国元证券 GUOYUAN SECURITIES

传媒行业

投资评级

推荐 维持

模型能力向上价格向下，应用繁荣

——AI行业深度报告

证券研究报告 2024年12月31日

证券分析师: 李典

邮箱: lidian@gyzq.com.cn

SAC执业资格证书编码:S0020516080001

联系人: 郜子娴

邮箱:gaozixian@gyzq.com.cn

报告要点

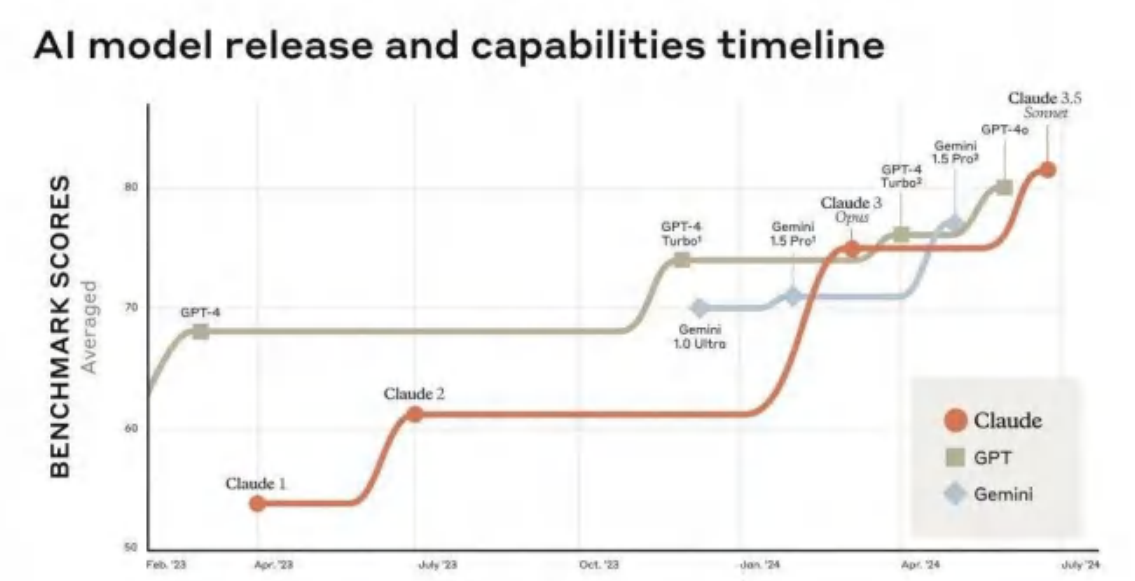
- **模型层：竞争格局收敛，o1引领大模型发展新范式。**海外头部大厂模型能力差距在2024年缩小，各巨头及其深度合作的厂商通过上游资本开支和技术人才优势已和其他玩家拉开身位差距，目前海外形成了五强格局，分别是以OpenAI、Anthropic以及谷歌为代表的第一梯队，以及x AI和Meta。国内大模型目前竞争格局相对分散，涵盖互联网和科技大厂、创业公司、传统技术类厂商这三类力量，其中互联网和科技大厂和云业务结合，综合布局。创业类厂商则依托不同资源禀赋进行差异化赛道聚焦。受到训练数据逐渐枯竭以及堆GPU卡模式所面临算力利用率降低的问题，传统Scaling Law即LLM性能与计算量、参数量和数据量三者呈现幂律关系受到挑战，OpenAI o系列提出推理侧的Scaling Law有望成为大模型发展的新驱动力，同时对于技术创新、工程能力和算力提出更高要求。
- **应用层：成本下行推动创新，应用端百花齐放。**我们重点看好2025应用端的投资机会，随着大模型竞争格局的逐步清晰，行业进入到价值实现和落地阶段。模型层能力向上调用成本向下降低应用端创新门槛，进一步促进应用端繁荣。交互方式上，AI产品逐渐从Copilot模式向Agent模式转变，C端AI Agent与AI端侧硬件相结合有望重塑流量入口；在B端则有望帮助AI加速落地行业场景。软件应用层面，企业可以通过本地部署、公有云、私有云、混合云等部署方式适配不同的规模和不同行业的企业，实现成本、私密安全性和大模型能力效果三者的平衡，企业端在大模型投入预算有望持续提升，同时企业主对于大模型投入ROI越来越重视。目前大模型在代码辅助、营销与客户管理、企业检索、办公软件等多场景落地较好，从行业上我们看好金融、政府服务、医疗等行业。C端软件应用方面，整体应用流量保持良好增长，ChatGPT周度活跃用户数突破3亿，web端流量较年初增长138%，AI ChatBots、AI内容生成与编辑、AI搜索、AI角色扮演是目前主流场景，我们看好AI搜索成为杀手级产品潜力。AI硬件方面，国内外AI+硬件的进程加快，小模型的发展推动AIPC、AI手机、AI眼镜、AI耳机等端侧硬件落地，Ray-Ban Meta成为首个爆款消费级产品，AI眼镜被视为AI端侧落地的关键硬件载体，引发广泛关注。
- **标的方面，**重点关注昆仑万维、视觉中国、恺英网络、神州泰岳、巨人网络、浙数文化、完美世界、吉比特、上海电影、中文在线、焦点科技、快手等。
- **风险提示：**技术进展不及预期的风险，大模型安全性风险，应用推广不及预期的风险。

目录

1.1 海外模型：竞争格局收敛，形成5家超级公司

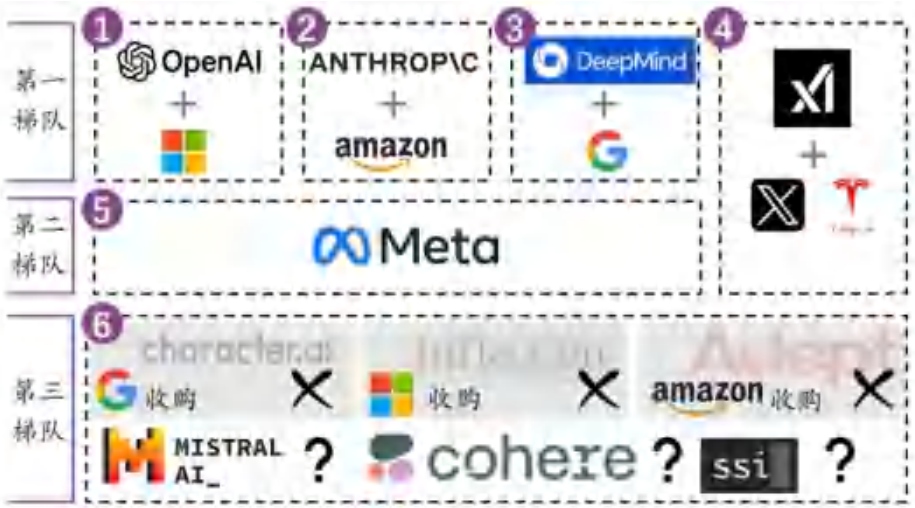
- **海外大模型玩家竞争格局收敛，OpenAI领跑，与Anthropic、谷歌厂商形成第一梯队。**2023年全球范围内OpenAI领先优势明显，但进入2024年，GPT-5进度延缓，随着Anthropic Claude3.5/谷歌Gemini1.5的推出，第一梯队模型厂商能力差距缩小，Anthropic旗舰模型如代码能力等方面超越GPT-4o。一梯队中，OpenAI具有先发、品牌认知、商业化能力等优势；Anthropic人才优势显著；而谷歌具备垂直整合能力更强，其TPU有望在2025年与英伟达GPU正面竞争；xAI算力和人才资源增长迅速，以创纪录的速度打造了拥有10万GPU的“Colossus”集群，成为数据中心扩张的标杆，是潜在的一梯队预备成员。第二梯队Meta采取差异化开源策略，Llama3.1有效缩小了开源模型和GPT-4差距。第三梯队Inflection、Adept、Character被巨头收购退出竞争，其他玩家还包括Mietral AI、Cohere、ssi等。以巨头及巨头深度合作的厂商为主的海外大模型竞争格局已经形成。

图：Claude、GPT、Gemini对比情况



资料来源：量子位，国元证券研究所

图：海外模型厂商竞争格局



资料来源：量子位《大模型落地与前沿趋势研究报告》，国元证券研究所

1.2 国内模型：格局有望进一步集中，创业公司差异化布局

- **国内模型层目前主要由互联网大厂、创业公司和传统技术大厂三类组成，未来竞争格局有望进一步集中。**2022年11月ChatGPT发布后，国内模型厂商逐渐崛起，创业公司和互联网头部公司纷纷入局，目前我国AI大模型主流厂商大致可以分为三类：1）互联网/科技大厂：以百度、阿里、腾讯、字节、华为等为代表，大模型驱动大厂云业务的增长，通常进行综合布局。2）技术类公司：以昆仑万维、科大讯飞、商汤科技为代表的专注于AI研发与应用的科技公司，主要在已有业务和渠道上做延伸；3）创业类公司：以智谱AI、MiniMax、阶跃星辰、百川智能、月之暗面和零一万物为代表的创业六小虎。创业类公司根据自生的资源禀赋形成差异化的布局，如智谱更加聚焦B端和G端，Minmax在C端影响力较强，百川智能聚焦医疗赛道。

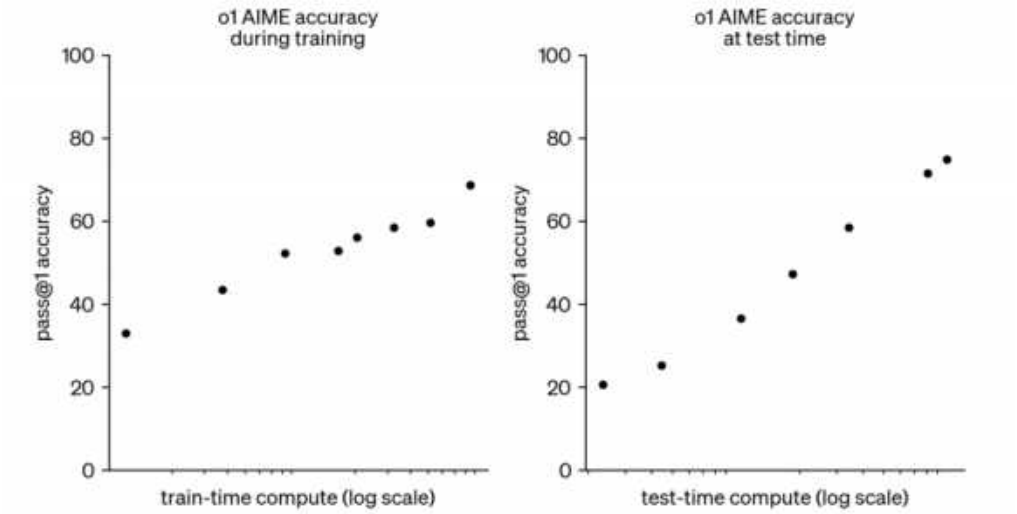
图：国内大模型竞争和发展情况

资料来源：SuperClue，量子位《大模型落地与前沿趋势研究报告》，国元证券研究所

1.3 o1引领AI模型迭代新范式，推理侧Scaling Law成为新驱动

- **OpenAI 发布推理模型o1，引领AI模型迭代新范式。**9月，OpenAI公布推理模型o1，o1推理模型重新分配计算资源，将更多资源从训练侧转移到推理阶段，显著提升模型在复杂推理上的能力；同时o1采用强化学习加思维链（CoT）的模式，将原有的人工思维链自动化，通过思维链把一个复杂问题拆解成若干简单步骤，这种链式推理过程帮助它在复杂任务中进行深度推理，有利于大模型解决复杂逻辑问题，并生成精准的答案。o1模型通过自我对弈强化学习和过程奖励模型的结合，使模型在推理能力和应用范围上实现了显著提升。12月，OpenAI正式发布o1满血版，相较o1-preview，o1数学性能提升近30%，代码能力提升27%，o1 Pro Mode的数学性能在o1基础上提升7.5%，几天后OpenAI发布了下一代大模型o3，目前仍在进行安全测试中，o3-mini版本将率先于25年1月底开放，在AIME 2024数学竞赛评测中，o3取得了96.7%的准确率，较o1提升13.4%；在博士级科学问答基准GPQA Diamond上，o3准确率为87.7%，相较于上一代o1提升9.7%。

图：o1性能随强化学习和思考时间的增长而提升



资料来源：OpenAI，国元证券研究所

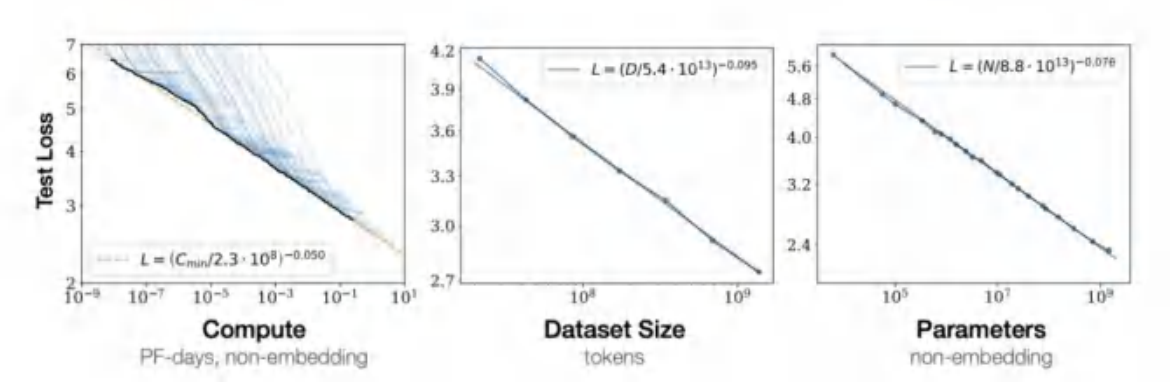
图：OpenAI o1和GPT-4o在推理密集型任务中成绩比较

资料来源：OpenAI，国元证券研究所

1.3 o1引领AI模型迭代新范式，推理侧Scaling Law成为新驱动

- **传统Scaling Law面临挑战，但仍是驱动模型能力的核心。**传统Scaling Law由OpenAI团队在2020年的论文《Scaling Laws for Neural Language Models》中最先提出，指LLM性能与计算量、参数量和数据量三者呈现幂律关系，随着大模型参数的增加，大模型性能不断提升，Scaling Law是过去几年大模型发展的核心驱动，但也面临挑战，主要原因包括：（1）训练数据逐渐枯竭，特别是高质量数据。（2）模型参数过大对于GPU训练集群的内存和通信要求较高，堆GPU卡的方式进行训练面临算力利用率降低问题。
- **推理侧的Scaling Law有望成为大模型发展的新驱动力，同时对于技术创新、工程能力和算力提出更高要求。**2024年OpenAI在o1模型的论文《On The Planning Abilities of OpenAI's o1 Models: Feasibility, Optimality, and Generalizability》提出全新的Scaling Law：当模型的推理时间越多，模型的推理能力越强。o系列模型重构了训练范式，将GPU资源在预训练、后训练和推理之间做出了更均衡的分配，开辟了模型发展的新范式。新的Scaling Law对于技术创新、工程能力和算力提出更高要求。OpenAI发布o1推理模型后，国内大模型厂商也纷纷推出了自己的推理模型。

图：LLM性能与计算量、参数量和数据量三者呈现幂律关系



资料来源：Jared Kaplan，Sam McCandlish《Scaling Laws for Neural Language Models》，国元证券研究所

图：部分国产推理模型

厂商	模型	时间
月之暗面Kimi	新一代强化模型k0-math	2024.11.16
DeepSeek	R1-Lite 推理模型	2024.11.21
昆仑万维	Skywork o1	2024.11.27
阿里云通义	QwQ-32B-Preview	2024.11.28

资料来源：腾讯网，国元证券研究所

2. 应用层：成本下行推动创新，应用端百花齐放

- **应用层百花齐放，2B和2C、软硬件各具亮点，关注可持续化商业模式的打造。**整个AIGC产业链可以分成基础设施、模型层以及应用层。模型层资源向头部玩家聚集，巨头24年加大资本开支投入算力，25年随着模型层格局的逐步清晰，行业进入到价值实现和落地阶段，25年我们认为更多的投资机遇集中于应用端。目前应用端按类型分AI原生和X+AI两种，根据量子位《中国AIGC应用全景报告》，前者占比接近57%。按产品形态分可分为软、硬件两种形态，目前国内90%+AI应用为软件形态，而AI硬件开始层出不穷，尚未迎来“iPhone时刻”。商业模式上，C端AIGC应用达到50%，但商业模式仍在探索，B端产品80%实现营收，商业模式更加清晰。按模态来分，目前44%的应用专注于文本生成，图像生成占比29%，音频占比15%，视频占比9%，3D生成规模较小占比2%。

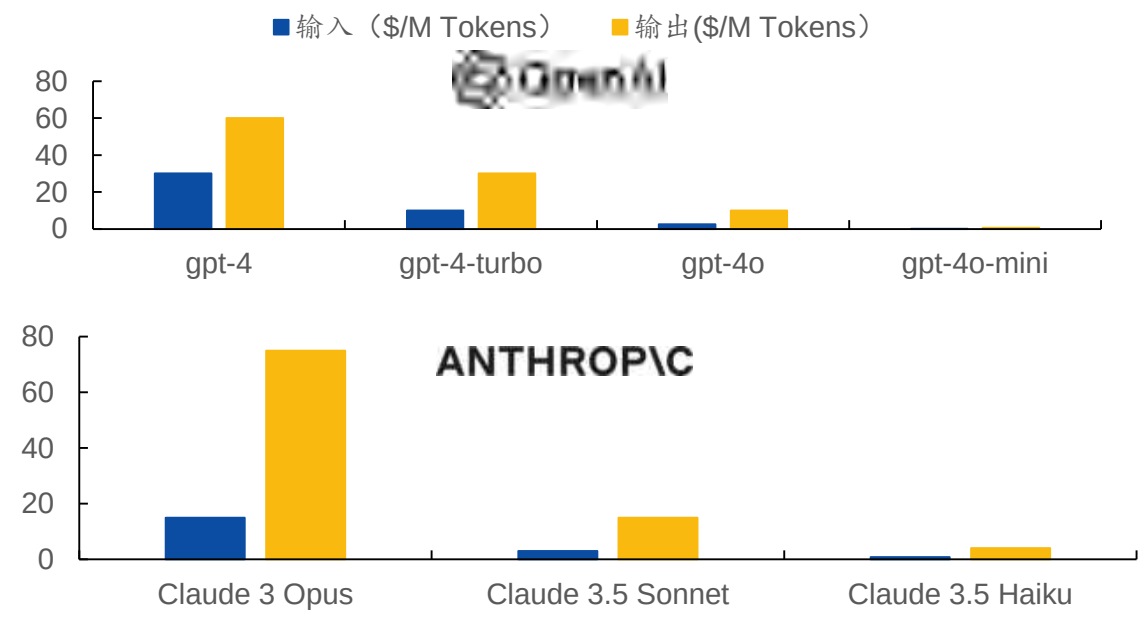
图：国内AIGC应用全景图谱

资料来源：量子位《中国AIGC应用全景报告》，国元证券研究所

2. 应用层：成本下行推动创新，应用端百花齐放

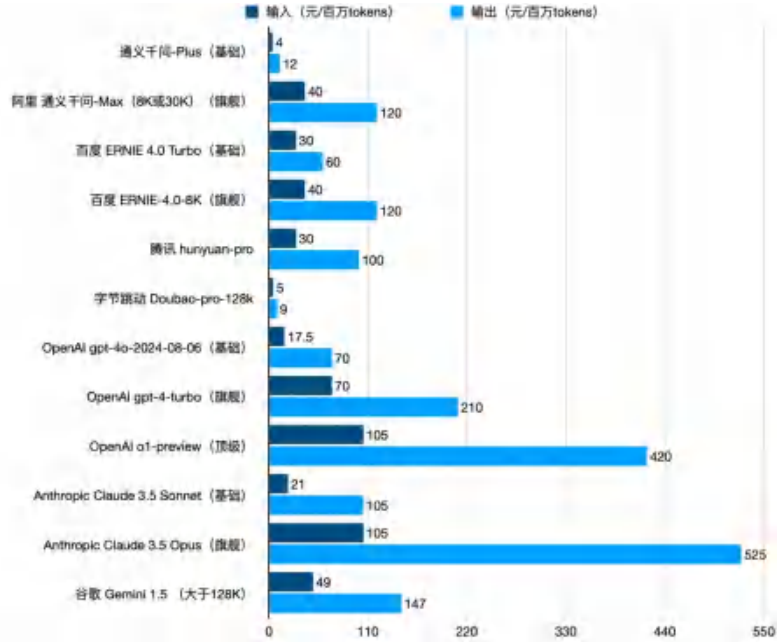
- **模型层能力向上成本向下推动应用创新。**相比于移动互联网时代应用端的投入，大模型时代应用厂商既需要承担获客成本还需要承担模型推理成本，厂商既需要关注PMF同时还需要考虑技术成本。今年以来，推理成本逐渐走低，我们认为明年这一趋势有望延续，降低应用创新门槛。海外方面，以OpenAI为例，4月，GPT4升级到了GPT-4-Turbo，输入输出价格却分别下降67%/50%，5月13日发布GPT-4o，模型性能升级，但输入输出价格较GPT-4-Turbo下降75%/67%。7月18日推出的小模型 GPT-4o mini输入输出价格分别为降低至0.15和0.6 美金1M tokens。国内模型方面，今年国内AI模型厂商也陆续降价，以阿里为例，5月21日，阿里云宣布其9款商业化及开源系列模型降价。其中，通义千问主力模型Qwen-Long的API输入价格降至0.0005元/千tokens，降幅达97%。9月19日阿里云在云栖大会上宣布通义千问三款主力模型再降价。现阶段国内模型价格普遍只有OpenAI的20%-50%。成本下降，有望带动开发者调用和探索，支撑应用侧的繁荣。

图：OpenAI及Anthropic的模型调用成本下移



资料来源：OpenAI、Anthropic官网，国元证券研究所

图：国内以及国际厂商主力模型推理价格对比

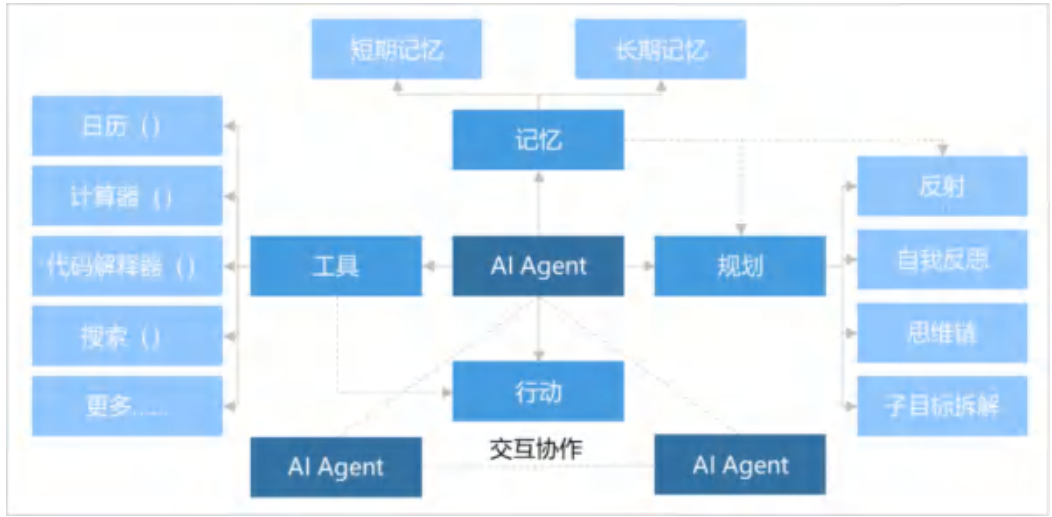


资料来源：各公司官网，《财经》，国元证券研究所
统计时间：2024年9月16日

2.1 AI Agent：交互方式升级，Agent带动应用繁荣

- **AI Agent带动交互方式升级，进一步促进AI应用端繁荣。**人类和AI交互按照自动化程度可以分为Embedding、Copilot和Agent模式，2024年AI Agent逐渐爆发，AI产品逐渐从Copilot模型向Agent模式转变，在大语言模型（LLM）驱动的Agent系统中，LLM充当Agent的大脑，并由Planning（规划）、Memory（记忆）、Tools（工具）、Action（动作）等几个关键组件补充。
- **国内外厂商加大力度开发布局AI Agent，C端流量入口有望重塑，B端加速AI行业场景落地。**国内外厂商加速布局AI Agent，根据彭博消息，OpenAI正在准备推出一款代号“Operator”的全新AI Agent产品，可以自动执行各种复杂操作，包括编写代码、预订旅行、自动电商购物等；Salesforce在9月12日推出自动化AI Agent产品——Agentforce，并在10月宣布Agentforce进入全面商用阶段，企业可构建定制AI Agent，连接企业数据并代表员工执行销售、服务、营销、商务等相关任务。苹果于10月发布Apple Intelligence。国内厂商方面，荣耀率先发布首个跨应用开放生态AI智能体，智谱于11月发布AutoGLM、AutoGLM-Web、AutoGLM-PC三款Agent产品。AI Agent改变AI交互方式，与AI端侧硬件相结合重塑流量入口。在B端方面，则有望帮助AI加速落地行业场景。

图：AI Agent的组成部分



资料来源：Lilian Wen《LLM Powered Autonomous Agents》，来觅数据，国元证券研究所

表：AI Agent国内外大厂落地情况

厂商	布局
OpenAI	根据彭博消息，OpenAI正在准备推出代号“Operator"全新的AI Agent产品，可以自动执行各种复杂操作，包括编写代码、预订旅行、自动电商购物等，预计将于2025年1月发布。
微软	10月微软宣布在Dynamics 365中集成10个自主AI Agent，并支持OpenAI最新模型o1；11月微软Ignite大会上，微软宣布已建立全球规模最大的企业级AI Agent生态系统。企业用户现在可以通过Azure AI目录访问超过1800个AI模型，用于支持各类AI Agent的部署和运行。
谷歌	12月11日，谷歌发布了Gemini 2.0并推出了三个基于 Gemini 2.0 架构的 AI 智能体原型，分别是 Project Astra、Project Mariner 和Jules。
Anthropic	10月Anthropic升级Claude3.5 Sonnet，公测版中引入了一项突破性的新功能：计算机使用。通过API接入，开发人员可以指导Claude使用计算机，就像人们看屏幕、移动光标、单击按钮和输入文本一样。
苹果	10月28日推出APP Intelligence
联想	10月推出AI Now PC个人助手智能体
荣耀	9月6日，荣耀在IFA 2024 上发布了行业首个跨应用开放生态智能体——荣耀 AI智能体（AI Agent），并宣布荣耀Magic7 系列将首发搭载荣耀AI智能体。
华为	推出Harmoney Intelligence
智谱	11月发布AutoGLM、AutoGLM-Web、AutoGLM-PC三款Agent产品，其中AutoGLM可模拟用户访问网页、点击网页的浏览器助手，与微信、淘宝、美团、小红书等8款知名应用软件合作，覆盖日常生活常用的线上聊天、网购、社交、地图、酒店火车订票等功能，同时支持用户通过语音进行交互。

资料来源：各公司官网，新浪新闻，36Kr，国元证券研究所

2.2 软件：B端看好金融政务医疗等领域，C端期待杀手级产品

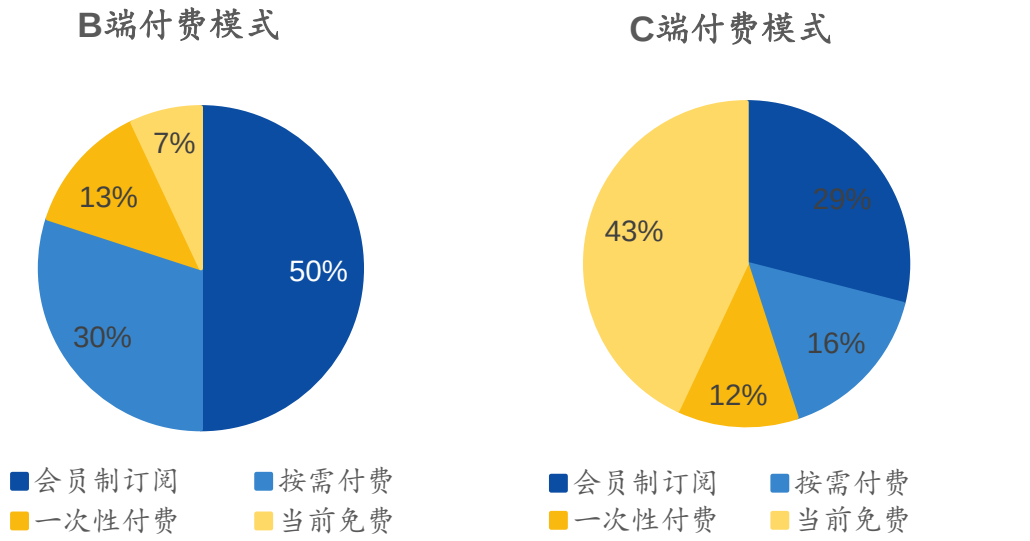
- **中美对比来看，美国市场AI应用当前更偏向于B端企业级服务，而在国内AI应用则更加侧重于从C端。**
- **商业化方面，B端变现模式更加清晰且更快盈利， C端产品商业化潜力有待进一步开发，期待杀手级产品的出现。**根据量子位智库，B端产品从通用场景到垂直赛道分布较均，商业模式较为清晰且80%以上的产品均实现营收，B端用户需求明确，与C端需求相比，指标更易量化。C端产品商业化潜力有待进一步开发，目前近50%的产品当前仍未有明确的收入模式，向公众免费开放。这是由于C端用户需求并不明确，往往是供给激发需求对产品本身的体验要求较高，强调“易用性”。未来商业化潜力有待进一步被开发。

图：中美AI应用层对比

	海外/美国市场	中国市场
应用层	● 海外AI应用，2B多于2C，这主要是源自于海外的企业软件基础设施与市场环境更完善，2B AI原生应用即便聚焦细分市场，仍然具备稳定的成长空间。 ● 原生AI应用跑得更快，无论是PMF还是商业化，跑得快的AI原生应用以自有模型，或者多模型调用为主	● 中国AI应用，2C AI原生应用更为广泛。 ● 现存公司/应用利用LLM实现能力增强，并依托其在用户/客户界面的占有和销售能力，仍然具备较强的领先优势

资料来源：易观《中国AI开发者应用生态调研报告》，国元证券研究所

图：国内2B/2C商业模式对比



资料来源：量子位《中国AIGC应用全景报告》，国元证券研究所

2.2.1 B端应用:企业投入预算加大，看好金融政务医疗等方向

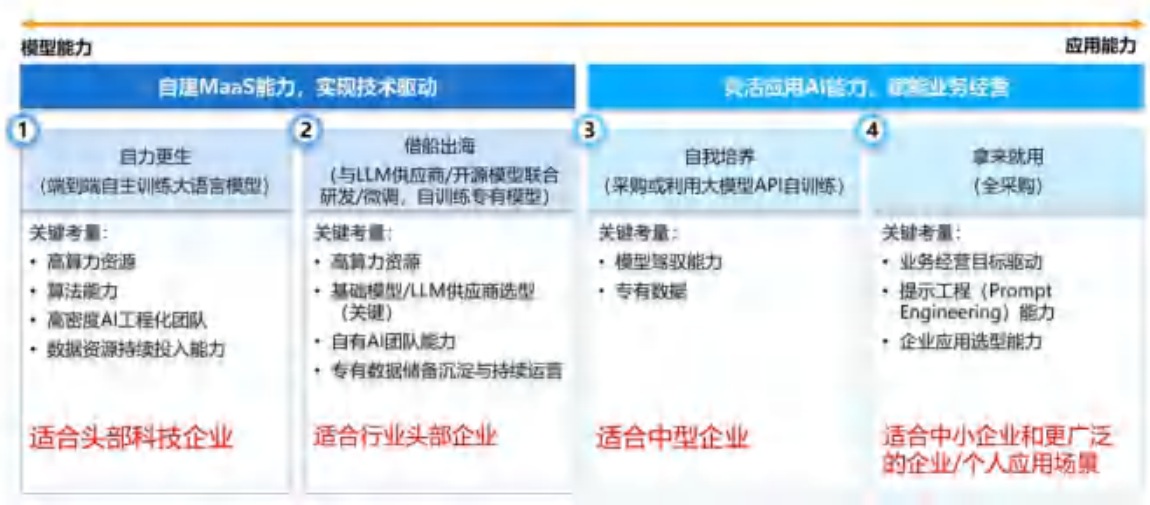
- **2B方面，当前大模型在企业可以通过本地部署、公有云、私有云、混合云等部署方式适配不同的规模和不同行业的企业，实现成本、私密安全性和大模型能力效果三者的平衡**，其中私有化部署适合对于数据私密性和控制性较高的企业，对于调用开源大模型或者大模型API服务的企业，可以通过微调、RAG（检索增强生成）、Prompt Engineering等方式实现场景落地。付费模式上包括了产品授权费用（按年/买断）、二次开发分成、订阅模式、按次数收费、广告收入等模式。
- **企业端大模型投入预算有望持续提升，同时企业主对于大模型投入ROI越来越重视。**根据a16z对于500强和行业头部企业今年3月的调研数据，被采访公司2023年在基础模型 API、模型托管和微调模型方面的平均支出为 700 万美元，几乎每家企业都看到了 Gen-AI 将对企业工作流程产生巨大效益，并计划在 2024 年将支出增加 2 倍至 5 倍，以支持将 AI 嵌入到更多业务生产环节中。另一方面，企业主也更加重视大模型投入的ROI，目前AI投入短期主要为企业带来效率提升和成本节约，中期（2~3年）看有助于带动收入增长、客户满意度提升等。

表 ：AI大模型行业应用商业模式分类

部署方式		大模型使用方式	收费模式	适用企业规模	典型行业/场景
本地部署		本地调用	产品授权费用(按年/买断)+人员服务费(人*天)	中大型企业	党政、工业
云部署	SaaS模式	APP/网页	订阅模式、广告收入、按次数收费	小微企业	知识搜索、内容生成
	PaaS模式	远程平台	订阅模式、二次开发分成	小微企业、初创企业	电商
	Maas模式	调用API	按流量计费、二次开发分成	中小企业	医疗、教育、文旅
混合部署		本地+云	产品授权费用(按年/买断)+人员服务费(人*天)+流量费用	中大型企业	金融、工业
AaaS模式		融合智能终端/APP	买断模式、订阅模式、广告收入	不限	不限

资料来源：前瞻研究院， 国元证券研究所

图：企业布局大模型的路径选择

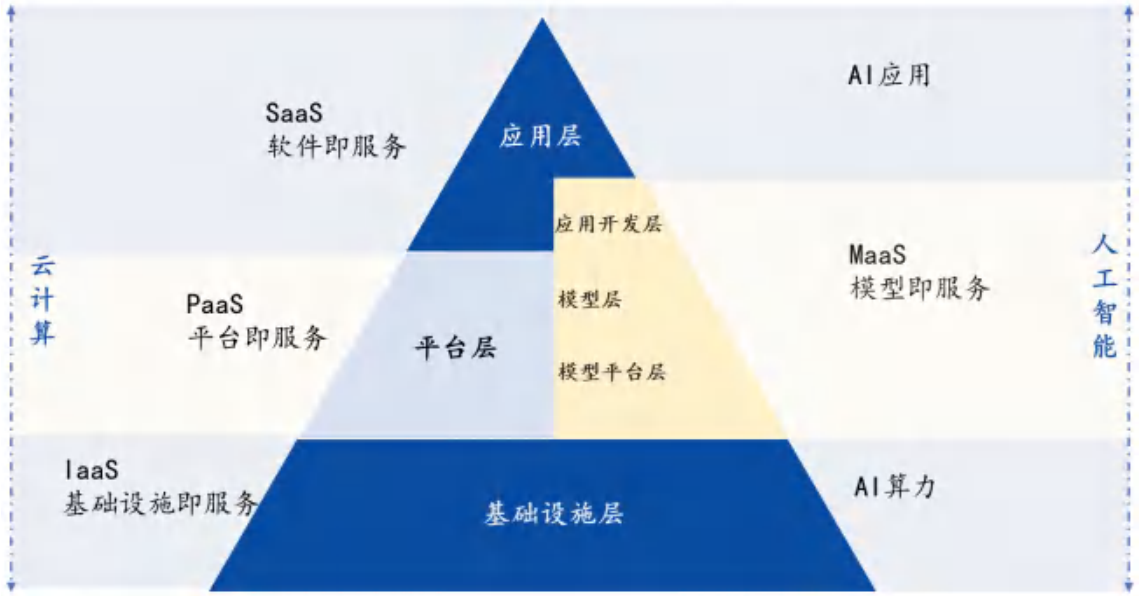


资料来源：易观，国元证券研究所

2.2.1 B端应用:企业投入预算加大，看好金融政务医疗等方向

- **MaaS（Model as a Service）成为主流商业模式，**MaaS指将AI模型及其相关能力打包成可重复使用的服务，使企业能够快速高效地构建、部署、监控、调用模型，无须开发和维护底层基础能力。MaaS主要提供三部分能力：一、提供包括模型训练、调优和部署在内的全栈平台型服务，二、大小模型及公私域数据集的丰富资产库服务；三、基于AI模型的应用开发工具服务。
- **AaaS（Agent as a Service）作为新兴服务模式，帮助大模型落地于多元应用场景中。**AaaS是将AI Agent作为一种服务来提供。AI Agent可以根据用户的需求和环境的变化，自主做出决策并执行相应的操作。通过AaaS平台，用户可以快速获取所需的AI功能和服务，而无需担心Agent的开发、部署和维护等复杂问题，灵活应对不同应用场景、提高资源利用效率。同时，AaaS平台还支持对Agent进行动态调整和优化，以适应不断变化的需求和环境。AaaS模式使得大模型能够广泛应用于各种行业场景中，推动其最终落地。

图：MaaS定位与比较示意图



资料来源：中国信息通信研究院，国元证券研究所

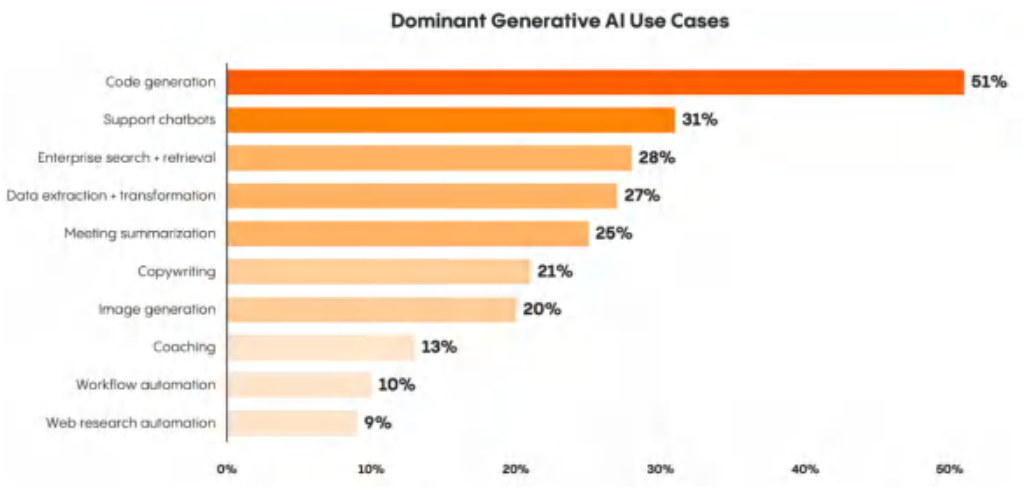
图：ModelScope模型层实践图

资料来源：阿里云，中国信息通信研究院，国元证券研究所

2.2.1 B端应用:企业投入预算加大，看好金融政务医疗等方向

- **海外市场方面，目前大模型在代码辅助、营销与客户管理、企业检索、办公软件等多场景落地较好，垂类行业中在医疗、法律、金融服务、媒体等垂直行业落地更快。**在企业工作场景和流程中，代码辅助、辅助聊天机器人、企业检索、数据提取转换、会议总结等功能使用占比最高。公司中，如Salesforce、Adobe、SAP、ServiceNow、Palantir、Workday等软件企业深度拥抱AI。垂类行业中，医疗领域如 Abridge 和 SmarterDx 这样的辅助工具正在自动化临床工作流程，包括分诊到收入周期管理的各个环节。法律方面，Everlaw 和 Harvey 等工具正在实现法律研究、合同审查和电子证据发现流程的自动化。金融服务方面， Numeric 和 Arkifi 这样的初创公司通过AI驱动的解决方案革新了会计和金融研究领域。

图：生成式AI在企业中应用场景的占比情况



资料来源：MENLO《2024: The State of Generative AI in the Enterprise》，国元证券研究所

图：企业级AI产品一栏

资料来源：SEQUOIA，国元证券研究所

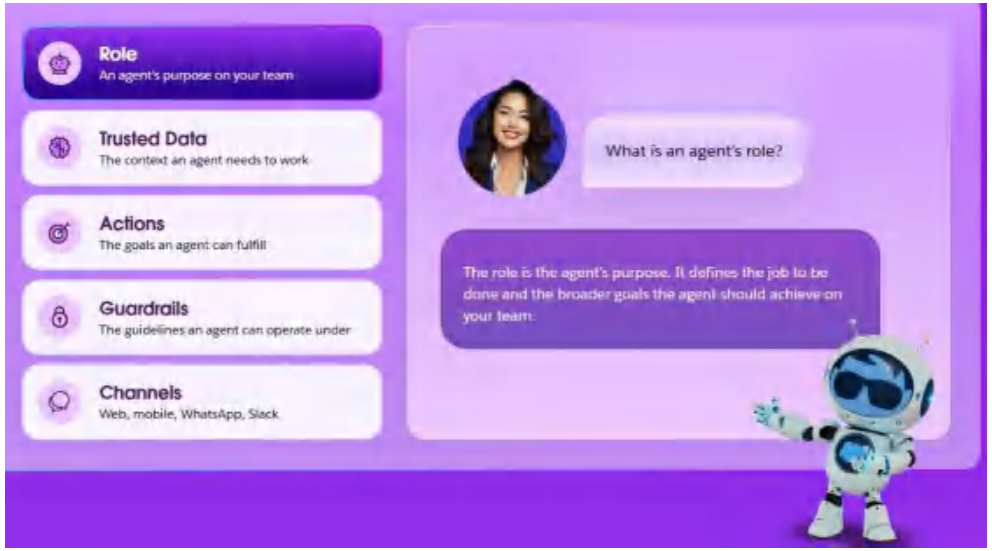
海外典型企业——Salesforce

- **Salesforce：Data+AI+CRM驱动，Agentforce平台革新SaaS平台新模式。**Salesforce是一家CRM软件服务提供商，以创新和前瞻性闻名。根据最新公布的2025财年第三季度的报告显示，Salesforce实现营收94.4亿美元，同比增长8%。更引人注目的是，公司的GAAP营业利润率达到了20.0%，创下历史新高。公司坚持Data+AI+CRM驱动，助推业绩稳定增长。
- **9月，Salesforce推出的自动化AI Agent产品Agentforce，**作为一款功能全面的客户服务解决方案，提供全渠道支持，允许代理从统一平台管理电子邮件、聊天、电话和社媒等多种交互方式，帮助客户解决服务、销售、营销等方面的任务。Agentforce拥有强大的知识库，为代理提供实时信息、文章和资源，可以快速准确地解答客户问题。在案例管理方面，Agentforce通过自动分配案例和提供清晰的工作流程，简化案例处理过程，提高解决效率，AI支持的自动化工具能够自动执行日常任务，减轻代理的工作负担，提升工作效率。Agentforce还提供高级分析和报告功能，帮助管理人员跟踪绩效指标，做出优化决策。

图：Salesforce的产品

资料来源：公司官网，国元证券研究所

图：Agentforce平台

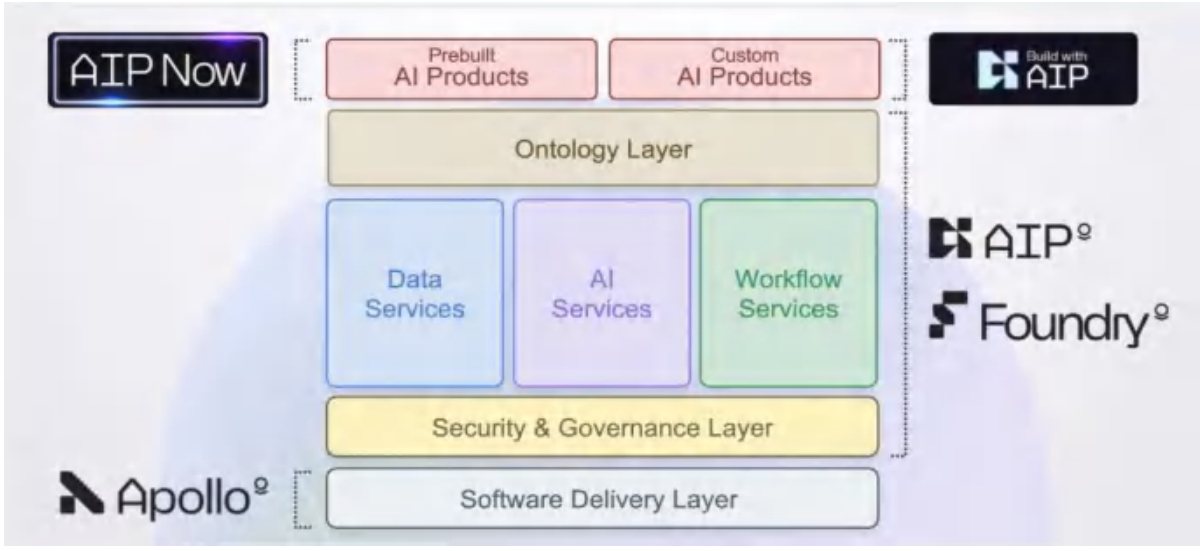


资料来源：公司官网，国元证券研究所

海外典型企业——Palantir

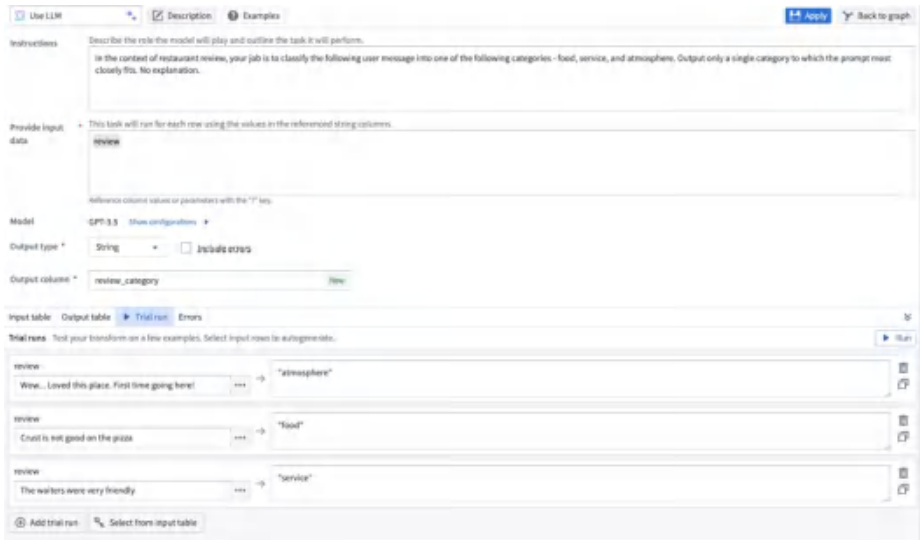
- **大数据分析公司Palantir建立AIP平台，实现AI赋能平台。**Palntir是一家大数据分析公司，旗下的产品主要分为Gotham、Foundry和Apollo三种，其中Gotham专为情报和国防部门设计，通过对复杂数据分析支持反恐和军事行动，主要客户为美国情报界和国防部，Foundry专为商业客户定制，可以实现数据的整合和分析，Apollo则支持在各种环境中的集成和交付，支持软件的无缝部署和更新。2024Q3，该公司美国市场收入同比增长44%，环比增长14%，达到4.99亿美元，美国商业收入同比增长54%，环比增长13%，达到1.79亿美元，高于上一季度的1.59亿美元。
- **Palantir2023年推出人工智能平台AIP，**该平台可以将AI技术与平台的日常经营相结合，核心功能可以分为3个：1）支持LLM和其他模型嵌入工作流，用户可以用自然语言向 AIP Assist 提问，并获得实时帮助，目前主流的OpenAI GPT-4 Turbo、Anthropic Claude2和Meta Llama3等大语言模型已经成功接入平台；2）实现自定义工作流程的AIP功能，允许开发人员构建私有的由LLM支持的工作流程或应用程序； 2）对AIP的操作可以实时监控，保证决策的合规性。AIP在支持私有化部署、实时响应的同时，重视对于数据隐私的保护和操作合规性，与军事和国防场景适配。

图：Palantir产品体系



资料来源：公司官网，国元证券研究所

图：Palantir AIP平台

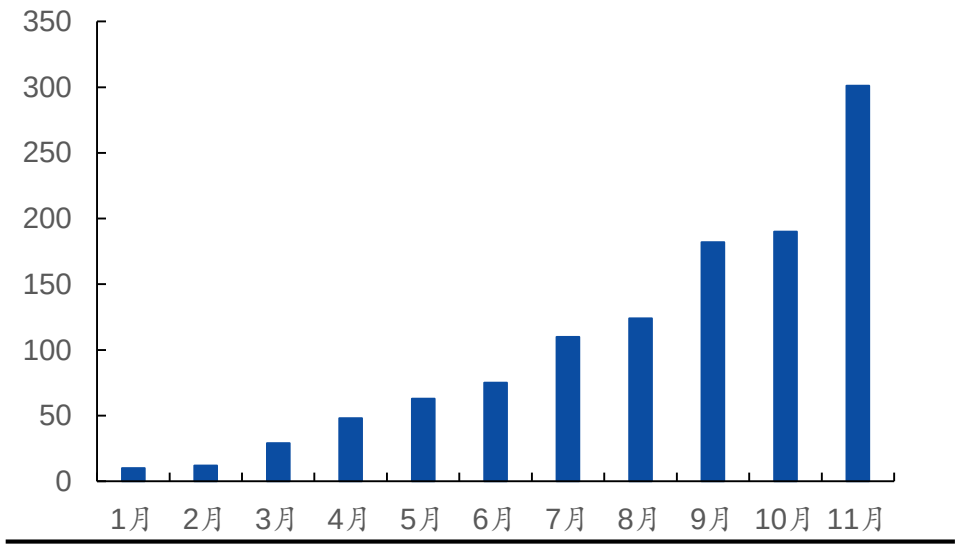


资料来源：公司官网，国元证券研究所

2.2.1 B端应用:企业投入预算加大，看好金融政务医疗等方向

- **国内大模型在企业端商业化进场提速，全年大模型项目中标数量超1000个。**根据智能超参数公众号统计，11月份公开渠道统计到的大模型相关中标项目301个，其中：有66个项目，未披露中标金额（为方便统计，金额以0计算），其余235个中标项目披露的金额达到约10.62亿元，创下2024年以来单月新高。1-11月可统计范围内的大模型相关中标项目达到1144个。
- **行业分布**今年前三季度来看，按照项目数量计算，教科、通信、能源、金融、政务是排名前五的行业；按照金额计算，技术服务、政务、能源、通信、教科是排名前五的行业。年初，能源、通信行业的采购较多，最近几个月，教科、金融、政务等行业的项目较多。
- **中标企业方面，科大讯飞、百度、智谱AI、腾讯云、阿里云、火山引擎表现领先。**

图：2024年1-11月中国大模型中标项目情况



资料来源：智能超参数，国元证券研究所

表 ：2024Q1-Q3大模型中标行业分布情况

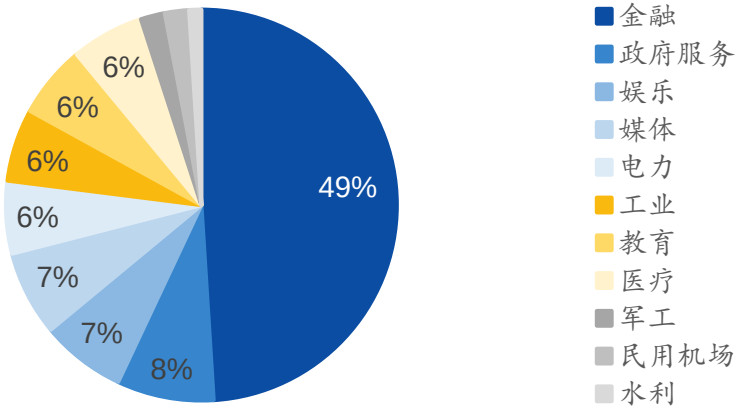
类别	项目数量	数量占比	披露项目金额（万元）	披露项目金额占比
教科	157	24.00%	19,810	9.50%
通信	150	23.00%	20,047	9.70%
能源	83	12.70%	31,010	14.90%
金融	66	10.10%	10,127	4.90%
政务	62	9.50%	48,434	23.30%
技术服务	53	8.10%	54,313	26.20%
大交通	33	5.10%	8,516	4.10%
其他	49	7.50%	15,282	7.40%
合计	653	100.00%	207,538	100.00%

资料来源：智能超参数，国元证券研究所

2.2.1 B端应用:企业投入预算加大，看好金融政务医疗等方向

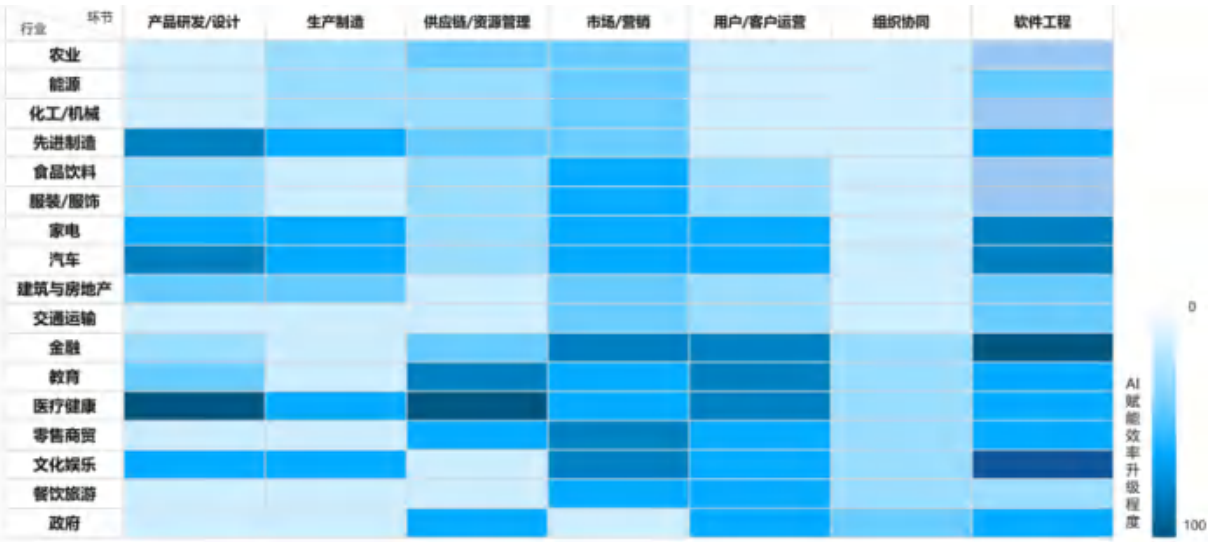
- **从行业分布上来看，我们看好金融、政府服务、电信、文娱、医疗等行业AI的渗透。**
- **场景上，AI在营销/客服、研发、OA、人力资源管理环节和场景有望快速渗透，**以金融为例，AI主要应用于营销、客服、风控等场景；电信行业则应用于客服服务、营销推广、网络运维、故障预测等场景。政务方面，主要应用于市民咨询、舆情分析等场景，解决基层工作强度大、响应速度等问题。

图：MaaS在各行业应用占比



资料来源：中国信息通信研究院，国元证券研究所

图：大模型和企业经营环节结合情况（上）
图：各个行业通用场景AI化的成熟度（下）

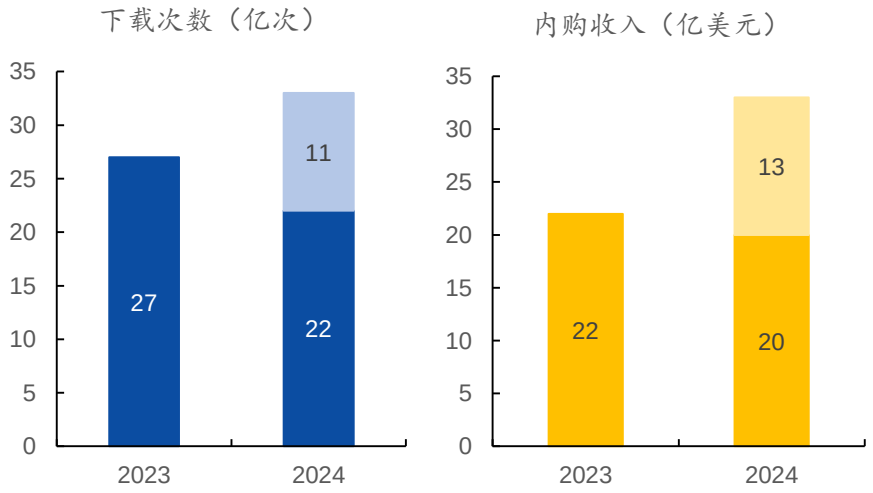


资料来源：易观，国元证券研究所

2.2.2 C端应用：商业化持续探索，期待杀手级产品落地

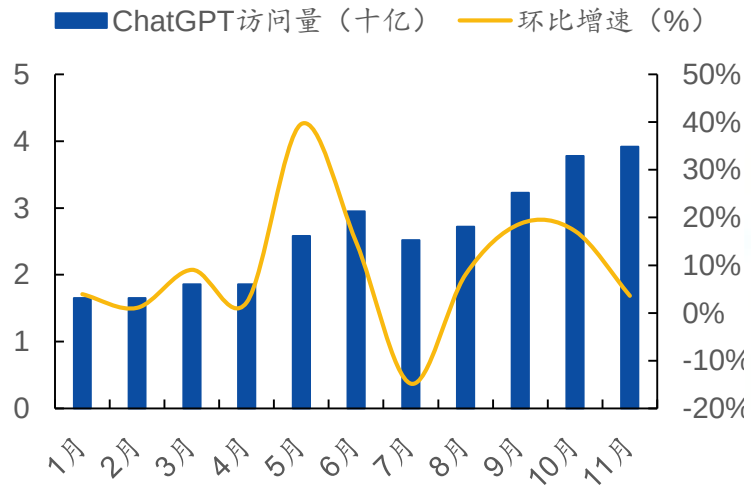
- **AI C端应用流量保持良好增长，ChatGPT周度活跃用户数突破3亿，web端流量较年初增长138%。**2023年全球AI应用爆发，2024年伴随AI在图片生成、视频生成、交互等能力持续突破，流量保持良好增长，根据Sensor Tower数据，2024年预计全球AI下载量同比增长22%，全球AI应用内购收入同比增长50%。头部产品ChatGPT流量保持稳步提升，虽然在7月份流量一度环比下滑，随着o1-preview和o1-mini的发布，但很快重返增长态势，月度访问量从年初的17亿次，提升到了11月接近40亿的月度水平，流量较年初提升138%，12月ChatGPT周度活跃用户数突破3亿。
- **分类别看AI应用分为：底层设计逻辑的的AI原生产品、在原有互联网产品上深度嵌入AI功能的AI+X产品、基于外接API微创新的套壳类产品和集合站类产品四大类。**

图：2024年1-8月全球AI应用下载量及内购收入



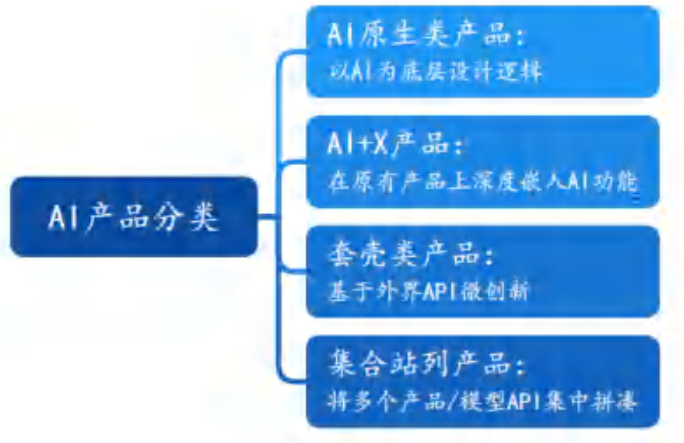
资料来源：SensorTower，国元证券研究所，注：浅色部分为SensorTower预估的四季度数据

图：Chatgpt Web端访问量



资料来源：AI产品榜aicpb.com，国元证券研究所

图：AI产品分类

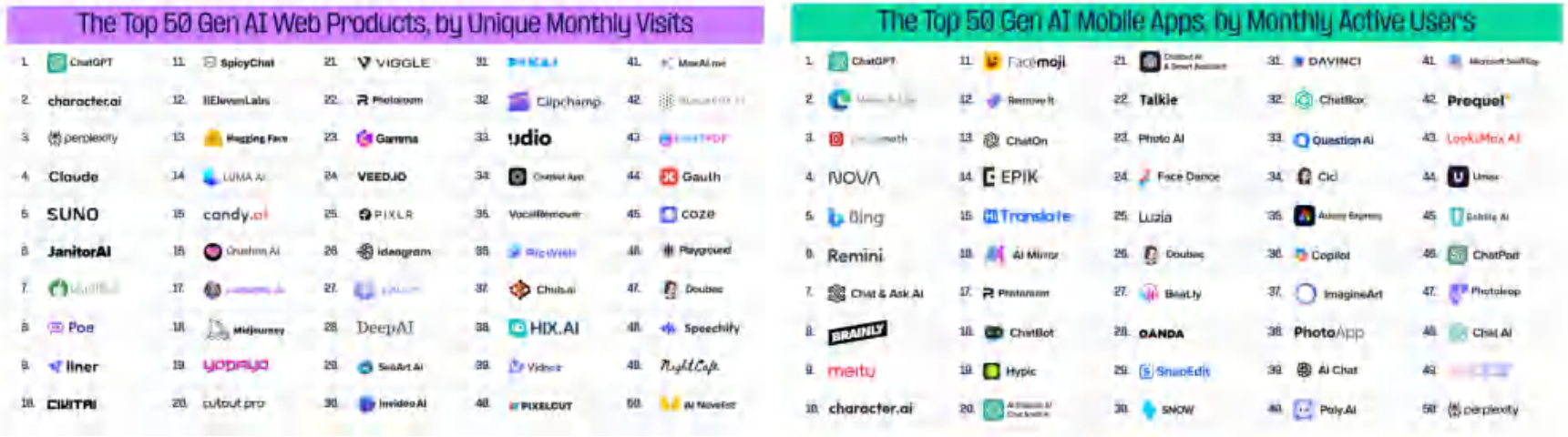


资料来源：量子位，国元证券研究所

2.2.2 C端应用：商业化持续探索，期待杀手级产品落地

- **全球应用产品来看，ChatGPT保持断层领先，TOP20中Suno、Claude、Perplexity增长较快。**根据AI产品榜1月及11月web端数据对比，ChatGPT以接近40亿的月度访问量保持断层第一，较年初增长138%，AI应用产品Suno、 Claude、AI搜索引擎Perplexity增速分别达到648%/342%/127%，增速领先。
- **AI ChatBots、AI内容生成与编辑、AI搜索、AI角色扮演是目前主流场景，看好AI搜索成为杀手级产品潜力。**AI搜索颠覆传统搜索引擎，ChatGPT外，Perplexity、Claude、Grok、Meta AI、Poe等AI产品持续瓜分搜索市场,专业领域如Causaly（科学）、Consensus（学术研究）、Harvey（法律）和Hebbia（金融服务）也在持续渗透。内容生产及编辑类工具产品呈现出多模态的趋势，过去以图片为主，今年来我们看到视频类、音频类的增长快速，典型的代表如音乐类Suno、Udio，视频类Luma、Viggle和Vidnoz。在AI角色扮演赛道，Character AI 11月web端访问量达到2.14亿，APP端MAU达到2688万，出海产品Talkie APP MAU也达到了2519万，Janitor AI、Crushon、Candy.AI等表现不俗。

图：全球AI产品榜



资料来源：a16z，国元证券研究所

表：11月全球AI产品web端访问量（百万次）以及相对1月增速（%）

排名	产品	11月访问量	VS 1月增速
1	ChatGPT	3920	137.58%
2	NewBing	1830	27.08%
3	CanvaTexttoImage	827.62	52.52%
4	Gemini	283.61	-16.60%
5	360AI搜索	282.73	-
6	Deepl	215.92	-19.14%
7	CharacterAI	213.6	2.99%
8	NotionAI	155.04	-15.13%
9	Shop	135.79	49.32%
10	Perplexity	107.84	126.75%
11	Q-Chat	103.96	-9.94%
12	SalesforceAI	99.97	8.26%
13	Claude	89.32	342.40%
14	Jambot	86.36	0.90%
15	copilot|微软	81.33	-
16	Quillbot	79.56	44.73%
17	JanitorAI	78.17	119.64%
18	Remove.bg	72.08	15.49%
19	Grammarly	62.83	-0.59%
20	Suno.com	55.81	648.12%

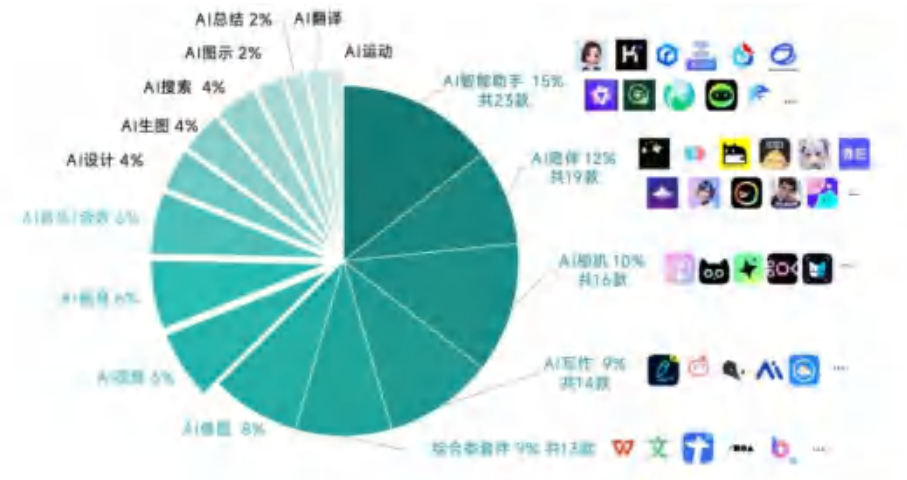
资料来源：AI产品榜aicpb.com，国元证券研究所

2.2.2 C端应用：商业化持续探索，期待杀手级产品落地

国元证券 GUOYUAN SECURITIES

- **国内市场方面，字节采取饱和式打法，豆包取得领先。**根据量子位智库，国内Web产品数量最多的赛道则依次为AI智能助手、AI写作、AI视频和综合类套件，APP端的产品数量最多的赛道依次为AI智能助手、AI陪伴、AI相机、AI写作和综合类套件。国内AI ChatBots在各细分品类中有绝对优势，目前形成了“1+1+6”的格局，其中豆包在规模、增长、活跃、留存等各项数据上均断层式领先，截止2024年10月，豆包累计下载量超过1.4亿。Kimi则表现次之，文小言、智谱清言、讯飞星火、天工AI、阿里通义、腾讯元宝和海螺AI六个产品位于第三梯队。AI陪伴赛道国内众多厂商布局，猫箱、星野、X Eva表现靠前，AI视频领域2024年受益于大模型多模态能力的提升，Vidu、PixVerse、可灵AI、智谱清言、即梦等代表性产品出现，有待引爆。厂商方面，字节采取饱和式打法，根据新皮层统计目前正常运营AI应用约有20款，在国内和出海方面都取得了不错成绩，字节在8月a16z的全球AI榜单中共有5个产品上榜，表现优异，分别是教育应用 Gauth、开发平台 Coze、通用助手豆包、豆包英文版 Cici、照片和视频编辑工具 Hypic。

图：国内AI产品APP端赛道分布



资料来源：量子位智库，国元证券研究所，注：产品统计截止至2024.10

图：11月国内AI APP端及web端TOP10产品

APP端产品	MAU	web端产品	访问量
豆包	59.98M	360AI搜索	282.73M
文小言	12.99M	百度文库	46.8M
Kimi	12.82M	Kimi|月之暗面	32.82M
智谱清言	6.37M	文心一言|百度	22.07M
讯飞星火	5.94M	豆包|抖音	21.43M
天工AI	5.78M	通义千问|阿里	10.65M
星野	5.25M	秘塔AI搜索	8.01M
猫箱	4.58M	AiPPT. cn	7.4M
通义	3.88M	百度搜索 AI助手	6.15M
光速写作	3.77M	天工AI|昆仑万维	5.19M

资料来源：AI产品榜aicpb.com，国元证券研究所

表：字节跳动AI布局

层级	研发团队	类型	产品
模型层	Seed	语言模型	Doubao-pro
			Doubao-lite
		语音模型	Seed-ASR（语音识别）
			Seed-TTS（语音生成）
		图片模型	SDXL-Lightning（文生图）
			Seed-Edit（图像编辑）
		音乐层	Seed-Music（音乐制作）
		视频模型	Boximator（视频编辑）
			MagicVideo-V2(文生视频)
			AnimateDiff-Lightning（文生视频）
			PixelDance（文生视频、图生视频）
			Seaweed（文生视频、图生视频）
		多模态模型	BuboGPT
		3D模型	MVDream
	ByteDance Research	具身智能模型	GR-1
			GR-2
应用层	Flow	智能体开发平台	扣子
			Coze(海外)
		智能助手	豆包
			CiCi（海外）
		工具集	小悟空
			ChitChop（已停止运营）
		社交	猫箱
			AnyDoor（海外）
		图像	星绘
			PicPic(海外)
	剪映	图片/视频生成	即梦AI
			Dreamina（海外）
		视频剪辑	剪映
			CapCut(海外)
		教育	豆包爱学
			Gauth（海外）
	大力教育	数字人	抖音AI分身(KOL内测)
			TikTok AI网红
	抖音/TikTok	抖音电商内容生成	即创
	字节跳动开发者服务团队	编程助手	豆包MarsCode
			MarsCode（海外）
		模型分享社区	炉米 Lumi
	其他	教育	识典古籍
		音乐生成	海绵音乐
智能硬件	Oladance+ Flow	智能体耳机	Ola Friend
	大力教育	智能台灯	大力智能学习灯
	FoloToy+火山引擎	智能玩偶	显眼包

资料来源：新皮层公众号，国元证券研究所

2.3 硬件：AI硬件加速落地，AI眼镜成为关键载体

- **2024年，国内外AI+硬件的进程加快，小模型的发展推动AIPC、AI手机、AI眼镜、AI耳机等端侧硬件落地。**2024年4月，Meta、微软、苹果等集中发布Llama-3、Phi-3、OpenELM的等模型，此外MobileLLM、Gemma-7B、Qwen-7B、MiniCPM、TinyLlama等一系列端侧模型及小模型相继问世。国内外各厂商对端侧布局的硬件布局也在2024年先后落地。在手机和电脑端，年初国内OPPO和vivo即推出AI手机，将部分AI功能集成到手机终端，三星也在最新的S24系列中加入GalaxyAI的功能，12月苹果推出Apple Intelligence。在AIPC方面，联想、戴尔为代表的头部电脑厂商开始在PC终端集成AI相关功能，配合全球芯片的发展，端侧计算和推理算力提升，联想PC端智能体AI Now的推出有望加速AIPC布局进程。Gartner预测，2024年AI PC的出货量将达到4300万台，较2023年增长99.8%，2025年将达到1.14亿台，占PC总出货量的43%。
- **AI眼镜被视为AI端侧落地的关键硬件载体，引发广泛关注。**去年9月，Meta发布Ray-Ban Meta，据The Verge统计，截至24年5月已经实现全球销量100万副，成为首个消费级爆款AI硬件产品。AI眼镜作为轻量级的可穿戴产品，通常嵌入了耳机、摄像头、WiFi蓝牙模块等相关硬件以及生成式人工智能模型应用，具备信息获取和交互的便捷性，相比于手机、电脑等传统电子终端，AI智能眼镜能够解放用户双手，提供更具沉浸式的交互体验，是未来发展下AI部署的关键硬件载体。目前AI眼镜主要有三种技术路径，分别是AI、AR及AI+AR，以Ray-Ban Meta、Meta Orion、Rokid Glasses为代表。

表：部分AI眼镜产品情况

产品	技术路线	企业	发布时间	产品特点	售价/元
Apple Vision Pro	AR	苹果	2024.01.19	混合现实头显设备，M2芯片，2300万像素屏幕，支持眼球追踪和手势操作	29999
OPPO Air Glass 3	AI	OPPO	2024.02.26	轻量级双目全彩AR眼镜，集成OPPO的AndesGPT大语言模型，树脂衍射光波导镜片，显示亮度高达1000尼特，支持触控手势操作，可通过软件更新获得导航、提词器、快速健康、健身信息预览等功能	4999
XREAL Air 2 Ultra	AR	XREAL	2024.05.30	提供高达330英寸虚拟屏幕，重量仅72克，兼容多种设备，支持大部分新款手机和游戏设备	3999
智能拍摄眼镜A1	AI	闪极科技	2024.05.31	采用紫光展锐旗舰级AI芯片和索尼1600万像素背照式摄像头，支持实时在线、听音和视觉体验	999
界环AI音频眼镜	AI	北京蜂巢	2024.08.08	内置先进的AI芯片与音频模块，实现语音命令控制音乐播放、接听电话或获取导航信息等智能功能	799
小度AI眼镜	AI	百度	2024.11.12	首款搭载中文大模型的原生AI眼镜，具备第一视角拍摄、边走边问、识物百科等六大功能	
Rokid Glasses	AI+AR	Rokid	2024.11.18	与BOLON眼镜合作，采用衍射光波导成像技术，接入通义千问大模型，兼具AR眼镜、耳机、AI助手和相机的多方面能力，支持物体识别、文字翻译、数学题解答等功能	2499

资料来源：量子位，国元证券研究所

3. 投资建议

- **展望明年，模型层面，**o1引领大模型发展新范式，新的Scaling Law有望驱动模型能力进一步提升，同时对于技术创新、工程能力和算力提出更高要求。**竞争格局方面，**海外头部大厂模型能力差距在2024年有所缩小，同时巨头及巨头深度合作的厂商通过上游资本开支和技术人才优势和其他竞争对手拉开身位，目前形成了五强格局，分别是OpenAI、Anthropic以及谷歌为代表的第一梯队，以及x AI和Meta。国内大模型目前竞争格局相对分散，包括互联网科技大厂、创业公司、传统技术类厂商为代表的三股力量，其中互联网科技大厂和自身云业务结合，综合布局；创业类厂商则依托不同资源禀赋进行差异化赛道聚焦。
- **应用层面来看，我们看好2025年应用端的投资机会，随着大模型竞争格局的逐步清晰，行业进入到价值实现和落地阶段。今年以来，模型调用成本逐渐走低，模型层能力向上成本向下降低应用端创新门槛，进一步促进应用端繁荣。**交互方式上，AI产品逐渐从Copilot模式向Agent模式转变，C端AI Agent与AI端侧硬件相结合有望重塑流量入口；在B端方面，则有助于AI加速落地行业场景。软件应用层面，企业可以通过本地部署、公有云、私有云、混合云等部署方式适配不同的规模和不同行业的企业，实现成本、私密安全性和大模型能力效果三者的平衡，企业端在大模型投入预算有望持续提升，同时企业主对于大模型投入ROI越来越重视。目前大模型在代码辅助、营销与客户管理、企业检索、办公软件等多场景落地较好，从行业上我们看好金融、政府服务、医疗等行业。C端软件应用方面，整体应用流量保持良好增长，ChatGPT周度活跃用户数突破3亿，web端流量较年初增长138%，AI ChatBots、AI内容生成与编辑、AI搜索、AI角色扮演是目前主流场景，我们看好AI搜索成为杀手级产品潜力。AI硬件方面，国内外AI+硬件的进程加快，小模型的发展推动AIPC、AI手机、AI眼镜、AI耳机等端侧硬件落地，Ray-Ban Meta成为首个爆款消费级产品，AI眼镜被视为AI端侧落地的关键硬件载体，值得关注。
- **标的方面，重点关注昆仑万维、视觉中国、恺英网络、神州泰岳、巨人网络、浙数文化、完美世界、吉比特、上海电影、中文在线、焦点科技、快手等。**

目录

4. 风险提示

- 风险提示：技术进展不及预期的风险，大模型安全性风险，应用推广不及预期的风险

投资评级说明

（1）公司评级定义		（2）行业评级定义	
买入	股价涨幅优于基准指数 15%以上	推荐	行业指数表现优于基准指数 10%以上
增持	股价涨幅相对基准指数介于 5%与 15%之间	中性	行业指数表现相对基准指数介于-10%~10%之间
持有	股价涨幅相对基准指数介于-5%与 5%之间	回避	行业指数表现劣于基准指数 10%以上
卖出	股价涨幅劣于基准指数 5%以上		

备注：评级标准为报告发布日后的6个月内公司股价（或行业指数）相对同期基准指数的相对市场表现，其中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指数或纳斯达克指数，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。

分析师声明

作者具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，以勤勉的职业态度，独立、客观地出具本报告。本人承诺报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业操守和专业能力，本报告清晰准确地反映了本人的研究观点并通过合理判断得出结论，结论不受任何第三方的授意、影响，特此声明。

证券投资咨询业务的说明

根据中国证监会颁发的《经营证券业务许可证》（Z23834000），国元证券股份有限公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询业务是指取得监管部门颁发的相关资格的机构及其咨询人员为证券投资者或客户提供证券投资的相关信息、分析、预测或建议，并直接或间接收取服务费用的活动。证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

法律声明

本报告由国元证券股份有限公司（以下简称“本公司”）在中华人民共和国境内（台湾、香港、澳门地区除外）发布，仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。若国元证券以外的金融机构或任何第三方机构发送本报告，则由该金融机构或第三方机构独自为此发送行为负责。本报告不构成国元证券向发送本报告的金融机构或第三方机构之客户提供的投资建议，国元证券及其员工亦不为上述金融机构或第三方机构之客户因使用本报告或报告载述的内容引起的直接或连带损失承担任何责任。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。本报告所载的信息、资料、分析工具、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投资标的的投资建议或要约邀请。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。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。在法律许可的情况下，本公司及其所属关联机构可能会持有本报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取投资银行业务服务或其他服务，上述交易与服务可能与本报告中的意见与建议存在不一致的决策。

免责条款

本报告是为特定客户和其他专业人士提供的参考资料。文中所有内容均代表个人观点。本公司力求报告内容的准确可靠，但并不对报告内容及所引用资料的准确性和完整性作出任何承诺和保证。本公司不会承担因使用本报告而产生的法律责任。本报告版权归国元证券所有，未经授权不得复印、转发或向特定读者群以外的人士传阅，如需引用或转载本报告，务必与本公司研究所联系并获得许可。

国元证券研究所

合肥	上海	北京
地址：安徽省合肥市梅山路 18 号安徽国际金融中心 A 座国元证券	地址：上海市浦东新区民生路 1199 号证大五道口广场 16 楼国元证券	地址：北京市东城区东直门外大街 46 号天恒大厦 A 座 21 层国元证券
邮编：230000	邮编：200135	邮编：100027