



行业评级 推荐（维持）

报告日期 2025 年 12 月 28 日

相关研究

【兴证计算机】从火山原动力大会看 AI 投资机会-2025.12.21

【兴证计算机】Robotaxi：进入商业化落地加速阶段-2025.12.14

【兴证计算机】结构致胜，创新为矛——计算机行业 2026 年度投资策略-2025.12.09

分析师：蒋佳霖

S0190515050002
jiangjialin@xyzq.com.cn

分析师：孙乾

S0190518110001
sunqian@xyzq.com.cn

分析师：陈鑫

S0190522030001
chenxin21@xyzq.com.cn

OpenAI 深度：大模型王者，引领 AGI 之路

投资要点：

- **OpenAI 作为大模型时代的全球领航者，经过 10 年发展，持续引领大模型产业风向。**对 OpenAI 的复盘和研究，有助于更好地理解 AI 大模型产业趋势。本篇报告从 OpenAI 的历史沿革、产品布局、未来展望等维度出发，对 AI 大模型行业做了前瞻剖析。
- **复盘：AI 新物种，大模型时代的全球领航者。**1) 团队激励显效，海纳顶尖人才：OpenAI 自 2015 年以非营利机构形式成立，凭借顶尖创始团队的技术积淀奠定领先地位。近年团队规模持续扩张，但仍保持扁平化架构，2025 年 4 月员工人数已超过 5000 人，公司薪酬中位数（63 万美元）领先全美，并通过 PPU 激励机制吸引 AI 人才，涵盖大量 AI 学术和实践领军人物。2) 收入增长强劲，结构趋向多元：公司预计 2025 全年营收 130 亿美元，同比增长 2.5 倍；并预计 2028 年营收超过千亿美元，收入结构从 ChatGPT 向 API、Agents 等领域多元发展。3) 估值方面：公司融资历程丰富，融资窗口多伴随重磅 AI 模型升级、产品发布，历经 10 余轮融资后，截至 2025 年底估值升至 7500 亿美元，微软、软银、英伟达等巨头持续注资。
- **展望：模型为基，多模态、AI 应用全面发力。**1) 大模型：OpenAI 保持年度迭代、季度更新的高频模型发布节奏，GPT-5 系列在数学推理、编程等领域领跑同行，助力产品从“工具型助手”到“生产力伙伴”的跨越。2) 多模态：Sora 2 性能跨阶质变，物理模拟精度、生成一致性显著提升。迪士尼作为首个内容授权伙伴，与公司达成 10 亿美元股权投资合作，200 余个版权角色将增加用户活跃度、留存率。3) AI 搜索：ChatGPT 多端月活已经超 12 亿，用户规模长期处于行业榜首。周活用户付费率约 5%，同样优于行业水平，考虑到生产力属性的提升，后续付费转化空间仍较大。4) AI 应用：公司模型为基拓宽产品线、完成多元应用布局，Atlas AI 浏览器重塑交互模式、Codex 代码工具性能先进，API 覆盖文本、视频、图像等多场景，AI 硬件进入规划阶段。
- **愿景：强强联合，共赴 AGI 造福人类新时代。**1) 平台：以 AGI 造福人类为核心使命。OpenAI 成立之初便以实现“让 AGI 惠及全人类”为核心使命，2025 年 10 月重组后形成 Foundation（26%控股）+PBC 的双轨架构，非营利主体掌控核心决策，确保商业发展与使命对齐。2) 生态：携手巨头，聚力驱动 AI 变革。相较于谷歌、阿里巴巴、字节等互联网厂商，公司依托技术先发优势、8 亿周活的用户规模，构建差异化竞争壁垒。①算力方面，与微软、英伟达、甲骨文、亚马逊 AWS 等达成巨额合作，启动“星际之门”项目打造 AI 基建网络，同时携手博通自研定制 AI 加速器，实现软硬件垂直整合。②应用方面，2025 年 12 月推出 ChatGPT 应用商店，集成 Photoshop、Apple Music 等多款应用，为其带来链接第三方应用的能力，推动 ChatGPT 向类操作系统转型。③公司重视 B 端用户，截至 2025 年 11 月全球企业客户突破百万，ChatGPT for Work 的商业用户席位数量已突破 700 万。
- **投资建议与公司梳理：**建议持续关注国内外相关大模型公司的模型、产品进展，包括 OpenAI、谷歌、阿里巴巴、字节跳动、DeepSeek、智谱华章、Minimax 等。
- **风险提示：**1) 行业竞争加剧的风险；2) 应用进展不及预期的风险；3) 核心技术人才流动的风险。

目录

1. 复盘：AI 新物种，大模型时代的全球领航者	4
1.1、概况：十年磨剑，引领全球大模型产业	4
1.2、团队：激励显效，海纳 AI 领域顶尖人才	4
1.3、估值：持续新高，最新估值 7500 亿美元	7
2. 展望：模型为基，多模态、AI 应用全面发力	9
2.1、大模型：高频迭代，GPT5 系列领跑同行	9
2.2、多模态：两年沉淀，Sora2 性能跨阶质变	11
2.3、AI 搜索：月活居首，用户付费率持续提升	13
2.4、AI 应用：点面结合，模型为基拓宽产品线	15
3. 愿景：强强联合，共赴 AGI 造福人类新时代	20
3.1、平台：从大模型顶流，到 AGI 造福全人类	20
3.2、生态：携手行业巨头，聚力驱动 AI 新变革	22
4. 风险提示	25

图目录

图 1、OpenAI 与微软公司的合作机制	4
图 2、OpenAI 发起人及创始人	4
图 3、2025Q3 OpenAI 人才流动情况	5
图 4、OpenAI 员工的院校结构	5
图 5、OpenAI 分级别软件工程师薪资结构（万美元）	6
图 6、PPU 激励机制采用 4 年归属期安排	6
图 7、OpenAI 2024-2030E 营收情况（10 亿美元）	7
图 8、OpenAI 2028、2030 年预计收入结构	7
图 9、OpenAI 2025 年 10 月重组完成后股权结构情况	9
图 10、GPT-5 在 AIME 2025 测试中得分位居第一	10
图 11、GPT-5 在 MMMU 测评中表现突出	10
图 12、GPT-5.2 在 GDPval 评测中以 70%+ 的胜率优于或持平于行业专家	11
图 13、GPT-5.2 长多轮任务工具调用在 Tau2-bench Telecom 测试中领先	11
图 14、Sora APP 发布后登顶 Apple store	12
图 15、Sora APP 上线首周下载量高速增长	12
图 16、Sora APP 应用页面	12
图 17、Sora APP 应用使用界面	12
图 18、OpenAI 与迪士尼达成合作	13
图 19、Sora 目前的用户留存率情况	13
图 20、ChatGPT 月活问鼎 2025M10 应用+网站双榜首	14
图 21、2024-2025M10 ChatGPT 用户增长超 8 倍	14
图 22、生成式 AI 模型市场份额统计	14
图 23、企业级大模型市场份额图	14
图 24、ChatGPT 分场景使用占比情况	15
图 25、消费级 AI 行业用户付费率情况梳理	15
图 26、GPT-5 API 价格下调，性价比提升	15
图 27、AI 大模型应用为多个应用场景带来收入增长	17
图 28、通过 Atlas 实现代理购物	

图 29、 ChatGPT Atlas 浏览器页面	18
图 30、 Codex 水平领先	19
图 31、 Codex 支持多种拓展形式	19
图 32、 2019-2025M10 OpenAI 股权架构变化情况	20
图 33、 OpenAI 重组后 OpenAI Foundation 使命	21
图 34、 奥特曼描述的 OpenAI 的结构体系	21
图 35、 ChatGPT 应用（BETA）页面	24
图 36、 用户可在 ChatGPT 对话过程中调用应用	24

表目录

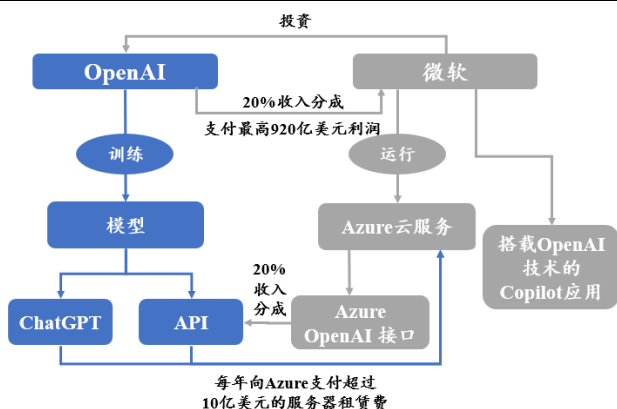
表 1、 OpenAI 核心员工离职后的创业情况	6
表 2、 OpenAI 历次融资情况梳理	8
表 3、 OpenAI 的关键模型迭代情况梳理	9
表 4、 AI 大模型颠覆多应用商业模式	16
表 5、 OpenAI 代表性 API 及应用场景	17
表 6、 AI 浏览器市场的主要产品梳理	18
表 7、 AI 编程领域的主要产品梳理	19
表 8、 国内外主流大模型厂商模型、应用及生态梳理	21
表 9、 主流 AI 大模型公司竞争力梳理	22
表 10、 OpenAI 在算力领域的重点合作梳理	23

1. 复盘：AI 新物种，大模型时代的全球领航者

1.1、概况：十年磨剑，引领全球大模型产业

OpenAI 是全球最具影响力的大模型公司之一。OpenAI 于 2015 年 12 月以非营利的研究机构形式成立，获得了包括埃隆·马斯克、里德·霍夫曼（LinkedIn 联合创始人）、彼得·蒂尔（PayPal 创始人）等多位科技及商业领袖的 10 亿美元捐赠承诺。创始人奥特曼、伊尔亚、格雷格等人具备顶尖学术背景与工程实践能力。2019 年，公司转向“封顶盈利”的商业实体模式，获得微软的 10 亿美元投资，并集成 Azure 云计算资源，为公司早期发展提供算力、资金支持。2022 年 11 月，ChatGPT 发布后 5 天内实现用户量增长 100 万，并随后快速成长为 AI 大模型领域的科技巨头，依托大模型能力构建应用生态体系。截至 2025 年 10 月，ChatGPT 周活跃用户规模攀升至 8 亿，覆盖全球成年人口的 10% 以上。

图1、OpenAI 与微软公司的合作机制



数据来源：APPSO，兴业证券经济与金融研究院整理

图2、OpenAI 发起人及创始人



数据来源：新智元，兴业证券经济与金融研究院整理

公司维持高密度的 AI 大模型、产品更新节奏。1) 模型方面：目前公司保持每 3-4 个月便有新模型发布或者升级的频率,2019 年 GPT-1 首次采用 Transformer 奠定大语言模型技术发展方向；2023 年 GPT-4 作为首个多模态大模型引领行业从单模态向多模态演进；2024-2025 年，公司 GPT 和 o 系列保持高频迭代，不断提升复杂推理、类人思考等能力。2) 产品方面：2025 年 9 月 30 日，发布 AI 视频社交应用 Sora，首周下载量高达 62.7 万次，开创 AI 社交和内容生成新范式；2025 年 10 月 22 日，发布 AI 浏览器 Atlas，以 ChatGPT 为流量分发渠道。OpenAI 多次引领大模型行业技术发展方向，并加速在多领域的商业化落地进程，成为全球 AI 发展的风向标。

1.2、团队：激励显效，海纳 AI 领域顶尖人才

OpenAI 汇聚硅谷顶尖 AI 人才，团队背景优势显著。公司早期便汇集 AI 行业精英，包括伊尔亚（Google Brain 负责人）、格雷格（Stripe CTO）等 50 多位技术专家。经过 7 年发展，OpenAI 经历了由非营利向营利模式的转变，依托具备市场吸引力的薪酬和奖励机制汇聚大量技术人才。具体来看：

1) 人数方面：由 GPT-1 的 4 名贡献者增加到 GPT-3 的 30 人，GPT-4、GPT-4o、o1 模型的贡献者合计进一步增加至 631 人；员工规模已从 2023 年的 770 人增长至 2025 年 4 月的 5328 人，主要分布于旧金山（研发中心）、纽约、伦敦（首个国际办事处）、新加坡（亚太地区战略枢纽）。已打造多个完善成熟的团队，包括后训练团队、OAI Labs、Agent 团队，以及消费级产品应用团队。

2) 背景方面：从早期多位 AI 大模型技术路线引领者至今已涵盖大量 AI 学术和实践人才，包括斯坦福大学、加州伯克利、MIT 等顶级高校，2023 年 OpenAI 的技术报告贡献者中超过 40% 有大厂工作经验。

3) 人才流动方面：2025Q3 吸纳来自传统科技巨头 Google、Meta、Apple 的精英，从 Meta 跳槽至 OpenAI 的人数多于反向流动，公司人才吸引力保持强劲。

图3、2025Q3 OpenAI 人才流动情况

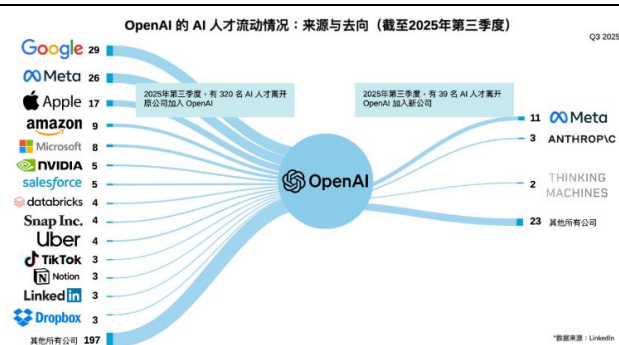
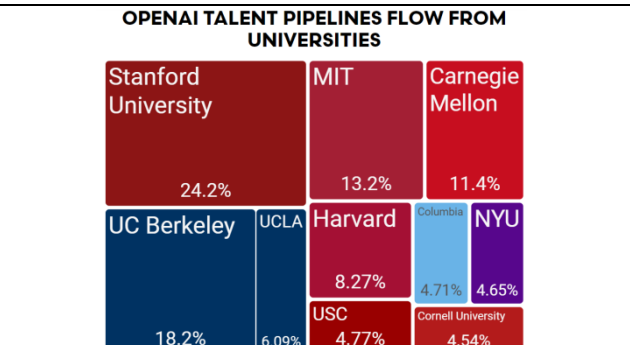


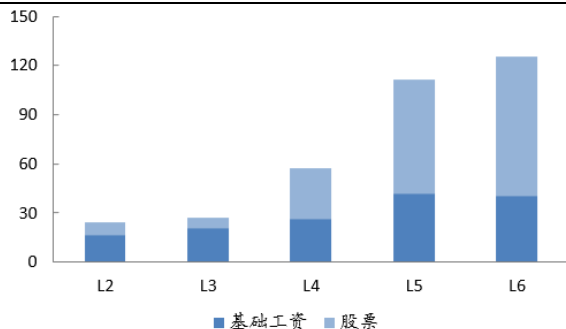
图4、OpenAI 员工的院校结构



OpenAI 的治理架构、激励机制较为特殊。1) 组织结构方面：公司技术团队采用高度扁平化的项目驱动型架构，不同项目从各部门抽调人员组成跨职能团队，减少传统的层级管理，自下而上的组织文化形成更加快速、灵活的决策和开发流程。2) 治理结构方面：2025 年 10 月 OpenAI 重组后形成了 Foundation（基金会，26%控股实体）+PBC（公益公司，负责研究、产品、商业化运营）的双轨架构，在制度层面与公司核心使命高度对齐；公司鼓励员工作为“自己的 CEO”发起项目，赋予自主权，加速最前沿创意转化。3) 激励机制方面：OpenAI 的软件工程师分为 L2-L6 共 5 个级别，每个级别工程师的薪酬都包括基本工资及股票，年薪范围从 L2 级的 24.5 万美元到 L6 级的 119 万美元不等，中位数为 63 万美元（截

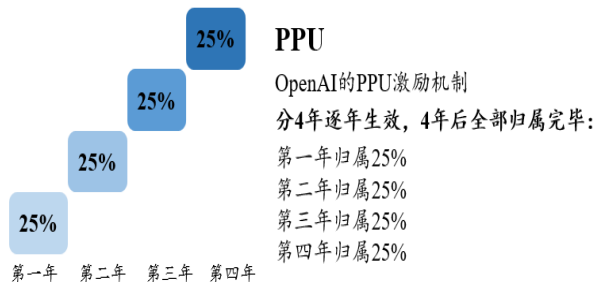
至 2025 年 12 月 27 日)，所提供薪酬居全美软件工程师前列；公司采用利润参与单位（PPUs）激励模式，将员工薪酬与公司长期盈利能力直接挂钩，L5 级 AI 工程师年化 PPUs 可达 50 万美元。

图5、OpenAI 分级别软件工程师薪资结构(万美元)



数据来源：Levels.fyi，兴业证券经济与金融研究院整理
注：数据截止 2025 年 12 月 27 日，OpenAI L2、L3、L5 级软件工程师薪资结构中有少部分奖金，金额分别为 625、1700、938 美元

图6、PPU 激励机制采用 4 年归属期安排



数据来源：Levels.fyi，兴业证券经济与金融研究院整理

“人才外溢”效应促进 AI 大模型行业共同繁荣。从 OpenAI 离职的核心员工创立了多个 AI 公司，已经有至少 7 家成为独角兽企业，聚焦 AI 安全、机器人、搜索、教育等细分领域，形成与 OpenAI 竞合交织的生态网络。例如，公司前副总裁达里奥于 2020 年底离职并携手多名 OpenAI 骨干共同创立 Anthropic，开发以安全性著称的 Claude 系列产品；前研究科学家阿拉文德在 OpenAI 工作一年后离职并于 2022 年 8 月创立 AI 搜索引擎 Perplexity；前 CSO 伊利亚于 2024 年 5 月离职后同年 6 月创立 SSI，聚焦开发“安全优先”的超级智能 AI 系统。OpenAI 的人才外溢彰显其作为 AI 产业创新巨头的辐射效应，带动行业百花齐放。

表1、OpenAI 核心员工离职后的创业情况

核心员工	在职时间	在职职位	创立公司	成立时间	公司概况
Ilya Sutskever	2015-2024	创始人、CSO	Safe Superintelligence, Inc. (SSI)	2024-06	致力于开发“安全超级智能”
Andrej Karpathy	2015-2017 & 2023-2024	创始人、研究科学家	Eureka Labs	2024-07	聚集 AI 教育
Dario Amodei	2016-2020	副总裁	Anthropic	2021	专注安全与对齐，核心产品包括 Claude 模型
Mira Murati	2018-2024	CTO	Thinking Machines Lab	2025-02	专注于构建“协作型通用智能”的 AI 研究和产品
Aravind Srinivas	2021-2022	研究科学家	Perplexity	2022-08	开发 AI 搜索引擎
David Luan	2017-2020	工程副总裁	Adept AI Labs	2022-01	开发 AI 智能体
Tim Shi	2017	技术研究员	Cresta	2017	专注于 AI 客服中心解决方案

数据来源：智东西，财联社 AI daily，IT 桔子等，兴业证券经济与金融研究院整理

1.3、估值：持续新高，最新估值 7500 亿美元

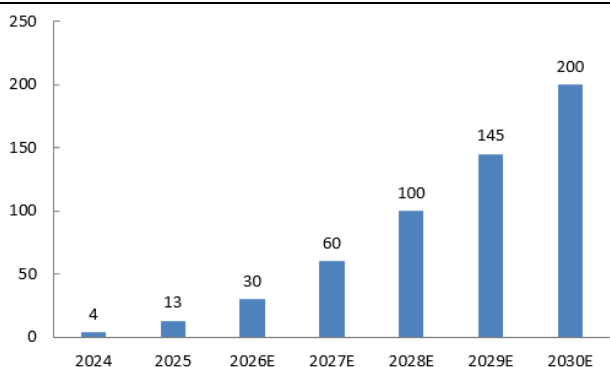
收入增长强劲，收入结构向多元化演进，预计 2028 年营收超千亿美元。具体看：

1) 营收增长迅速：公司 2023、2024 年分别营收 16、37 亿美元；同比分别增长 5714.29%、231.25%；2025H1 营收 43 亿美元，已经超过 2024 年全年营收，ChatGPT 订阅和 API 收费为核心增长引擎。

2) 业绩预期向好：据 OpenAI 预测，2025 年营收将达到 130 亿美元，为 2024 年的 3.5 倍，2028 年营收有望突破 1000 亿美元，2030 年营收将增长至 2000 亿美元。2024–2030 年复合年均增长率（CAGR）接近 92%。

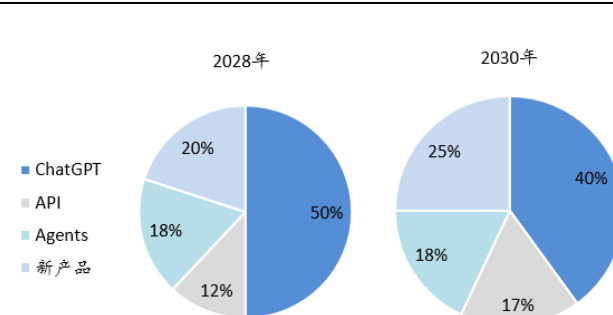
3) 收入结构多元化：目前“C 端订阅+B 端服务”结构下，公司预测 2025 年 ChatGPT、Agents、API 将分别贡献 100、20、10 亿美元收入。公司开拓以 Atlas、Apps、Sora 为入口的“免费用户变现”模式，通过将“高价值意图”（购物、法律、出行等）直连交易与抽佣闭环，让免费流量具备现金流。根据 The Information 2025Q3 预测，2028 年公司新产品业务将放量，收入占比可达 20%；2030 年 ChatGPT、API、Agents、新产品四大业务将实现均衡发展，ChatGPT 收入占比将下降至 40% 左右，新产品、API、Agents 份额预计分别为 25%、17%、18%。

图7、OpenAI 2024-2030E 营收情况（10 亿美元）



数据来源：DataLearner，兴业证券经济与金融研究院整理

图8、OpenAI 2028、2030 年预计收入结构



数据来源：DataLearner，兴业证券经济与金融研究院整理

OpenAI 经历多轮融资，估值规模提升至 7500 亿美元。从 2015 年创立至今，公司共完成 10 余轮融资，2025 年 3 月的 400 亿美元融资使估值从 1570 亿美元提升至 3000 亿美元，10 月公司完成重组后估值增长至 5000 亿美元，12 月公司估值实现 50% 跃升达到 7500 亿美元。1) 从融资节奏上看：公司的融资窗口多伴随重磅 AI 模型的升级或产品发布。2) 投资机构方面：公司与算力及硬件供应商、资本巨头整合形成良好的合作生态，微软作为 OpenAI 独家云服务与算力提供方总计投资超 130 亿美元，英伟达作为 OpenAI 核心硬件供应商成为一轮 66 亿融资的重要出资方，软银多次领投融资成为 2025 年重要大规模资本投资方。

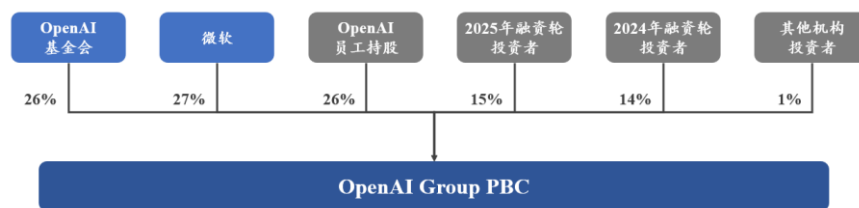
表2、OpenAI 历次融资情况梳理

时间	融资轮次/性质	估值规模 (亿美元)	窗口期内重要模型/产品	融资金额 (美元)	出资方
2015-12	非营利启动金	-	-	10 亿	马斯克、Reid Hoffman、AWS、Infosys 等
2016-08	Pre-seed	-	2016/12 开源 AGI 测试训练平台 Universe; 2018/6 发布 GPT-1	12 万	Y Combinato (著名创业孵化器和 VC)
2019-03	Seed	-	2019/2 发布 GPT-2; 3 月开源多智能体博弈环境 Neural MMO	-	Reid Hoffman、科斯拉风投
2019-04	Angel	-	2019 年 4 月发布 MuseNet	1000 万	-
2019-07	Angel+	-	2020/6 发布 GPT-3 并开放 API	10 亿	微软
2021-01	Pre-A	-	2021/1 发布多模态模型 CLIP & DALL·E	-	-
2022-01	A 轮	-	2022/1 发布 InstructGPT; 5 月 GPT-3; 11 月推出 ChatGPT	2.5 亿	-
2023-01	B 轮	-	2023/3 推出大模型 GPT-4;	100 亿	微软
2023-04	B+轮	270-290	2023/7 全面开放 GPT-4 API	3 亿	老虎全球管理、红杉资本、A16z、Thrive Capital、K2 Global、Founders Fund 等
2024-10	一轮融资	1570	2024/10 发布 sCM; 免费开放 ChatGPT Search	66 亿	Thrive Capital (领投)、微软、英伟达、软银、老虎全球基金、AltimeterCapital 等
2024-11	战略融资	-	2024/12 开放 Canvas, 发布 o1 完整版、ChatGPT Pro o1 正式版	15 亿	软银
2025-04	战略融资	3000	2025/4 发布 o3 & o4-mini, 支持图像深度思考	400 亿	软银 (75%)、微软、Coatue Management、Altimeter Capital、Thrive Capital 等
2025-08	战略融资	3000	2025/8 发布 GPT-5; 9 月发布 Sora2 以及 Sora APP; 10 月发布 AI 浏览器 ChatGPT Atlas	83 亿	老虎全球管理 (领投)、黑石、红杉资本、A16z、Thrive Capital 等

数据来源: 36 氪, PE 星球, 界面新闻, 科创板日报等, 兴业证券经济与金融研究院整理

OpenAI 基金会掌握控制权, 利益分配模式利于吸引人才。2025 年公司完成重组后, OpenAI 基金会持有营利实体 OpenAI Group PBC 26%的股权, 并获得控制权; 微软由重组前的 32.5%降至 27%, 享有分红权, 不参与公司日常决策, 并继续拥有 OpenAI 的独家知识产权和 Azure API 独家经营权; 另外 47%的股权由员工与投资者合计持有, 其中, 软银作为 2025 年主要出资方持有 11%的股份。同时, 重组后 OpenAI 取消“封顶盈利”模式, 不再设置分红上限。

图9、OpenAI 2025 年 10 月重组完成后股权结构情况



数据来源：腾讯科技，兴业证券经济与金融研究院整理

注：因数值四舍五入，百分比之和不为 100%

2. 展望：模型为基，多模态、AI 应用全面发力

2.1、大模型：高频迭代，GPT5 系列领跑同行

OpenAI 保持高频迭代，模型能力稳步提升。随着参数规模扩大和技术架构精进，OpenAI 通过高频次迭代升级优化模型能力。在迭代速度方面，从 2018 年的 GPT-1 到 2025 年 8 月的 GPT-5，公司完成了 5 次基础模型重大升级，其中衍生版本（GPT-4 Turbo、GPT-4o 等）平均以 0.5 年频次推出更新，精细化版本（GPT-4.1、GPT-5.1 等）以季度级频次实现优化升级。

表3、OpenAI 的关键模型迭代情况梳理

模型名称	发布时间	参数规模	模型特点
GPT-1	2018.06	1.17 亿	开创生成式预训练 Transformer 架构
GPT-2	2019.02	15 亿	监督多任务学习，文本生成质量提升
GPT-3	2020.05	1750 亿	重点突破小样本学习
GPT-3.5	2022-11	-	将人类反馈（RLHF）用于微调
GPT-4	2023.03	-	采用专家混合模型（MoE）技术，作为首个大规模多模态模型支持图像和文本输入
GPT-4 Turbo	2023.11	-	支持 128k 上下文窗口，开发者可通过微调创建定制版 ChatGPT
Sora	2024-02	1800 亿	初步具备物理世界模拟能力，实现文生视频
GPT-4o	2024.05	2000 亿	音频输入响应时间仅 232 毫秒（平均 320 毫秒）
o1-preview	2024-09	3000 亿	具备链式思维推理能力，擅长处理多步骤、高难度的推理问题
o1-mini	2025.09	-	输入和输出成本降低 80%，推理速度更快
o1-pro	2024-12	-	支持视觉处理、函数调用、结构化输出，兼容响应式和批量 API
o3-mini	2025-01	-	STEM 领域优化，高推理强度模式下准确率显著提升，响应速度更快
GPT-4.5	2025-02	-	对世界的理解更深入，幻觉率降至 37.1%，情商更高
o3	2025-04	-	200k 上下文窗口，支持更长的对话和复杂任务处理，深度推理能力强
o4-mini	2025-04	-	专为快速经济高效的推理而优化，擅长数学、编程和视觉任务，适合高吞吐量场景
GPT-4.1	2025-04	-	上下文窗口扩展至 100 万 tokens，在编程、指令遵循等能力上超越前代 GPT-4o
GPT-5	2025-08	52 万亿	首次集成多模态与推理能力于一体化系统，幻觉率显著降低至 o3 的六分之一；内置“实时路由器”，为不同任务匹配合适模型

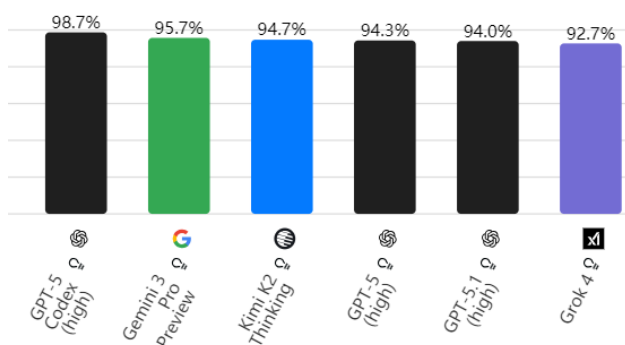
Sora2	2025-09	5000 亿	能精准模拟物体运动和物理规律，首次实现视频画面、环境音效、动作声效及人物口型与对白的精准同步，无需后期配音
GPT-5.1	2025-11	-	强化情绪价值供给与个性化交互体验，引入自适应推理功能
GPT-5.2	2025-12	-	专业能力大幅提升，在金融、法律、医疗、工程等专业任务评估中胜率达 70% 以上，效率快于人类专家 11 倍，幻觉率较前代减少 38%

数据来源：OpenAI 官网，新智元，第一财经等，兴业证券经济与金融研究院整理

在参数规模和模型能力方面：1) 参数规模：GPT-3 从 1.17 亿参数 (GPT-1) 增长至 1750 亿参数，深度语义理解能力提升；GPT-4 参数高达 1.76 万亿，实现多模态能力；GPT-5 参数跃升至 52 万亿，泛化和推理能力增强；2) 模型能力方面：从初代的通用生成逐步演进为具备深度推理、智能交互、多模态等能力的模型，在专业性、准确性、速度、幻觉率等方面有所优化。

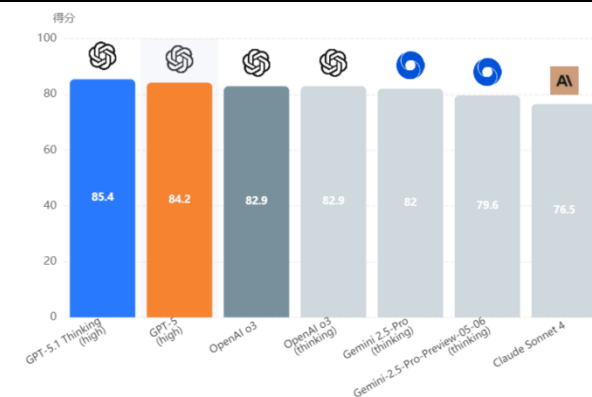
GPT-5 领跑多项榜单。GPT-5 首次将多模态与推理能力集成于一体化系统，实现多维性能突破。1) 数学推理能力：2025 年 8 月，GPT-5-pro 在 AIME2025 (竞赛数学) 基准测试中获得 100% 成绩，在 FrontierMath - Tier 4 测评中以 13% 的得分问鼎；2) 编程能力：2025 年 9 月，GPT-5 在软件工程基准 SWE-BENCH PRO 测试中以 23.3% 的解决率登顶，并在网页开发测试中超越 Gemini-2.5-pro 和 Claude-Opus-4 成为冠军；3) 多模态能力：GPT-5 以 84.2% 和 84.6% 的得分登顶 MMMU (学院水平视觉问题) 和 VideoMMMU (基于视频的多模态推理) 榜单，相较 o3 (2025 年 4 月, 82.9%、83.3%)、Gemini-2.5-pro (2025 年 6 月, 82.0%、83.6%) 能力突出；4) 幻觉率方面：在开放题 prompt、医学对话、ChatGPT 实际用户问题三组场景中，GPT-5 的错误率大幅降低，其中医疗对话错误率下降至 1.6%，日常问答场景错误率从 22% 降到 4.8%。

图10、GPT-5 在 AIME 2025 测试中得分位居第一



数据来源：Artificial Analysis，兴业证券经济与金融研究院整理

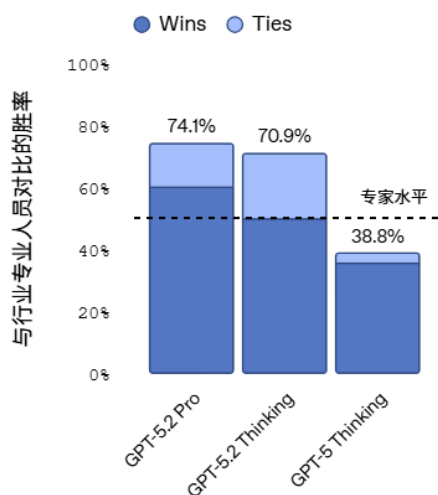
图11、GPT-5 在 MMMU 测评中表现突出



数据来源：DataLearner，兴业证券经济与金融研究院整理

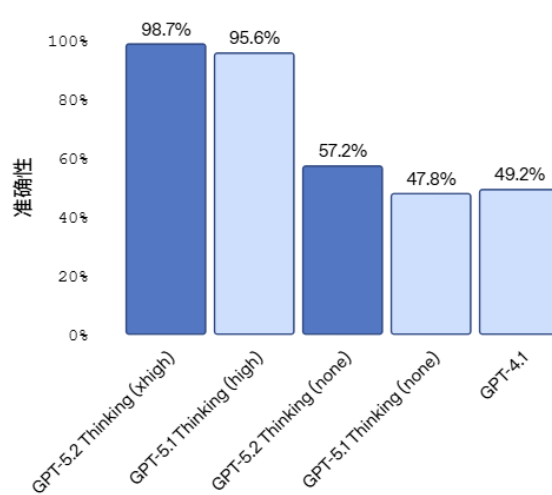
GPT-5.2 发布，重点提升专业知识能力。OpenAI 于 12 月推出为专业知识型工作打造的 GPT-5.2 系列模型，该模型在通用智能、超长文本理解、Agent 工具调用以及视觉能力上有所升级。**1) 专业任务处理能力：**在 GDPval 任务测评中，GPT-5.2 Thinking 成为 GPT 系列首个达到人类专家水平的模型，在 70.7% 的高难度知识型工作任务上表现优于或持平于行业顶尖专家，且速度 11 倍快于专业人士，而成本仅约为 1%；模型从工具型辅助向智能协作伙伴转变，为各行业的工作流程优化、生产力提升和人机协作提供支持。**2) 更类人的知识泛化能力：**GPT-5.2 Pro 成为首个突破通用推理能力的基准测试 ARC-AGI-1 90% 门槛的模型，在抽象思维、模式识别和泛化能力方面更加接近人类在类似任务中的表现水平。

图12、GPT-5.2 在 GDPval 评测中以 70%+ 的胜率优于或持平于行业专家



数据来源：OpenAI 官网，兴业证券经济与金融研究院整理

图13、GPT-5.2 长多轮任务工具调用在 Tau2-bench Telecom 测试中领先



数据来源：OpenAI 官网，兴业证券经济与金融研究院整理

2.2、多模态：两年沉淀，Sora2 性能跨阶质变

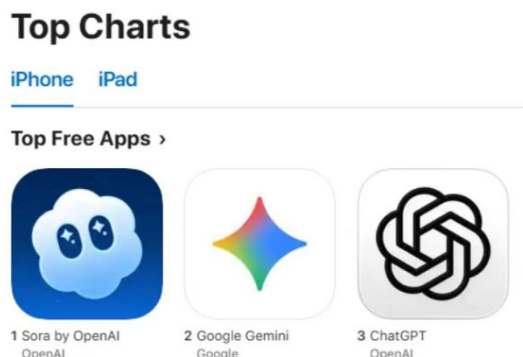
两年沉淀多模态能力，**Sora2 实现跨阶提升。**OpenAI 自 2024 年 2 月正式披露初代文生视频模型 Sora 后，历经近 2 年沉淀，于 2025 年 9 月发布全新模型 Sora2，并同步推出 AI 社交应用 Sora App。具体来看：

1) 模型能力方面：**①遵循物理规律：**Sora 2 能模拟刚体碰撞、流体力学及布料动力学等复杂物理现象，物理模拟精度有所提升；**②生成一致性提升：**Sora 2 可建立帧间关联，动作连贯性得到突破，物理模拟的一致性水平较高；**③可生成同步音频：**Sora 2 可以基于文本指令直接生成同步音频。

2) 应用流量方面：Sora App 作为全部由 AI 生成视频内容的社交平台，创作门槛大幅降低。Sora 于 10 月 3 日登顶苹果美国“热门免费应用”榜单，首周收获 62.7

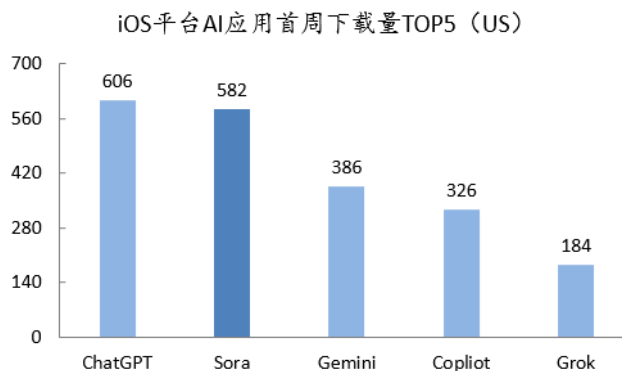
万下载量，超过 ChatGPT 首周表现（60.6 万下载量）。截至 10 月 31 日，Sora 下载量已攀升至 400 万次，日均生成数百万条 10 秒 AI 视频。

图14、Sora APP 发布后登顶 Apple store



数据来源：财联社，兴业证券经济与金融研究院整理

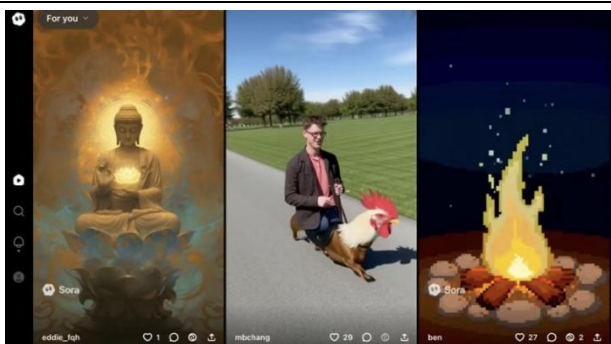
图15、Sora APP 上线首周下载量高速增长



数据来源：量子位，兴业证券经济与金融研究院整理
注：下载量数据单位为千次

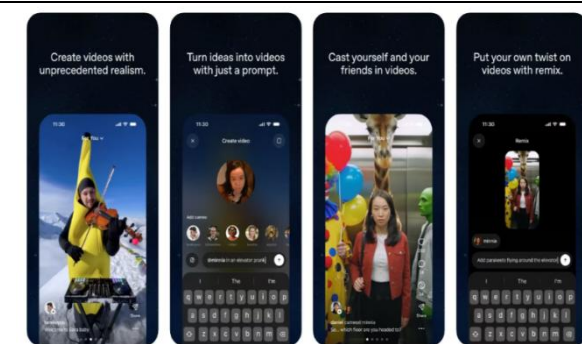
Sora 具备强社交属性，加速 C 端应用场景落地。1) 功能上：Sora 提供 Cameo（客串）、Character Cameo（角色客串）、二次创作（Remix）和社区排行榜功能，从传统文生视频模型升维至创意短视频社交应用，以更具创意和想象力的社交理念激活用户创作与互动的意愿，市场潜力巨大；2) 商业化模式上：Sora 与 ChatGPT 收费结构相似，C 端采取“免费生成+付费额度”模式，同时面向企业和创作者开放 API，从生产工具转向 AI 社交平台，推动 AI 视频生成行业从生产力属性升级至视频方案产品，加速 C+B 端应用商业化。

图16、Sora APP 应用页面



数据来源：量子位，兴业证券经济与金融研究院整理

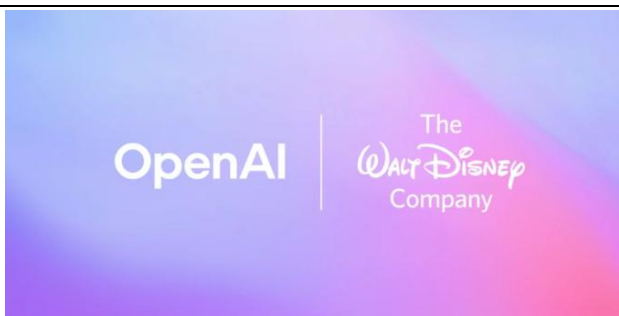
图17、Sora APP 应用使用界面



数据来源：AI 寒武纪，兴业证券经济与金融研究院整理

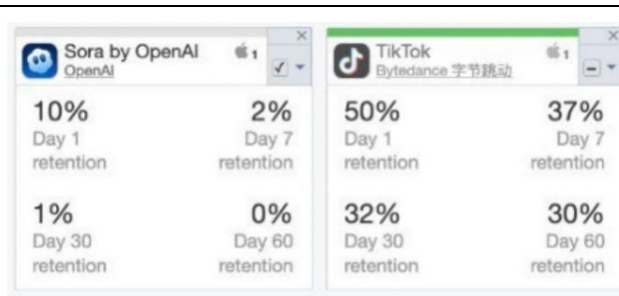
牵手迪士尼，Sora 创作生态的商业价值提升。据量子位，迪士尼计划向 OpenAI 进行 10 亿美元的股权投资，成为 Sora 的首个内容授权伙伴。预计从 2026 年初开始，Sora 用户可基于迪士尼旗下 200 余个版权角色生成创意视频内容。此次授权远超简单的素材库价值，迪士尼的热门 IP 将为 Sora 平台注入创作吸引力，满足用户对喜爱角色的互动需求，加强 Sora 作为社交平台的情感连接，有望提升 Sora 的用户活跃热度和留存率。

图18、OpenAI 与迪士尼达成合作



数据来源：量子位，兴业证券经济与金融研究院整理

图19、Sora 目前的用户留存率情况



数据来源：量子位，兴业证券经济与金融研究院整理
注：数据为 2025 年 10 月 1 日至 11 月 14 日期间统计的 Sora vs Tiktok 用户留存

2.3、AI 搜索：月活居首，用户付费率持续提升

ChatGPT 月活持续快速增长。2025 年 10 月，ChatGPT 应用端以超过 8 亿的月活跃用户稳居全球第一，加上网站端用户，ChatGPT 已拥有超过 12 亿的月活用户。在 token 调用量方面，据官方介绍，2025 年 10 月初，公司 API 每分钟处理近 60 亿 token；2025 年 ChatGPT token 消耗达 6425 万亿，同比 2024 年增长超 16 倍。具体来看：

1) ChatGPT 用户规模长期占据榜首：2023 年 1 月，ChatGPT 月活突破 1 亿，成为历史上增长最快的 C 端 app，后续快速拉开与相关竞品用户规模的差距，截至 2025 年 10 月，ChatGPT 应用端月活已破 8 亿，相较于 Gemini（应用端 3.3 亿）、Perplexity（网站端 5 亿）等公司具有规模优势。

2) 模型升级吸引用户使用并提升粘性：2024 年 3 月，多模态模型 GPT-4 发布当月月活增长 54.21%，月独立访客数量增加 8.48 亿；5 月 GPT-4o 实现多模态无缝交互能力革新后月活从 4 月的 3.87 亿升至 4.72 亿，2024 全年月活规模增长 103.88%；2025 年 4 月 GPT-4.1 凭借性价比优势吸引价格敏感和轻量级用户，月活从 3 月的 7.88 亿突破至 4 月的 8.9 亿；GPT-5 多榜单领跑的强劲技术优势驱动下，8 月双端月活突破 11 亿。

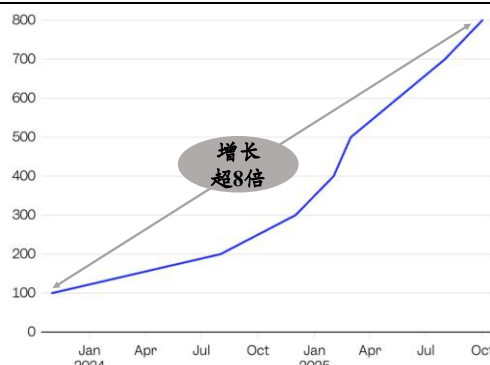
3) 潜在的用户规模仍然广阔：谷歌多个巨大流量入口依托 Gemini 嵌入 AI 功能，Chrome 凭借浏览器市场长期近 70% 的强势主导份额，2025 年 7 月 AI Overviews 功能月活已达 20 亿；Workspace 全球用户超 30 亿，已全面集成 Gemini AI 助理；多渠道推动下，Gemini 用户增速超过 ChatGPT。OpenAI 当前用户增长主要依赖 ChatGPT 这一核心平台，后续将向更多应用场景发力，公司预测 2030 年周活可达 26 亿，随着公司加速拓展社交、搜索等入口，将从单点应用升级为“超级助手”平台，整体用户规模增长潜力依旧强劲。

图20、ChatGPT 月活问鼎 2025M10 应用+网站双榜首

应用端				
排名	产品	分类	月活 (万人)	环比
1	ChatGPT	聊天机器人	80,282	1.36%
2	百度AI搜索	智慧搜索	63,470	3.05%
3	Gemini	聊天机器人	32,557	4.90%
4	豆包	聊天机器人	22,622	5.17%
5	Deepseek	聊天机器人	15,719	1.85%
网站端				
排名	产品	分类	月活 (万人)	环比
1	ChatGPT	聊天机器人	48,377	5.43%
2	Gemini	聊天机器人	20,643	-0.61%
3	Perplexity	智慧搜索	4,945	62.88%
4	Deepseek	聊天机器人	4,179	2.68%
5	Grok	个人助理	3,207	27.58%

数据来源：AI 百强榜，兴业证券经济与金融研究院整理

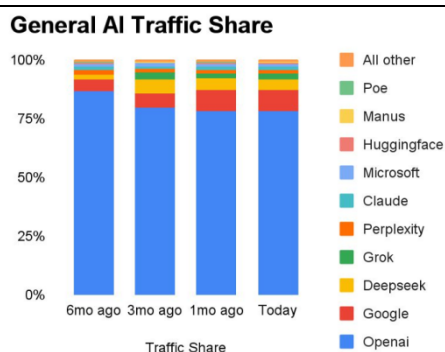
图21、2024-2025M10 ChatGPT 用户增长超 8 倍



数据来源：GPLP 科技说，兴业证券经济与金融研究院整理

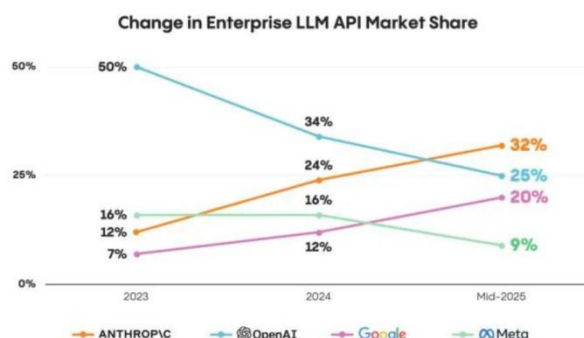
大模型行业竞争激烈,OpenAI 具有领先优势。1) 生成式 AI 市场: 根据 Similarweb 数据, 截至 11 月 13 日, ChatGPT 以 72.3% 的流量份额强势占据全球生成式 AI 模型市场, 相比 2024 年减少 14.3pcts, 主要由于用户需求从“聊天式交互”向多维场景 (写作、低成本 API、搜索) 迁移, 细分需求紧密增加下流量有所分散; 2) 企业级 API 市场: 根据 Menlo Ventures 数据, 公司市占率由 2024 年的 34% 下降至 2025 年中期的 25%, 位居第二, 仅次于 Anthropic, 随模型设计和定价策略优化, B 端 API 市场份额有望回升; 3) 用户粘性方面: 根据 BOND 2025 年 5 月的研究, ChatGPT 每天搜索量达到 10 亿次, 用户每天在 ChatGPT 应用上花费的时间比 21 个月前增加 3 倍, 用户粘性维持在较高水平。

图22、生成式 AI 模型市场份额统计



数据来源：Similarweb，兴业证券经济与金融研究院整理

图23、企业级大模型市场份额图

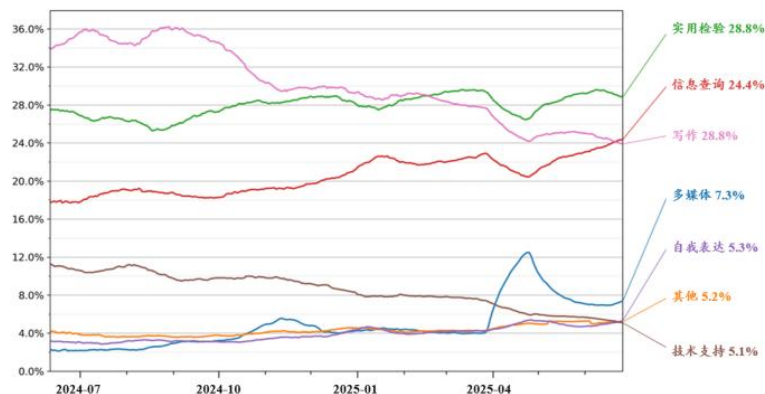


数据来源：Menlo Ventures，兴业证券经济与金融研究院整理

使用维度方面, OpenAI 非工作类信息处理需求广泛。据 2025 年 9 月数据, ChatGPT 处理的非工作类信息由 2024 年 6 月的 53% 增长至 2025 年 6 月的 70%,

其中实用建议、信息查询、文书写作等日常场景占比 80%，ChatGPT 在改善日常生活方面的作用日益凸显。

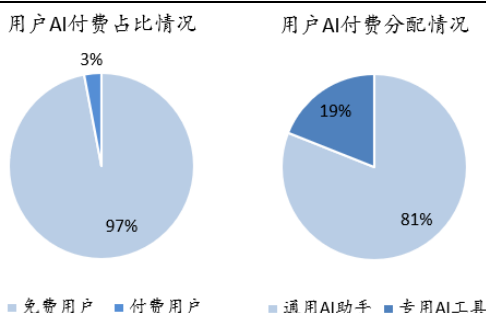
图24、ChatGPT 分场景使用占比情况



数据来源：OpenAI、杜克大学、哈弗大学：《How People Use ChatGPT》，兴业证券经济与金融研究院整理 注：按照 2025 年 7 月截止的调查数据

用户付费意愿稳步提升，潜在付费转化空间较大。1) 个人付费：目前，OpenAI 8 亿周活跃用户中约 5%为付费用户，优于消费级 AI 行业水平（3%付费率），公司预计 2030 年付费率可达 8.5%；与微软 Office 365 的 9 个应用组成生产力套件的 4.5 亿付费用户数相比，ChatGPT 凭借单个应用在 3 年内积累超 3500 万付费用户，随着 OpenAI 横向应用拓展以及实务功能的增加，订阅用户转化空间仍较大。2) 企业付费：截至 2025 年 6 月，OpenAI 付费企业用户已突破 300 万，相比 2025 年 2 月的 200 万数据实现快速增长。3) 性价比提升：GPT-5 实时路由器设计降低推理成本，主模型输入端价格由 2.5 美元/百万 token（GPT-4o）下调至 1.25 美元/百万 token，定价较 Gemini 3 Pro（2 美元/百万 token）等模型有一定优势。新的定价策略下，用户付费意愿预计将继续提升。

图25、消费级 AI 行业用户付费率情况梳理



数据来源：Menlo Ventures，兴业证券经济与金融研究院整理 注：数据为 Menlo Ventures 2025 年中期统计数据

图26、GPT-5 API 价格下调，性价比提升

MODEL	INPUT	CACHED INPUT	OUTPUT
gpt-5.1	\$1.25	\$0.125	\$10.00
gpt-5	\$1.25	\$0.125	\$10.00
gpt-5-mini	\$0.25	\$0.025	\$2.00
gpt-5-nano	\$0.05	\$0.005	\$0.40
gpt-5.1-chat-latest	\$1.25	\$0.125	\$10.00
gpt-5-chat-latest	\$1.25	\$0.125	\$10.00
gpt-5.1-codex	\$1.25	\$0.125	\$10.00
gpt-5-codex	\$1.25	\$0.125	\$10.00
gpt-5-pro	\$15.00	-	\$120.00
gpt-4.1	\$2.00	\$0.50	\$8.00
gpt-4.1-mini	\$0.40	\$0.10	\$1.60
gpt-4.1-nano	\$0.10	\$0.025	\$0.40
gpt-4o	\$2.50	\$1.25	\$10.00

数据来源：OpenAI 官网，兴业证券经济与金融研究院整理

2.4、AI 应用：点面结合，模型为基拓宽产品线

AI 大模型正在重构传统应用及其商业模式。1) **浏览器领域**: AI 大模型直接提供结构化回答并完成任务辅助, 替代传统浏览器繁琐的查找与验证流程; 2) **视频、图像编辑领域**: AI 大模型的加入降低了 PS、AE、PR 应用的专业化门槛并提升处理质量; 3) **电商购物领域**: AI 大模型的引入压缩传统电商漫长比价链路, 流量匹配效率、推荐信息流点击率得以提升; 4) **客服领域**: AI 智能客服已经大规模替代传统人工客服, 降低运营成本并提高客户满意度。AI 大模型已从“工具性赋能”迈向“结构性替代”, 商业应用价值尽显。

垂类 Agent 正进入加速商业化阶段。大模型已经在多维日常领域展现 Agent 能力, Perplexity 的 Buy with Pro、亚马逊的 Buy for Me、OpenAI 的 Operator Agent 通过完全 AI 托管的服务实现自动加购和付款的全套购物操作。在传统需求端, AI 大模型正在深刻重构交互范式与决策模式, 开辟全新商业空间。

表4、AI 大模型颠覆多应用商业模式

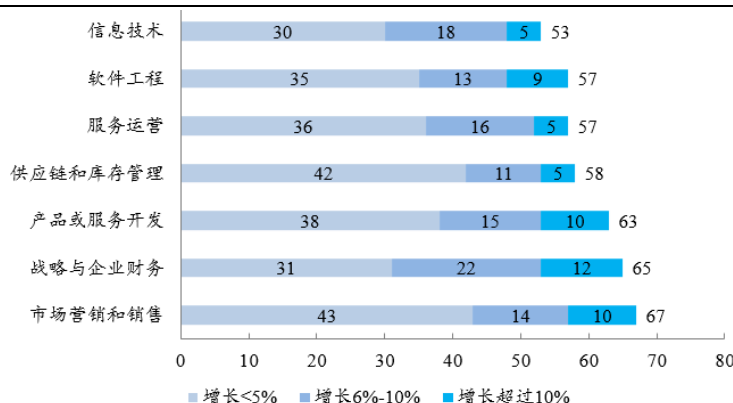
应用	大模型赋能	相关案例
浏览器	直接提供结构化的答案、完成任务代理或辅助, 跳过传统浏览器繁琐的搜索、筛选与验证环节, 重塑信息获取的场景	Fellou 处理复杂任务的速度相比手动操作提升 5.2 倍; Atlas 浏览器“代理模式”功能自动完成购物任务; 谷歌 AI Overviews 曝光量增长 49%, 同时网站点击量下降 30%
视频/图像编辑	将创意思图转化为视觉呈现, 摆脱传统编辑依赖人工逐帧操作与专业软件技巧	Adobe 自研 Firefly Image5 模型可生成 400 万像素高清图像并支持分层式提示编辑, ElevenLabs 模型可生成高质量旁白
电商购物	提供个性化的商品推荐和“购物代理”, 为商家提供流量输送和转化, 将传统电商中“搜索-比价-决策”的链路压缩为自然对话	阿里巴巴引入大模型后商品推荐信息流点击率增长 10%+, 带动购买效率提升 25%, 广告环节商家 ROI 提升 12%; “万相台 AI 无界”为帮助品牌实现 ROI 提升 15%+
客服系统	精准识别用户意图、处理复杂多轮对话, 提供个性化解决方案, 解决传统客服有限知识库与被动响应问题, 降低人力依赖	淘宝店小蜜 AI 客服使店铺的用户满意度从 90%左右提升到 96%+, 转化率比人工提升 4 个百分点; 京小智 5.0 使店铺智能客服的解决率提升 14.8%, 满意度提升 5.7%
编程开发	根据自然语言描述快速生成代码片段、函数、完整模块, 降低编程门槛、提高开发效率	84% 的开发者已在日常工作中使用 AI 编码助手, Copilot 使开发者工作满意度提高 75%, 代码编写效率提升 55%
语言翻译	通过海量语料训练所获得的深层语义理解与生成能力, 突破传统翻译对规则与语法的依赖	DeepL 采用深度神经网络技术, 服务全球超过 10 万家企业客户, 平均翻译正确率超过 95%

数据来源: 新智元, 电商派, 电商头条, 微软科技等, 兴业证券经济与金融研究院整理

AI 大模型的生产力工具属性持续强化, 并加速渗透。1) **生产能力方面**: 根据 OpenAI 研究, AI 大模型在政府部门、零售和批发上的能力已经达到或超越人类水平, 同时显著提升使用者的工作速度或体验, 帮助用户每天节省 40-60 分钟, 深度使用者每周节省可超 10 小时。2) **生产应用方面**: 大模型行业已然技术验证阶段迈向规模化价值创造阶段, 据 Artificial Analysis 调研显示, 45%的企业已将 AI 大模型部署到生产环境中, 活跃支持工程研发、客户支持和营销, 并为多应用

场景带来收入增长；**3）生产需求方面**：2025 年上半年中国大模型累计中标项目达 1810 个，总金额突破 64 亿元，从辅助工具跃升为核心生产力，深度渗透政务、金融、制造、医疗等实体经济领域。

图27、AI 大模型应用为多个应用场景带来收入增长



数据来源：麦肯锡，兴业证券经济与金融研究院整理

注：图中数据为麦肯锡 2025 年 6 月 25 日至 7 月 29 日间，对共计 1993 名参与者调查统计得出的，各职能因使用 AI 而产生不同幅度营收增长的受访者比例；因四舍五入，各百分比之和可能不为 100%

API：OpenAI 持续升级 API，重视开发者生态建设。目前公司 API 模型矩阵已经实现文本（GPT 系列）、视频（Sora 2）、图像（GPT-Image）、Agent（AgentKit）、代码（CodeX）、语音（GPT-Realtime-Mini）等多维能力覆盖，为企业和应用生态提供了更完整的 AI 基础设施。完善 API 布局是 AI 原生应用商业化落地的重要抓手，办公领域涌现自动文档生成、客服行业被语音级 Agent 重塑、影视行业因 Sora 出现低成本专业级视频制作、编程行业因 CodeX 能力实现自动化软件工程，OpenAI 多 API 模型能力渗透到多行业信息工作流，加速大模型商业化落成。

表5、OpenAI 代表性 API 及应用场景

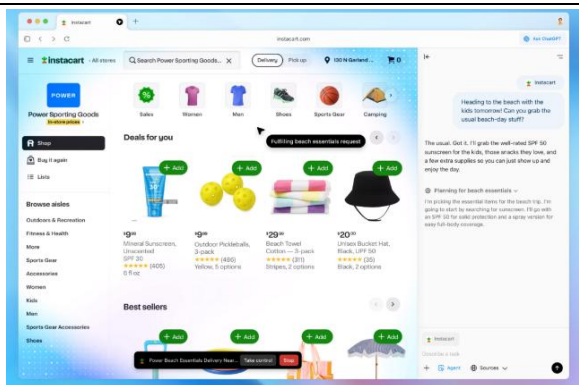
模型类别	代表性 API	典型应用场景
文本推理	GPT-5 Pro、GPT-4.0 Turbo、GPT-4o	高级智能问答、知识助手、复杂逻辑推理、代码解释、数学推导、大规模自动化内容生产
视频生成	Sora2	广告、短视频、影视分镜生成、产品展示与营销素材自动化、教学/培训视频、场景概念设计
图像生成	GPT-image-1	平面设计、品牌视觉设计、电商商品图、模型图、场景图生成
Agent	Agent Builde、ChatKit	智能化工作流、自动化智能体助手、搭建业务代理流程
代码生成	CodeX	自动生成代码、重构、注释添加、快速构建原型应用、软件开发
语音生成	GPT-Realtime	MCP 服务器、SIP 电话呼叫、实时语音对话助手、内容创作者语音合成、AI 语音智能体

数据来源：36 氪，机智流，Founder Park 等，兴业证券经济与金融研究院整理

AI 浏览器：发布 ChatGPT Atlas，布局 AI 浏览器赛道。2025 年 10 月 22 日，OpenAI 正式发布 Atlas，深度内嵌 ChatGPT 能力，融合 ChatGPT 侧边栏、浏览

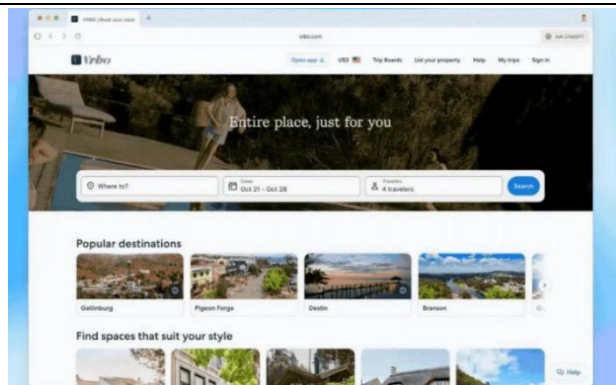
器记忆、智能体代理等诸多 AI 功能。以“AI 优先”的理念将浏览器本身重塑为 ChatGPT 的延伸和操作界面，改变传统浏览器“搜索框+地址栏”的交互模式，转向以对话和任务为导向的代理工具。

图28、通过 Atlas 实现代理购物



数据来源：第一财经，兴业证券经济与金融研究院整理

图29、ChatGPT Atlas 浏览器页面



数据来源：腾讯科技，兴业证券经济与金融研究院整理

商业模式上，浏览器核心功能免费，发布初期“代理模式”仅限于付费的 ChatGPT Plus、Pro 和 Business 订阅用户，未来有望依托 ChatGPT 巨大分发渠道实现用户规模转化。11 月 19 日，Atlas 进行九项升级，包括支持 iCloud 通行密钥、简化用户浏览器迁移插件流程、新增垂直标签等实用属性，降低传统浏览器的迁移门槛，AI 能力与日常浏览深度融合，提升用户使用舒适度。

表6、AI 浏览器市场的主要产品梳理

产品	公司	近期发布&更新	AI 搜索功能特点
ChatGPT Atlas	OpenAI	2025-10-21	深度整合 ChatGPT，侧边栏实时总结网页、对比产品、分析数据；“代理模式”可代表用户自主完成复杂任务。
Gemini in Chrome	谷歌	2025-9-19	地址栏直接进行对话式搜索；跨标签页操作，可对比和总结多个网页信息；深度集成谷歌生态（如 YouTube、日历）。
Edge Copilot	Edge Bing	2025-7-29	长期记忆（Memory）：记住聊过的事实，并在未来对话中调用；真实对话模式（Real Talk）：富有同理心和支持性的 AI 助手。
Perplexity Comet	Comet	2025-7-9	标签自动整理：AI 可根据内容、类别自动整理浏览器标签；自动生成内容：可根据需求生成网站、写邮件、制定学习计划等。
豆包	字节跳动	2024-12-14	将 AI 分析功能内嵌入 b 站等视频网站，实现“边看边解读”的 AI 沉浸式分析视频场景。网页阅读时可以随时询问 AI，写作方面主打人机协作模式，支持以对话的方式来修改文章。
百度	百度	2025-7-2 改版	侧重通用知识检索和多模态交互（文本、图像、语音），服务于大众信息获取需求。
Fellou	Fellou AI	2025-5-11	Deep Search（深度搜索）+Visual Report（可视化报告生成）；任务自主规划与执行，深度集成与自动化 workflow。

数据来源：科创板日报，硅星人 Pro，极客科技 GeekTech 等，兴业证券经济与金融研究院整理

AI 编程: Codex 作为云端软件工程智能体, 团队管理与协作工具成熟。早期 CodeX 基于 GPT-3 演化为代码补全工具, 2025 年 5 月 OpenAI 发布研究预览版 CodeX, 以 o3 为底层模型成为云端软件工程智能体; 2025 年 10 月, Codex 正式发布, 开放 Slack 集成、Codex SDK、管理 Codex 环境等功能, 覆盖软件开发全生命周期。思科、乐天已经将 Codex 引入其开发工作流。

表7、AI 编程领域的主要产品梳理

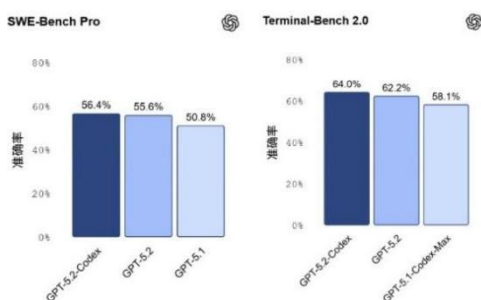
产品名称	公司	产品/模型特点	性能指标
Codex	OpenAI	任务在云端沙盒环境中运行; 集成 Codex CLI、IDE 插件、云端服务等多种工具	GPT-5-Codex SWE-bench 评分 74.5%; 可独立连续工作 7 小时并产生自动化 PR
GitHub Copilot	GitHub	与 VS Code 等 IDE 深度集成, 插件生态丰富, 支持 30+ 语言, 代码补全响应迅速	平均响应时间<200ms, 单元测试生成成功率 92%
Cursor	AnySphere	支持最多 8 个智能体并行工作, Composer 模型生成速度极快, 长上下文处理能力强	Composer 模型 30 秒可完成复杂任务 (比同类快 400%), 支持百万行级项目
Claude Code	Anthropic	终端 AI 编程工具, 内置强大的工具系统 (如文件操作、执行 Shell 命令), 代码简练	Claude Sonnet 4.5 SWE-bench 评分 77.2%, 代码准确率 82%
Qoder	阿里巴巴	中文理解和生成能力出色, 在电商场景接口开发中优势明显	检索召回率领先业界标杆 12%, 代码生成准确率领先 13%
CodeBuddy	腾讯	集成微信小程序开发和腾讯云 API, "Craft 智能体"可将 PRD 转为多模块代码	设计稿转 React 组件还原度 95%+, 等保三级 2.0 认证

数据来源: 量子位, Qoder 等, 兴业证券经济与金融研究院整理

注: GPT-5-Codex SWE-bench 评分 74.5%为 2025 年 9 月 16 日 OpenAI 官方发布数据, Claude Sonnet 4.5 SWE-bench 评分 77.2%为 2025 年 9 月 29 日 Anthropic 官方发布数据

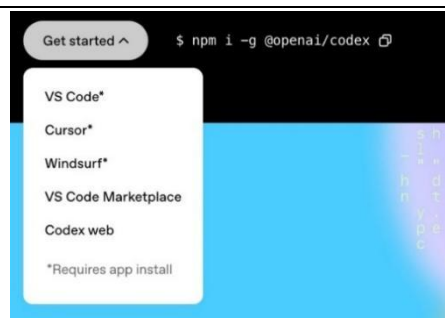
Codex 模型性能提升, 驱动份额向上。GPT-5-Code 在 SWE-benchVerified 评分高达 74.5%, 驱动 CodeX 在 AI 编程工具市场份额较模型发布前提升 5pcts (占据 24.9%)。GPT-5.2-Codex 较前代实现约 6%的性能提升, 在长时间、跨步骤的编程任务中实现稳定的端到端工作流, 对大规模代码库的重构、迁移和功能构建支持更加成熟, 推动 AI 编程“软件工程”能力增强。

图30、Codex 水平领先



数据来源: OpenAI 官网, 兴业证券经济与金融研究院整理

图31、Codex 支持多种拓展形式



数据来源: 新智元, 兴业证券经济与金融研究院整理

AI 端侧: 战略布局智能硬件, 打开新增长空间。1) 产品规划: 公司硬件产品将定位于强调语音交互和情境感知的 AI 伴侣, 计划涵盖无显示屏智能音箱、可穿戴徽章 (AI Pin)、智能眼镜和数字录音笔等, 首批产品预计在 2026 年末或 2027 年

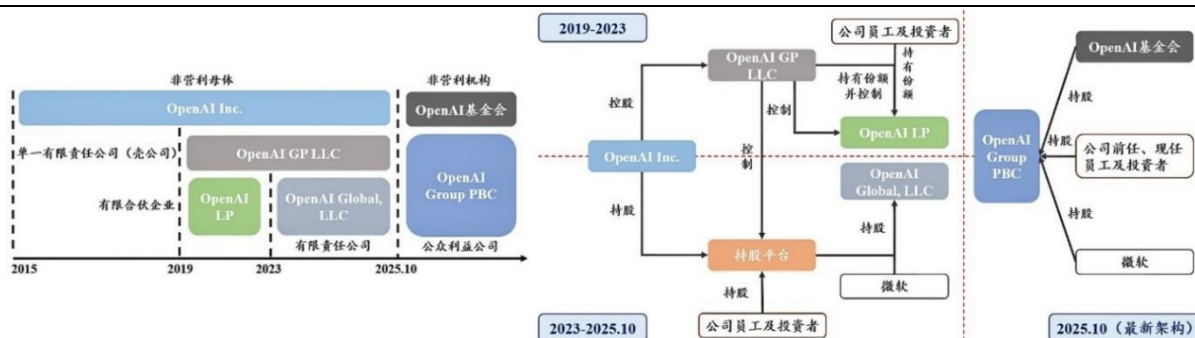
初发布。**2) 合作伙伴**：2025 年 5 月，OpenAI 以 65 亿美元收购前苹果首席设计官创立的 io Products 公司，获得工业设计领域核心能力；9 月，OpenAI 宣布将与立讯精密共同开发消费级 AI 硬件产品；同月，OpenAI 联合歌尔股份以寻求供应扬声器等零部件。2025 年 11 月 18 日，首批 AI 硬件已成功完成原型设计，OpenAI 有望依托强软件优势，抢占端侧 AI 流量入口，打造新的成长曲线。

3. 愿景：强强联合，共赴 AGI 造福人类新时代

3.1、平台：从大模型顶流，到 AGI 造福全人类

OpenAI 坚持“AI 赋能于人”的理念，并基于此构建模型及产品的底层逻辑。2015 年公司成立起便将“实现并安全推动通用人工智能（AGI）惠及全人类”设定为根本目标，2025 年 10 月完成公司组织结构重组，成立 OpenAI Foundation（非营利基金会）持有 OpenAI Group（公共利益公司，PBC）约 26% 股权，以“公共事业体”形式，通过资源分配推动“AI 造福全人类”的目标。在此愿景下，OpenAI 垂直整合 AGI 战略，工作围绕 AI 产品服务、前沿模型研究实验室、超大规模基础设施三大核心支柱展开，打造超级平台，并相信超级智能体将于 10 年内到来。

图32、2019-2025M10 OpenAI 股权架构变化情况



数据来源：36 氪，兴业证券经济与金融研究院整理

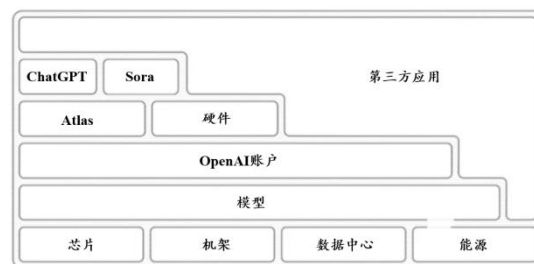
非营利主体绝对控制权一以贯之，践行 AGI 使命。2015 年 OpenAI 成立至 2025 年 10 月公司资本结构重组前，非营利组织 OpenAI Inc. 始终担任管理角色；2019 年起，非营利组织通过普通合伙人（GP）的角色控制营利性实体，不直接参与利润分配，对公司的治理、决策、战略规划有最终决定权；公司完成重组后，非营利主体改名为 OpenAI Foundation（基金会），依旧不以获取经济回报为目的继续掌控营利实体，设立 5000 万美元的“以人为本人工智能基金”，旨在推动 AI 技术发展的同时优先考虑人类福祉。另外，重组后员工持股比例约为 26%，成为公司重要的利益相关方，员工原有的 PPU 权益预计将逐步转化为普通股形式，与公司价值增长实现更紧密管理，共同承担推动 AI 技术发展的责任。

图33、OpenAI 重组后 OpenAI Foundation 使命



数据来源：硅星 GenAI，兴业证券经济与金融研究院整理

图34、奥特曼描述的 OpenAI 的结构体系



数据来源：AI 寒武纪，兴业证券经济与金融研究院整理

在 AGI 使命指引下，打造 OpenAI 应用生态。1) 新产品推出、新版本迭代的速度快：以 AGI 为目标下，技术团队工作节奏快且氛围严肃，Codex 仅用 7 周便完成了从第一行代码到上线，实现模型及产品的高质量高频迭代；2) 模型开放程度高：公司最先进的模型并不局限于某个需要年度协议的企业，API 效率高，供开发者和初创企业使用；3) 公司积极打造 AI 用户生态：公司以通用 AI 平台为目标打造塑造其应用生态，使得各类用户可以通过 OpenAI 的消费产品、API、专用硬件等使用 OpenAI 的技术及服务，形成开放的 AI 应用生态。

表8、国内外主流大模型厂商模型、应用及生态梳理

公司	核心模型	API 定价	核心产品	应用生态
OpenAI	GPT 系列、o 系列、Sora、GPT-oss 系列	GPT-5: 1.25 美元/百万 token (输入)、10 美元/百万 token (输入); o3-pro: 2 美元/百万 token(输入)、8 美元/百万 token(输出)	ChatGPT、Sora、Atlas、CodeX、Agents、APPsSDK、API	构建以大模型能力为中枢的生态平台，依托 ChatGPT 分发渠道，提供开发者工具，构建搜索、社交、视频、Agent 等多元产品矩阵
谷歌	Gemini 系列、Gemma 系列、Veo、Imagen、	gemini-3-pro-preview: 2 美元/百万 token (输入)、12 美元/百万 token (输出) (<=20 万 token)	GeminiAI、Chrome (AI Mode、AI Overviews)、Veo3、Flow、Jules、谷歌云、地图、支付	利用其云基础设施和产品线，将大模型能力集成到自身生态中，围绕 Google 自有产品扩张迭代，并提供云调用
Meta	LLama 系列、SAM 系列、CM3leon	LLaMA3.1: 70B 版 0.35 美元/百万 token(输入)、0.40 美元/百万 token (输出)、405B 旗舰版 2.7 美元/百万 token (输入)	MetaAI、Oakley 智能眼镜、MetaQuest 系列 (VR/MR 头显)、Facebook、Instagram	社交媒体平台矩阵庞大，Meta AI 深度集成，模型开源在研究与产品层面建立生态，AI 智能硬件发展迅速
阿里巴巴	Qwen 系列、Wan 系列、Fun 系列	Qwen3-Max 最低 0.0032 元/千 token(输入), 0.0128 元/千 token (输出)	淘宝、夸克、通义千问、千问 APP、钉钉、Agent(企业级)	AI 技术深度赋能其核心电商业务，强大开源策略构建繁荣生态，开源 Agent、GPU 算力套餐、智能应用方案等 AI 业务条线丰富，带动云服务增长
字节	文本模型 (Doubao-1.6)、视频生成模型 (Seedance 系列)、语音合成模型、声音复刻模型	Doubao-seed-1.6: 0.8 元/百万 token (输入)、2 元/百万 token (输出)	豆包 AI 助手、浏览器、扣子、即梦、星绘、醒图	AI 功能深入内嵌字节系产品，依托大规模投流迅速积累大规模用户，产品体系量大

DeepSeek	DeepSeek Chat 系列、Coder 系列、VL 系列、Reasoning 系列	DeepSeek-V3.2: 2 元/百万 tokens(输入)、3 元/百万 tokens(输出)	DeepSeekAI、API、企业级解决方案	API 平台和开发者社区庞大，涵盖企业、硬件厂商、云商等多元生态，开放+高性价比使其在 API 市场占据较大份额，与多家知名政企达成合作
----------	--	--	------------------------	--

数据来源：AI 寒武纪，Llama 官网，Gemini API 官网，火山引擎官网，阿里云官网等，兴业证券经济与金融研究院整理

3.2、生态：携手行业巨头，聚力驱动 AI 新变革

AI 大模型竞赛已从单维技术评比转向生态体系的较量。主流厂商依托自身资源禀赋，构建差异化竞争优势：1) OpenAI：技术先发优势突出，依托 8 亿周活用户构建强大流量能力，并通过入口平台化加速生态延展；2) 谷歌：充分整合 Google 搜索、安卓等生态系统，形成规模效应与闭环模型体验；3) 字节：豆包深度集成进抖音、今日头条、飞书、智能耳机等软硬件，串联软硬件生态；广告投放和字节系平台投流优势明显，旗下豆包为 2025Q3 国内月活与下载量双冠；4) 阿里巴巴：淘宝、支付宝等 B 端/C 端用户生态庞大，强化眼镜、手机等端侧应用落地。同时云服务能力强，提升 B 端大模型 API 体验。5) DeepSeek：技术领先+开源生态+政企合作，引领大模型技术创新和开源发展。如下，对目前主流 5 家公司进行相关竞争力梳理：

表9、主流 AI 大模型公司竞争力梳理

公司	技术研发	生态赋能
OpenAI	模型快速迭代，核心模型的迭代周期大致为每年 1-2 次，中间版本季度/月度级迭代；强大推理、任务迁移和多模态处理能力，高效跨领域完成复杂任务	周活已达 8 亿，目标年底达到 10 亿，分发渠道巨大；“入口+生态”战略，ChatGPT、Sora、Atlas 布局下平台化应用仍将拓宽
谷歌	原生集成 Google Search、代码执行等工具，可调用工具数量多，Gemini-3 在推理和多模态等方面显著升级；Gemini-3-Pro 系列能力领先	2025 年 10 月，谷歌搜索占据超过 90% 的全球市场份额，每日处理超过 150 亿次搜索，用户基础巨大、使用率高频；Gemini 2025Q3 月活超 6.5 亿，深度整合 Google 生态
阿里巴巴	模型迭代加速，Qwen3-Max MMLU-Pro 得分为 84.1%；开源效率高且能力强，开源模型全球下载量超过 6 亿次	通义千问开源 300+款模型，通义万相连续开源 20+款模型；用户生态庞大，强化眼镜、手机等端侧 AI
字节	豆包大模型覆盖文本、视觉、语音、交互等多维度任务，多模态能力强；Doubao-Seed-Code 在 MMLU-Pro 测评中得分为 85.4%	豆包深度集成进抖音、头条、飞书、智能耳机等软硬件；2025 年 10 月抖音 APP 月活达 9.48 亿，豆包月活达 1.6 亿
DeepSeek	支持最长 128k 上下文窗口；DeepSeek-V3.2、DeepSeek-R1 在 MMLU-Pro 测试中分别取得 86.2%、84.9%得分；技术创新性强	DeepSeek-R1 在 AI 社区平台 Hugging Face 上累计下载超过 1090 万次，开源社区庞大；与云服务商及智算企业深度合作

数据来源：阿里研究院，量子位，新智元，Artificial Analysis，QuestMobile 等，兴业证券经济与金融研究院整理

注：相关模型测试打分成绩数据按照模型发布时间为准

在算力端：OpenAI 持续壮大生态联盟。OpenAI 成立早期便与英伟达和微软深度合作，在芯片、算力资源、云服务等方面获得支持，后续算力联盟逐步扩张，并启动芯片自研。具体来看：

1) 芯片合作方面：公司成立之初便获得英伟达超级计算机支持，后续基于英伟达芯片进行训练（GPT-3 使用英伟达 V100、A100）并展开紧密合作。2025 年 5 月 OpenAI 首次大规模租用谷歌 TPU 以降低算力成本。2025 年 10 月，OpenAI 与博通达成数十亿美元芯片开发协议，计划联合博通自研芯片，OpenAI 负责芯片设计，与博通共同设计机架系统，并将于 2026 年下半年开始部署 10GW 定制 AI 加速器与网络系统架构，目标在 2029 年底前完成全部部署。

2) 算力建设方面：随着模型快速迭代，公司对算力的需求日益庞大，OpenAI 于 2025 年 1 月启动“星际之门”（Stargate）项目，与多家科技公司达成算力合作，包括与微软及甲骨文等签署 5000 亿美元 10GM AI 大模型骨干基建网络项目。

3) 云基础设施方面：早期微软与 OpenAI 在云计算领域合作紧密；2025 年 9 月，OpenAI 与甲骨文签订合同，公司计划将在未来五年内向甲骨文采购价值 3000 亿美元的云计算算力服务；2025 年 11 月，亚马逊 AWS 与 OpenAI 共同宣布签署价值 380 亿美元的七年算力协议。

表10、OpenAI 在算力领域的重点合作梳理

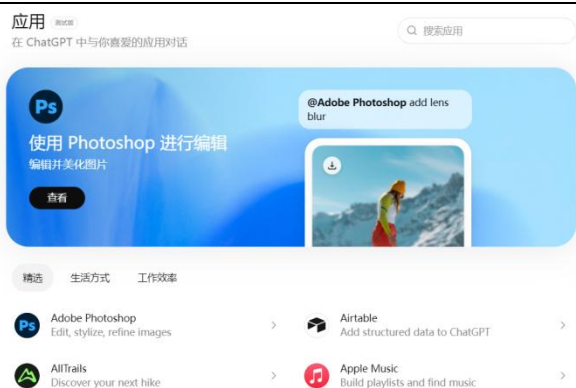
时间	合作方	合作内容
2016	英伟达	英伟达将首台 NVIDIA DGX-1 AI 超级计算机送至 OpenAI；此次合作标志着双方首次深度绑定，后期 OpenAI 依赖英伟达 A100、H100 等芯片进行模型训练
2019-07	微软	微软投资启动深度算力合作，成为其独家云提供商
2019	微软	为 OpenAI 建设超大规模数据中心集群
2020-05	微软	为 OpenAI 构建 Azure 超级计算机，专用于训练大型 AI 模型，拥有超 28.5 万个 CPU 核心和 1 万个 GPU，网络带宽达 400 Gbps
2025-01	微软、甲骨文、MGX 等	启动“星际之门”（Stargate）项目，计划投资 5000 亿美元打造 10GM AI 大模型骨干基建网络
2025-03	CoreWeave	签署价值 119 亿美元、为期五年的合作协议，CoreWeave 将为 OpenAI 提供算力支持
2025-05	谷歌	OpenAI 首次大规模租用谷歌 TPU，以降低算力成本
2025-07	Nscale、Aker	挪威星际之门 0.29GW 项目，投入 10 亿+美元，至 2027 年配备 10 万+英伟达 GPU
2025-09	甲骨文	签订五年期总价值 3000 亿美元的云计算合同，年均支出约 600 亿美元
2025-09	CoreWeave	扩大算力供应协议，新增合同金额达 65 亿美元，总金额已累计达约 224 亿美元
2025-09	英伟达	OpenAI 购买英伟达 10GW 芯片，英伟达投资 1000 亿美元系统建设 AI 数据中心，周期 5 年
2025-10	AMD	AMD 向 OpenAI 发行 1.6 亿股普通股认股权证，计划部署 6GM 的 AMD GPU 算力

2025-10	博通	双方计划从 2026 年下半年开始部署 10GW 定制 AI 加速器与网络系统架构，目标在 2029 年底前完成部署
2025-11	亚马逊	签署为期 7 年、价值 380 亿美元的云服务协议，用于运行和训练其人工智能模型

数据来源：第一财经，科创板日报，华尔街见闻等，兴业证券经济与金融研究院整理

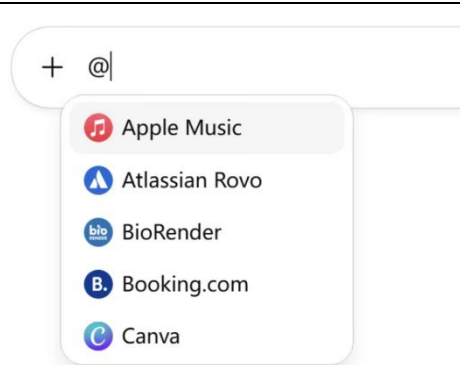
在应用端：**OpenAI 推出应用商店，基于 ChatGPT 搭建新的插件、应用生态。**应用产品维度，OpenAI 依托 ChatGPT 分发渠道铺开多产品线，包括 Agent、社交媒体（Sora）、浏览器（Atlas）、购物、音乐、API、定制化模型等，并延伸至机器人、AI 手机、AI 端侧设备等硬件。OpenAI 于 2025 年 12 月正式推出 ChatGPT 版应用商店，该功能与 2025 年 10 月推出的 ChatGPT Apps SD 相结合，使开发者可以将自有服务封装为在 ChatGPT 内可调用的应用，目前包括 Photoshop、Apple Music、Canva、Figma 等多款应用。公司探索将 ChatGPT 从对话型 AI 工具向类操作系统转型，构建长期生态竞争力。

图35、ChatGPT 应用（BETA）页面



数据来源：APPSO，兴业证券经济与金融研究院整理

图36、用户可在 ChatGPT 对话过程中调用应用



数据来源：世界互联网大会，兴业证券经济与金融研究院整理

OpenAI 与头部应用企业广泛建立合作关系。OpenAI 从模型层、行业应用层、平台层形成了高度开放、可扩展的协同体系。公司 GPT、Sora、GPT-Image、Audio 等系列模型已成为通用 AI 能力的基础设施，嵌入到海量 C 端与 B 端产品中，与 Adobe、Bloomberg、HubSpot、Stripe 等科技公司建立稳定合作，加速模型产品在创意软件、办公协作、金融服务、SaaS 等领域渗透。同时通过 API、Assistants、GPTs、代理框架与插件生态，为各类开发者和企业留下自主空间。

OpenAI 重视 B 端客户市场，打造更多的行业增长点。公司通过高度开放的 AI 部署平台获 B 端企业认可。**1）模型和产品方面：**OpenAI 推出 Codex 代码工具、AgentKit 企业智能体功能、Image Generation API 等模型或工具帮助企业客户快速部署 AI 大模型。**2）客户规模方面：**截至 2025 年 11 月 5 日，公司全球企业客户突破百万，澳大利亚联邦银行、摩根士丹利等知名企业已采用 OpenAI 平台，

ChatGPT for Work 的商业用户席位数量已突破 700 万。**3) 生态建设方面:** Figma、Spotify 等科技公司已直接接入 ChatGPT 平台; 沃尔玛、Paypal、Shopify 通过代理商务协议向 AI 工具用户提供实时购物选项。

4. 风险提示

1) 行业竞争加剧的风险

大模型作为高投入的科技领域, 代表公司普遍处于亏损状态, 而在行业利润暂未转正的情况下, 行业内模型调用价格普遍下探、竞争趋烈。

2) 应用进展不及预期的风险

公司除了在 AI 大模型领域大幅投入, 同时布局 Sora、AI 浏览器等应用产品, 产品线趋向丰富。对应领域存在大模型应用落地进展不及预期的可能。

3) 核心技术人才流动的风险

AI 大模型行业的竞争升温, 核心技术人才成为了各公司争夺的焦点。在外部高薪岗位吸引下, 公司的核心技术人才存在流动的风险。

分析师声明

本人具有中国证券业协会授予的证券投资咨询执业资格并登记为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

投资评级说明

投资建议的评级标准	类别	评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后的 12 个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅。其中：沪深两市以沪深 300 指数为基准；北交所市场以北证 50 指数为基准；新三板市场以三板成指为基准；香港市场以恒生指数为基准；美国市场以标普 500 或纳斯达克综合指数为基准。	股票评级	买入	相对同期相关证券市场代表性指数涨幅大于 15%
		增持	相对同期相关证券市场代表性指数涨幅在 5% ~ 15% 之间
		中性	相对同期相关证券市场代表性指数涨幅在 -5% ~ 5% 之间
		减持	相对同期相关证券市场代表性指数涨幅小于 -5%
		无评级	由于我们无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使我们无法给出明确的投资评级
	行业评级	推荐	相对表现优于同期相关证券市场代表性指数
		中性	相对表现与同期相关证券市场代表性指数持平
		回避	相对表现弱于同期相关证券市场代表性指数

信息披露

本公司在知晓的范围内履行信息披露义务。客户可登录 www.xyzq.com.cn 内幕交易防控栏内查询静默期安排和关联公司持股情况。

本公司为寒武纪(688256)做市商。但上述持仓不曾、不会、不将对研究业务的独立性、客观性产生影响。

使用本研究报告的风险提示以及法律声明

兴业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。

，本公司不会因接收人收到本报告而视其为客户。本报告中的信息、意见

等均仅供客户参考，不构成所述证券买卖的出价或征价邀请或要约，投资者自主作出投资决策并自行承担投资风险，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，任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以本公司向客户发布的本报告完整版本为准。该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本报告所载资料的来源被认为是可靠的，但本公司不保证其准确性或完整性，也不保证所包含的信息和建议不会发生任何变更。本公司并不对使用本报告所包含的材料产生的任何直接或间接损失或与此相关的其他任何损失承担任何责任。

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、价值及投资收入可升可跌，过往表现不应作为日后的表现依据；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；本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现。过往的业绩表现亦不应作为日后回报的预示。我们不承诺也不保证，任何所预示的回报会得以实现。分析中所做的回报预测可能是基于相应的假设。任何假设的变化可能会显著地影响所预测的回报。

本公司的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

本报告并非针对或意图发送予或为任何就发送、发布、可得到或使用此报告而使兴业证券股份有限公司及其关联子公司等违当地的法律或法规或可致使兴业证券股份有限公司受制于相关法律或法规的任何地区、国家或其他管辖区域的公民或居民，包括但不限于美国及美国公民（1934 年美国《证券交易所》第 15a-6 条例定义为本「主要美国机构投资者」除外）。

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。未经授权的转载，本公司不承担任何转载责任。

特别声明

在法律许可的情况下，兴业证券股份有限公司可能会持有本报告中提及公司所发行的证券头寸并进行交易，也可能为这些公司提供或争取提供投资银行业务服务。因此，投资者应当考虑到兴业证券股份有限公司及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。

兴业证券研究

上海	北京	深圳
地址：上海浦东新区长柳路 36 号兴业证券大厦 15 层	地址：北京市朝阳区建国门外大街甲 6 号世界财富大厦 32 层 01-08 单元	地址：深圳市福田区皇岗路 5001 号深业上城 T2 座 52 楼
邮编：200135	邮编：100020	邮编：518035
邮箱：research@xyzq.com.cn	邮箱：research@xyzq.com.cn	邮箱：research@xyzq.com.cn