

从Tokens角度跟踪AI应用落地进展 ——计算机行业大模型及AI应用专题

评级：推荐(维持)

刘熹(证券分析师)

S0350523040001

liux10@ghzq.com.cn

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

最近一年走势



相关报告

《计算机事件点评：甲骨文RPO增至4550亿美元，AI算力强力增长（推荐）*计算机*刘熹》——2025-09-12

《液冷：AI算力新一极——AI算力“卖水人”专题系列（6）（推荐）*计算机*刘熹》——2025-08-17

《计算机专题报告：OpenAI发布GPT-5，AI应用业绩提速——模型及AI应用月报（202507）（推荐）*计算机*刘熹》——2025-08-09

相对沪深300表现

表现	1M	3M	12M
计算机	7.8%	20.6%	89.7%
沪深300	9.1%	16.2%	42.5%

◆ 本报告主要解决：1) tokens调用量概念界定、2) tokens调用量应用端拆解、3) tokens调用量展望三方面问题。

◆ **Tokens：模型定价的主要单位，与收入同步高增**

- ✓ Tokens 是大语言模型用来切割自然语言文本的基本单位，本质反映大语言模型计算量。在推理过程中，token是计算的基本单位，模型对每个token的处理都需要消耗计算资源，处理的token数量越多，计算量会呈平方级增长。
- ✓ OpenAI、Anthropic、字节跳动等基模厂商的主流商业模式包括C端订阅/付费/间接和B端行业客户/API调用。同时，tokens是模型变现中的最小单位，厂商主要围绕tokens定价。
- ✓ 模型厂商以tokens为主要定价单位的底层逻辑是算力投入，模型调用时的tokens消耗量与相应算力投入存在强关联性。
- ✓ 模型厂商营收正与其tokens调用量呈现显著同步的高增趋势，特别是对于OpenAI、Anthropic，因为其商业模式为典型的C端订阅/付费和B端行业客户/API调用，收入与tokens调用量关系紧密。

◆ **Tokens调用量应用端拆解：B+C双端**

- ✓ 近年来，受C端AI原生应用/AI功能等用户基础扩张及B端行业客户加速渗透驱动，Google、字节跳动、OpenAI等海内外头部模型厂商日均Tokens调用量持续增长。拆解来看：

- C端驱动：包括原C端产品内部的附加AI功能、聊天助手、新兴应用（图像、视频、陪伴、办公、教育赛道）；
- B端驱动：包括行业企业客户、直接API调用。

◆ **展望：技术迭代解锁应用需求，算力投入加大模型厂预期乐观**

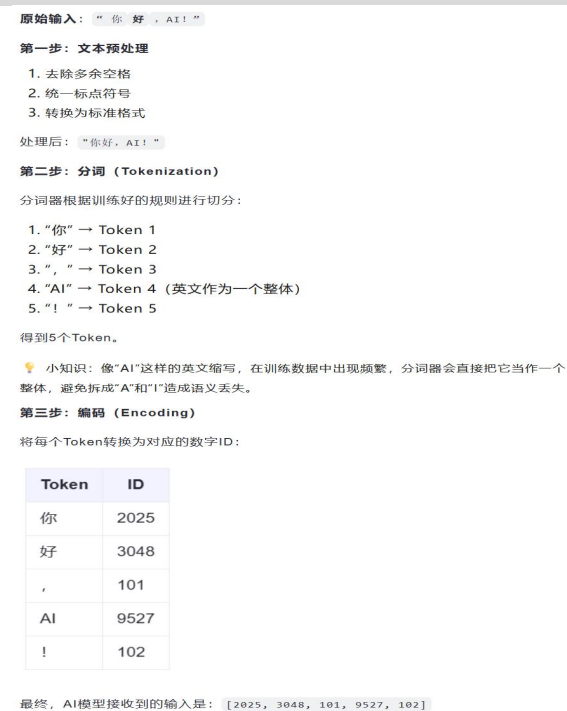
- ✓ 大模型的技术迭代方向为推理增强+多模态+Agent化+长上下文，这样的技术迭代趋势将增强AI在更复杂、更具备“生产力”的关键场景下的实用性、准确性，使得AI应用加速落地，并且带来存量场景和增量场景的tokens增长空间。
- 存量：“推理增强”将模型的一次性回答拆解为多步并进行多次调用；“多模态”需增加对图片、语音、视频等的编码解码；“Agent化”会在模型生成的过程中引入工具链；“长上下文”使得大型文档/项目可被全量处理而无需截断。
- 增量：大量原本因“不准、不全、不落地”而被搁置的需求有望被解锁。当准确率、可控性跨过可行性线后，用户特别是B端企业（有生产力场景需求）或将从观望转为批量采购。
- ✓ 使用成本侧，模型厂进行了算力成本优化、多价格策略的尝试以压降大模型使用成本，企业/个人模型调用ROI得以提高。
- ✓ 2025年及以后，模型厂算力投入仍在加码并有望成为支撑千亿美元营收预期的核心。
- ◆ 投资建议：大模型技术迭代不断加速，B/C双端大量新应用场景将被解锁；国内AI应用公司成长主线明确，业绩提速在即；维持对AI应用方向“推荐”评级。
- ◆ 风险提示：宏观经济影响下游需求、人工智能产业发展不及预期、技术突破放缓、商业化进度不及预期、市场竞争加剧、中美博弈加剧、相关公司业绩不及预期。

1、Tokens: 模型定价的主要单位，与收入同步高增

1.1、Tokens调用量本质：反映大语言模型计算量

- ◆ Tokens 是大语言模型（LLM）用来切割自然语言文本的基本单位，可以直观的理解为“字”或“词”。各大模型的切割规则不完全相同，通常1个中文词语、1个英文单词、1个数字或1个符号计为1个 token，并赋予相应的数字编号。一般情况下， token 和字数的换算比例大致为——1个英文字符≈0.4个token、1个中文字符≈0.5至1个token，可以近似地认为一个汉字就是一个 token。
- ◆ Tokens调用量本质反映大语言模型计算量。大语言模型将输入文本切割为tokens，并转换为相应数字编码，经模型内部推理计算后，最终以一连串tokens的形式进行文本输出。在这个过程中，token是计算的基本单位，模型对每个token的处理都需要消耗计算资源，处理的token数量越多，计算量会呈平方级增长。因此，tokens调用量本质反映了模型推理过程的计算量。

图：大语言模型输入过程：将输入文本切割为token，再转化为向量进行处理



图：DeepSeek V3.1的输出：最大输出长度为8k/64k tokens

型		深度搜索聊天	深度搜索推理器
型号版本		DeepSeek-V3.1（非思考模式）	DeepSeek-V3.1（思考模式）
上下文长度		128K	
最大输出		默认值：4K 最大值：8K	默认值：32K 最大值：64K
特征	Json 输出	✓	✓
	函数调用	✓	✗ ⁽¹⁾
	聊天前缀完成（测试版）	✓	✓
	FIM 完成（测试版）	✓	✗

1.2、模型厂主流商业模式：C端订阅/付费+B端行业客户/API调用

- ◆ OpenAI、Anthropic、字节跳动等基模厂商的主流商业模式包括C端订阅/付费/间接和B端行业客户/API调用：1）C端：原生聊天助手、工具类原生应用（影视、图片、编程等）的订阅收入、付费功能以及与内部C端产品整合后的间接收入（如Google Chrome）；2）B端：为大客户落地AI应用、企业直接API调用。
- ◆ Tokens是模型变现中的主要单位，厂商主要围绕tokens定价：无论是C端打包收费还是B端按量计价，模型厂商定价都以tokens为核心单位。

图：基模厂商主流商业模式



图：围绕tokens的定价方式，以OpenAI为例

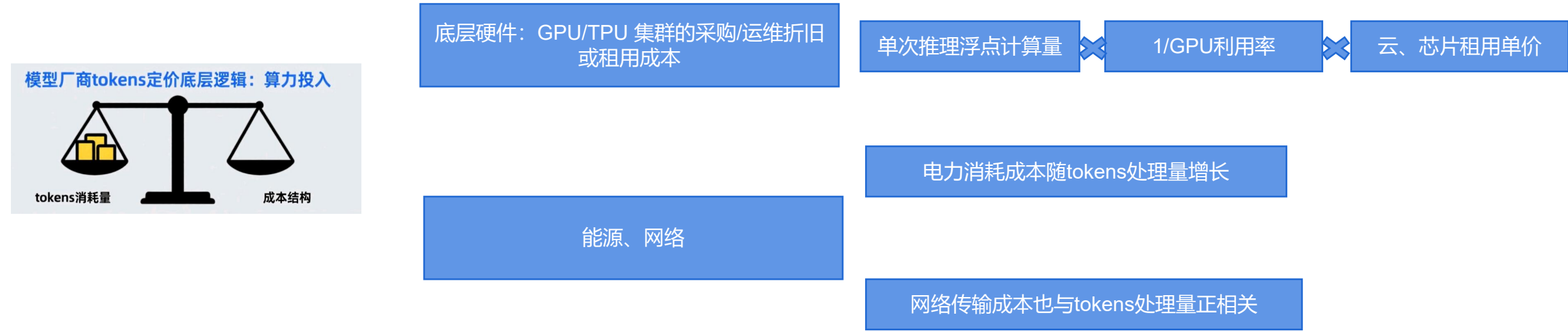
C端打包收费，有使用限制		
免费版 了解 AI 如何帮助处理日常任务 US\$0 /月 获取免费版 ✓ 可有限的访问我们的旗舰模型 GPT-5 ✓ 可通过搜索从网络获取实时数据 ✓ 使用文件上传、语音和图片工具（有使用限制） ✓ 使用 macOS 桌面应用编辑代码 ✓ 可使用自定义 GPT 已有套餐？请参见结算帮助	Plus 版 利用扩展访问权限提升效率和创意 US\$20 /月 获取 Plus 版 ✦ 可使用免费版套餐中的所有功能，并且： ✓ 扩展使用限制 ✓ 可使用带有视频和屏幕共享功能的语音模式 ✓ 可访问深入研究和多个推理模型 ✓ 可创建和使用项目、任务及自定义 GPT ✓ 试用新功能的机会 有使用额度限制	Pro 版 利用更高访问权限充分体验 OpenAI 的功能 US\$200 /月 获取 Pro 版 ✦ 可使用 Plus 版套餐中的所有功能，并且： ✓ 扩展使用限制 ✓ 可使用带有视频和屏幕共享功能的语音模式 ✓ 可访问深入研究和多个推理模型 ✓ 可创建和使用项目、任务及自定义 GPT ✓ 抢先体验新功能
GPT-5 我们最先进的跨行业编码与智能体任务模型 价格 输入：US\$1.250 / 1M 令牌 缓存输入：US\$0.125 / 1M 令牌 输出：US\$10.000 / 1M 令牌	GPT-5 mini 适用于明确任务的更快速、更实惠 GPT-5 版本 价格 输入：US\$0.250 / 1M 令牌 缓存输入：US\$0.025 / 1M 令牌 输出：US\$2.000 / 1M 令牌	GPT-5 nano 适用于摘要生成与分类任务的最快速、最实惠 GPT-5 版本 价格 输入：US\$0.050 / 1M 令牌 缓存输入：US\$0.005 / 1M 令牌 输出：US\$0.400 / 1M 令牌

B端API按量计价，并设置多样化策略，轻量级的mini、nano每百万tokens输入/输出价格更便宜

1.3、以tokens为定价单位的底层逻辑：算力投入

- ◆ **模型厂商以tokens为主要定价单位的底层逻辑是算力投入：**模型调用时的tokens消耗量与相应算力投入存在强关联性。大模型处理tokens的计算过程需调用底层硬件（如 GPU/TPU）的算力/显存和能源、网络等，并直接反映在成本结构中。
- ✓ 1）底层硬件的算力/显存：对应GPU/TPU 集群的采购/运维折旧或租用成本。以租用为例，单张GPU 处理tokens计算时占用显存，相应拉长任务时间。租金成本可近似量化为 $\text{FLOPs（即单次推理浮点计算量）} \times 1/(\text{GPU利用率}) \times 1/(\text{云、芯片租用单价/小时})$ 。
- ✓ 2）能源、网络：电力消耗随计算量（即tokens处理量）增长，网络传输成本亦与输入输出 tokens 量正相关。

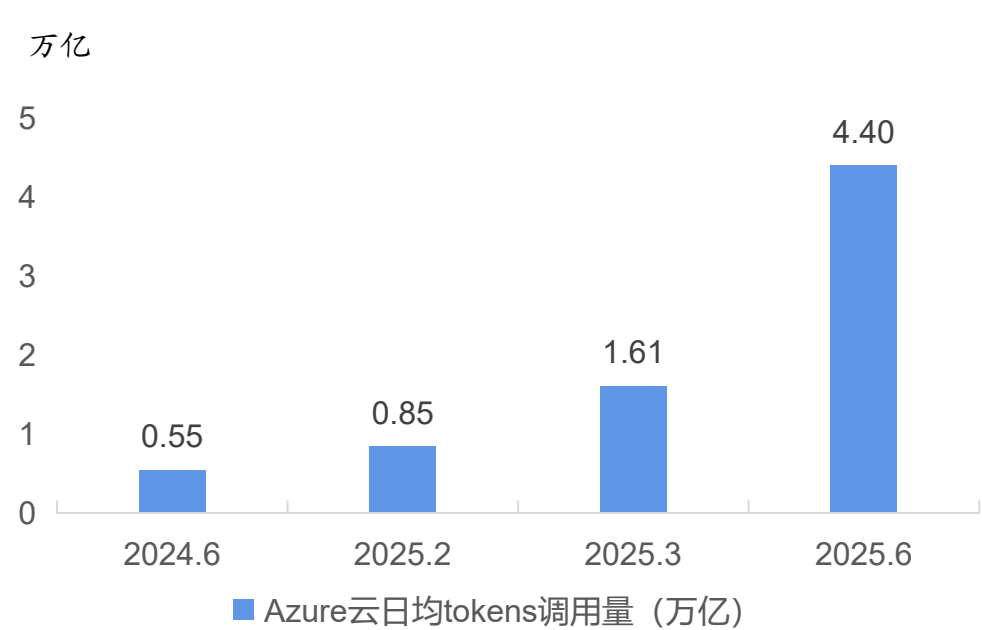
图：以tokens为定价单位的底层逻辑是算力投入



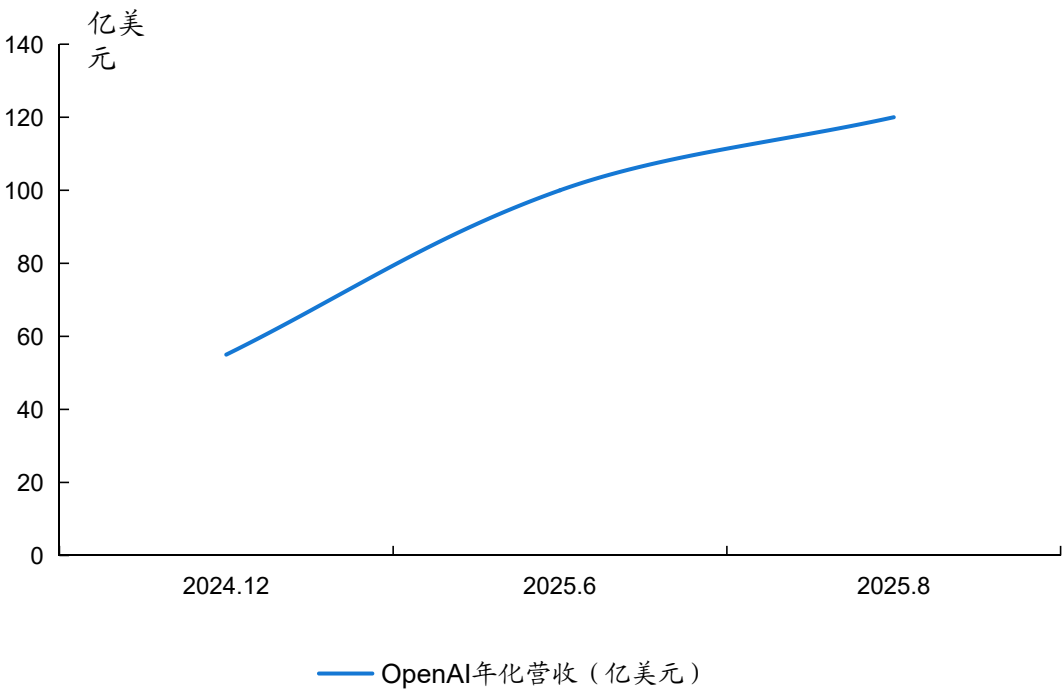
1.4、Tokens调用量：与收入同步高增

◆ 模型厂商营收正与其tokens调用量呈现显著同步的高增趋势，特别是对于OpenAI、Anthropic。其商业模式为典型的C端订阅/付费和B端行业客户/API调用，收入与tokens调用量关系紧密。2024年6月至2025年6月，OpenAI大模型基础设施——微软Azure云的日均tokens调用量从0.55万亿上涨至4.40万亿，与此同时，OpenAI年化营收（ARR）从2024年12月的55亿美元增长至2025年6月的突破100亿美元，并在2025年8月达到120亿-130亿美元。

图：Azure云日均tokens调用量



图：OpenAI年化营收



2、Tokens调用量应用端拆解：B+C双端

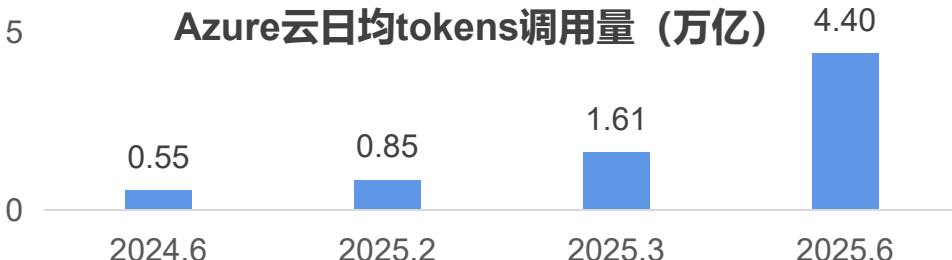
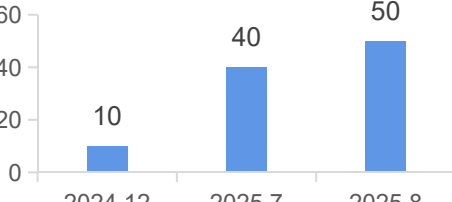
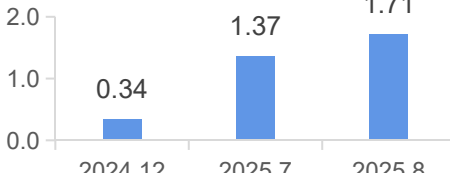
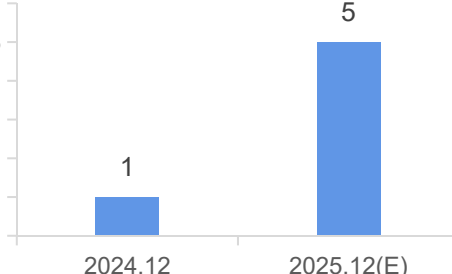
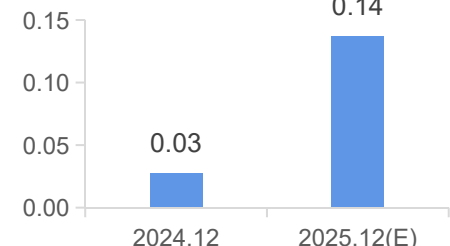
2.1、Google、字节C端应用布局丰富，间接推动B端商业落地

表：Google、字节跳动tokens调用量应用端拆解

模型厂商	大模型日均Tokens调用量持续增长	Tokens调用量应用端拆解															
Google	Google日均tokens调用量（万亿） <table><tr><th>时间</th><th>调用量 (万亿)</th></tr><tr><td>2024.4</td><td>0.3</td></tr><tr><td>2024.12</td><td>2.9</td></tr><tr><td>2025.2</td><td>5.7</td></tr><tr><td>2025.3</td><td>9.7</td></tr><tr><td>2025.4</td><td>16</td></tr><tr><td>2025.7</td><td>31.6</td></tr></table>	时间	调用量 (万亿)	2024.4	0.3	2024.12	2.9	2025.2	5.7	2025.3	9.7	2025.4	16	2025.7	31.6	C端：AI原生应用、整合于谷歌原C端产品内部	B端
时间		调用量 (万亿)															
2024.4		0.3															
2024.12	2.9																
2025.2	5.7																
2025.3	9.7																
2025.4	16																
2025.7	31.6																
典型模型： Gemini1.5Flash/Pro → Gemini2.0Flash/FlashLite/Pro → Gemini2.5Flash/Pro	✓ GeminiAPP/网页端（接入Gemini、Veo、Imagen系列） ✓ 谷歌Flow、谷歌Whisk（AI影视工具、AI图片工具，分别接入多模态模型Veo系列、Imagen系列） ✓ 谷歌Music AI Sandbox工具集（接入Lyria系列）	✓ 行业企业客户：定制化行业AI解决方案，我们认为多为接入Gemini系列。截至2025年4月，谷歌发布的全球领先企业客户已扩展至601个，截至2025Q2企业客户总数已超过8.5万家															
	✓ 谷歌搜索内置AI overview功能及AI mode（接入Gemini系列） ✓ 谷歌Workspce办公软件内置AIAgent、AI写作、表格智能分析等功能（接入Gemini系列）	✓ 直接API调用：调用Gemini等API接口，其中截至2025年5月开发者人数超过700万															
字节跳动	字节跳动日均tokens调用量（万亿） <table><tr><th>时间</th><th>调用量 (万亿)</th></tr><tr><td>2024.5</td><td>0.12</td></tr><tr><td>2024.12</td><td>4</td></tr><tr><td>2025.3</td><td>12.7</td></tr><tr><td>2025.5</td><td>16.4</td></tr></table>	时间	调用量 (万亿)	2024.5	0.12	2024.12	4	2025.3	12.7	2025.5	16.4	C端：AI原生应用、整合于字节跳动原C端产品内部	B端				
时间		调用量 (万亿)															
2024.5		0.12															
2024.12	4																
2025.3	12.7																
2025.5	16.4																
典型模型： 豆包1.5/1.5Pro → 豆包1.6	✓ 豆包App/网页端、CiCi（接入豆包大语言模型、视觉理解vicon系列、语音合成voice系列、图像视频生成seed系列） ✓ 猫箱（AI陪伴应用，可对标海外Character.AI，根据其语音交互、对话等功能，我们认为或为接入豆包大语言模型、语音合成voice系列） ✓ 即梦、醒图/星绘（AI影视工具、AI图片工具，接入图像视频生成seed系列） ✓ 海绵音乐（AI音乐创作） ✓ 豆包爱学（AI学习应用，接入豆包大语言模型、视觉理解vision系列）	✓ 行业企业客户：定制化行业AI解决方案，接入豆包多系列模型或基于火山引擎定制Agent等。截至2025年6月，豆包已接入对话助手类、教育类、投融资顾类等多个企业案例															
	✓ 抖音及Tiktok内置AI创作、互动、搜索等功能（直接接入豆包APP） ✓ 剪映/CapCut内置文案生成、数字人、自动剪辑、一键成片等功能(接入豆包大语言模型、图像视频生成seed系列) ✓ 飞书内置会议纪要生成、文档续写、表格智能填充、智能客服等功能(接入豆包大语言模型) ✓ 今日头条内置AI新闻总结功能(接入豆包大语言模型)	✓ 直接API调用：调用豆包系列大模型，包括大语言、图片、视频、语音等															

2.1、C端聊天助手构成OpenAI重要来源，B端API调用成Anthropic收入主力

表：OpenAI、Anthropic、xAItokens调用量应用端拆解

模型厂商	大模型日均Tokens调用量持续增长		Tokens调用量应用端拆解																	
OpenAI 典型模型： GTP系列：GPT-4/Turbo→GPT-4o/4omini→GPT-4.1→GPT-4.5→GPT-5 O系列：o1→o3	Azure云日均tokens调用量（万亿）  <table><tr><th>时间</th><th>调用量 (万亿)</th></tr><tr><td>2024.6</td><td>0.55</td></tr><tr><td>2025.2</td><td>0.85</td></tr><tr><td>2025.3</td><td>1.61</td></tr><tr><td>2025.6</td><td>4.40</td></tr></table>		时间	调用量 (万亿)	2024.6	0.55	2025.2	0.85	2025.3	1.61	2025.6	4.40	C端：由聊天助手带动,C端为OpenAI重要收入来源，2025年7月C端ARR收入占比约达46% ✓ chatgptAPP/网页端（2025年7月APP端月活5.77亿人，网页端月活4.38亿人；7月消费者订阅ARR收入达55亿美元，占比45.8%；月付20至200美元，估算付费用户数量为229万至2292万）	B端：行业企业客户及直接API调用，2025年7月B端ARR收入占比约达54% ✓ 行业企业客户：2025年7月收入达36亿美元，占比达30%。客户付费形式包括chatgpt企业/团队/教育版等B端账号（2025年6月，B端付费账号数达300万）以及定制项目，如Microsoft365 Copilot ✓ 直接API调用：2025年7月ARR收入达29亿美元，占比24%。直接接入API的B端客户包括Basis、Notion Invideo等等，在对o3,GPT-4.1,GPT-5等调用的基础上，开发、运营自己的AI应用产品						
时间	调用量 (万亿)																			
2024.6	0.55																			
2025.2	0.85																			
2025.3	1.61																			
2025.6	4.40																			
Anthropic 典型模型： Claude2.0→Claude3.0Haiku/Sonnet/Opus→Claude3.5 Haiku/Sonnet→Claude4.0Sonnet/Opus	Anthropic ARR（亿美元）  <table><tr><th>时间</th><th>ARR (亿美元)</th></tr><tr><td>2024.12</td><td>10</td></tr><tr><td>2025.7</td><td>40</td></tr><tr><td>2025.8</td><td>50</td></tr></table> 假设每百万Tokens8美元 Anthropic日均tokens调用量（估算值/万亿）  <table><tr><th>时间</th><th>调用量 (万亿)</th></tr><tr><td>2024.12</td><td>0.34</td></tr><tr><td>2025.7</td><td>1.37</td></tr><tr><td>2025.8</td><td>1.71</td></tr></table>		时间	ARR (亿美元)	2024.12	10	2025.7	40	2025.8	50	时间	调用量 (万亿)	2024.12	0.34	2025.7	1.37	2025.8	1.71	C端：AI原生应用，包括聊天助手与编程工具 ✓ ClaudeAPP/Claude网页端（2025年7月APP月活700万，网页端月活1600万；7月消费者订阅收入ARR达7亿美元，占比14%） ✓ Claude code（AI编程工具，2025年7月ARR达4亿美元，占比8%；月付20至200美元，估算用户数量为17万至167万）	B端：行业企业客户、直接API调用，2025年7月B端ARR收入占比达80% ✓ 行业企业客户：2025年7月ARR达9亿美元，占比18%；客户包括辉瑞、诺和诺德、Zoom等多家行业巨头，使用场景多为聊天对话、文档内容撰写生成、数据分析等 ✓ 直接API调用：2025年7月ARR收入达31亿美元,占比62%，Anthropic预计至2027年API调用收入将升至230亿美元，成为收入增长主力
时间	ARR (亿美元)																			
2024.12	10																			
2025.7	40																			
2025.8	50																			
时间	调用量 (万亿)																			
2024.12	0.34																			
2025.7	1.37																			
2025.8	1.71																			
xAI 典型模型： Grok →Grok-1.5 →Grok-2 →Grok-3 →Grok-4	xAI ARR（亿美元）  <table><tr><th>时间</th><th>ARR (亿美元)</th></tr><tr><td>2024.12</td><td>1</td></tr><tr><td>2025.12(E)</td><td>5</td></tr></table> 假设每百万Tokens10美元 xAI日均tokens调用量（估算值/万亿）  <table><tr><th>时间</th><th>调用量 (万亿)</th></tr><tr><td>2024.12</td><td>0.03</td></tr><tr><td>2025.12(E)</td><td>0.14</td></tr></table>		时间	ARR (亿美元)	2024.12	1	2025.12(E)	5	时间	调用量 (万亿)	2024.12	0.03	2025.12(E)	0.14	C端：AI原生应用或整合于x平台内部 Grok APP/网页端（2025年7月APP月活2135万，网页端2797万） x平台内置Grok入口（x平台2025年5月月活5.2亿）	B端：行业企业客户、直接API调用 ✓ 行业企业客户：1）与Oracle云联合，面向Oracle的企业客户，首批签单的行业客户包括stream等。2）接入特斯拉车端、机器人端，实现语音助手、FSD支持等功能。2025年一季度特斯拉汽车交付量33.67万辆，并计划年内生产人形机器人5000台（2026年产量升至5万台，2029年升至100万台） ✓ 直接API调用				
时间	ARR (亿美元)																			
2024.12	1																			
2025.12(E)	5																			
时间	调用量 (万亿)																			
2024.12	0.03																			
2025.12(E)	0.14																			

2.2、C端Tokens调用量拆解总结——看好大流量池产品内部附加AI功能

◆ 驱动一：大流量池C端产品内部附加AI功能

✓ 以谷歌搜索为例：2024年5月，谷歌搜索上线AI Overview功能，至2025年二季度，AI Overview功能月活已超20亿。分别假设单用户AI Overview使用次数（被动）/日、单次tokens消耗量为1至3次（据我们估算Chrome浏览器单用户日均搜索次数约为4.67次）、800tokens至1600tokens，可以得出AI Overview功能单日tokens消耗量在1.6至9.6万亿区间内，在2025年7月Google日均tokens调用量中的占比为4.9%至29.4%。

表：AI Overview tokens调用量估算

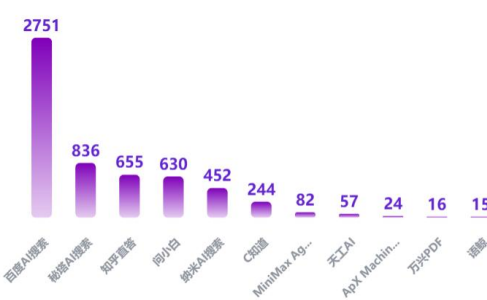
AI Overview月活（亿）	单用户单日 AI Overviews使用次数	单次tokens消耗量 AI Overviews单日调用总次数（亿次）	800	1200	1600
20	1	20	1.6	2.4	3.2
	2	40	3.2	4.8	6.4
	3	60	4.8	7.2	9.6

✓ 抖音、剪映、今日头条等同样为大流量池C端产品，月活量级已达到10亿（2025年3月）、7亿（2025年7月）、2.6亿（2024年下半年月均）。

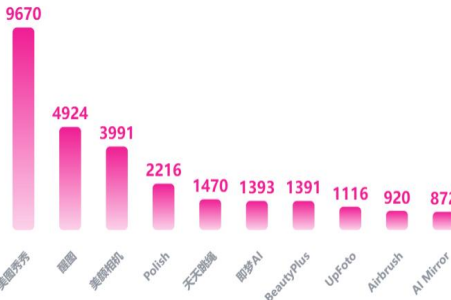
✓ 类比百度之于搜索、美图秀秀之于图像，我们看好大流量C端应用的AI改造长期逻辑。据非凡产研，2025年7月百度AI搜索访问量居国内智慧搜索品类第一、美图秀秀的国内访问量/存量月活、新增下载量依旧居图像品类第一，且月度收入仍在环比提升。

图：百度之于搜索，美图秀秀之于图像

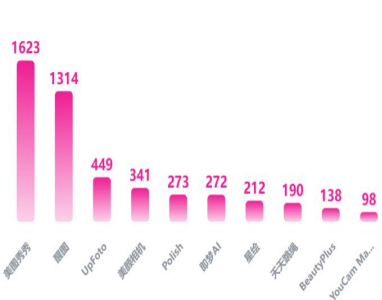
2025年7月国内智慧搜索TOP10访问量（万次）



中国图像与设计APP TOP10活跃用户（万人）



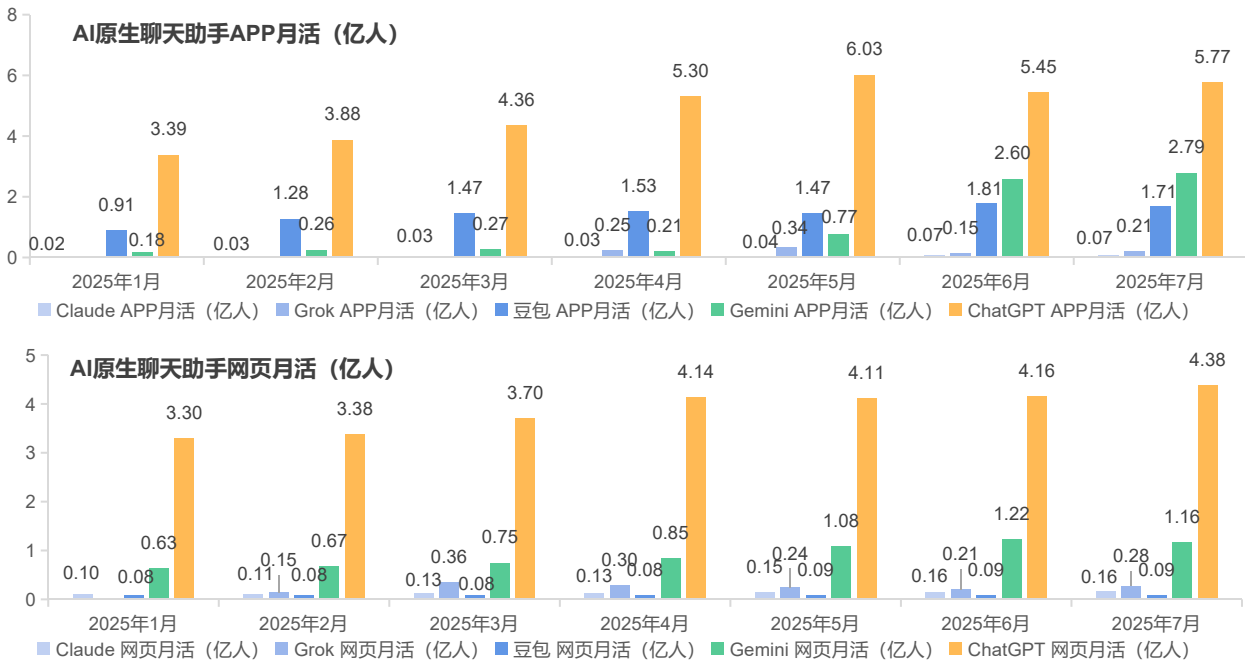
中国图像与设计APP TOP10下载量（万次）



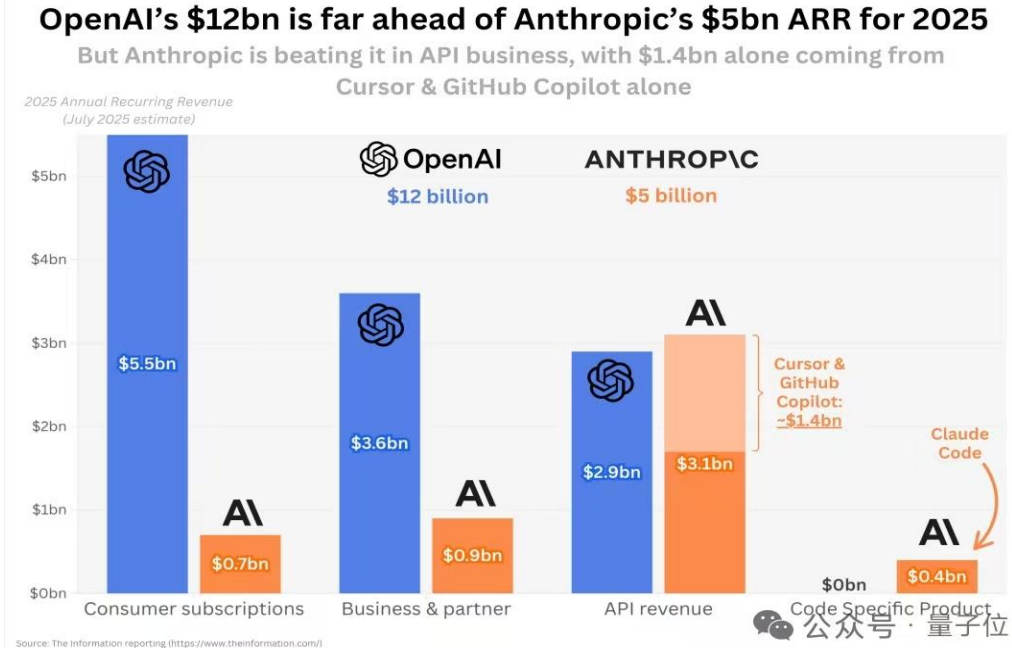
2.2、C端Tokens调用量拆解总结——聊天助手是OpenAI重要tokens调用量驱动因素

- ◆ **驱动二：聊天助手——OpenAI重要Tokens调用量驱动因素**
- ✓ **即模型厂AI原生聊天助手。存量方面：**2025年7月ChatGPT、Gemini、豆包、Grok、Claude在APP端的月活分别为5.77亿、2.79亿、1.71亿、0.21亿、708万；网页端月活分别为4.38亿、1.16亿、942万、0.28亿、0.16亿。**增量方面：**GeminiAPP和豆包APP月活明显增长，7月月活用户数为1月的15.5倍、1.9倍。
- ✓ **ChatGPT聊天助手保有较大C端用户规模，2025年7月APP+网页端合计月活达10.15亿，是OpenAI重要Tokens调用量驱动因素。**对比OpenAI、Anthropic收入结构，2025年7月，OpenAI120亿美元ARR收入中，ChatGPT聊天助手（APP/网页端）带来的C端消费者订阅收入达55亿美元，占比45.8%；而Anthropic的消费者订阅ARR收入（来自ClaudeAPP/网页端）为7亿美元，占比14%。
- ✓ 2025年6月，微软日均tokens调用量为4.46万亿，对应C端9.61亿月活及B端行业企业客户和直接API调用。结合Gemini、豆包、Grok、Claude聊天助手月活量级，我们推算由AI原生聊天助手带来的日均tokens调用量不会超过1万亿tokens，在Google及字节日均tokens调用量中份额均较小，为3%和6%。

图：五大模型厂AI聊天助手APP端及网页端月活



图：OpenAI、Anthropic收入结构



2.2、C端Tokens调用量拆解总结——图像、视频赛道中高频多模态需求带来更多tokens消耗



- ◆ 驱动三：新兴应用——图像、视频赛道拥有较大用户基础，高频多模态需求带来更多tokens消耗
- ✓ 除产品内置AI功能、聊天助手外，图像、视频、陪伴、办公、教育赛道内均出现了有较大潜力的C端新兴AI应用。
- ✓ 字节跳动进行多维度布局，推出醒图/星绘（图像）、即梦（视频）、猫箱（陪伴）、豆包爱学（教育）等AI应用。其中醒图、即梦7月月活达到4924万（当月收入59万美元）、1393万（当月收入58万美元），已成为图像、视频赛道内拥有较大用户量级的产品；猫箱7月月活794万，当月收入达112万美元（qoq+71%），商业转化效率较高。
- ✓ 图像、视频类产品对多模态输入输出有高频需求，相应带来更多tokens消耗；例如Canva接入gpt-image-1、Leonardo.AI，用于文生图、文生视频、图像补全等除文本模态外的编辑、生成场景。根据Gemini、Kimi等大模型的折算口径，单张图片的输出（输入）tokens消耗量在1024（kimi）—1290（Gemini）之间;我们假设5–15秒视频短片按4fps计算，输出（输入）tokens消耗量在6000—18000之间，文生图、图像补全、文生视频的单次tokens消耗量估算见下表。

表：图像、视频、陪伴、办公、教育赛道内C端AI应用			
赛道	AI功能	Top产品月活(2025年7月)	Top产品收入
图像	AI图片编辑、生成；可用于平面、形象、室内设计	Canva、PicsartAI：2.39亿、1.28亿	Canva2025年7月ARR突破30亿美元
视频	AI视频、数字人剪辑、生成	剪映、Inshot：7.06亿、1.18亿；Loom、keling：563万、451万	剪映2024年收入接近百亿人民币
陪伴	虚拟AI形象情感陪伴	Character.AI：3126万	Character.AI 2025年8月ARR约3000万美元
办公	文档、PPT、表格、会议等助手	Notion、讯飞听见（6月）：约4951万、1073万	Notion2024年收入突破4亿美元
教育	语言、课内多阶段AI助教	Duolingo、作业帮：1.34亿、1.02亿	Duolingo2025年全年收入预期超10亿美元

表：文生图、图像补全、文生视频的单次tokens消耗量估算				
使用场景	输入		输出	合计消耗
	文本输入	图像输入		
文生图	40–80tokens	/	1024–1290tokens	1064–1370tokens
图像补全/扩展	30–60tokens	原图/模板共1024–1290tokens	1024–1290tokens	2078–2640tokens
文生视频（5–15s）	60–100tokens	/	4000–10000tokens（假设视频帧数为4fps）	6060–18100okens

2.3、B端Tokens调用量拆解总结——B端AI应用已经在头部模型大厂tokens调用量中占据重要份额

- ◆ 对于Google、字节、OpenAI、Anthropic、xAI五家模型厂，我们认为B端AI应用（包括行业企业客户、直接API调用）均已构成其重要AI收入来源，并在每日Tokens调用量中占据较重份额。

图：我们认为B端AI应用是OpenAI、Anthropic、Google、字节跳动、xAI五家模型厂的重要收入来源，并在日均tokens调用量中占据重要份额

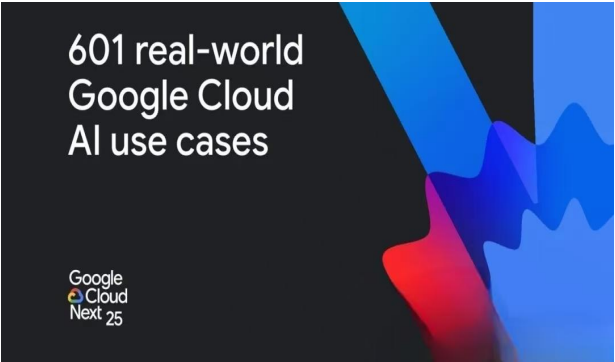
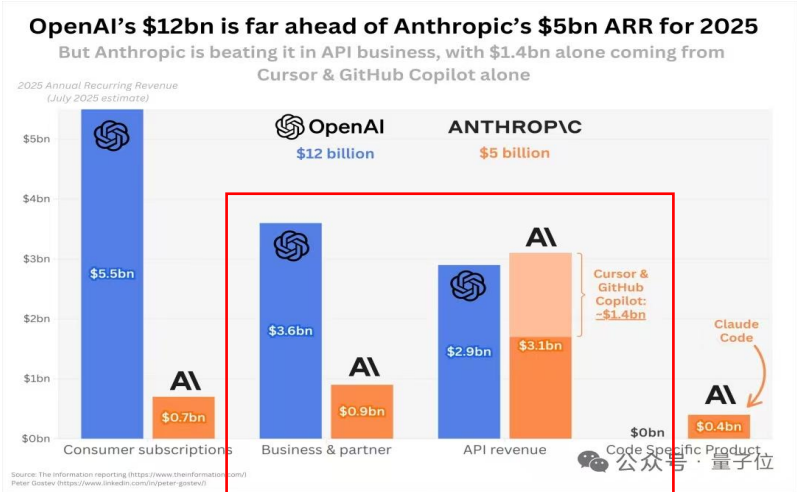
OpenAI

Anthropic

Google

字节跳动

xAI



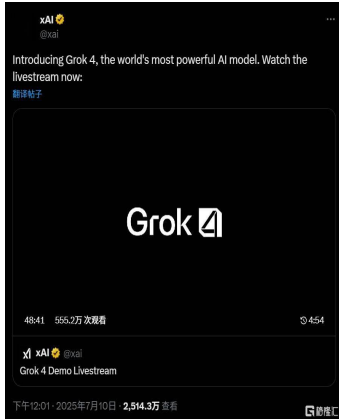
豆包大模型的产业落地正在加速

全球 Top10 手机厂商有 9 家和火山引擎深度合作

八成主流车企接入豆包大模型

已服务 70% 的系统重要性银行、数十家券商和基金公司

五成 985 高校用豆包大模型做教学、科研创新



- ✓ 2025E，OpenAI B端ARR收入占比达54%；Anthropic B端ARR收入占比达80%；
- ✓ OpenAI、Anthropic C端产品形式简洁，主要为聊天助手APP/网页端；我们认为此类个人需求tokens调用量份额与对应收入贡献份额偏离程度不大

- ✓ Gemini企业客户超过8.5万家，推动调用量同比增长35倍；
- ✓ 对于B端，火山引擎CEO指出：2024年，大模型收入在火山引擎所有业务中增速最快；而且在国内公有云市场中（2024），火山引擎的份额排名第一，占比达46.4%（外部使用量，不包括豆包等内部APP）。

- ✓ Space X、特斯拉、X平台等马斯克旗下企业或与xAI有业务合作空间，并加速细分行业标杆AI应用的打造

- ◆ 驱动一：行业企业客户
- ✓ Google发布“全球601个领先企业AI应用案例”报告：各大规模的企业已开始尝试将生成式AI投入生产，涉及汽车与物流、商业与专用服务、金融服务、医疗与生命科学、酒店与旅游、制作、工业与电子、媒体、营销与游戏、零售、科技、通信、公共部门与非盈利组织”11大行业，每个行业中又细分为六种AI智能体类型：客户、员工、创意、代码、数据和安全。
- ✓ Google指出2025年AI商业化的五大趋势为：多模态AI、AIAgent、辅助搜索、赋能客户体验以及AI安全防护；1）多模态AI推动上下文感知精度提升，Google预测至2037年末多模态AI市场规模将扩展至989亿美元；2）AI应用产品形式演进的下一个阶段是从聊天助手至多Agent系统，据凯捷对1100家大型企业高管的调查，其中82%计划在未来三年内集成AIAgent；3）AI使得搜索展现新的发展方向，赋能企业构建先进搜索系统，截至2031年末，全球企业搜索市场规模预计将达到129亿美元；4）AI聊天客服机器人仅为过渡方案，终极形态是可个性化解决客户需求，现已在多个行业内兴起；5）AI在安全领域的应用场景包括创建规则、模拟攻击和检查违规，2024年，AI安全产品在企业中的采用率还低于5%，Google认为2028年该比例将提升至50%。

图：Google 601个领先企业应用案例，以其中的汽车&物流行业为例

汽车与物流



面向客户的智能体

- ✓ Continental 正在利用谷歌的数据与AI技术开发安全、高效且以用户为中心的汽车解决方案。此次合作的初步成果之一，是将 Google Cloud 的对话式AI技术集成到 Continental 的 Smart Cockpit HPC 中，这是一款车载语音指令解决方案。
- ✓ 通用汽车（General Motors）旗下的 OnStar 已新增多项AI功能，包括一个由 Google Cloud 对话式AI技术支持的虚拟助手，该助手能更准确地识别用户的意图。
- ✓ 梅赛德斯-奔驰（MercedesBenz）在新款CLA系列车型中，使用 Google Cloud 面向汽车行业优化的 Automotive AI Agent 实现了语音搜索与导航功能。

面向员工的智能体

- ✓ 704 Apps是一家服务于“最后一公里”交通领域的应用开发商，每天连接成千上万的司机与乘客。在行程中，车内对话的音频内容会被发送至 Gemini，用于测量情绪“温度”。当检测到“抢劫”、“袭击”、“绑架”等特定词汇时，系统会将其归类为敌意行为，并发出警报，从而提前预防潜在风险。
- ✓ Oxa 是一家自动驾驶软件开发商，正在使用 Gemini for Google Workspace 来创建用于指标报告的营销模板、撰写社交媒体内容、编写职位描述以及跨团队校对文本，节省了大量时间和资源。
- ✓ Rivian使用搭载 Gemini的Google Workspace提升技术和市场团队之间的沟通与协作效率，显著加快了工作进度并提升了质量。
- ✓ 丰田（Toyota）利用 Google Cloud 的AI基础设施部署了一个AI平台，使工厂员工能够开发和部署机器学习模型，每年减少了超过1万个工时，同时提升了效率和生产力。
- ✓ 优步（Uber）正在使用AI代理帮助员工提升工作效率、节省时间、增强执行力。针对客服代表，公司推出了新工具，可自动总结与用户的交流内容，还能提取过往互动的上下文，使一线员工更具针对性地提供帮助。

代码智能体

- ✓ 雷诺集团（Renault Group）旗下2023年成立的电动汽车及软件子公司 Ampere，正在使用企业版的 Gemini Code Assist，专为开发团队打造，能够理解公司的代码库、规范和编码习惯。

数据智能体

- ✓ 宝马集团（BMW Group）与Monkeyway 合作开发了 AI 解决方案 SORDI.ai，利用生成式AI优化工业规划流程和供应链。该方案通过扫描资产并使用Vertex AI 创建3D模型，作为数字孪生体进行数千次模拟，从而提升配送效率。
- ✓ Dematic 正在利用 Vertex AI 和 Gemini 的多模态功能，为电商和全渠道零售商构建端到端的履约解决方案。
- ✓ Geotab（全球领先的车联网数据公司）使用BigQuery和Vertex AI每天分析来自460多万辆汽车的数十亿条数据点，实现车队优化、驾驶安全、交通脱碳和宏观层面的城市交通分析，推动更安全、可持续的城市发展。

2.3、B端Tokens调用量拆解总结——火山引擎发布多行业企业AI应用实例

- ◆ **驱动一：行业企业客户**
- ✓ 经我们梳理，自2024年5月豆包大模型首次发布至今，火山引擎发布了多个企业AI应用实例，现有企业客户来自智能终端、汽车、零售消费、教育科研、金融、游戏、互联网、医药健康等多个行业。
- ✓ 2024年在火山引擎所有业务中，大模型收入增速最快；且对于国内云端大模型调用量，火山引擎的份额达46.4%，排名第一（2024年数据，不包括豆包等内部APP在内）。火山引擎总裁谭待指出，当前AI应用的智能化程度仅“L1”水平，而AI应用的终极形态是建立在大模型深度思考能力和多模态能力上的综合Agent而非单点应用；未来2-3年，若模型能力若能在视觉推理、Agent协作等领域取得突破，tokens调用量将有百倍增长空间。

表：火山引擎AI应用实例

企业客户行业	企业客户范围	代表性案例
智能终端	已接入4亿台智能终端设备	- OPPO、荣耀、小米、vivo、三星、努比亚：AI助手 - 荣耀、努比亚、三星：智能手机图片编辑 - 联想：AI桌面助手
汽车	八成主流车企接入豆包应用于智能驾驶、智能座舱、智能营销等场景	- 宝马：市场营销、供应链管理 - 奔驰、上汽：车机助手
零售消费	零售消费客户应用环节及场景较为广泛	- 瑞幸：lucky智能体为客户推荐咖啡（售前客服） - 飞鹤奶粉：利用豆包构建近百个智能体用于品宣、营销、办公等内部场景
教育科研	覆盖学前至中小学全阶段，并与高校合作渗透科研场景	- 学前：儿歌点点、叫叫；AI伙伴 - 中小学：上海市虹口区智能体应用系统，供区域内教师使用 - 高校科研：北大、浙大等超五成985高校；渗透教学、教务、科研、校园等场景
金融	招商、浦发、北京银行等70%系统性重要银行以及包括国泰海通、华泰等在内的数十家券商、基金	- 北京银行：智能客服 - 国信证券、华泰证券：面向普通投资者的投资服务智能体
游戏	/	巨人网络：接入豆包大模型，实现玩家挑战AI到玩家操控AI对抗的玩法跃迁
互联网	/	- 微博：推出微博智搜功能 - 元客视界：打造数字人生成系统
医药健康	/	美中爱瑞：接入豆包大模型搭建AI预问诊智能体

- ◆ 驱动二：直接API调用
- ◆ 各模型厂商的API开放平台已汇聚大量开发者：模型厂商的API开放平台将大模型能力封装为标准接口，大幅降低开发门槛，使企业无需自建算力即可快速集成智能功能，推动了大量第三方AI应用产品的涌现。
- ✓ Cursor、GithubCopilot等AI编程工具接入Claude大模型API接口，对Anthropic的收入贡献比例已达28%（2025年7月ARR），直接印证了AI编程工具的成熟度。同时2025年7月ARR中，Anthropic API调用收入达31亿美元，占比62%，其预计至2027年API调用收入将升至230亿美元，成为收入增长主力。
- ✓ 火山方舟API开放平台设置应用广场页面，展示了联网增强、智能体、视频生成器等多种应用模板，便于企业/开发者直接复用。

表：AI编程工具	
国外热门AI编程工具	工具特点
Cursor (来自Anysphere)	AI 重构的 VS Code 编辑器，特点是对项目的理解能力和对话式交互界面；2025年3月用户达100万，收入达2亿美元
GitHub Copilot (来自微软)	集成 GitHub 生态，截至2025年7月用户超 2000 万。支持 GPT-5/Claude sonnet 4等多模型切换，PR 自动生成与代码审查功能成熟
Claude Code (来自Anthropic)	擅长复杂算法逻辑处理。处理复杂算法、调试深层逻辑等方面表现突出



3、展望：

技术迭代解锁应用需求，算力投入加大模型厂预期乐观

3.1、大模型技术迭代：推理增强+多模态+Agent化+长上下文

- ◆ 大模型不再只靠“更大参数”取胜，而是向推理增强、多模态、Agent化、长上下文转型。
- ◆ 以GPT-5和Grok4为例：
 - ✓ GPT-5把“更强的推理能力（通过引入test-timecompute）+多模态+更长上下文+更严格的安全控制等”置于产品默认层面；
 - ✓ Grok4核心升级则是把“原生工具调用+多代理协同推理+超长上下文等”做成一个可商用产品。
- ◆ 推理增强、多模态、Agent化、长上下文转型的技术迭代趋势也将增强AI在更复杂、更具备“生产力”的关键场景下的实用性、准确性，并且使得AI应用加速落地。

图：大模型技术迭代



3.2、大模型技术迭代：存量场景tokens增长空间

◆ 大模型技术迭代将带来存量AI应用场景的tokens调用量增长。例如“推理增强”将模型的一次性回答拆解为多步并进行多次调用；“多模态”需增加对图片、语音、视频等的编码解码；“Agent化”会在模型生成的过程中引入工具链；“长上下文”使得大型文档/项目可被全量处理而无需截断。

表：大模型技术迭代带来存量AI应用场景tokens调用量增长

技术趋势	Tokens调用量放大逻辑	存量场景举例	升级后
推理增强 (Chain-of-Thought/深度思考)	一次回答→多步拆解→多次调用	客服问答（假设原来1轮客服对话服务消耗200tokens）	我们认为升级后客服问答场景中的大模型推理过程将扩展成： 客户意图澄清+内部知识库检索+逻辑校验+答案润色4个环节，即4轮内部推理，每轮150~200tokens，最终消耗600至800tokens
多模态 (文本+图片+语音+视频)	纯文本→多模态编码、解码	电商商品详情生成（假设原来上传1张图片+300字文案共消耗约1590tokens）	我们认为升级后完全多模态场景下的电商商品详情生成场景可能变为： 用户上传5张商品图片（1290tokens/张）后；模型输出3段不同风格的图文视频脚本（每段脚本总长度约4000tokens），并输出5s视频；总计消耗约25000tokens
Agent化 (工具调用+循环执行)	单轮回答→多轮工具链	周报自动生成（假设原来用户粘贴文本和大模型总结生成共消耗1000tokens）	我们认为升级后周报自动生成场景可能变为： 1）读取邮件（API检索，假设消耗600tokens） 2）本地代码diff（4000tokens） 3）生成3套大纲+1轮自洽检查（每轮800tokens） 总计消耗7800tokens
长上下文 (128K→1M+)	截断→全量	合同审查（假设原来将50页PDF截断为前8页，消耗12000tokens）	我们认为升级后可直接塞入50页（≈75000tokens）+输出5000tokens的风险评估；总计消耗80000tokens

3.3、大模型技术迭代：准确率、可控性“跨线”，更多应用需求解锁

- ◆ 我们认为：随着推理增强、多模态、Agent、长上下文等技术趋势的推进，大量原本因“不准、不全、不落地”而被搁置的需求将被解锁。当准确率、可控性跨过可行性线后，用户特别是B端企业（有生产力场景需求）或将从观望转为批量采购。
- ✓ 推理增强：把能用变成敢用
 - 场景解锁：金融报告合规检查、医药注册文档、司法判例检索等“零容忍错误”场景
 - 举例：制药企业用大模型生成可达到FDA注册要求的临床试验报告。
- ✓ 多模态：把单点工具变成端到端工作流
 - 场景解锁：制造业设备巡检、保险定损、医疗影像诊断
 - 举例：保险公司用“无人机拍照+视觉大模型”完成车险查勘，并由大模型生成带图定损报告，减少现场理赔员需求。
- ✓ Agent化：把对话变成可审计的业务系统
 - 场景解锁：供应链采购、ERP数据填报、跨境关务
 - 举例：跨境卖家使用Agent同时对接海关、物流、支付等API，自动生成报关单、箱单、发票，缩短跨境业务报单时间。
- ✓ 长上下文：把项目级任务放进模型
 - 场景解锁：工程审图、专利撰写、影视剧分镜脚本
 - 举例：设计院把项目工图一次性喂给模型，自动完成碰撞检测、规范校核、材料清单，提高错误检出率。

图：火山引擎白皮书



国海证券
SEALAND SECURITIES

- 图：技术途径压降使用成本**

图：大模型降价

模型名/提供方 □		推理(输入)定价 ①	推理(输出)定价 ①
 Doubao-pro-4k 字节跳动	5	0.0008 元/千tokens	0.0020 元/千tokens
 Doubao-pro-32k 字节跳动	5	0.0008 元/千tokens	0.0020 元/千tokens
 Doubao-pro-128k 字节跳动	5	0.0050 元/千tokens	0.0090 元/千tokens
 Doubao-lite-4k 字节跳动	5	0.0003 元/千tokens	0.0006 元/千tokens
 Doubao-lite-32k 字节跳动	5	0.0003 元/千tokens	0.0006 元/千tokens
 Doubao-lite-128k 字节跳动	5	0.0008 元/千tokens	0.0010 元/千tokens
 Doubao-embedding 字节跳动	5	0.0005 元/千tokens	0.0005 元/千tokens

通义千问2024.05降价公告

模型规格	input 价格 (元/千 tokens)		output 价格 (元/千 tokens)			
	按秒算	按分钟	按秒算	按分钟		
Qwen-Turbo	0.008	0.002	75% ↓	0.008	0.006	25% ↓
Qwen-Plus	0.02	0.004	80% ↓	0.02	0.012	40% ↓
Qwen-Long	0.02	0.005	97% ↓	0.02	0.002	90% ↓
Qwen-Max	0.12	0.04	67% ↓	0.12	0.12	

Qwen1.5-72B	0.006	0.001	83% ↓	0.006	0.002	67% ↓
Qwen1.5-148B	0.008	0.002	75% ↓	0.008	0.004	50% ↓
Qwen1.5-32B	0.0035	七天 按时计费		0.007	七天 按时计费	
Qwen1.5-72B	0.02	0.005	75% ↓	0.02	0.01	50% ↓
Qwen1.5-110B	0.007	七天 按时计费		0.014	七天 按时计费	

5月21日-6月21日，上述所有模型，阿里云新购用户每单输入tokens输个模型4000tokens免费额度

模型规格	调整前价格 元/千tokens		调整后价格 元/千tokens	
	输入	输出	输入	输出
混元-lite	0.008	0.008	★ 免费	★ 免费
混元-standard	0.01	0.01	0.0045 ↓ 下降65%	0.005 ↓ 下降50%
混元-standard-256k	0.12	0.12	0.015 ↓ 下降87.5%	0.06 ↓ 下降50%
混元-pro	0.1	0.1	0.03 ↓ 下降70%	0.1
元甄-API分装 (其他模型)	100万免费tokens		★ 1亿免费tokens	
元甄-API分装 (腾讯元宝)	全量免费支持			

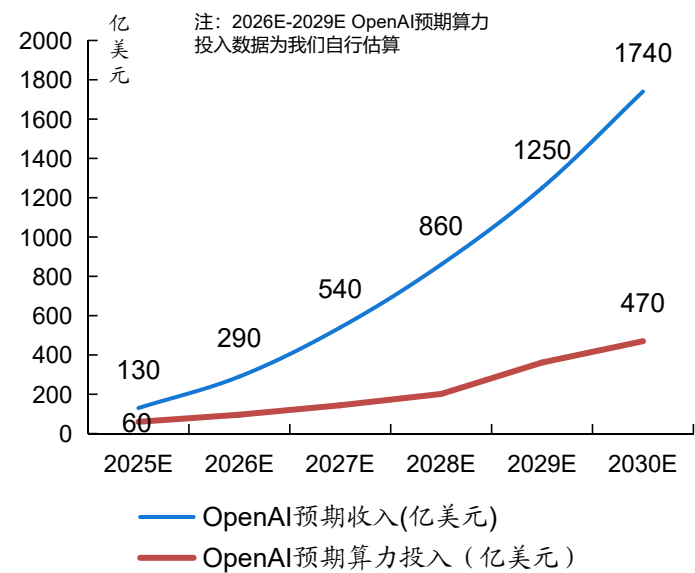
模型名称	上下文长度	输入	输出
ERNIE Speed	8K、128K	免费	免费
ERNIE Lite	8K、128K	免费	免费

Spark Lite 轻量级大语言模型	Spark Pro 专业级大语言模型	Spark 3.5 Max 最强的大模型版本
- 支持在线联网搜索功能 - 响应快速、便捷 - 适用于场景推理与模型精准定制化场景	- 数学、代码、医疗、教育等场景有专项优化 - 支持联网搜索与数据等内置插件 - 覆盖知识问答、语义理解、文本控制等多场景	- 核心能力全面升级，持续应用效果优化 - 支持联网搜索与数据等内置插件 - 支持System角色人设FunctionCall函数调用
免费可用	0.21-0.30元/千tokens	0.21-0.30元/千tokens

3.5、加大算力投入，模型厂收入预期乐观

- ◆ 2025年及以后，模型厂算力投入仍在加码，我们认为有望成为支撑千亿美元营收预期的核心：
- ✓ 2025 年，OpenAI 计划投入模型训练/运营成本 60 亿美元，并指出至2030 年，相关支出升至近470亿美元。持续加码算力投入的另一头是乐观的收入预期。OpenAI预计至2029年，其营业收入将增长至1250亿美元，现金流转正实现净流入近120亿美元，毛利率提升至69%（新技术带动模型效率提升）。
- ✓ Anthropic 2025年现金消耗及营收预计分别为30亿美元和37亿美元，并且重点聚焦企业API市场。公司预计乐观情况下，2027年将实现企业API收入230亿美元，并由此带动营收增长至345亿美元。
- ✓ xAI为打造GPU超算集群，月均支出高达10亿美元，2025年总支出预计达127亿美元（营收仅5亿美元），旨在“以资金换时间”用自建算力壁垒缩短与OpenAI的技术代差。
- ◆ 我们认为：算力投入与营收增长的飞轮效应正在形成，模型厂商以基模、生态平台、企业工具、智能体等为驱动，万卡集群或可带来同样广阔的市场空间。

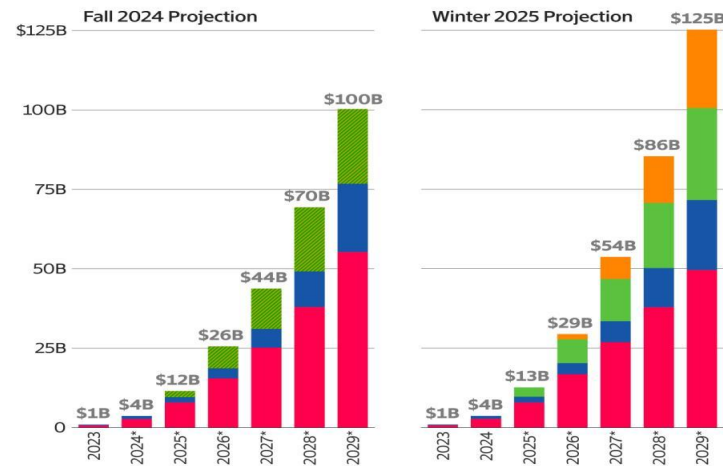
图：OpenAI预期收入与预期算力投入



Rising Revenues

OpenAI expects agents and new products to be key portions of future revenue.

● ChatGPT ● API ● Other ● Agents ● New products (including free user monetization)



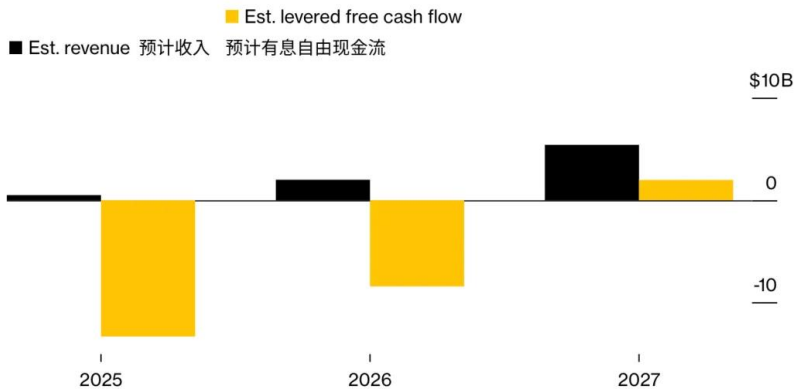
图：xAI收入、现金流预测

xAI Is Burning Through Cash

xAI 正以每月 10 亿美元的速度烧钱

Forecast of annual revenue and cash flow

年度收入和现金流预测



4、投资建议及风险提示

■ 大模型技术迭代不断加速，B/C双端大量新应用场景将被解锁；国内AI应用公司成长主线明确，业绩提速在即；维持对AI应用方向“推荐”评级：

AI应用相关公司：

- AI企业服务：金蝶国际、鼎捷数智、赛意信息、浪潮数字企业、用友网络、泛微网络、致远互联；
- AI办公/设计：金山办公、万兴科技、福昕软件、合合信息；
- AI财税：博思软件、税友股份、航天信息；
- AI营销：第四范式、迈富时，新大陆、明源云、焦点科技、光云科技；
- AI教育：科大讯飞、新开普；
- AI医疗：卫宁健康、嘉禾美康；
- AI军工：中科星图、莱斯信息、能科科技、索辰科技；
- AI出行：捷顺科技、道通科技、锐明技术、慧瀚股份；
- AI泛政务：中科曙光、太极股份、金桥信息、盛视科技、托普云农；

- 1、宏观经济影响下游需求：宏观经济可能影响下游企业对IT支出需求。
- 2、人工智能产业发展不及预期：产业发展不及预期，可能影响相关产品推进节奏。
- 3、技术突破放缓：基座模型技术突破放缓可能影响AI应用产品的功能体验及推广进度。
- 4、商业化进度不及预期：商业化进度不及预期可能影响相关公司的短期业绩。
- 5、市场竞争加剧：市场竞争加剧可能影响相关企业盈利能力。
- 6、中美博弈加剧：中美博弈加剧可能影响整体产业发展进程。
- 7、相关公司业绩不及预期：相关公司可能存在自身经营等因素导致业绩不及预期。

计算机小组介绍

刘熹，计算机行业首席分析师，上海交通大学硕士，多年计算机行业研究经验，善于把握由技术、政策驱动的科技产业新趋势，致力于进行前瞻重磅推荐。2024年金牌分析师，2024年同花顺最受欢迎分析师。

唐锦珂，计算机行业研究助理，中山大学岭南学院硕士，主要覆盖AI芯片、服务器、算力租赁、AIDC。

分析师承诺

刘熹，本报告中的分析师均具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度，独立，客观的出具本报告。本报告清晰准确的反映了分析师本人的研究观点。分析师本人不曾因，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取到任何形式的补偿。

国海证券投资评级标准

行业投资评级

- 推荐：行业基本面向好，行业指数领先沪深300指数；
- 中性：行业基本面稳定，行业指数跟随沪深300指数；
- 回避：行业基本面向淡，行业指数落后沪深300指数。

股票投资评级

- 买入：相对沪深300指数涨幅20%以上；
- 增持：相对沪深300指数涨幅介于10%~20%之间；
- 中性：相对沪深300指数涨幅介于-10%~10%之间；
- 卖出：相对沪深300指数跌幅10%以上。

免责声明

本报告的风险等级定级为R3，仅供符合国海证券股份有限公司（简称“本公司”）投资者适当性管理要求的客户（简称“客户”）使用。本公司不会因接收人收到本报告而视其为客户。客户及/或投资者应当认识到有关本报告的短信提示、电话推荐等只是研究观点的简要沟通，需以本公司的完整报告为准，本公司接受客户的后续问询。

本公司具有中国证监会许可的证券投资咨询业务资格。本报告中的信息均来源于公开资料及合法获得的相关内部外部报告资料，本公司对这些信息的准确性及完整性不作任何保证，也不保证其中的信息已做最新变更，也不保证相关的建议不会发生任何变更。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。报告中的内容和意见仅供参考，在任何情况下，本报告中所表达的意见并不构成对所述证券买卖的出价和征价。本公司及其本公司员工对使用本报告及其内容所引发的任何直接或间接损失概不负责。本公司或关联机构可能会持有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提供投资银行、财务顾问或者金融产品等服务。本公司在知晓范围内依法合规地履行披露义务。

风险提示

市场有风险，投资需谨慎。投资者不应将本报告为作出投资决策的唯一参考因素，亦不应认为本报告可以取代自己的判断。在决定投资前，如有需要，投资者务必向本公司或其他专业人士咨询并谨慎决策。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议。投资者务必注意，其据此做出的任何投资决策与本公司、本公司员工或者关联机构无关。

若本公司以外的其他机构（以下简称“该机构”）发送本报告，则由该机构独自为此发送行为负责。通过此途径获得本报告的投资者应自行联系该机构以要求获悉更详细信息。本报告不构成本公司向该机构之客户提供的投资建议。

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。本公司、本公司员工或者关联机构亦不为该机构之客户因使用本报告或报告所载内容引起的任何损失承担任何责任。

郑重声明

本报告版权归国海证券所有。未经本公司的明确书面特别授权或协议约定，除法律规定的情况外，任何人不得对本报告的任何内容进行发布、复制、编辑、改编、转载、播放、展示或以其他方式非法使用本报告的部分或者全部内容，否则均构成对本公司版权的侵害，本公司有权依法追究其法律责任。

国海证券 · 研究所 · 计算机研究团队

心怀家国，洞悉四海



国海研究上海

上海市黄浦区绿地外滩中心C1栋
国海证券大厦

邮编：200023

电话：021-61981300

国海研究深圳

深圳市福田区竹子林四路光大银
行大厦28F

邮编：518041

电话：0755-83706353

国海研究北京

北京市海淀区西直门外大街168
号腾达大厦25F

邮编：100044

电话：010-88576597