

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

■ **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。



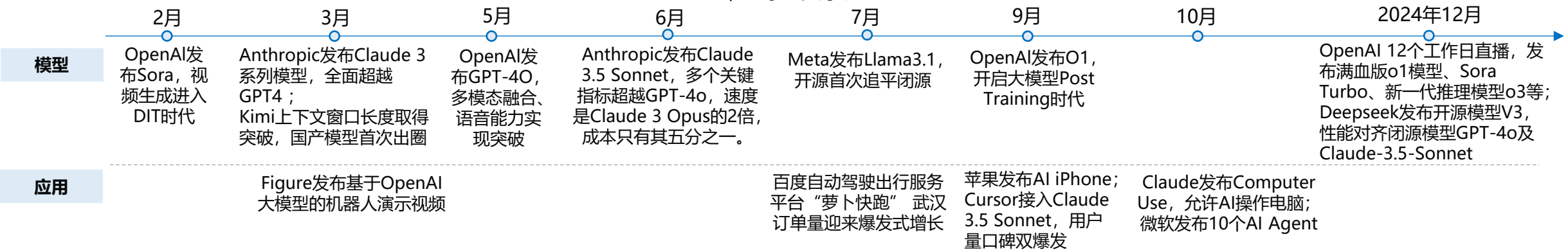
- 1、国内AI产业正进入加速时刻
- 2、为什么说字节链是最值得重视的AI产业机会之一
- 3、字节AI生态及产业机会框架初探
- 4、投资建议
- 5、风险提示

1、国内AI产业正进入加速时刻

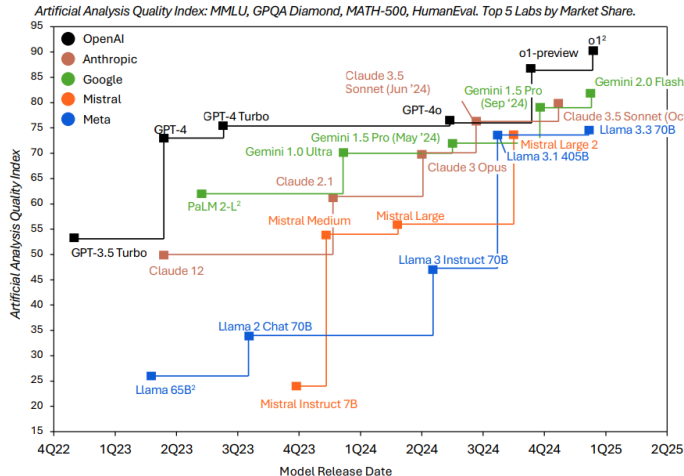
1.1 2024是重塑认知的一年：大模型间技术和份额差距逐步缩小

- 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）。

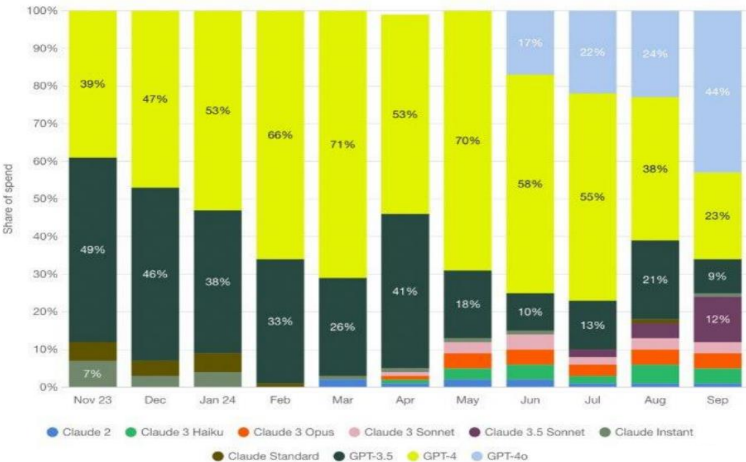
2024年全球AI大事记



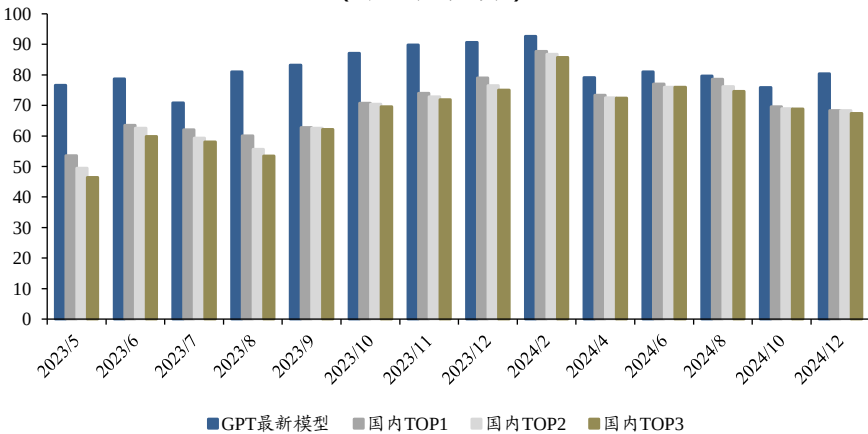
2024年其他大模型玩家与OPENAI的差距缩小



海外大模型token消耗量份额变化



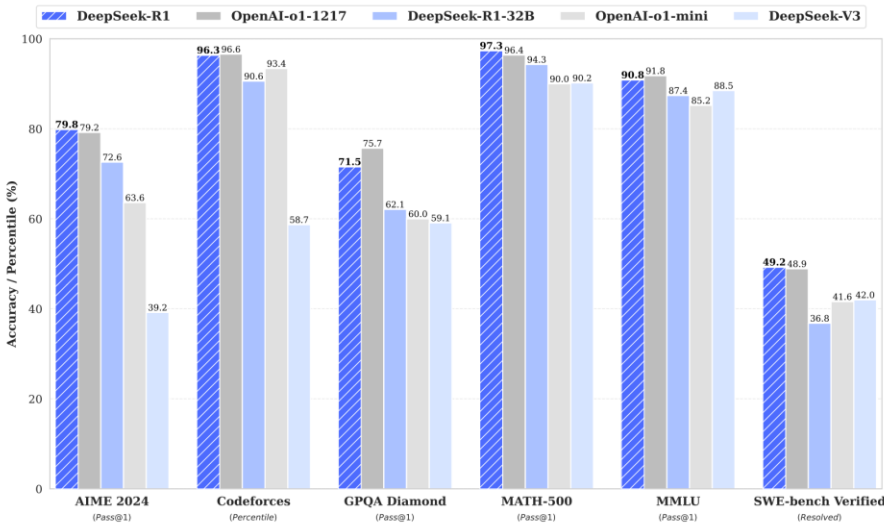
SUPERCLUE基准：国内TOP大模型与GPT最新模型的性能差距缩小（纵轴为分数）



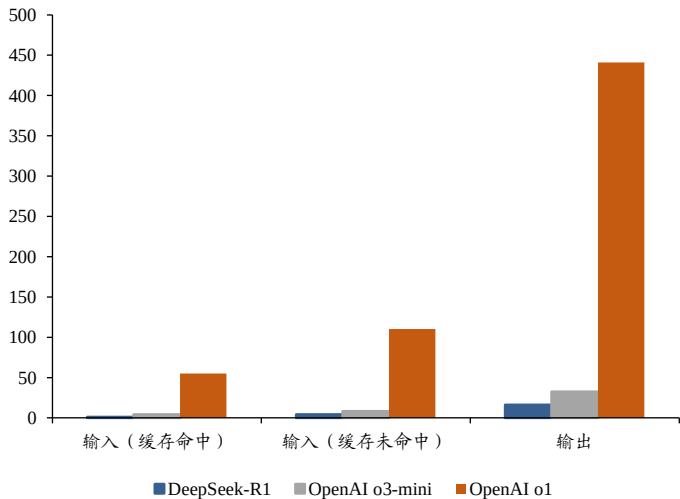
1.2 算力受限下找更优解，2025将是国产AI进击的一年

- **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。**

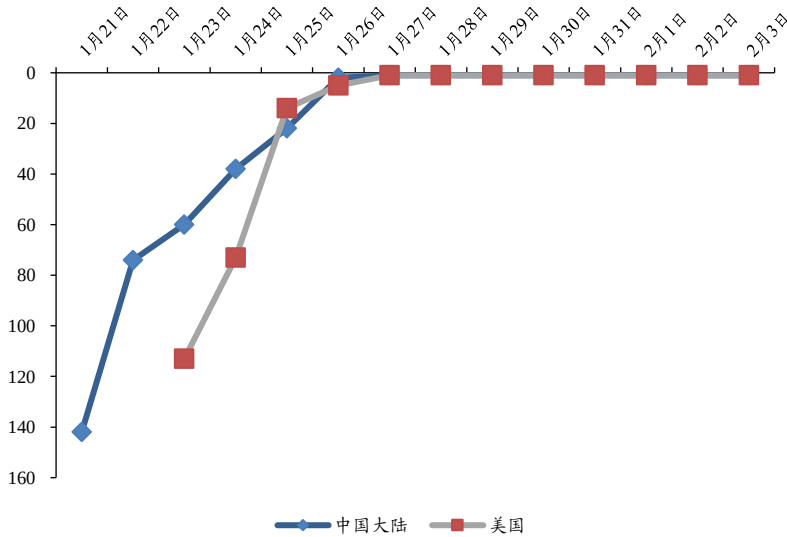
DeepSeek-R1基准测试结果对比openai模型



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）



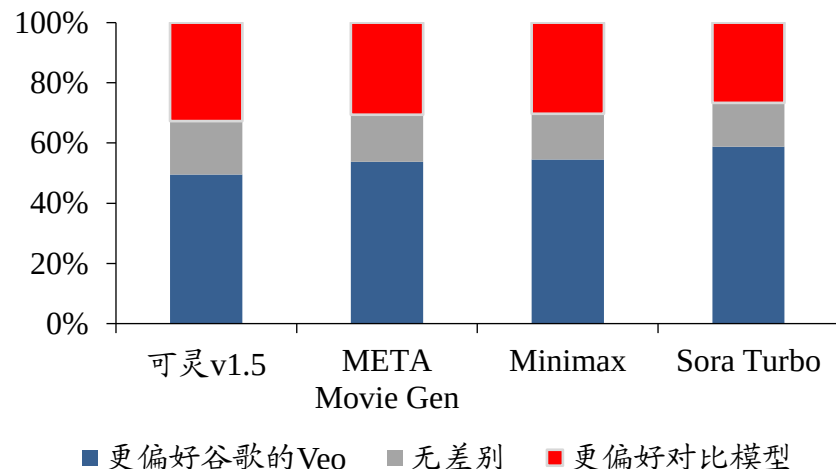
DeepSeek APP在中国大陆和美国iphone免费榜总榜排名



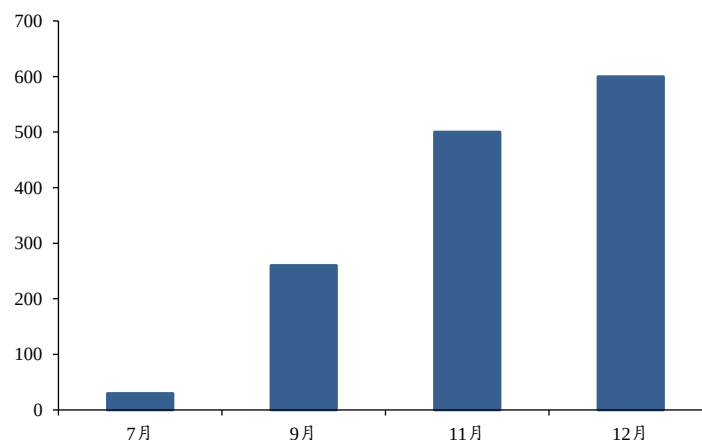
1.3 视频生成模型国内逐渐追平，期待2025年国内进展

- **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。**

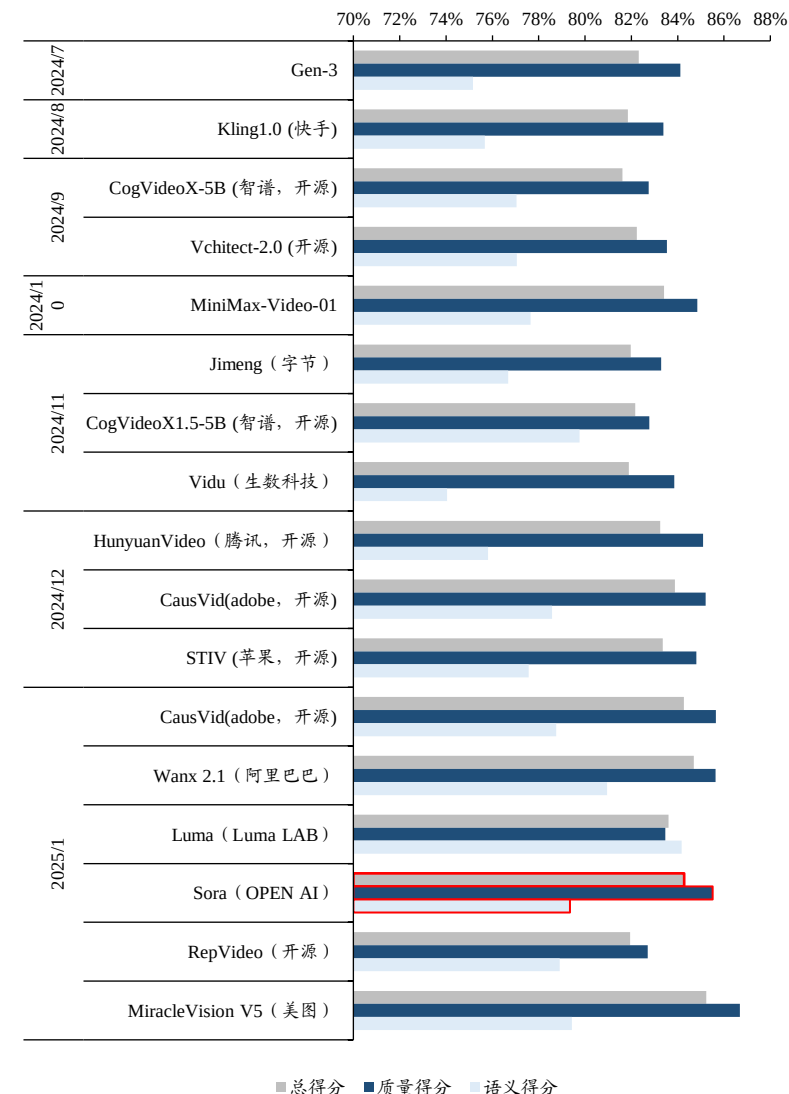
DEEPMIND测试用户在VEO模型和其他视频生成模型之间的整体偏好



2024年快手可灵用户数增长趋势 (万人)

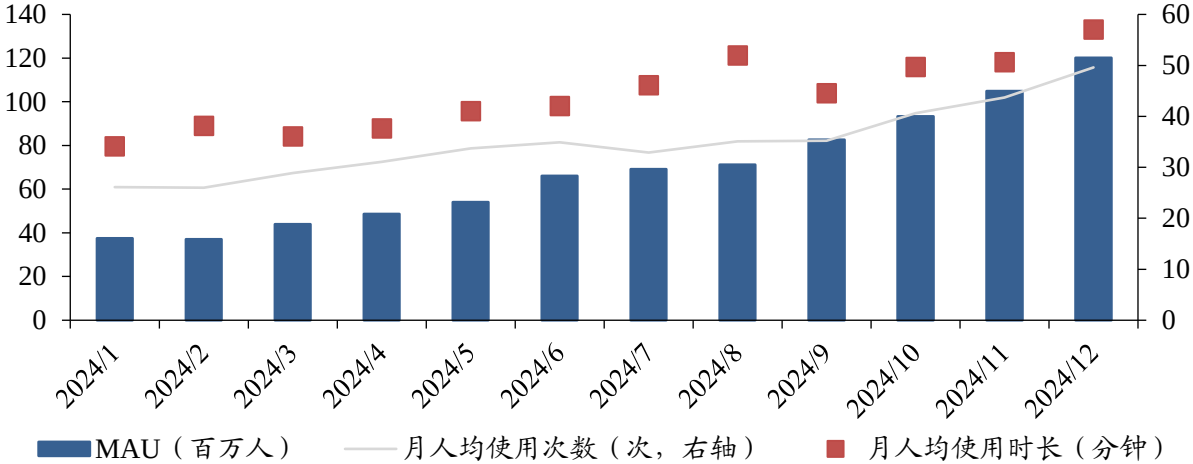


视频生成模型VBENCH得分对比

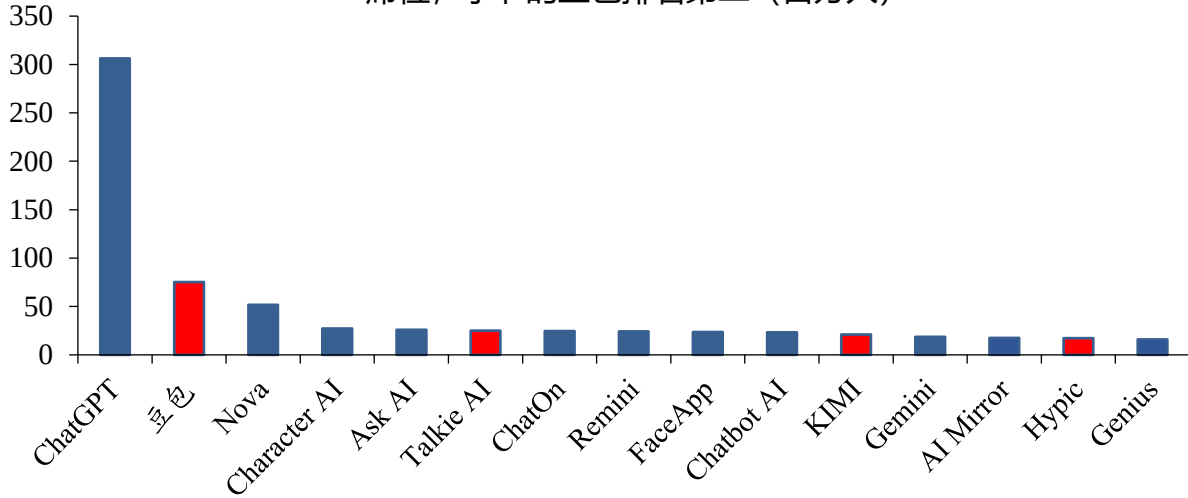


1.4 国内AI原生应用在网民中渗透率已超过10%，看好2025年加速

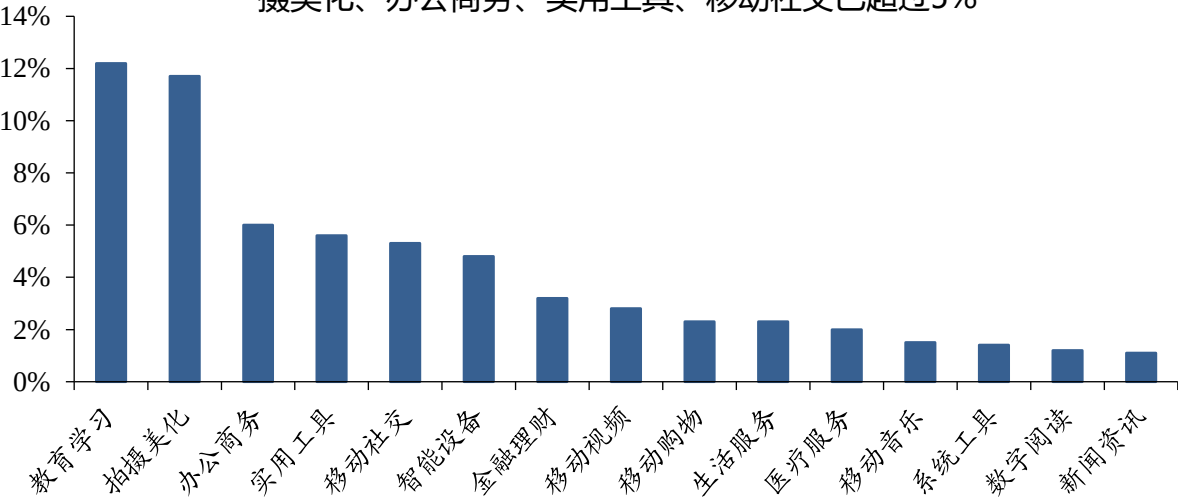
国内AI原生APP用户规模及粘性正快速提升，MAU在网民中的渗透率从24年1月的3%提升到12月的11%



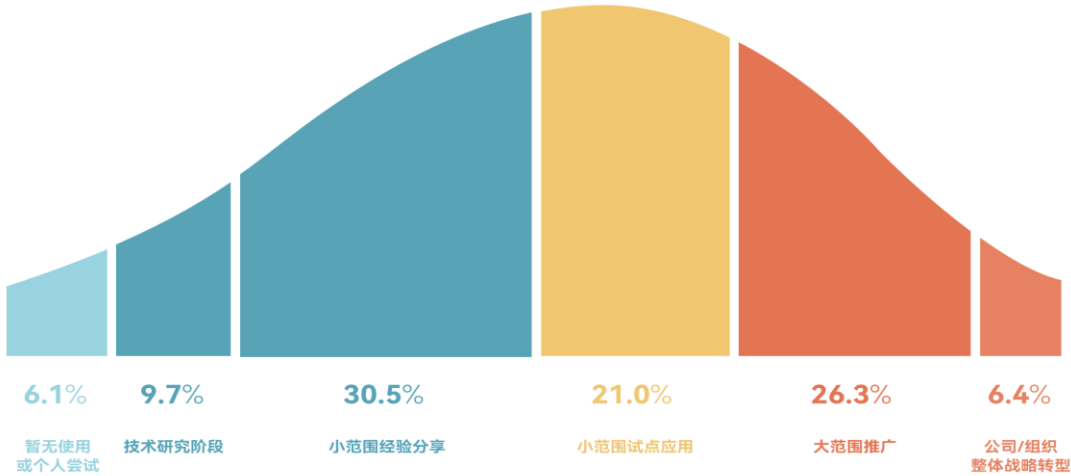
2024年12月全球MAU TOP15的AI APP 中，中国公司产品占据4个席位，字节的豆包排名第二（百万人）



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）



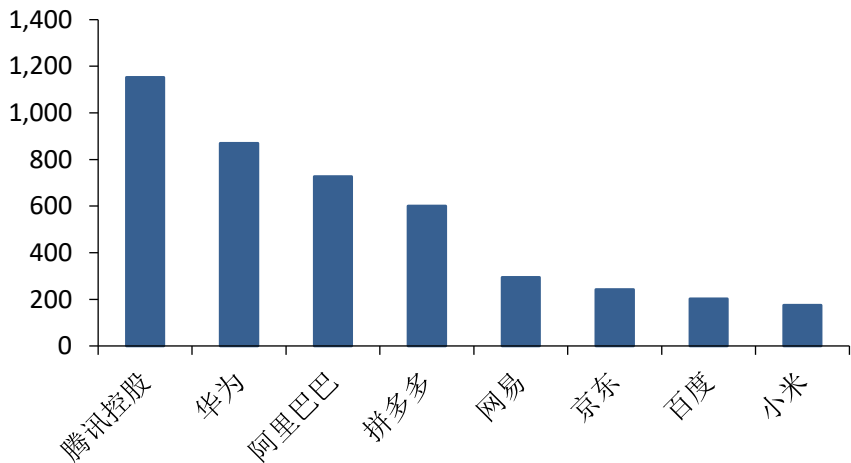
2、为什么说字节链是最值得重视的AI产业机会之一

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



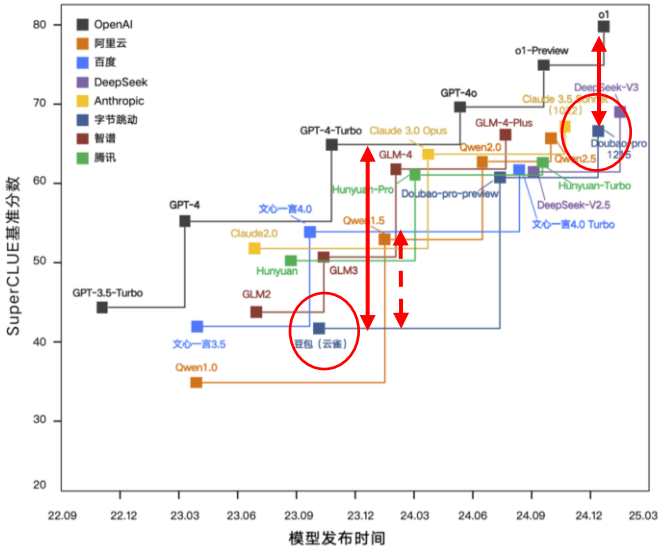
2.3 字节产业链充满机会：豆包大模型虽非顶尖，但快速追赶

- 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。

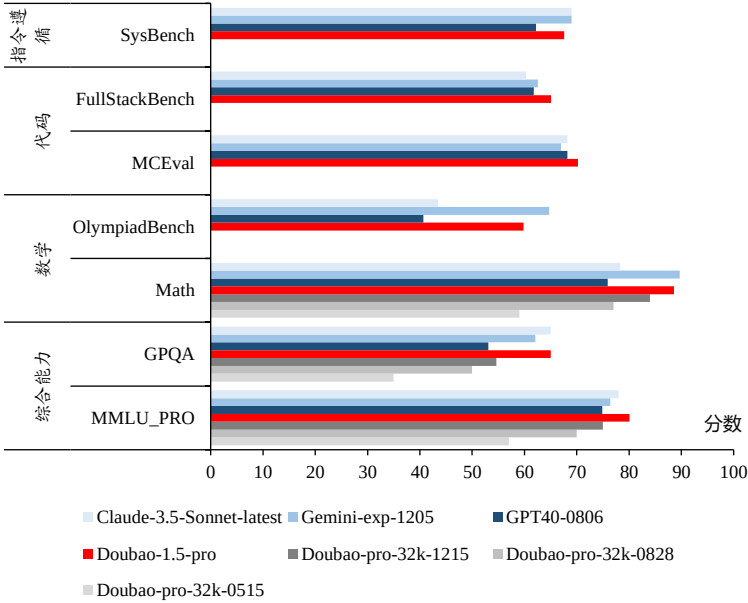
字节大模型布局全面



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，虚线表示与国内领先模型间的能力差距



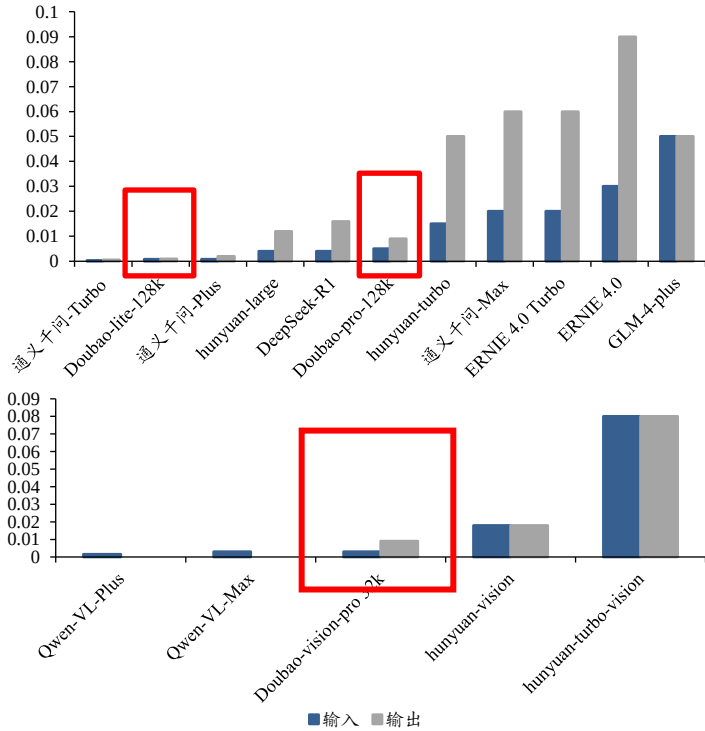
2.4 字节产业链充满机会：降低模型成本，推动应用加速

- 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）

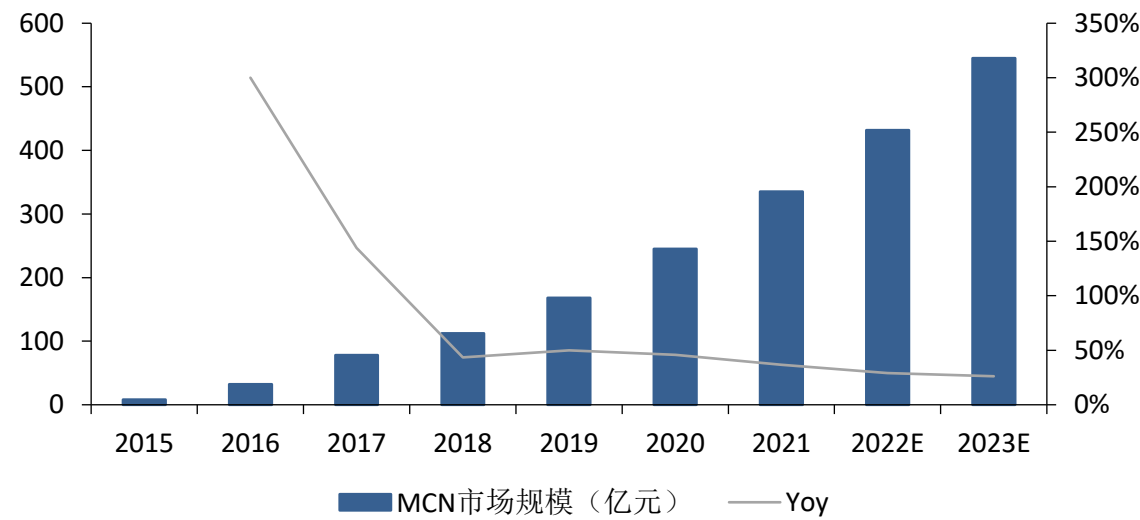


资料来源：云厂商官网，东吴证券研究所。注：DEEPMSEEK R1输入价格为缓存未命中的输入价格

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)

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/898036030033007030>