技术人才投入及资源投入，顶尖技术人才带队做技术突破，而字节目前更偏向于提升基座模型能力和全方位布局，在视频生成这个需要高人才密度做前沿技术研究的方向难以很好兼顾，也无法简单复用其在通用大模型上的技术路径（可参考海外GOOGLE 的VEO系列模型表现优于OPENAI的SORA），因而在技术上相对落后，产品可用性相对较差，对创作者吸引力不如头部模型。根据aipcb.com，即梦AI 2024年12月网页访问量219万次，APP MAU 275万人，体量较小。
- 但基于模型亮点、生成效率和高效产品迭代等原因，我们看好其在创作者端占据一定心智及长期发展潜力，后续建议关注行业开源技术进步、团队技术人才变动及模型重大迭代。即梦在某些能力上表现优秀，如生成的单帧图像质量（部分创作者会先用即梦生图，再用其他视频生成模型图生视频）、24年12月领先同行推出的中文文字生成能力等；生成效率上亦有较为明显的优势，排队和生成时间均较短（1分钟左右即可生成视频，竞对2-5分钟）。

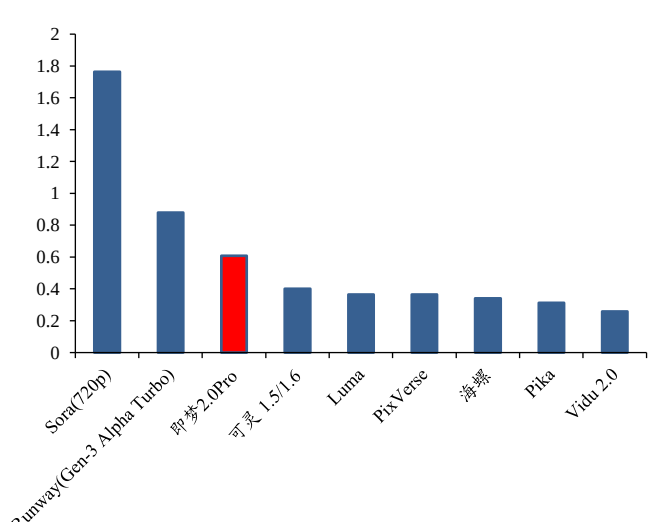
即梦视频生成模型VBENCH得分情况（测试时间为2024/11）



可灵1.0视频生成模型VBENCH得分情况（测试时间为2024/7）



单月购买模式下视频生成模型定价对比（元/秒）



3.7 AI终端：百花齐放、持续迭代

- 随着字节豆包、DeepSeek、阿里通用等国产大模型生态共荣，我们看好端侧赛道的投资机会。举例来看，在AI玩具、AI宠物、AI眼镜、AI耳机、AI音箱等智能硬件中，豆包实时语音大模型能够为其赋予更强大的语音交互功能，使这些硬件不再仅仅是简单的播放设备或执行单一任务的工具，而是可以与用户进行深度互动的智能伙伴，拓展了智能硬件的应用场景和功能边界。

Ola Friend 是字节跳动旗下豆包推出的首款 AI 智能体耳机。
发布时间：2024 年 10 月 10 日发布，10 月 17 日正式发货。
产品定位：一款更轻更舒适的开放式智能体耳机，定位为用户身边的 AI 伙伴。



显眼包 AI 玩具是字节跳动在 2024 年中秋推出的一款 AI 陪伴玩偶。

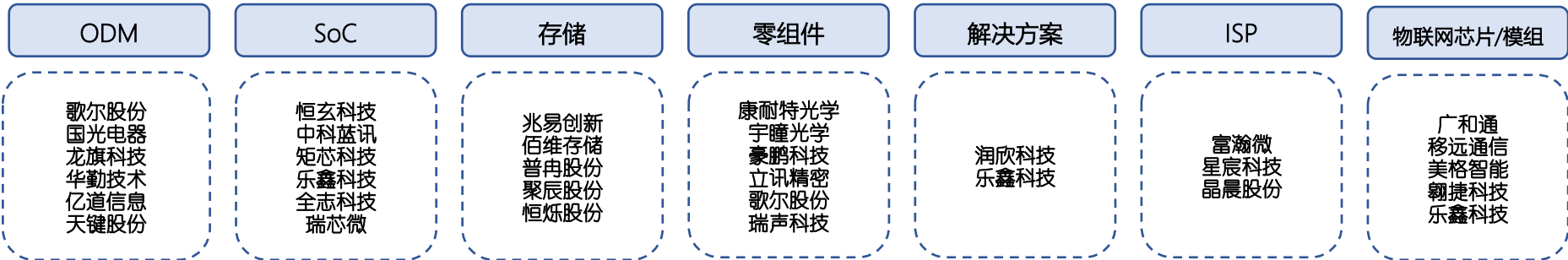
内嵌 FoloToy 的大模型 AI 机芯 Magicbox（魔匣），可搭载并呈现豆包大模型和扣子专业版的能力，为其提供强大的智能交互基础。



展望未来，豆包大模型的持续升级有望持续赋能AI眼镜、AI耳机、AI玩具、AI音箱等新型智能硬件。

除了推出自有品牌产品、投资相关品牌外，TOKEN分成模式也持续打开硬件公司想象空间。

- 大模型的发展持续赋能端侧硬件产业机遇，建议关注1) 端侧芯片及模组:恒玄科技、全志科技、移远通信、乐鑫科技、润欣科技、中科蓝讯、星辰科技、兆易创新、瑞芯微、广和通、移远通信、美格智能、翱捷科技等；2) 端侧零部件及组装：歌尔股份、亿道信息、天键股份、龙旗科技、康耐特光学、华勤科技、瑞声科技等



4、投资建议

4 投资建议：推荐算力及下游应用合作伙伴，关注AI应用板块性机会

- 投资机会上，我们认为字节的投入加速了国内AI大模型及应用的进展，未来有望给算力、数据/IP提供商及下游细分链条上的合作伙伴（营销、内容安全、多模态技术、AI玩具、AI眼镜、AI耳机等）带来增量和投资机会；国内AI应用亦有望迎来更多落地。具体标的上：
 - 数据：建议关注掌阅科技、海天瑞声、视觉中国、丝路视觉等。
 - AI玩具及情感陪伴方向：推荐昆仑万维、建议关注实丰文化、乐鑫科技、盛天网络、汤姆猫、奥飞娱乐等。
 - AI硬件：1) 端侧芯片及模组建议关注恒玄科技、全志科技、移远通信、乐鑫科技、润欣科技、中科蓝讯、星辰科技、兆易创新、瑞芯微、广和通、移远通信、美格智能、翱捷科技等；2) 端侧零部件及组装环节建议关注歌尔股份、亿道信息、天键股份、龙旗科技、康耐特光学、华勤科技、瑞声科技等。
 - AI工具：推荐美图公司、万兴科技、创业黑马等。
 - AI游戏：推荐巨人网络、恺英网络等。
 - AI电商：推荐焦点科技、值得买、华凯易佰、吉宏股份、建议关注返利科技、光云科技等。

5、风险提示

- 模型能力提升不及预期。若国内外AI模型迭代不及预期，可能会导致算力及应用落地不及预期。
- 字节AI业务进展不及预期。若字节AI业务进展不及预期，可能会导致合作伙伴相关业务进展低预期。
- 竞争格局变化。AI产业仍处在发展早期，若竞争格局发生较大变化，可能对字节AI业务发展产生不确定性。

东吴证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本研究报告仅供东吴证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。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，本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。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。

在法律许可的情况下，东吴证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

市场有风险，投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息，本公司力求但不保证这些信息的准确性和完整性，也不保证文中观点或陈述不会发生任何变更，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。

本报告的版权归本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。经授权刊载、转发本报告或者摘要的，应当注明出处为东吴证券研究所，并注明本报告发布人和发布日期，提示使用本报告的风险，且不得对本报告进行有悖原意的引用、删节和修改。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

东吴证券投资评级标准

投资评级基于分析师对报告发布日后6至12个月内行业或公司回报潜力相对基准表现的预期（A股市场基准为沪深300指数，香港市场基准为恒生指数，美国市场基准为标普500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），北交所基准指数为北证50指数），具体如下：

公司投资评级：

买入：预期未来6个月个股涨跌幅相对基准在15%以上；

增持：预期未来6个月个股涨跌幅相对基准介于5%与15%之间；

中性：预期未来6个月个股涨跌幅相对基准介于-5%与5%之间；

减持：预期未来6个月个股涨跌幅相对基准介于-15%与-5%之间；

卖出：预期未来6个月个股涨跌幅相对基准在-15%以下。

行业投资评级：

增持：预期未来6个月内，行业指数相对强于基准5%以上；

中性：预期未来6个月内，行业指数相对基准-5%与5%；

减持：预期未来6个月内，行业指数相对弱于基准5%以上。

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应当充分考虑自身特定状况，如具体投资目的、财务状况以及特定需求等，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

东吴证券研究所
苏州工业园区星阳街5号

邮政编码：215021

传真：(0512) 62938527

公司网址：http://www.dwzq.com.cn

东吴证券 财富家园