



# AI应用：商业化初露锋芒

——AI应用深度之二暨GenAI系列报告之三十九



# 主要内容

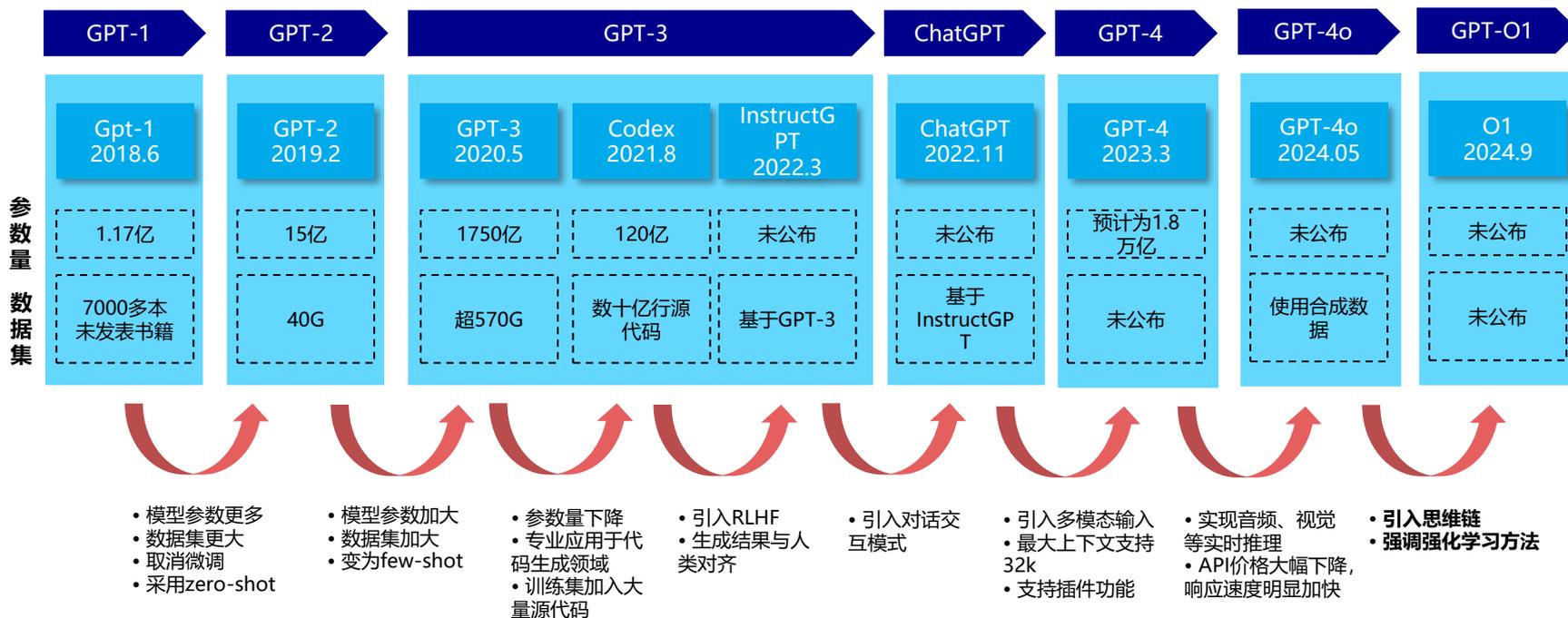
---

1. 大模型及AI应用进展
2. C端：AI广告/AI教育/AI陪伴/AI视频/AI搜索五大场景
3. B端：Agent嵌入企业 workflow，数据+定制化是关键
4. 估值表及风险提示

# 1.1 大模型：RL推动性能提升，算法优化带动价格下降

- **大模型性能仍在迭代提升：**由预训练带来的Scaling Laws进步速度有所放缓，强化学习（RL）带来的推理侧优化，OpenAI、Anthropic、谷歌等均压宝强化学习为技术进步迭代路径，有望在推理侧实现Scaling Laws，大模型能力仍将持续提升。
- **算法工程优化进展明确，大模型价格保持快速降低趋势：**各大模型厂商的API价格下降趋势明显，将推动AI应用落地。24年11月GPT4o 输入API价格为2.5美元/百万Tokens，而2023年3月GPT4推出时输入API价格为10美元/百万Tokens，已大幅下降。

图：强化学习流程，通过奖励机制增强大模型推理侧能力



# 1.1 大模型：正进入AI Agent爆发阶段

- **AI Agent时代到来：**根据OpenAI，迈向AGI共有5个阶段，当前正处于推理阶段，下一阶段即为智能体（AI Agent）。相比推理器阶段，AI Agent可实现与周边环境更强的交互能力，理解复杂指令并做出多步任务规划，完成难度更高的任务。
- **有助于AI应用爆发：**AI Agent将进入更多领域，包括办公、业务工作流程处理等，实现独立完成部分任务，进一步节省人力。

表：OpenAI提出的AI的五个阶段，当前正迈向AI Agent

AI发展阶段	所处阶段AI特性
Level1	<b>聊天机器人：</b> AI实现日常对话功能
Level2 (当前阶段)	<b>推理器：</b> AI具备基本的逻辑推理能力，能够分析复杂信息并进行推断
Level3	<b>智能体 (AI Agent)：</b> 理解复杂指令的能力，并能够在多任务环境中自主决策和灵活应对
Level4	<b>创新者：</b> 具备创新和创造的能力。它能够独立提出新的概念、假设和解决方案
Level5	<b>组织者：</b> 具备协调和管理庞大系统、资源和团队的能力

图：大模型RL路径下，Agent规划、执行等能力提升，可实现复杂工作流程操作



# 1.1 大模型：新一代大模型及AI Agent密集发布期将至

- **大厂Capex大幅提升下，各企业有望将迎来AI新进展：**24年底-25年初或即将发布重磅大模型，性能有望大幅提升。随着资本的持续投入、核心AI人才的流动，大模型你追我赶局面仍在继续，闭源模型TOP3厂商-OpenAI、谷歌、Anthropic的新一代大模型仍可能搅动现有大模型格局。
- **其次，期待AI Agent陆续成熟，大模型在推理侧使用强化学习的进步路径陆续清晰：**复杂任务规划等能力大幅提升，拓宽了AI Agent的能力边界。苹果端侧AI助手、谷歌已储备的Project Jarvis、Project Astra两大AI Agent项目、META AI助手等产品进展值得期待。

表：未来主要AI大事件催化

预计发布时间	公司	产品类型	潜在的重要AI内容
24年底-25年	苹果	通用AI助手	向用户陆续开放苹果手机上的端侧AI助手Apple AI Intelligence的更多功能
24年底-25年初	OpenAI	大模型	下一代大模型，性能有望相比前代明显提升
24年12月	谷歌	通用AI助手	通用助手Project Jarvis，实现Chrome浏览器网页任务自动化
25年	谷歌	通用AI助手	多模态的通用助手Project Astra，可实现视觉模态的交互
24年底-25年初	谷歌	大模型	新一代大模型Gemini 2.0，性能有望相比前代明显提升
24年底-25年初	Anthropic	大模型	新一代大模型的高性能版本Claude 3.5 Opus，此前已发布中等参数规模版本3.5 Sonnet
2025年	META	大模型	新一代大模型Llama4，参数规模将为Llama3的十倍

# 1.2 初创公司：大模型格局陆续清晰，各大巨头争夺人才



- **基础大模型公司多接受了科技巨头的投资，格局初定。**
- **顶级AI人才重要性凸显：**1) Ilya创办SSI、李飞飞创办Worlds Labs获得较高的初始估值；2) 谷歌、微软、亚马逊分别收购初创公司，也是为了储备和招揽AI人才，人才争夺进入白热化。
- **AI创意工具竞争加剧，投资重点转向AI视频：**AI图像竞争激烈；AI视频迎来拥有数据和生态优势的短视频巨头。

**表：主要大模型及创意工具初创公司估值及变动**

公司	前期估值	最新估值	类型	AI应用产品概况及落地时间	代表性投资人
OpenAI	290亿美元 (2023.04)	1570亿美元 (2024.10)	基础大模型	根据The Information, OpenAI 2024年收入预计为37亿美元, 其中ChatGPT收入预计为27亿美元	微软
Anthropic	40亿美元 (2023.03)	计划以400亿美元估值融资	基础大模型	根据CNBC, Anthropic 2024年收入预计将达到40亿美元	亚马逊, 谷歌, Salesforce, SAP
xAI	240亿美元 (2024.05)	计划以400亿美元估值融资	基础大模型	推出Grok聊天机器人, 已开启订阅制收费	埃隆马斯克
Mistral AI	20亿美元 (2023.12)	64亿美元 (2024.06)	基础大模型	推出Mistral系列大模型	微软, 英伟达, A16z
SSI	-	50亿美元 (2024.09)	基础大模型+AI安全	成立仅数月, 暂无产品发布, 将面向超对齐等AI安全方向开发产品	A16z等
月之暗面	25亿美元 (2024.05)	33亿美元 (2024.08)	基础大模型	Kimi智能助手月访问量曾超2000万	腾讯, 阿里巴巴
智谱 AI	-	投前估值200亿人民币 (2024.09)	基础大模型	推出智谱GLM系列大模型、智谱清言助手、端侧AI助手AutoGLM	腾讯, 阿里巴巴
Inflection (微软)	40亿美元 (2023.06)	微软6.5亿美元收购核心团队 (2024.03)	大模型+情感陪伴	推出以聊天陪伴为导向的AI聊天机器人Pi; 暂未商业化	谷歌, 英伟达, 微软
Character.AI (谷歌)	10亿美元 (2023.03)	谷歌25亿美元收购核心团队 (2024.08)	情感陪伴	推出可创建聊天角色的AI平台, 用户可定制对话式AI机器人; 23年6月开始收费, 但商业化进展较慢	A16z等
MiniMax	10亿美元 (2023.04)	预计约25亿美元	基础大模型+AI应用	旗下拥有情感陪伴应用talkie, MAU达到1100万, 推出AI视频工具海螺AI	米哈游, IDG资本, 高瓴资本
Synthesia	10亿美元 (2023.06)	-	AI创意工具	旗下数字媒体平台Synthesation可实现AI生成视频创作	谷歌, 英伟达
Stability.ai	10亿美元 (2022.10)	联合创始人辞职	AI创意工具	推出AI文生图工具Stable Diffusion及模型集; 定制化提供收费服务	Coatue
Lightricks	18亿美元 (2022.02)	-	AI创意工具	推出Photoleap、Videoleap等AI内容工具	Insight Partners
Runway	15亿美元 (2023.06)	计划以40亿美元估值融资	AI创意工具	AI视频生成工具, 至2023年底ARR达到2500万美元	谷歌, 英伟达, Salesforce
Midjourney	预计约10亿美元 (2023)	-	AI创意工具	推出AI文生图工具Midjourney; 2022年已开启收费	-
Worlds Labs	-	亿美元 (. )	D/VR/AR 智能工具	成立仅数月, 暂未推出产品, 旨在开发具备空间感知能力的AI产品	Az 等
Canva	130亿美元 (2023.03)	260亿美元 (2024.05)	传统设计+AI创意工具	推出平面设计工具Canvass AI, 24年收购AI图像公司Leonardo AI	Coatue

## 1.2 初创公司：编程、搜索、办公和法律场景商业化顺利

- **编程、搜索、办公、法律等场景初创企业估值提升明显，初步商业化。**办公、法律领域的初创公司主要为企业提供定制化AI服务，例如为企业内部构建企业级检索等，以Cohere、Glean为代表的AI+办公初创企业年化收入已达到数千万美元。

**表：主要效率工具、垂类场景AI初创公司估值及变动情况**

公司	前期估值	最新估值	类型	AI应用产品概况及落地时间	代表性投资人
HuggingFace	45亿美元 (2023.08)	-	开源社区	AI大模型开发社区，提供预训练的开源模型、数据集和各种工具	谷歌，英伟达
Adept (亚马逊)	10亿美元 (2023.03)	亚马逊收购核心团队 (2024.06)	开发工具	推出ACT-1大模型，可通过训练使用软件工具、API和Web应用程序	Coatue
Anysphere	4亿美元 (2024.07)	25亿美元 (2024.11)	编程工具	推出AI编程工具Cursor，24年11月ARR已达到5000万美元	Insight Partners
Cognition	-	20亿美元 (2024.04)	编程工具	推出AI编程工具Devin	Founders Fund
Augment	-	10亿美元 (2024.04)	编程工具	推出了AI编程工具	Sutter Hill Ventures
Magic	5亿美元 (2024.02)	15亿美元 (2024.07)	编程工具	专注于GenAI编程工具	谷歌、Jane Street
Replit	11.6亿美元 (2023.04)	-	开发工具	推出AI辅助编程工具Ghostwriter；2022年已开始收费	谷歌，英伟达，Salesforce
Perplexity	5.4亿美元 (2024.01)	正计划以80亿美元估值融资	搜索引擎	推出AI搜索引擎Perplexity，目前正寻求通过广告实现商业化	
Notion	103亿美元 (2021.10)	-	办公+AI	个人/企业的办公协作平台；23年3月Notion AI开始收费	红杉资本
Cohere	21亿美元 (2023.06)	55亿美元 (2024.07)	办公+AI	提供企业定制生成式AI服务，包括文本生成、文本分类和文本检索领域	英伟达，SAP
Glean	10亿美元 (2022.05)	46亿美元 (2024.09)	办公+AI	企业级搜索和知识管理平台，可通过AI帮助企业员工完成工作；根据WSJ，至24年8月ARR已达到5500万美元	红杉资本
Scale AI	73亿美元 (2021)	140亿美元 (2024.05)	AI数据标注工具	根据The Information，Scale AI在24年9月年化收入约为10亿美元	英伟达，亚马逊，META
Sierra	10亿美元 (2024.01)	45亿美元 (2024.10)	AI Agent	路透社报道24年10月报道ARR为2000万美元	红杉资本
Jasper	15亿美元 (2022.10)	联合创始人辞职	企业服务	推出了AI内容写作工具，面向营销文案等领域；2022年已开始收费	Foundation Capital
英矽智能	亿美元 ( )	港股寻求上市	医药+AI	推出AI药物研发平台PHARMA.AI	药明康德，淡马锡，高瓴资本
Harvey	1.5亿美元 (2023.04)	15亿美元 (2024.07)	法律+AI	推出法律行业的辅助办公对话式AI机器人	OpenAI，红杉资本

# 1.3 美股：云厂商体现AI收入，AI广告等应用陆续落地



- 以谷歌、微软、亚马逊、Oracle为代表的公有云厂商正在兑现AI收入，体现公有云客户对于算力的需求以及对AI IT支出的提升。
- AI广告进展相对较快：以META为代表的广告公司，通过升级AI推荐系统提升用户使用时长、广告主投放ROI，对广告业务形成正面拉动。

表：主要云计算与C端AI应用AI业务进展梳理

公司	AI收入来源类型	代表性AI应用	AI收入当前进展
微软	公有云+AI应用	Microsoft 365 Copilot、GitHub Copilot	AI云：24Q3，AI贡献Azure云收入12%的增长，仅推理，预计数十亿美元体量； AI应用：M365 Copilot、GitHub Copilot
谷歌	公有云+AI应用	Duet AI、NotebookLM、AlphaCode	AI云：24Q2为云客户提供的AI服务实现数十亿美元的收入； AI应用：已商业化的Gemini、Duet AI助手；C端应用NotebookLM、AI搜索功能AI Overview
亚马逊	公有云+AI应用	AmazonQ	AI云：24Q2为云客户提供的AI服务实现数十亿美元的收入； AI应用：B端助手AmazonQ；C端购物助手Rufus
Oracle	公有云+AI应用	OCI内集成的AI工具	AI云：截至24年6月的FY24财报，公司FY24签订了价值170亿美元的AI合同
Meta	广告/内容推荐	META AI助手，AI推荐系统迭代	AI推荐系统：24Q3，AI驱动信息流和视频推荐系统在24年已推动Facebook/Insgram 用户使用时间增加了 8%/6%。 AI应用：24Q3，META AI助手MAU已达到5亿
Duolingo	C端教育AI应用	Duolingo Max	AI应用：24年起从内测开始陆续向用户开放并推广，至24Q3，公司Duolingo Max功能已覆盖一半的DAU用户
APPLovin	广告营销	Axon AI算法	AI广告引擎：推出的Axon AI算法提升了广告匹配准确度，进一步提升广告主ROI

# 1.3 美股：企业服务、网络安全及数据分析进展迅速

- **软件公司AI应用重要落地方向有：** 1) 企业服务：以ServiceNow为代表，提升企业生产流程效率。2) 网络安全：AI主要用于加强威胁感知等能力，以及通过语言助手加强与客户之间的交互。3) 数据分析：LLM进一步加强了对数据的理解能力。

**表：主要软件公司估值以及AI业务进展梳理（单位：亿美元）**

公司	AI收入来源类型	市值	FY25营收预期	25年Forward PS	AI产品形态	季度法说会定期披露的AI营收或订单数据
Salesforce	CRM软件	3325	413.0	8.1	嵌入到Salesforce软件的助手Einstein Copilot；推出B端AI Agent平台Agentforce	暂未披露
Sap	ERP软件	2739	397.0	6.9	将数百个AI功能嵌入到Joule助手中	暂未披露
ServiceNow	IT软件	2204	133.0	16.6	推出Now Assist助手，为企业生产流程提供数百种AI功能	至24Q3，AI解决方案已有44位客户的ACV超过100万美元，其中6位超过500万美元，2位超过1000万美元
Atlassian	办公协同软件	683	51.0	13.4	1) Atlassian Intelligence：集成在产品中，提供LLM加持的交互、数据分析等；2) 24年5月推出Rovo，AI增强型协作工具	暂未披露
Palantir	数据分析	1473	35.0	42.0	推出AIP平台，包括LLM加持的数据分析及预测功能	截至FY24Q2，AIP Bootcamp已经培训超过1000家企业
Datadog	数据分析	524	32.4	16.2	旨在增强云应用程序的可观察性和性能监控。1) LLM Observability于2024年6月推出，旨在帮助客户监控、改进和保护LLM应用程序。2) Bits AI：基于GenAI的DevOps Copilot	24Q3 AI原生客户占ARR的6%
Snowflake	数据分析	574	43.4	13.2	1) Snowflake Cortex：利用大型语言模型(LLM)来分析非结构化数据、回答问题并在Snowflake生态系统中提供智能帮助。	暂未披露
Mongodb	数据库	248	23.1	10.7	24年5月发布MongoDB AI应用程序(MAAP)，旨在帮助组织快速构建和部署富含生成式AI技术的现代应用程序。	暂未披露
Crowdstrike	网络安全	893	47.6	18.8	1) Charlotte AI于2023年5月推出，通过自然语言命令协助安全分析师自动执行复杂的工作流程并加快任务完成速度；2) AI-SPM于2024年9月发布，可通过统一平台为云基础设施、应用程序、数据和AI模型提供全面的可视性和保护	暂未披露
Cloudflare	网络安全	351	20.9	16.8	1) Workers AI：AI开发平台；2) 与网络安全业务结合，包括AI防火墙、AI网关、API异常检测等	FY24Q3，获得了首份超过百万、价值700万美元的Workers AI合同
Palo Alto	网络安全	1301	91.5	14.2	1) Precision AI，旨AI加强实时安全性和主动威胁检测；2) Gen AI：与客户进行交互，并增强安全系统的预测能力	截至FY24Q4，AI ARR已超过2亿美元
Cyberark Software	网络安全	139	13.0	10.6	CORA AI：该产品旨在通过跨各种身份域（包括特权访问、端点、机密和云身份）提供高级检测和响应功能来增强安全措施。	暂未披露
Zscaler	网络安全	323	31.6	10.2	1) AI数据保护平台；2) 内联威胁防护	AI分析方案为FY24Q4的新业务等增长贡献3pct为财年贡献pct
Synopsys	芯片设计	851	68.8	12.4	1) 为AI芯片提供设计软件；2) 通过AI优化芯片设计流程	截止FYQ 的一年内，公司来自AI芯片设计的收入为5亿美元

# 1.3 北美科技巨头：算力、大模型和AI应用布局



表：美股核心科技巨头AI布局及禀赋

	AI芯片布局	大模型及开发框架	AI研发布局模式	已推出的核心AI应用	现有业务生态协同
微软	<p><b>外购：</b>根据Omdia，23年购买约15万张H100芯片</p> <p><b>自研：</b>2023年11月发布Maia 100芯片</p>	<p><b>大模型：</b>OpenAI推出GPT系列模型，2023年3月推出GPT-4，24年5月推出GPT-4o，24年9月推出GPT-o1</p> <p><b>开发平台：</b>Azure AI Studio，包括GPT系列独家模型及第三方大模型</p>	<p><b>大比例持股体外公司+深度合作。</b>2023年向OpenAI投资100亿美元，为OpenAI主要的算力提供商</p> <p><b>自研：</b>招揽Inflection AI核心团队，布局大模型</p>	<p><b>办公：</b>推出Microsoft 365 Copilot</p> <p><b>CRM/ERP：</b>推出Dynamic 365 copilot</p> <p><b>编程工具：</b>Github Copilot</p> <p><b>搜索引擎：</b>必应集成ChatGPT</p>	<p><b>云计算：</b>Microsoft Azure</p> <p><b>办公软件：</b>Microsoft 365、Office</p> <p><b>操作系统：</b>Windows</p> <p><b>浏览器：</b>Edge</p> <p><b>搜索引擎：</b>Bing</p>
谷歌	<p><b>外购：</b>根据Omdia，23年购买约5万张H100；</p> <p><b>自研：</b>2016年推出第一代TPU芯片，已发布第六代TPU Trilium，性能出色，可基本支撑自研大模型的训练和推理</p> <p><b>通信：</b>自研OCS通信系统，通信性能出色</p>	<p><b>大模型：</b>2023年12月推出首个多模态大模型Gemini，预计24年底-25年初发布Gemini 2.0</p> <p><b>深度学习框架：</b>TensorFlow（两大主流框架之一）、JAX</p> <p><b>开发平台：</b>Vertex AI</p>	<p><b>旗下部门自研：</b>此前有Google Brain、Deepmind等多个AI研发部门/全资子公司，分立运营；2023年4月起整合为单一AI研发部门Google Deepmind</p>	<p><b>办公：</b>推出Duet AI，定价30美元/月</p> <p><b>搜索：</b>AI搜索功能AI Overview，至24年10月，已覆盖10亿用户</p> <p><b>应用：</b>NotebookLM</p> <p><b>其他：</b>编程工具Alphacode等</p>	<p><b>云计算：</b>Google Cloud</p> <p><b>办公软件：</b>Workspace</p> <p><b>操作系统：</b>安卓</p> <p><b>浏览器：</b>Chrome</p> <p><b>搜索引擎：</b>Google</p> <p><b>应用矩阵：</b>谷歌地图、Youtube、Play store、Gmail</p>
Meta	<p><b>外购：</b>根据Omdia，2023年购买约15万张H100芯片。计划2024年底拥有35万颗H100 GPU，总共60万颗H100等效算力；</p> <p><b>自研：</b>2024年发布MTIA v2芯片，陆续投入使用</p>	<p><b>大模型（开源）：</b>2023年7月开源Llama2，2024年推出Llama3，Llama4正在10万卡集群上训练</p> <p><b>深度学习框架：</b>Pytorch（两大主流框架之一）</p>	<p><b>旗下部门自研：</b>AI业务均由旗下AI部门进行研发，为直属部门模式</p>	<p><b>推荐系统全面升级：</b>AI推截至24年10月，今年已提升Facebook/Instagram使用时长8%/6%</p> <p><b>AI助手：</b>META AI助手已集成于社交软件中，至24Q3 MAU超5亿</p> <p><b>广告创意及投放：</b>推出辅助广告内容生成工具，AI广告投放工具</p>	<p><b>社交应用：</b>Facebook，全球最大社交软件；以及Instagram</p> <p><b>元宇宙：</b>旗下VR设备品牌Quest以及内容平台</p>
亚马逊	<p><b>外购：</b>根据Omdia，2023年购买约5万张H100</p> <p><b>自研：</b>2020年推出Trainium，23年11月推出Trainium2，用于训练侧；2023年4月推出Inferentia2，用于推理侧</p>	<p><b>自研大模型：</b>2023年12月推出Titan系列AI模型</p> <p><b>大模型（Anthropic）：</b>24年6月推出Claude3.5</p> <p><b>开发平台：</b>Bedrock AI搭载自研及第三方模型</p>	<p><b>旗下部门自研+持股重点公司：</b>旗下AI部门完成自研大模型研发；重点投资Anthropic，2023年9月注资40亿美元；谷歌也参与Anthropic多轮投资</p>	<p><b>电商：</b>为电商运营提供一系列AI功能支持，以及导购助手Rufus；</p> <p><b>云端AI助手：</b>面向企业端的AmazonQ；</p> <p><b>广告：</b>辅助广告内容生成工具；通过AI实现广告智能投放提升效率</p>	<p><b>云计算：</b>AWS</p> <p><b>电商平台：</b>亚马逊商城</p>
苹果	<p><b>自研：</b>根据华尔街日报，公司正推进自研数据中心芯片Project AIDC项目</p>	<p><b>大模型：</b>2024年7月推出自研大模型AFM</p>	<p><b>旗下部门自研+引入外部大模型：</b>AI业务均由旗下AI部门进行研发，为直属部门模式</p> <p>未来可能将OpenAI、谷歌、Anthropic等引入苹果生态中</p>	<p><b>端侧助手：</b>苹果 AI Intelligence 助手</p>	<p><b>硬件：</b>手机、PC、VR设备等</p> <p><b>操作系统：</b>iOS</p> <p><b>应用商店：</b>Apple Store</p>

# 1.4 国内科技巨头：算力、大模型和AI应用布局

表：国内科技巨头AI布局及禀赋

	AI芯片布局	大模型及AI开发框架	已推出的核心AI应用	现有业务生态协同	投资公司
字节跳动	外购：根据Omdia，2023年公司购买了2万片H100	大模型：豆包；多模态BuboGPT	AI视频工具：即梦 AI Chatbot：豆包 AI Agent平台：Coze、小悟空	应用矩阵：抖音、TikTok、剪映、今崧智能、云鲸智能等，对日头条等 办公：飞书	外部AI大模型公司投资较谨慎
阿里巴巴	自研AI芯片：12nm 含光800（推理） 自研CPU：倚天系列 外购：根据Omdia，2023年公司购买了2.5万片H100	大模型：24年5月发布通义千问2.5 开发平台：百炼AI平台	AI Chatbot：通义 电商助手：淘宝问问（ToC）、AI生意助手（ToB） 开源大模型社区：魔塔社区	云计算：阿里云，FY25Q1法说会披露，预计FY25H2云业务两位数增长，大部分由AI产品推动 电商平台：淘宝 办公：钉钉	智谱AI，零一万物，百川智能，月之暗面，MiniMax等
腾讯	自研AI芯片：紫霄（推理） 外购：根据Omdia，2023年公司购买了5万片H100	大模型：24年11月推出Huanyuan large 389B MoE 开源模型 开发平台：腾讯云AI平台	AI Chatbot：混元助手、腾讯元宝 推荐系统：广告及内容推荐升级 AI视频平台：腾讯智影 AI Agent平台：腾讯元器 AI笔记：Ima copilot	云计算：腾讯云 应用矩阵：QQ、微信及视频号、QQ邮箱、游戏、腾讯视频等 办公：腾讯会议	MiniMax、智谱AI、百川智能、深言科技和无问芯穹等
百度	自研AI芯片：7nm 昆仑芯二代 外购：根据Omdia，2023年公司购买了3万片H100	大模型：24年6月发布文心4.0 Turbo 深度学习框架：飞桨 开发平台：千帆	AI搜索：百度AI智能问答，24年8月18%的搜索包含AI生成内容，24年5月为11%。 AI Chatbot：文心一言，24年8月API日均调用6亿次（24年5月为2亿次），生成1万亿Tokens内容 AI Agent平台：文心智能体 自动驾驶：萝卜快跑，24Q2提供89.9万次乘车服务，YoY+26%	云计算：百度云，FY24Q2 Gen AI相关收入占云总收入的9%，FY24Q1为6.9% 搜索引擎：百度搜索 应用矩阵：百度地图、爱奇艺	生数科技、中科加禾、巨存科技、无问苍穹、西湖星辰等
华为	自研AI芯片：昇腾310及昇腾910等	大模型：24年6月发布盘古大模型5.0 深度学习框架：MindSpore 开发平台：ModelArts	自动驾驶：ADS 3.0	云计算：华为云 硬件：手机、PC、智能家具等 操作系统：鸿蒙	面壁智能、生数科技等

# 1.4 国内科技巨头：云计算、AI广告商业化取得初步进展



表：国内互联网公司各季度财报、业绩会对于AI效果的表态

公司	业务	进展
腾讯	广告	<p>23Q2：广告业务实现了显著的快速增长，得益于在广告平台上应用了机器学习能力以及视频号的商业化。</p> <p>23Q4：升级了AI驱动的广告技术平台，显著提升了精准投放效果，从而增加了广告收入。</p> <p>24Q1：网络广告.....乃受微信视频号、小程序、公号及搜一搜的增长拉动，乃由用户参与度的提升以及我们不断升级的AI驱动的广告技术平台所推动。</p> <p>24Q2：视频号总用户使用时长同比显著增长，得益于推荐算法的增强和更多本地化内容。</p>
阿里巴巴	云业务	<p>24Q1：AI相关收入增长加快，并持续录得三位数的同比增长。AI相关收入来自基础模型公司、互联网公司，以及如金融服务、汽车等不同行业客户。</p> <p>24Q2：AI相关产品收入持续录得三位数同比增长。使用阿里云AI平台（百炼）的付费用户数量季度环比增长超过200%。</p>
	淘天集团	<p>24Q2：通过在例如推荐和匹配算法等技术上的投入，以及通过由多元化供应商提供更多有价格竞争力的供给，改善了产品与用户需求的匹配度，提升了用户流量转化为购买力。今年4月，我们推出一款新的AI驱动的全平台市场营销工具“全站推广”，看到使用此产品的商家数稳步提升。</p>
百度	云业务	<p>23Q4：生成式AI和大模型相关的收入在智能云的收入占比4.8%。</p> <p>24Q1：生成式AI和大模型相关的收入在智能云的收入占比6.9%。</p> <p>24Q2：生成式AI和大模型相关的收入在智能云的收入占比9%。</p>
哔哩哔哩	广告	<p>24Q1：哔哩哔哩创意中心上线，通过AIGC技术实现广告素材的批量化处理，提高效率。与AI相关的日常活跃内容创作者和视频浏览量分别同比增长了60%和80%，推动AI相关广告激增，使其成为哔哩哔哩增长最快的广告垂直领域之一。</p> <p>24Q2：新推出的创意中心和积极应用AIGC技术帮助我们的广告商在优化后诊断和改进他们的广告材料，CVR实现了超过30%的同比增长。</p>
爱奇艺	内容制作	<p>24Q1：丰富AI工具赋能内容从策划到宣发全流程。IP评估，阅读效率增9倍；声音修复，改善声音效果；营销助力，丰富宣发物料；智能搜索，内容推荐更精准。</p> <p>24Q2：AI助力项目评估、创意、营销和分发方面的使用范围进一步扩大，有效提升内容生产和运营效率。例如，AI在项目立项初期即可辅助收入预测。</p>
	广告	<p>24Q1：生成式AI赋能素材制作，提升广告主效率及投资回报率，游戏、短剧等行业广告主投资回报率提升超%。</p> <p>24Q2：AI赋能的效果广告收入环比增长150%。</p>

# 1.4 国内科技巨头：字节AI产品-Chatbot、视频、教育



- **字节在AI领域的战略较为清晰：**1) 对标优质竞品推出产品，覆盖聊天助手、情感陪伴、教育等全领域。2) 产品同步在国内海外推出；3) 坚持大模型到应用的端到端布局。
- **目前，已有部分AI应用产品关注度较高：**1) 字节持续导流的豆包助手；2) 依托自身丰富短视频数据的即梦AI视频生成平台；3) AI教育领域的AI答题助手Gauth近期流量实现了较快增长。

**表：字节跳动AI布局**

	团队	类型	产品	对标产品	
模型层	大模型团队	模型	豆包大模型、其他多模态模型	ChatGPT, LLaMa等	
	Flow	Agent定制	Coze (海外) /扣子 (国内)	Azure AI Agent、Dify等	
应用层	Flow	聊天	<b>豆包 (国内)</b> /CICI (海外)	ChatGPT, Claude等	
		虚拟陪伴	猫箱 (国内) /Anydoor (海外)	Character AI, Talkie等	
		图像	星绘 (国内) /Pic (海外)	Remini, 妙鸭相机等	
	剪映	图像/视频	即梦 (国内) /Dreamina (海外)	Runway, Midjourney, Stable Diffusion等	
			剪映 (国内) /CapCut (海外)		
	应用层	头条	办公	醒图 (国内) /Hypic (海外)	ChatGPT, Claude等
				小悟空 (国内) /Chipshop (海外)	
大力教育		教育	豆包爱学 (国内) <b>Gauth (海外)</b>	Question AI, Answer AI等	
巨量引擎		电商内容	即创 (国内)	Runway, Midjourney, Stable Diffusion等	
其他		音乐	海绵乐队 (国内)	Suno等	
	代码生成	MarsCode	GitHub Copilot等		
	AI模型社区	炉来 Lumi (国内)	Liblib等		
硬件	Oladance	耳机	Ola Friend		

# 主要内容

---

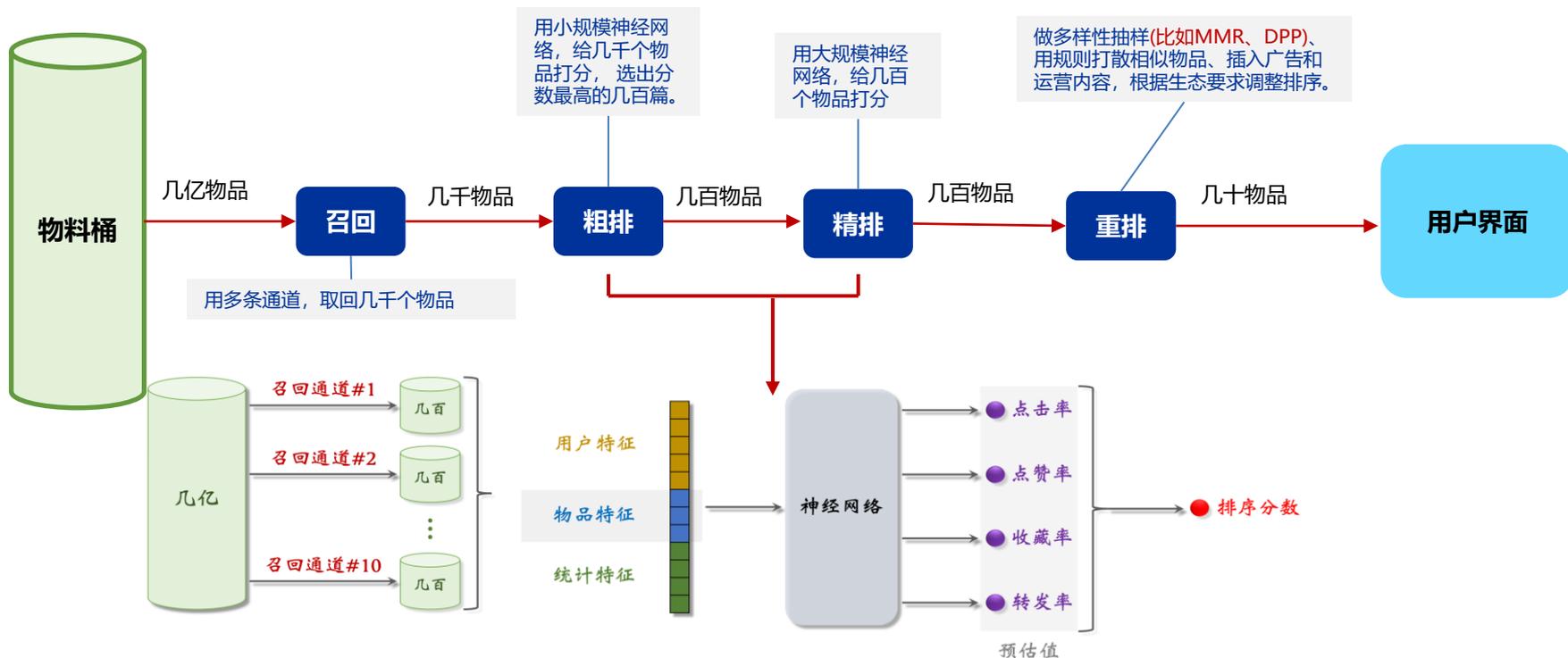
1. 大模型及AI应用进展
2. C端：AI广告/AI教育/AI陪伴/AI视频/AI搜索五大场景
3. B端：Agent嵌入企业 workflow，数据+定制化是关键
4. 估值表及风险提示

## 2.1 AI广告：当前主流推荐系统仍为顺序推荐架构

■ **目前主流推荐系统由召回-粗筛-预测-排名四个步骤组成：**

- 1) 召回阶段（离线计算）：**从整个内容库中快速筛选出一部分与用户可能相关的候选内容，通常是数量较多的粗筛结果；
- 2) 粗排阶段：**使用轻量级模型对召回阶段产生的候选物品进行初步排序，将不太相关的内容剔除，保留较相关的一部分；
- 3) 精排阶段：**对粗排后的少量候选内容进行精细化排序，形成可用的推荐结果；
- 4) 重排阶段：**在精排的基础上，根据广告投放的具体目标，对最终排序结果进行进一步的优化和调整。

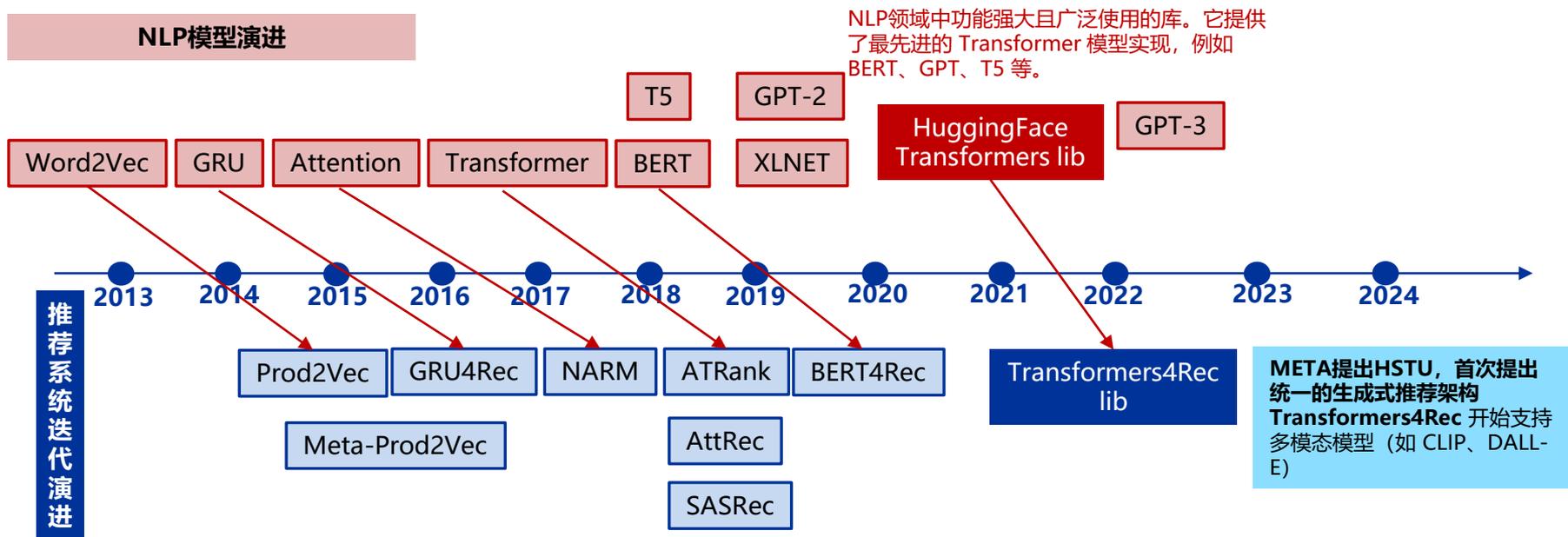
图：推荐系统包括召回、粗排、精排、重排等阶段



## 2.1 AI广告：LLM有望进一步赋能推荐系统

- **业界正探索将LLM用于推荐系统中。** 1) **直接使用LLM强化推荐系统各环节性能：** 包括在数据采集阶段、特征工程、特征编码、打分排序、流程控制等阶段集成LLM大模型能力，利用LLM的理解能力等完成推荐系统中的一些环节； 2) **借鉴大语言模型的架构设计：** 由于数学抽象式相对一致，许多架构设计对于推荐系统的进步有着重要意义，可从架构设计上提升推荐系统的性能。

图：推荐系统演进实际上跟随自然语言模型的技术演进

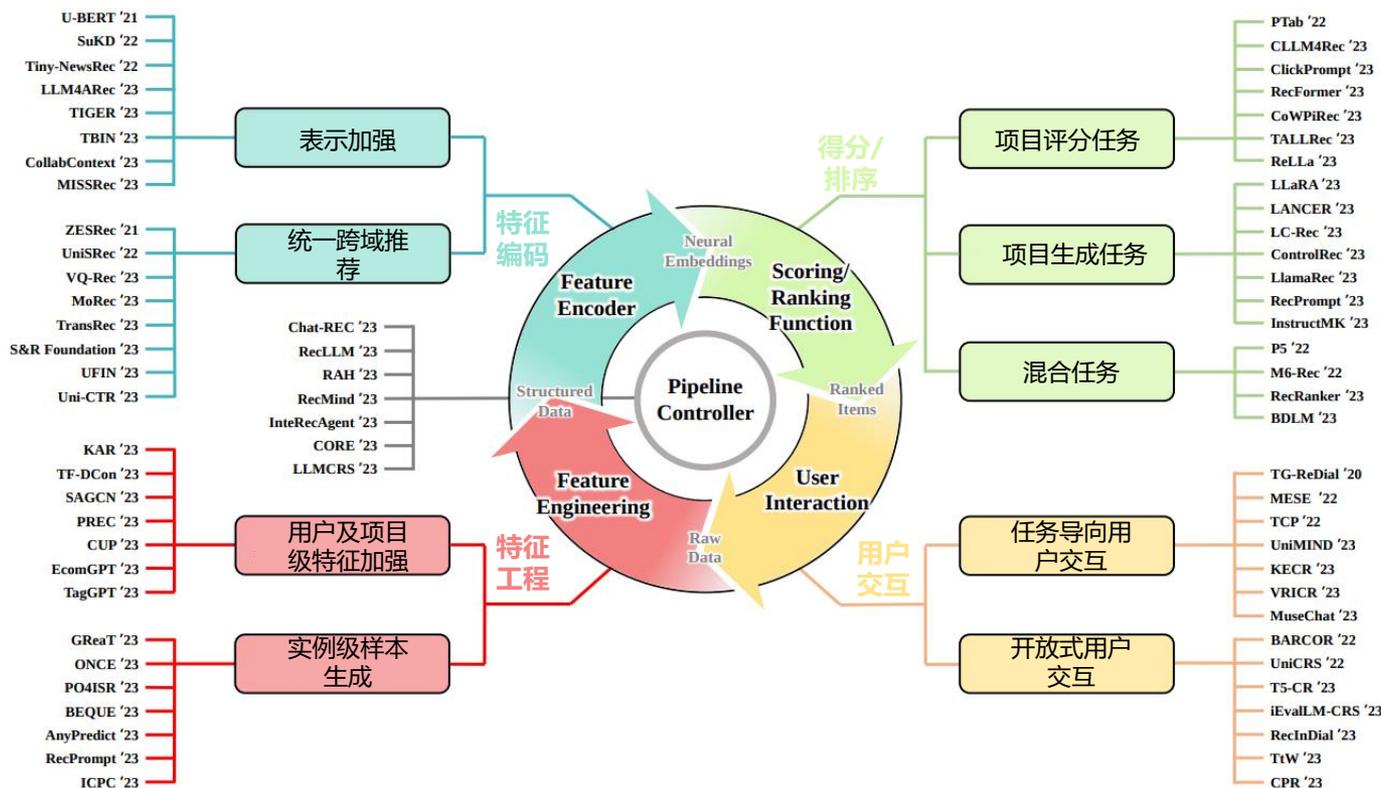


资料来源：《TransformersRec: Bridging the Gap between NLP and Sequential/Session -Based Recommendation》，申万宏源研究

## 2.1 AI广告：LLM强化推荐系统各环节

- 自22年以来，业界已有多个将LLM应用于推荐系统的研究，覆盖特征工程、特征编码、用户交互等领域，通过LLM替代或升级各环节的原有模型，以强化推荐系统的效果。

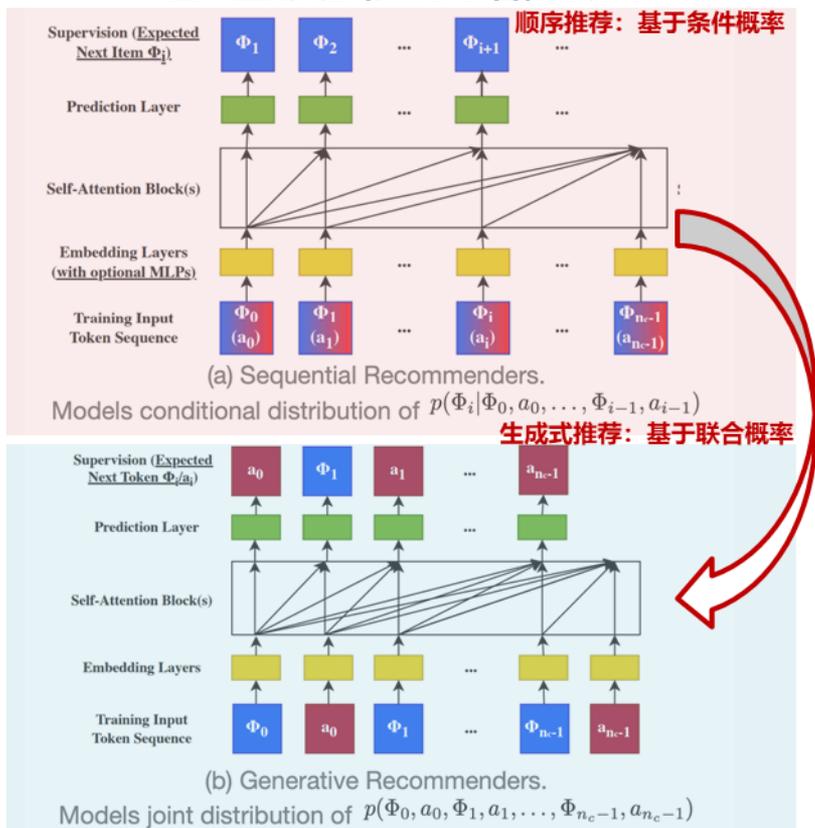
图：LLM 已应用于特征工程、特征编码等多个领域



## 2.1 AI广告：基于LLM的生成式推荐架构

- **Meta 的生成推荐系统解决了传统推荐系统的几个局限性：**
  - 1) 统一时间序列：**GR 将项目和操作交错在统一的时间序列中，将其他分类特征编码为缓慢变化的时间序列。
  - 2) 联合分布建模：**与对条件分布进行建模的顺序推荐器不同，GR 对用户操作的联合分布进行建模。
  - 3) 可扩展性：**GR 可以扩展至 1.5 万亿个参数，明显大于传统模型。**META验证了基于HSTU架构的生成式推荐模型，随着计算量增加性能可持续改善的Scaling Laws。**

图：生成式推荐基于联合概率分布



表：生成式推荐旨在学习用户的兴趣分布进行生成物品/内容推荐

	判别式推荐	生成式推荐
<b>核心目标</b>	判别式推荐系统直接预测用户对物品的偏好，而不关心数据的生成过程。	生成式推荐系统旨在学习用户和物品的潜在分布，从而能够生成或预测用户对物品的偏好。
<b>数学模型</b>	条件概率 $p(y x)$ , $x$ 是输入特征（例如用户、项目的特征）， $y$ 是目标（例如是否点击或评分）	联合概率分布 $p(x, y)$ , $x$ 为用户特征， $y$ 为生成的推荐项目
<b>优化目标</b>	最大化条件概率	模型从 $p(x, y)$ 中采样生成推荐结果，最大化对数似然函数
<b>常见模型</b>	逻辑回归、矩阵分解、深度神经网络 (DNN)、图神经网络(GNN)	变分自编码器(VAE)、生成对抗网络(GAN)、Transformer-based models
<b>适用场景</b>	分类、回归任务，基于已有数据进行推荐	生成新的推荐内容或项目
<b>预测能力</b>	对现有数据进行精确匹配预测	根据用户偏好生成新的推荐项
<b>模型训练方式</b>	基于用户和项目特征进行监督学习	通过生成和判别同时训练，或通过采样生成数据

- **AI推荐系统推动META应用使用时长每年提升6%~8%，带来广告供给增加。** META通过提升AI推荐系统推荐内容的占比以及迭代推荐系统，在更少的数据的条件下，完成更为精确的内容推荐，进一步增加用户在应用中的使用时长。

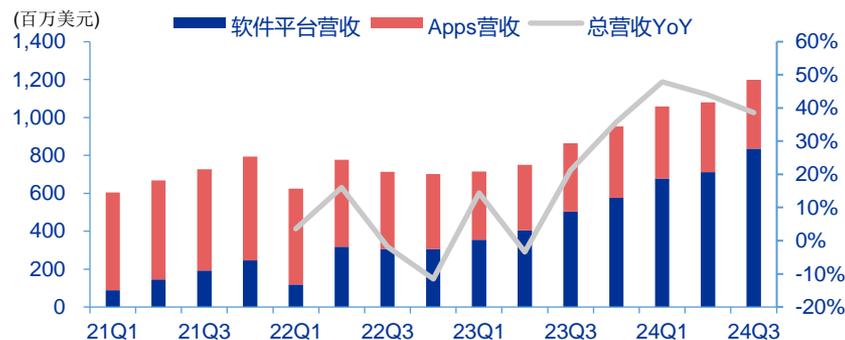
**表：META 各季度法说会对于AI推荐系统与Advantage+的表态，AI正深入META广告产品线**

财季	AI推荐系统	Advantage+
22Q2	AI会推荐大约15%的Facebook动态内容，以及略多于15%的 Instagram动态内容，这些内容来自用户不关注的人、群组或账户。我们预计到23年年底，这些数字将增加一倍以上。	
22Q3		最近对一批广告商进行的一项测试发现，使用 Advantages+ Shopping广告系列的广告商的广告支出回报率提高了32%。
23Q1	Facebook动态消息中超过20%的内容是由AI从用户不关注的人群或帐户推荐的。在整个Instagram上，这大约占您看到内容的40%。 自我们推出 Reels 以来，AI 推荐已使Instagram用户使用时间增加了24%以上。我们的AI工作也提高了盈利能力。与上一季度相比，Reels在 Instagram上的盈利效率提高了30%以上，在Facebook上的盈利效率提高了40%以上。	过去六个月，Advantage+购物活动的每日收入增长了7倍。
23Q2	自从推出AI推荐以来，它们已经 <b>推动了平台总使用时间增加7%。</b> <b>我们还部署了Meta Lattice</b> ，这是一种新的模型架构，可以学习预测各种数据集和优化目标中的广告效果。	
23Q3	仅今年一年，由于推荐改进，我们使用户在Facebook上花费的时间就增加了7%，在Instagram上花费的时间增加了6%。	我们为广告商提供的AI工具推动了Advantage+购物广告系列的成果，达到了100亿美元的年化收入，超过一半的广告商使用我们的 Advantage+创意工具来优化其广告创意中的图片和文字。
23Q4	我们看到Reels和视频整体持续增长，因为在第四季度，所有视频类型的每日观看时间同比增长超过25%，这得益于排名的持续改善。	
24Q1	<b>Facebook信息流中约30%的帖子是由我们的AI推荐系统提供的。Instagram上人们看到的50%以上的内容都是AI推荐的。</b> 最近，我们一直在 <b>开发一种新的模型架构</b> ，旨在为多种推荐产品提供支持。23年，我们开始部分验证该模型，使用它来支持 Facebook Reels，部署该模型后， <b>观看时间增加了8%到10%。</b>	根据我们进行的测试，使用 Advantage+受众定位的广告系列的每次点击费用或每个目标费用平均下降了28%。
24Q2	在Facebook上，我们于6月在全球推出统一的视频播放器和排名系统，看到了令人鼓舞的早期成果。 META AI助手MAU已经达到亿月活用户数；	今年进行的一项研究表明，美国广告客户在采用 Advantage+购物广告系列后，广告支出回报率提高了22%。 仅年月，超过一百万广告商使用我们的 Gen AI 工具制作了超过 万条广告，我们估计使用图像生成的公司的广告转化率增加了 7%。
Q	AI 驱动信息流和视频推荐仅在今年就使Facebook上的用户使用时间增加了 %，Instagram上的用户使用时间增加了6%。	

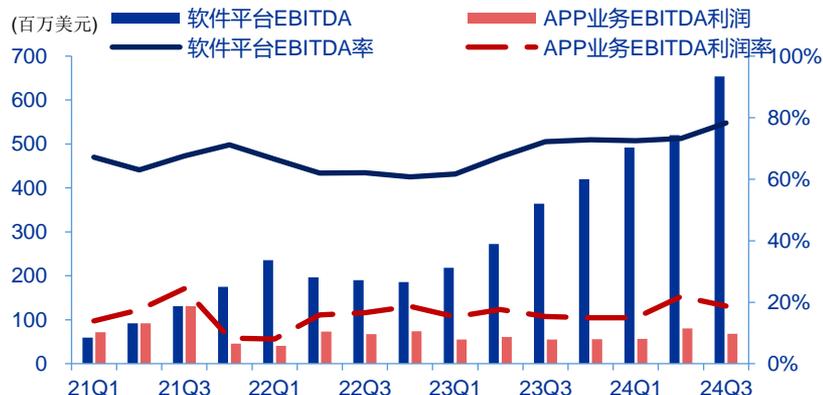
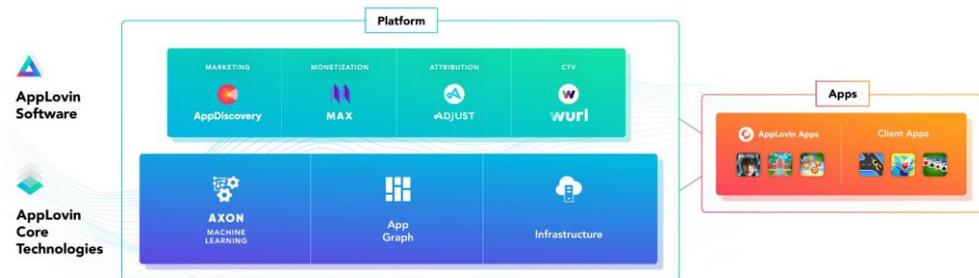
## 2.1 AI广告：AppLovin-营销投放算法升级，提升投放效率

- Axon2.0算法能力进步提升客户广告投放ROI，进一步拉动客户广告投放支出，对应公司软件平台业务EBITDA利润大幅增长，盈利能力明显提升。
- 2023年6月，AppLovin推出由AI驱动的Axon 2.0广告引擎（首个版本发布于2021年），支持AppDiscovery、MAX等工具，发布后对于广告主的投放效率及回报率有着较为明显的提升。Axon 2.0能够在几微秒内完成大规模的广告拍卖，可以识别可能下载客户应用的用户以及持续并加深与这些应用互动的信息。

图：AI加持下AppLovin业绩明显增长



图：AppLovin主要产品矩阵，AXON机器学习广告引擎加持软件平台各产品

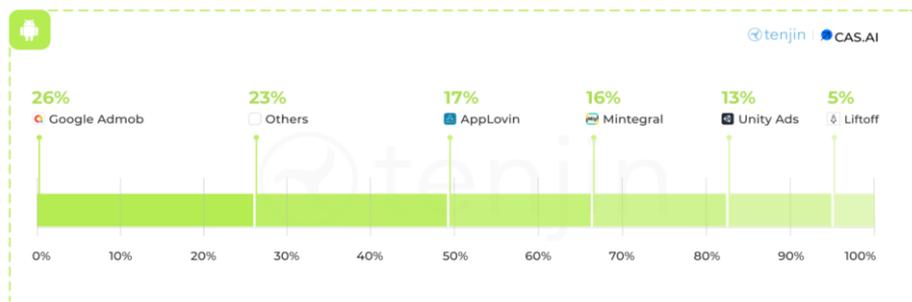


## 2.1 AI广告：AppLovin-构筑数据壁垒，为算法迭代奠基

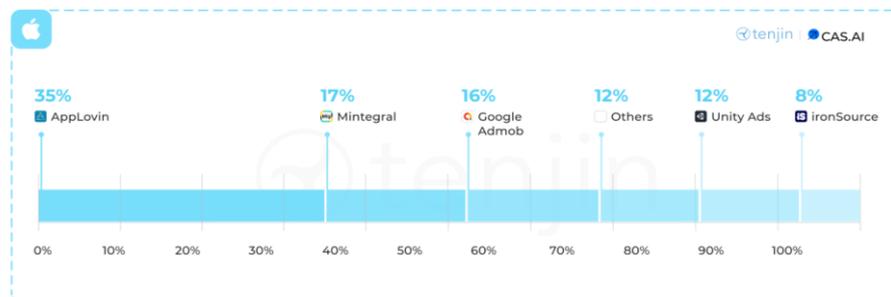
- **为什么AI进展顺利？领先来自数据优势+算法技术优势。** AppLovin通过收购广告聚合平台与归因工具积累了数据优势，构成Axon算法进步的基础。AI在广告领域的落地，数据的积累至关重要。
- **MAX广告聚合平台：** Applovin于2018年收购MAX广告聚合平台，MAX可以实时竞拍广告资源，实现开发者和流量匹配的同时，实现收入最大化，并收取一定服务费（5%）。我们认为，聚合平台掌握的竞价数据是关键数据之一，规模也至关重要，类似双边网络效应。根据Applovin公众号，聚合竞价产品MAX市场占有率在不断扩大，截至24Q1，全球已有130多个国家的开发者接入MAX，头部产品总体量已超14万。
- **Adjust第三方广告监测工具：** Adjust是海外较常用的第三方广告监测工具（归因工具），且安装量较高，归因是指广告主了解哪个投放组合（哪个渠道的哪个广告组+广告）带来的下载，以及后续转化行为情况。

图：24Q2 AppLovin市场份额

2024年第二季度各广告渠道的Android广告收入份额



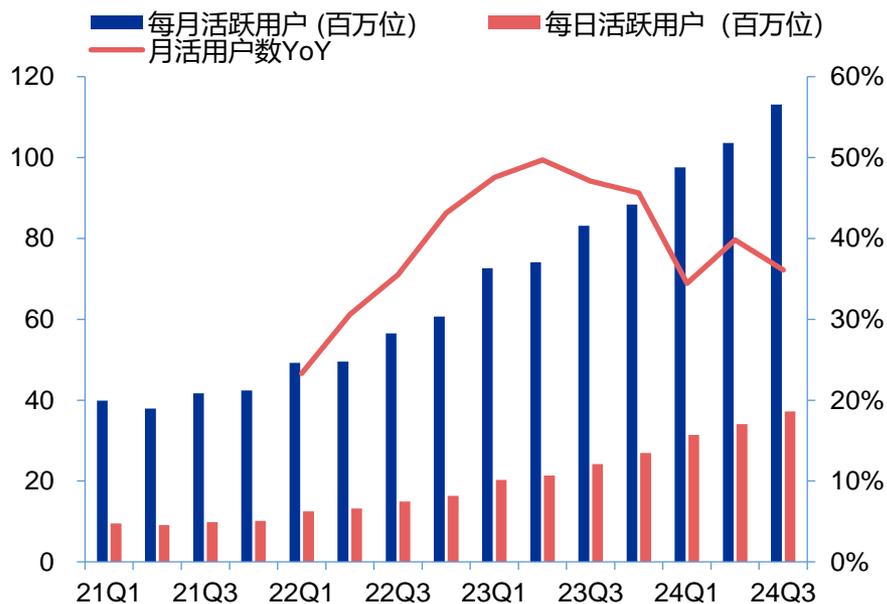
2024年第二季度各广告渠道的iOS广告收入份额



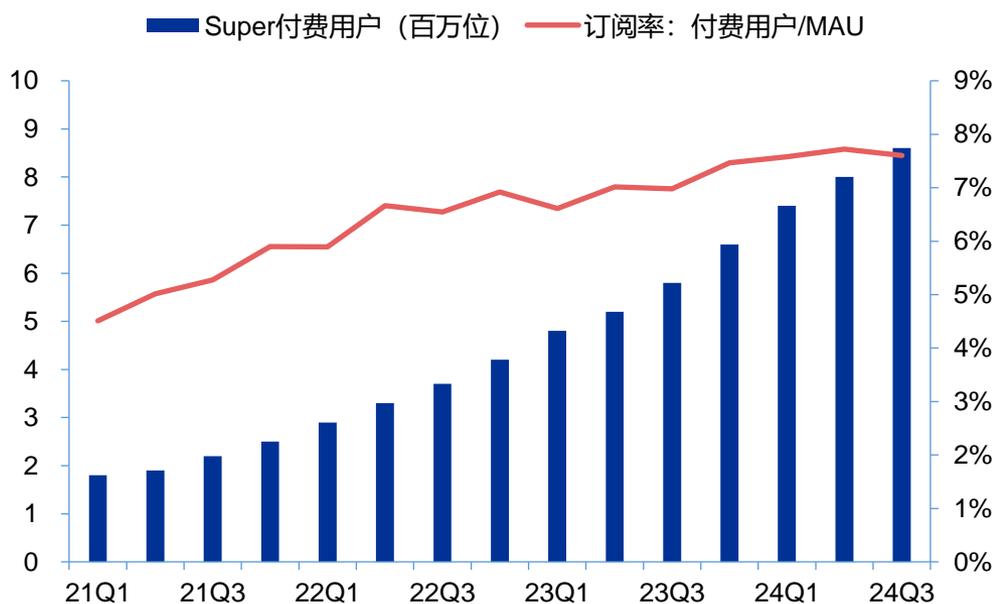
## 2.2 AI教育：Duolingo-Super付费率提升，Max即将贡献AI收入

- **Duolingo会员分为：**1) **Super：**常规会员，定价12.99美元/月，包括无广告、不间断课程、练习中心等升级服务；2) **Max：**AI功能会员，定价29.99美元/月，通过AI提供个性化教学，包括解释答案、角色扮演教学、AI视频通话功能。
- **Duolingo股价近期强势，市场关注后续AI收入进展：**1) 月活仍然保持较快增速，常规会员付费率稳中有升，未来用户数拓展仍有空间；2) Max功能在24Q1后陆续向日活用户放开，有望在未来季度陆续体现AI收入。

图：Duolingo 活跃用户数



图：Duolingo 付费率



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/968110125031007006>