传媒
行业专题报告

2026 年 05 月 12 日

Token 经济学新模式是怎样运行的？

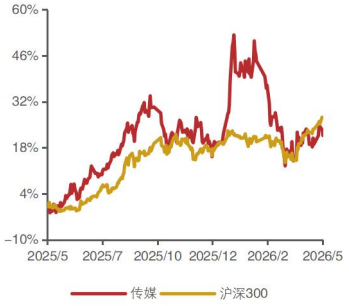
——AI 应用追寻系列报告（五）

投资评级：看好（维持）

证券分析师
陈良栋
SAC：S1350524100003
chenliangdong@huayuanstock.com

联系人
魏桢
weizhen@huayuanstock.com

板块表现：



投资要点：

Token 的运营开始形成一个新的中间层市场。Token 运营开始出现类分销模式，即批量采购 AI 厂商 API 额度，并加价转售给终端用户。Token 分销模式核心三方角色包括：1）模型方：如字节跳动（Seedance 系列）、阿里巴巴（Qwen 系列）、智谱（GLM 系列）、月之暗面（Kimi 系列）、深度求索（DeepSeek 系列）等，是 Token 源头供给方。2）代理平台：承接上游模型资源并分销给终端用户，是 Token 分销中转与统一服务的中间枢纽。3）终端用户：实际付费购买并调用消耗 Token，包含个人、开发者、企业等。

中国日均 Token 调用量显著增长，国产大模型能力实现跃迁。2024 年年初，中国日均 Token 调用量为 1000 亿；至 2025 年底，跃升至 100 万亿；2026 年 3 月，已突破 140 万亿，两年增长超千倍。OpenRouter 数据显示，截至 2026 年 5 月 7 日的最新一周榜单中，腾讯 Hy3 preview (free)位居调用量榜首，前五、前十、前二十名分别有 2 款、6 款、9 款国产大模型。

全球 AI 算力、模型、能源等核心资源在各国各地区呈现不均衡的分布格局。一方面，海外头部大模型如 OpenAI、Anthropic，受地域访问限制、合规规则与支付门槛制约，无法直接触达包括中国大陆开发者在内的海量全球用户。另一方面，以 DeepSeek 为代表的国产优质大模型，在走向国际市场时，又面临海外本地化适配、渠道铺设与用户获客的天然壁垒。催生 AI Token 的跨境流转、聚合路由与分层分销的市场需求。

Token 分销的盈利模式。1）Token 转售利差：即在模型厂商（上游）和 AI 需求方（下游）之间做流量聚合与分发，赚取差价。2）技术溢价：主要是通过自研推理加速引擎，依靠算力效率差获取超额技术毛利。3）企业增值服务：面向企业端提供 AI 相关的技术支持，如 prompt 工程、Agent 编排、模型选型等。

关注场景 Token 下的企业增值服务，有望成为 Token 分销的服务溢价来源。如硅基流动建立了企业级 MaaS 平台架构，面对企业用户提供模型训练调优、部署推理、应用开发支撑三层能力，包括数据处理、模型微调、Prompt 工程和 RAG 等，最终以标准化 API 形式交付给能源、金融、政府等多行业。国内营销公司的客户资源涉及短剧、漫剧、游戏和电商等领域，场景 Token 消耗需求真实且持续。

投资分析意见。顺应国产大模型出海浪潮，相关方向梳理：1）具备优质模型能力的公司：阿里巴巴、腾讯控股、快手、昆仑万维、智谱、MiniMax 等；2）具有强 Token 场景和优质客源的相关公司，例如部分营销广告公司拥有优质海外客户资源和营销场景，可以将 Token 融入客户场景中消耗。我们建议关注 AI 营销、AI 视频化中愿意积极布局相关业务尝试的公司，包括易点天下、蓝色光标等。

风险提示。同业竞争风险，垫资与坏账风险，模型厂商政策变动风险

更多免费行业报告 并购家 www.ipoipo.cn

内容目录

1. Token 运营正在形成一个新的中间层市场 4
1.1. 新中间层的核心价值 4
1.2. Token 运营新模式的运作机制 4
1.3. Token 分销的产业进度加快 5
2. Token 运营新模式产生的背景 7
2.1.1. 中国日均 Token 调用量显著增长 7
2.1.2. 国产大模型能力实现跃迁 8
2.1.3. 全球 AI 资源的非均匀分布 10
2.1.4. 企业的 AI 能力建设门槛较高 11
2.1.5. 模型从稀缺走向饱和，渠道匹配成为瓶颈 11
3. Token 运营盈利模式梳理 13
3.1. Token 转售利差 13
3.2. 自研加速引擎下的技术溢价 13
3.3. 场景 Token 下的企业增值服务 14
4. 投资建议 16
5. 风险提示 17

更多免费行业报告 并购家 www.ipoipo.cn

图表目录

图表 1： Token 分销的核心价值 4
图表 2： Token 分销的核心三方角色 5
图表 3： OpenRouter 平台内对应处理的 Token 量级增长显著 5
图表 4： 硅基流动模型广场 6
图表 5： 特朗普家族入局 AI 模型路由赛道 6
图表 6： WorldRouter 平台提供高达 7 折的优惠 6
图表 7： 各垂类 Agent 全面落地，中国日均 Token 调用量显著增长 7
图表 8： 2026-2031 年中国企业活跃智能体关键数据预测 8
图表 9： OpenClaw 在 OpenRouter 的周度 Token 消耗量及占比 8
图表 10： 2026 年 3 月 SuperCLUE 通用测评总排行榜 8
图表 11： 2026 年 5 月 4 日至 5 月 10 日 OpenRouter 平台中各模型 Token 调用量 9
图表 12： 全球上榜大模型每周调用量趋势 9
图表 13： 上榜中国大模型企业 Token 调用量趋势 9
图表 14： 模型基准测试和价格对比 10
图表 15： 截至 2024 年 3 月，各国 ChatGPT 总访问量 10
图表 16： 中国大陆用户无法访问国外顶级大模型 10
图表 17： 除了可见成本外，企业 AI 部署有潜在的结构性成本 11
图表 18： n1n.ai 聚合平台将全球 500+ 模型全部清洗并封装成了标准的 OpenAI 接口格式 12
图表 19： OpenRouter 在供应商成本上加收约 5.5%的溢价 13
图表 20： 硅基流动自研 OneDiff 技术，提高推理效率 14
图表 21： 通过 OneDiff 优化，消费级显卡 RTX 3090 的 SDXL 推理速度，能追平甚至超越未优化的专业卡 A100 14
图表 22： 硅基流动自研 SiliconLLM 与 OneDiff 技术，将语言模型推理速度提升 10 倍，文生图效率提高 3 倍 14
图表 23： 部分行业智能转型的场景成熟度 15
图表 24： 受访企业企业认为目前企业 AI 变革所面临的挑战 15
图表 25： 硅基流动企业级 MaaS 平台架构 15

请务
更多免费行业报告 并购家 www.ipoipo.cn

1. Token 运营正在形成一个新的中间层市场

Token 运营正在形成一个新的中间层市场，即探索 Token 分销模式，连接上游大模型厂商与下游开发者、企业和个人，本质是全球 Token 的批发到零售网络的流动性基础设施。

1.1. 新中间层的核心价值

Token 分销即批量采购 AI 厂商的 API 额度，并加价转售给终端用户。分销商会在网关层面，将各类模型（如 Gemini、Claude、Kimi 等）的接口协议，转换成统一的 API 标准格式。这使得下游在部署私有化 AI 智能体平台、接入各种通信渠道时，只需通过一个 API Key，就能实现多模型的无缝切换。Token 分销核心价值在于解决终端用户的支付门槛、网络限制和技术门槛等。

图表 1：Token 分销的核心价值

核心价值	说明
网络优化	国内直连，无需代理，低延迟
接口统一	一套代码适配所有模型，降低维护成本
支付合规	支持个人支付、对公支付等
成本更低	批量采购折扣，比官方直连更加便宜
多模型聚合	一个平台调用 GPT、Claude、DeepSeek 等数种模型

资料来源：华源证券研究所整理

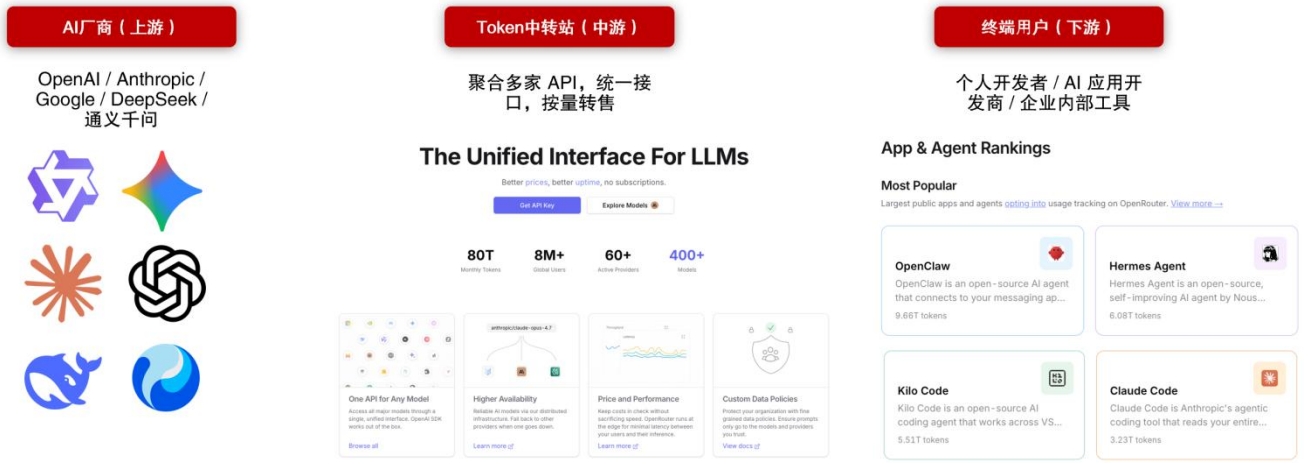
1.2. Token 运营新模式的运作机制

核心三方角色包括：**1）模型方：**如字节跳动（Seedance 系列）、阿里巴巴（Qwen 系列）、智谱（GLM 系列）、月之暗面（Kimi 系列）、深度求索（DeepSeek 系列）等，是 Token 源头供给方。**2）代理平台：**承接上游模型资源并分销给终端用户，是 Token 分销中转与统一服务的中间枢纽。**3）终端用户：**实际付费购买并调用消耗 Token，包含个人、开发者、企业及下级分销从业者等。

我们认为，Token 分销模式是轻资产的生意，其无需投入服务器集群、模型研发等重资产成本，而通过搭建 API 中转调度系统，依托上游模型方现成的 Token 算力资源，通过整合、拆分、定价和渠道分发即可开展业务，主要靠信息差、渠道差和服务差价盈利。

更多免费行业报告 并购家 www.ipoipo.cn

图表 2：Token 分销的核心三方角色

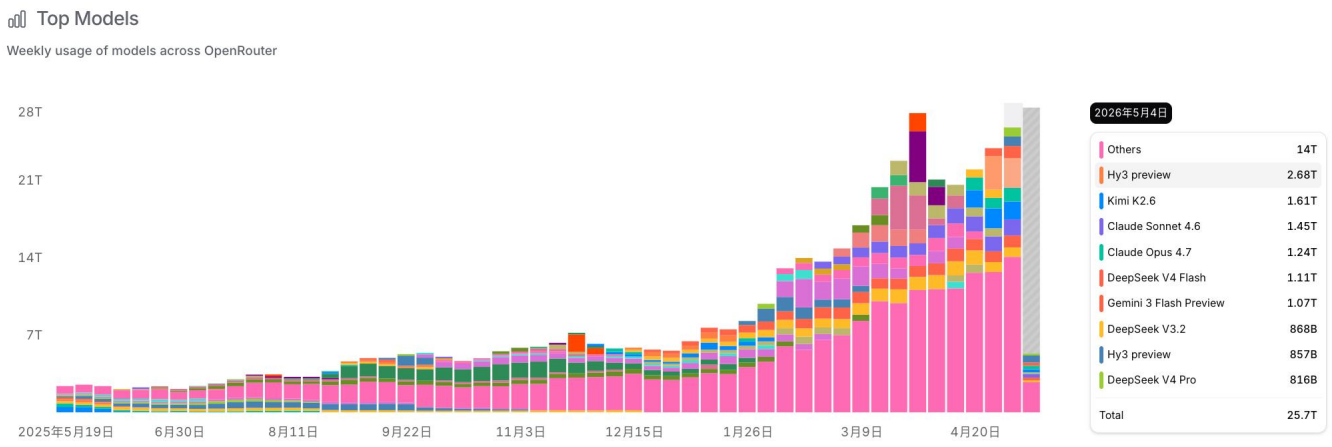


资料来源：OpenRouter，腾讯云，华源证券研究所整理

1.3. Token 分销的产业进度加快

据华尔街见闻、科技媒体 The Information，AI 模型聚合平台 OpenRouter 正在洽谈一轮新融资。在营收层面，OpenRouter 在 2026 年的年化收入已超过 5000 万美元，较 2025 年 10 月披露的逾 1000 万美元年化收入实现了约五倍增长。OpenRouter 数据显示，平台内对应处理的 Token 量级已从 2025 年的每周 5-7 万亿增长至 2026 年 4 月的每周超 20 万亿。

图表 3：OpenRouter 平台内对应处理的 Token 量级增长显著

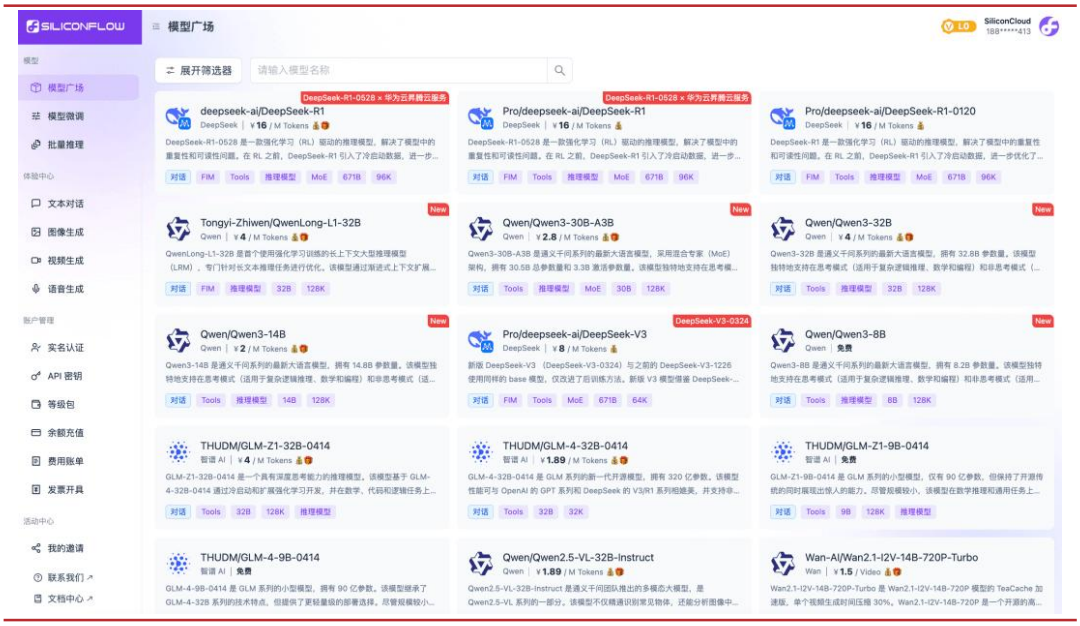


资料来源：OpenRouter，华源证券研究所

硅基流动（SiliconFlow)是一站式大模型云服务平台，基于自研推理引擎，实现大模型高效推理加速。此外，硅基流动拥有企业级大模型服务平台，为 B 端客户提供模型微调、推理部署和场景落地等解决方案。截至 2025 年 12 月，平台注册用户超 900 万，服务企业用户超 10000 位，上线模型超过 150 个。

请务
更多免费行业报告 并购家 www.ipoipo.cn

图表 4：硅基流动模型广场



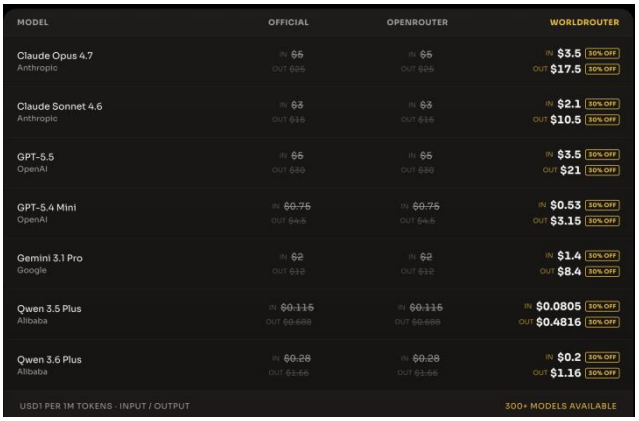
资料来源：硅基流动官网，华源证券研究所

特朗普家族入局 AI 模型路由赛道。2026 年 5 月 5 日，与美国总统特朗普及其家族有密切联系的加密货币公司 WLFI 宣布，携手 WorldClaw 推出 AI 模型统一入口 WorldRouter，整合包括 Claude 系列、GPT 系列和 Gemini 系列等超过 300 款主流模型。支付方式以 USD1 结算，定价比官方的公开费率低约 30%，以 Claude Opus 4.7 为例，官方 Input 费用为每百万 Token 5 美元，WorldRouter 仅收 3.5 美元。

图表 5：特朗普家族入局 AI 模型路由赛道

资料来源：X，华源证券研究所

图表 6：WorldRouter 平台提供高达 7 折的优惠



MODEL	OFFICIAL	OPENROUTER	WORLDROUTER
Claude Opus 4.7 Anthropic	IN $5 OUT $25	IN $5 OUT $25	IN $3.5 30% OFF OUT $17.5 30% OFF
Claude Sonnet 4.6 Anthropic	IN $3 OUT $15	IN $3 OUT $15	IN $2.1 30% OFF OUT $10.5 30% OFF
GPT-5.5 OpenAI	IN $5 OUT $30	IN $5 OUT $30	IN $3.5 30% OFF OUT $21 30% OFF
GPT-5.4 Mini OpenAI	IN $0.75 OUT $4.5	IN $0.75 OUT $4.5	IN $0.53 30% OFF OUT $3.15 30% OFF
Gemini 3.1 Pro Google	IN $2 OUT $12	IN $2 OUT $12	IN $1.4 30% OFF OUT $8.4 30% OFF
Qwen 3.5 Plus Alibaba	IN $0.115 OUT $0.688	IN $0.115 OUT $0.688	IN $0.0805 30% OFF OUT $0.4816 30% OFF
Qwen 3.6 Plus Alibaba	IN $0.28 OUT $1.65	IN $0.28 OUT $1.65	IN $0.2 30% OFF OUT $1.16 30% OFF

USD1 PER 1M TOKENS · INPUT / OUTPUT　　300+ MODELS AVAILABLE

资料来源：blocksummary，华源证券研究所

更多免费行业报告 并购家 www.ipoipo.cn

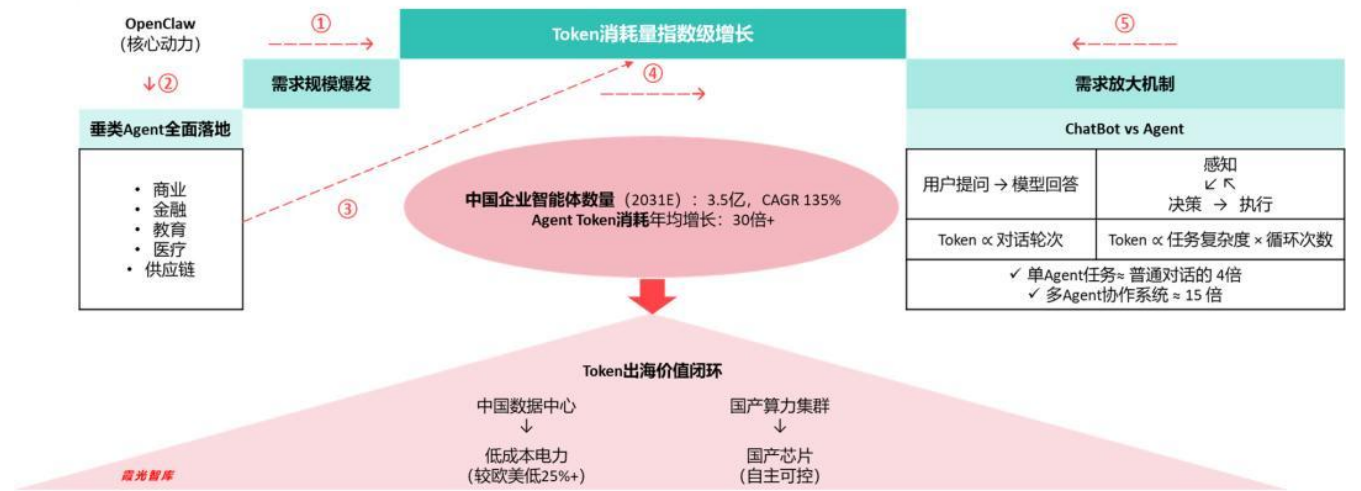
2. Token 运营新模式产生的背景

2.1.1. 中国日均 Token 调用量显著增长

2025 年 3 月 23 日，在中国发展高层论坛 2026 年年会上，国家数据局局长刘烈宏表示："Token'词元'不仅是智能时代的价值锚点，更是连接技术供给与商业需求的'结算单位'，为商业模式的落地提供了可量化的可能"。

中国日均 Token 调用量显著增长。2024 年年初，中国日均 Token 调用量为 1000 亿；至 2025 年底，跃升至 100 万亿；2026 年 3 月，已突破 140 万亿，两年增长超千倍。一套以 Token 计费为基础的新型商业逻辑正在加速演进。Token 是大模型处理信息的最小信息单元，具有智能时代可计量、可定价、可交易的特征。围绕 Token 的调用、分发与结算，一套新的价值体系正在加速演进形成，并成为人工智能产业可能变现的重要路径。

图表 7：各垂类 Agent 全面落地，中国日均 Token 调用量显著增长



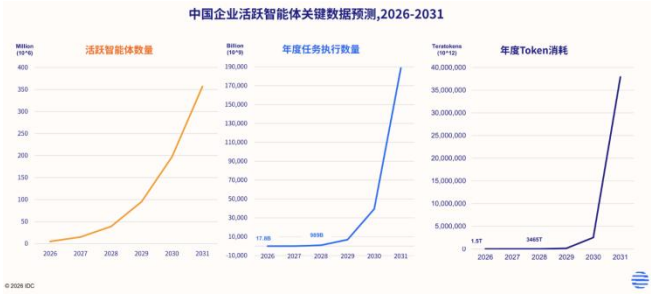
资料来源：36 氪，华源证券研究所

Token 需求加速，各垂类 Agent 全面落地。IDC 数据显示，中国企业活跃智能体数量将在 2031 年突破 3.5 亿规模，年复合增长率达到 135%以上，这一增速将领先全球主要市场。同时由于智能体任务执行密度的增长和任务复杂度的提升，也将带来智能体 Token 消耗年均超 30 倍的大幅跃升。

2026 年以来，Openclaw 等执行类智能体的发展带来了 Token 消耗量的快速增长。2026 年 2 月 2 日至 3 月 16 日，OpenClaw 在 OpenRouter 平台的周度 Token 消耗量从最初的 0.81T 快速提升至 4.97T，Token 消耗占比从 8.31%提升至 24.36%。

更多免费行业报告 并购家 www.ipoipo.cn

图表 8：2026-2031 年中国企业活跃智能体关键数据预测



资料来源：IDC，华源证券研究所

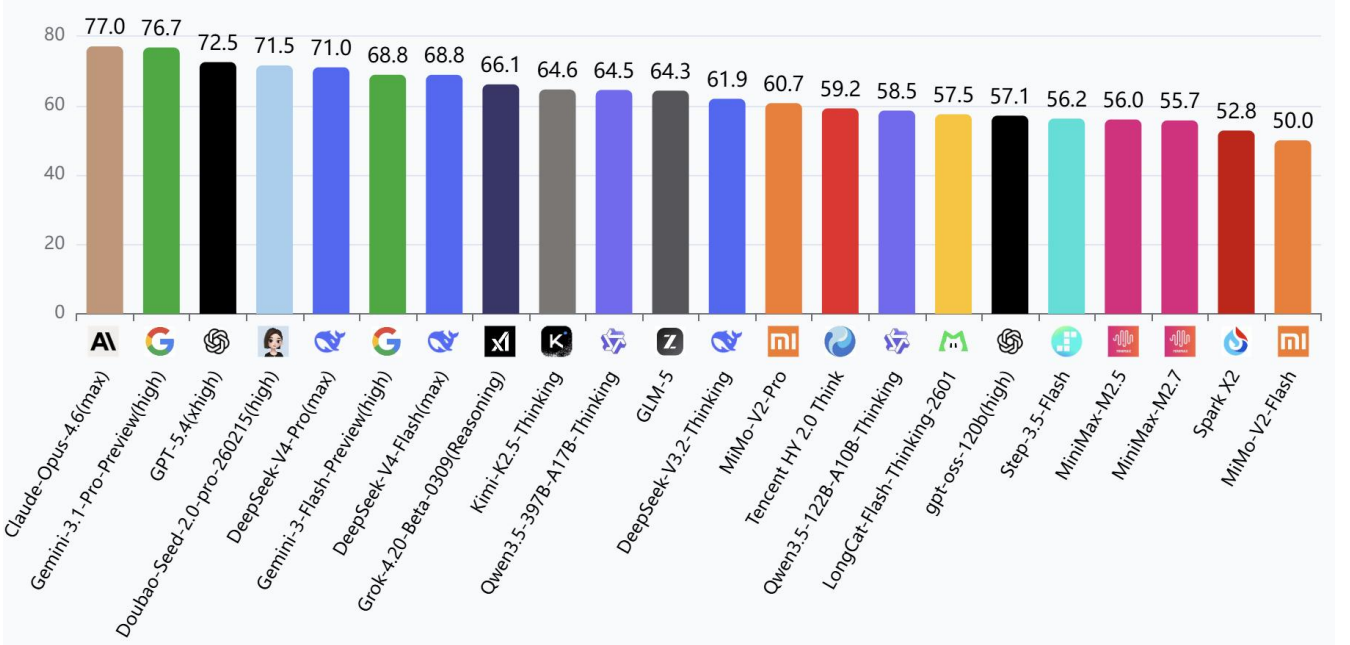
图表 9：OpenClaw 在 OpenRouter 的周度 Token 消耗量及占比

资料来源：36 氪，华源证券研究所

2.1.2. 国产大模型能力实现跃迁

国产大模型能力跃迁，Token 出海需求增长。SuperCLUE 数据显示，以字节豆包、DeepSeek 系列为代表的国产模型，综合评分已突破 70 分大关，与 GPT-5.4、Gemini 等海外头部模型的差距缩小，跻身全球第一梯队，通义千问、Kimi、智谱 GLM 等模型也表现亮眼，形成了梯队分明的竞争格局。OpenRouter 数据显示，截至 2026 年 5 月 10 日的最新一周榜单中，腾讯 Hy3 preview (free)位居调用量榜首，前五、前十、前二十名分别有 2 款、6 款、9 款国产大模型。

图表 10：2026 年 3 月 SuperCLUE 通用测评总排行榜



资料来源：SuperCLUE，华源证券研究所

更多免费行业报告 并购家 www.ipoipo.cn

图表 11：2026 年 5 月 4 日至 5 月 10 日 OpenRouter 平台中各模型 Token 调用量

LLM Leaderboard

Compare the most popular models on OpenRouter

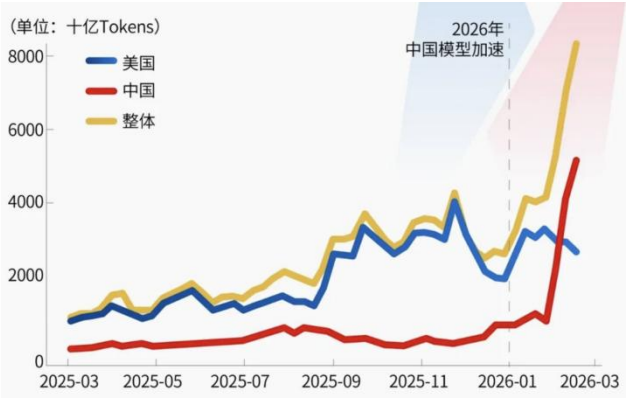
This Week

1.	Hy3 preview (free) by tencent	3.75T tokens	↑74%
2.	Kimi K2.6 by moonshotai	1.72T tokens	↓9%
3.	Claude Sonnet 4.6 by anthropic	1.43T tokens	↑5%
4.	Claude Opus 4.7 by anthropic	1.14T tokens	↑21%
5.	Gemini 3 Flash Preview by google	1.06T tokens	↑8%
6.	DeepSeek V4 Flash by deepseek	975B tokens	↑58%
7.	DeepSeek V3.2 by deepseek	836B tokens	↓15%
8.	MiniMax M2.7 by minimax	766B tokens	↑8%
9.	DeepSeek V4 Pro by deepseek	763B tokens	↑187%
10.	Grok 4.1 Fast by x-ai	730B tokens	↑13%
11.	Gemini 2.5 Flash Lite by google	655B tokens	↑9%
12.	Gemini 2.5 Flash by google	586B tokens	↓3%
13.	Step 3.5 Flash by stepfun	559B tokens	↓31%
14.	Claude Opus 4.6 by anthropic	546B tokens	↓5%
15.	Nemotron 3 Super (free) by nvidia	520B tokens	↓19%
16.	gpt-oss-120b by openai	397B tokens	↑10%
17.	Ling-2.6-1T (free) by inclusionai	364B tokens	↓32%
18.	GLM 5.1 by z-ai	355B tokens	↓3%
19.	Gemini 3.1 Flash Lite Previ... by google	330B tokens	↑17%
20.	GPT-5.5 by openai	322B tokens	↑43%

资料来源：OpenRouter，华源证券研究所

2026 年一季度，在全球 AI 模型 API 聚合平台 OpenRouter 中，中国大模型 Token 调用量首次超越美国。OpenRouter 数据显示，2026 年 2 月 9 日至 15 日，中国模型以 4.12 万亿 Token 的调用量，首次超过同期美国模型的 2.94 万亿 Token。2 月 16 日至 22 日，中国模型的周调用量进一步增长至 5.16 万亿 Token，平台调用量排名前五的模型中有四款来自中国厂商，分别为 MiniMax M2.5、Kimi K2.5、智谱 GLM-5 及 DeepSeek V3.2，合计贡献 Top5 总调用量的 85.7%。

图表 12：全球上榜大模型每周调用量趋势



资料来源：每经网，华源证券研究所

图表 13：上榜中国大模型企业 Token 调用量趋势

资料来源：每经网，华源证券研究所

更多免费行业报告 并购家 www.ipoipo.cn

对比海外头部模型，国产大模型性能接近，成本更低。OpenRouter 数据显示，性能上，MiniMax M2.5 在 SWE-Bench Verified、BFCL Multi-Turn 两项指标上接近或超过 Claude Opus 4.6，GLM 5 则在 SWE-Bench Multilingual 中大幅领先。成本上，国产模型更具优势，输入价格约为 Claude 的 1/17，输出价格最高仅为 Claude 的约 1/10。

我们认为，国产模型在 AI Agent、代码开发等高 Token 消耗场景中具备性价比竞争力，从而推动 Token 出海的发展。Token 出海本质上是中国本土 AI 模型依托标准化 API 通道，面向全球市场交付按 Token 计费的“推理即服务”。随着全球 AI 算力需求的膨胀，中国有望凭借低成本优势，实现算力资源的高效跨境流动与变现。

图表 14：模型基准测试和价格对比

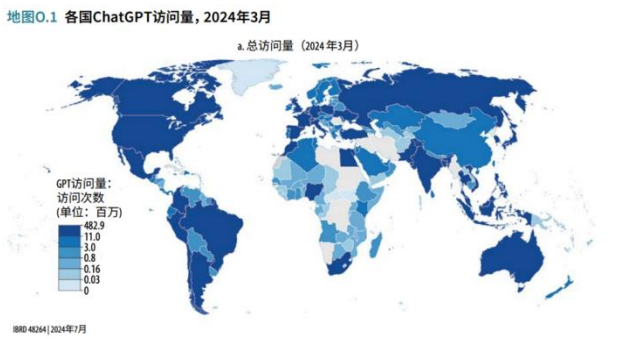
评测项	MiniMax M2.5	GLM 5	Claude Opus 4.6
SWE-Bench Verified	0.802	0.778	0.808
BFCL Multi-Turn	0.768	—	0.633
SWE-Bench Multilingual	—	0.733	0.562
Terminal-Bench 2.0	—	0.562	0.654
输入价格（每百万 Token）	**0.3 美元**	**0.3 美元**	**5 美元**
输出价格（每百万 Token）	**1.1 美元**	**2.55 美元**	**25 美元**

资料来源：每经网，OpenRouter，华源证券研究所

2.1.3. 全球 AI 资源的非均匀分布

全球 AI 算力、模型、能源等核心资源在各国各地区呈现不均衡的分布格局。一方面，海外头部大模型如 OpenAI 等，受地域访问限制、合规规则与支付门槛制约，无法直接触达包括中国大陆开发者在内的海量用户。另一方面，以 DeepSeek 为代表的国产优质大模型，在走向国际市场时，又面临海外本地化适配、渠道铺设与用户获客的天然壁垒。我们分析，这种全球 AI 资源的非均匀分布或催生了 AI Token 的跨境流转、聚合路由与分层分销的市场需求。

图表 15：截至 2024 年 3 月，各国 ChatGPT 总访问量



资料来源：世界银行，华源证券研究所

图表 16：中国大陆用户无法访问国外顶级大模型

资料来源：OpenAI 官网，华源证券研究所

更多免费行业报告 并购家 www.ipoipo.cn

2.1.4. 企业的 AI 能力建设门槛较高

除了基础的 Token 分销外，AI 相关的增值服务有望形成更为广阔的溢价空间。随着 DeepSeek 这类开源、低成本国产大模型快速普及，市场 AI 调用的基础 Token 单价被大幅拉低。但对企业客户而言，AI 使用成本更多来自于隐形成本，包括专业 Prompt 工程调试、不同业务场景下的多模型对比选型、模型与自身业务系统深度对接集成、工作流编排、运维调度以及员工 AI 应用能力搭建等一系列 AI 能力建设投入。

图表 17：除了可见成本外，企业 AI 部署有潜在的结构性成本

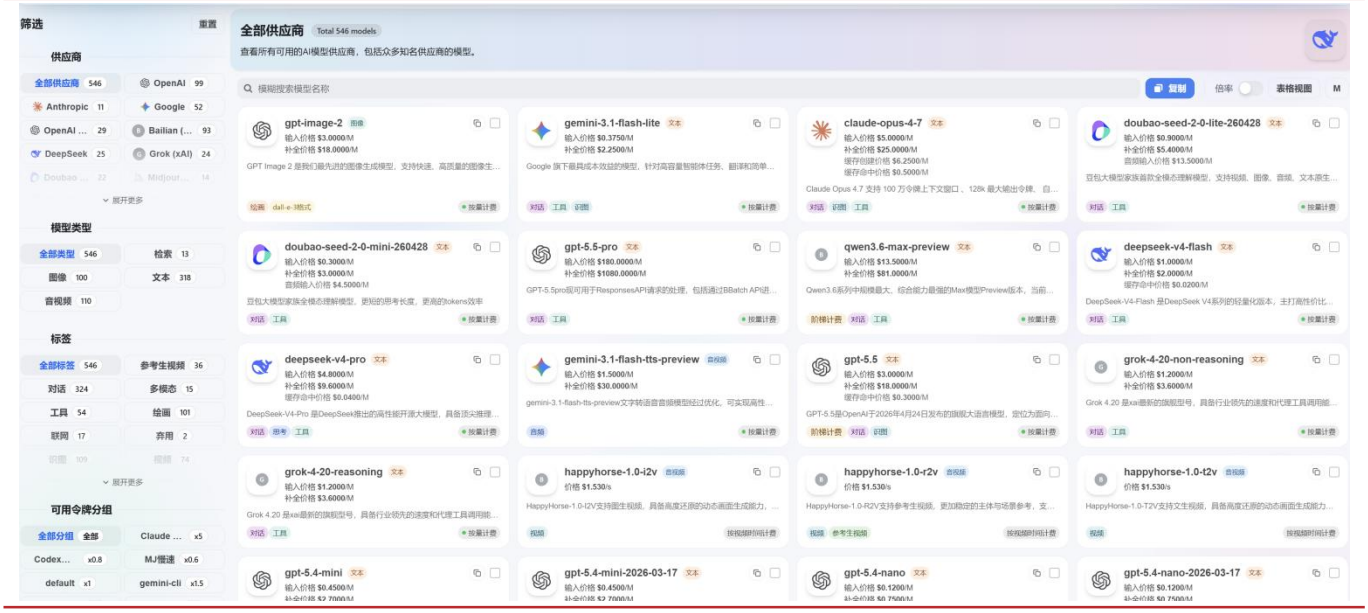
资料来源：truefoundry，华源证券研究所

2.1.5. 模型从稀缺走向饱和，渠道匹配成为瓶颈

当下 AI 市场缺乏统一的行业标准。随着 DeepSeek 等开源大模型、国产自研模型批量落地，叠加海外主流模型持续迭代上线，市场已从大模型稀缺供不应求，进入到模型供给或趋于饱和的新阶段。我们认为，当下真正的痛点变成供需两端的对接通路不畅。一方面海量模型与闲置推理算力找不到精准业务场景和终端用户，另一方面企业、开发者、AI 应用方又难以快速筛选适配自身业务的模型、拿到稳定合规的调用渠道与合理定价。因此，高效的资源聚合、智能路由、分级分销、场景匹配与落地服务等渠道能力，反而成为打通算力上下游的关键瓶颈。不同模型供应商的并发配额、请求格式、速率限制和错误处理代码各不相同。Token 分销商通过把复杂的后端交互封装在标准接口上，让应用层只需对接一套 API，就能具备调用各家模型的能力。

更多免费行业报告 并购家 www.ipoipo.cn

图表 18：n1n.ai 聚合平台将全球 500+ 模型全部清洗并封装成了标准的 OpenAI 接口格式



资料来源：n1n.ai，华源证券研究所

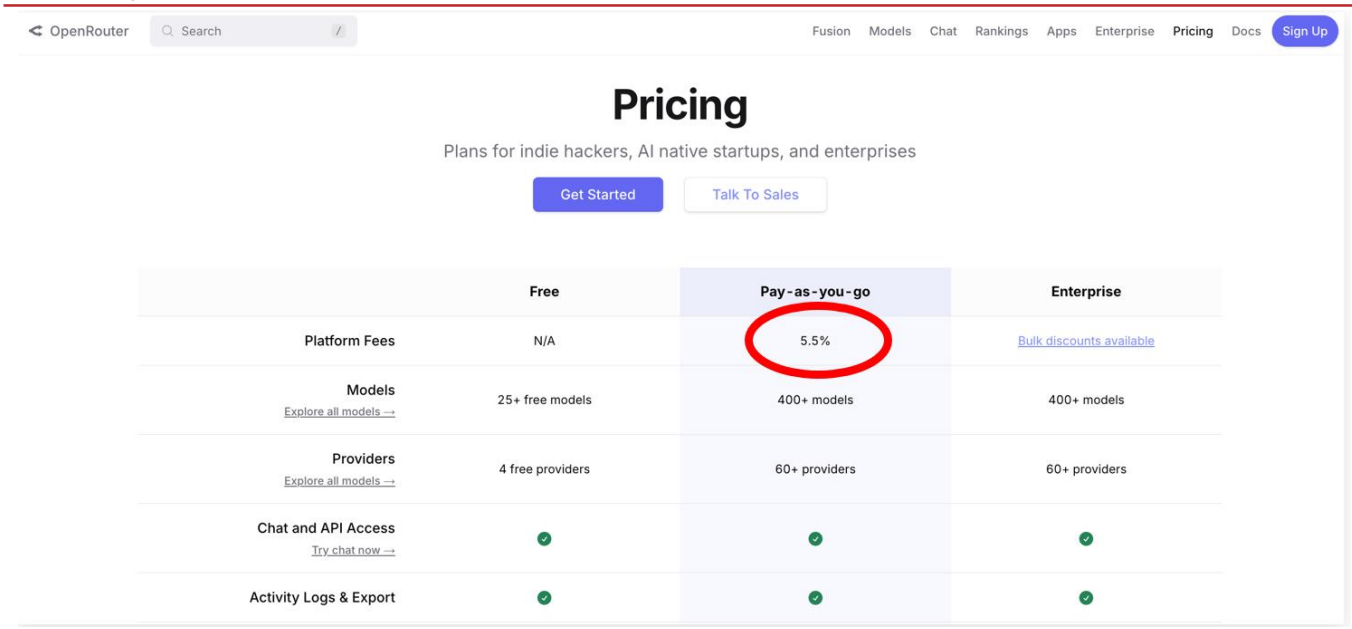
更多免费行业报告 并购家 www.ipoipo.cn

3. Token 运营盈利模式梳理

3.1. Token 转售利差

Token 转售利差即在模型厂商（上游）和 AI 需求方（下游）之间做流量聚合与分发，赚取差价。上游批量采购有折扣价，下游聚集大量客户统一消耗，中间加收一定比例的通道费。如 OpenRouter 在供应商成本上加收约 5.5%的溢价。国内代理商通过整合国产大模型，打包成可直接调用的 API 服务，卖给海外客户，即 Token 出海。

图表 19：OpenRouter 在供应商成本上加收约 5.5%的溢价



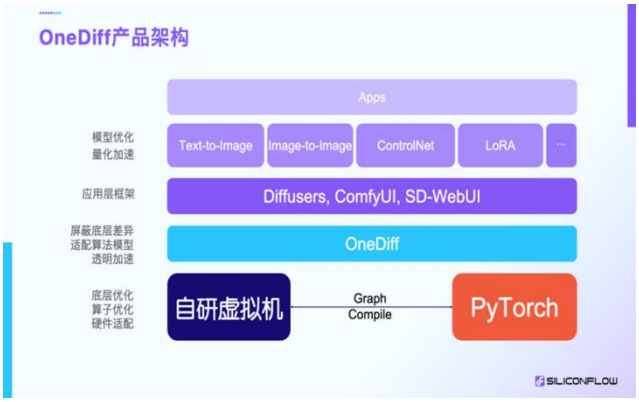
资料来源：OpenRouter，华源证券研究所

3.2. 自研加速引擎下的技术溢价

Token 分销的技术溢价，主要是通过自研推理加速引擎，在不依赖上游低价货源的前提下，将模型推理效率大幅提升，把单 Token 运行成本降低，从而在保持与上游官方相近甚至更低售价的同时，依靠算力效率差获取超额技术毛利。如硅基流动自研 SiliconLLM 与 OneDiff 技术，将语言模型推理速度提升 10 倍，文生图效率提高 3 倍，使得大模型 API 调用成本低至行业的 1/10。

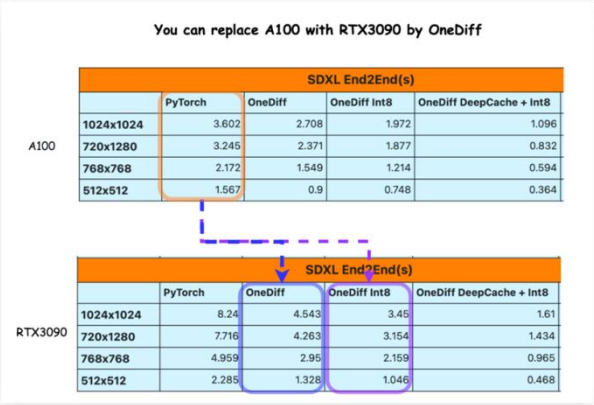
更多免费行业报告 并购家 www.ipoipo.cn

图表 20：硅基流动自研 OneDiff 技术，提高推理效率



资料来源：硅基流动，知乎，华源证券研究所

图表 21：通过 OneDiff 优化，消费级显卡 RTX 3090 的 SDXL 推理速度，能追平甚至超越未优化的专业卡 A100



You can replace A100 with RTX3090 by OneDiff

A100

SDXL End2End(s)	PyTorch	OneDiff	OneDiff Int8	OneDiff DeepCache + Int8
1024x1024	3.602	2.708	1.972	1.096
720x1280	3.245	2.371	1.877	0.832
768x768	2.172	1.549	1.214	0.594
512x512	1.567	0.9	0.748	0.364

RTX3090

SDXL End2End(s)	PyTorch	OneDiff	OneDiff Int8	OneDiff DeepCache + Int8
1024x1024	8.24	4.543	3.45	1.61
720x1280	7.716	4.263	3.154	1.434
768x768	4.959	2.95	2.159	0.965
512x512	2.285	1.328	1.046	0.468

资料来源：Github，华源证券研究所

图表 22：硅基流动自研 SiliconLLM 与 OneDiff 技术，将语言模型推理速度提升 10 倍，文生图效率提高 3 倍

资料来源：硅基流动官网，华源证券研究所

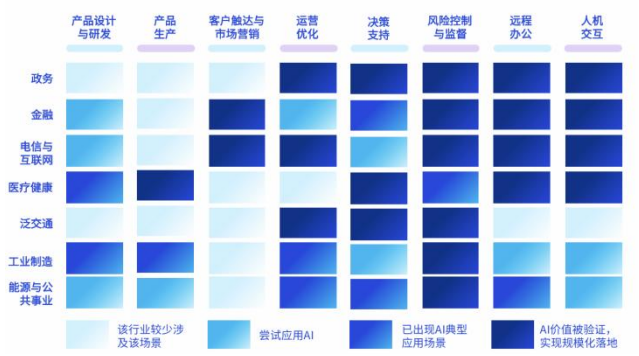
3.3. 场景 Token 下的企业增值服务

生成式 AI 的应用场景逐渐铺开，包括医疗健康、泛交通、工业制造等多个行业。业务层面来看，生成式 AI 愈发深入参与到企业核心业务流程中，满足决策支持、战略管理等场景需求。然而，部分企业面临智能化转型基础薄弱、数据资产积累不足 、算力投入有限等短板。

除了基础的 Token 分销外，关注场景 Token。即面向企业端提供 AI 相关的技术支持，如 prompt 工程、Agent 编排、模型选型等。例如，硅基流动建立了企业级 MaaS 平台架构，面对企业用户提供模型训练调优、部署推理、应用开发支撑三层能力，包括数据处理、模型微调、Prompt 工程和 RAG 等，最终以标准化 API 形式交付给能源、金融、政府等多行业。

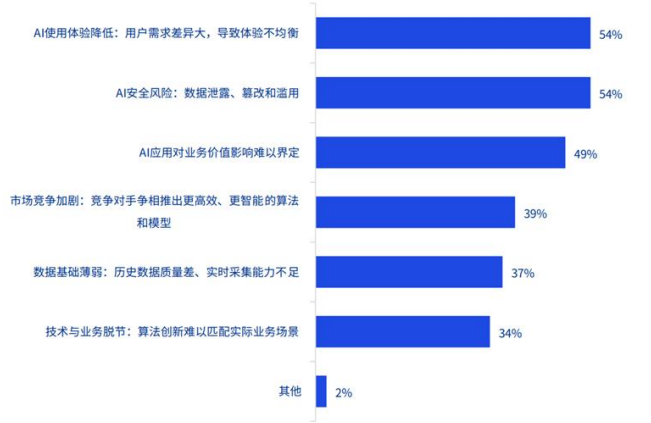
更多免费行业报告 并购家 www.ipoipo.cn

图表 23：部分行业智能转型的场景成熟度



资料来源：毕马威&思科《人工智能就绪度白皮书》，华源证券研究所

图表 24：受访企业企业认为目前企业 AI 变革所面临的挑战



资料来源：毕马威&思科《人工智能就绪度白皮书》，华源证券研究所

图表 25：硅基流动企业级 MaaS 平台架构

资料来源：硅基流动官网，华源证券研究所

更多免费行业报告 并购家 www.ipoipo.cn

4. 投资建议

顺应国产大模型出海浪潮，相关方向梳理：

1）具备优质模型能力公司：阿里巴巴、腾讯控股、快手、昆仑万维、智谱、MiniMax 等；

2）具有强 Token 场景和优质客源的相关公司，例如部分营销广告公司拥有优质海外客户资源和营销场景，可以将 Token 融入客户场景中消耗。

我们建议关注 AI 营销、AI 视频化中愿意积极布局相关业务尝试的公司，包括易点天下、蓝色光标等。

请务
更多免费行业报告 并购家 www.ipoipo.cn

5. 风险提示

1）同业竞争风险：Token 分销业务技术门槛较低，头部代理商可能凭借资金、客户与渠道优势随时入场，快速复制分销模式并挤压利润空间。

2）垫资与坏账风险：分销商对下游客户普遍采用月结或季结模式，但向上游模型厂商采购 API 时往往需垫资。随着 Token 消耗量的增长，垫资规模也会同步放大。一旦客户经营不善或拖欠，可能会有坏账风险。

3）模型厂商政策变动风险：大模型厂商掌握着定价权和接入规则的最终决定权。它们可能随时调整 API 价格，也可能直接收紧第三方接入政策。

更多免费行业报告 并购家 www.ipoipo.cn

证券分析师声明

本报告署名分析师在此声明，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注册为证券分析师，本报告表述的所有观点均准确反映了本人对标的证券和发行人的个人看法。本人以勤勉的职业态度，专业审慎的研究方法，使用合法合规的信息，独立、客观的出具此报告，本人所得报酬的任何部分不曾与、不与，也不将会与本报告中的具体投资意见或观点有直接或间接联系。

一般声明

华源证券股份有限公司（以下简称“本公司”）具有中国证监会许可的证券投资咨询业务资格。
本报告是机密文件，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本公司客户。本报告是基于本公司认为可靠的已公开信息撰写，但本公司不保证该等信息的准确性或完整性。本报告所载的资料、工具、意见及推测等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邀请或向人作出邀请。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、财务状况和特殊需求，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。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。
本报告所载的意见、评估及推测仅反映本公司于发布本报告当日的观点和判断，在不同时期，本公司可发出与本报告所载意见、评估及推测不一致的报告。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。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，过往的业绩表现不应作为日后回报的预示。本公司不承诺也不保证任何预示的回报会得以实现，分析中所做的预测可能是基于相应的假设，任何假设的变化可能会显著影响所预测的回报。本公司不保证本报告所含信息保持在最新状态。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。
本报告的版权归本公司所有，属于非公开资料。本公司对本报告保留一切权利。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修改、复制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。如征得本公司许可进行引用、刊发的，需在允许的范围内使用，并注明出处为“华源证券研究所”，且不得对本报告进行任何有悖原意的引用、删节和修改。本公司保留追究相关责任的权利。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。
本公司销售人员、交易人员以及其他专业人员可能会依据不同的假设和标准，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或交易观点，本公司没有就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。

信息披露声明

在法律许可的情况下，本公司可能会持有本报告中提及公司所发行的证券并进行交易，也可能为这些公司提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。本公司将会在知晓范围内依法合规的履行信息披露义务。因此，投资者应当考虑到本公司及/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一参考依据。

投资评级说明

证券的投资评级：以报告日后的 6 个月内，证券相对于同期市场基准指数的涨跌幅为标准，定义如下：

买入：相对同期市场基准指数涨跌幅在 20%以上；
增持：相对同期市场基准指数涨跌幅在 5%～20%之间；
中性：相对同期市场基准指数涨跌幅在－5%～＋5%之间；
减持：相对同期市场基准指数涨跌幅低于–5%及以下。
无：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。

行业的投资评级：以报告日后的 6 个月内，行业股票指数相对于同期市场基准指数的涨跌幅为标准，定义如下：

看好：行业股票指数超越同期市场基准指数；
中性：行业股票指数与同期市场基准指数基本持平；
看淡：行业股票指数弱于同期市场基准指数。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议；
投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者应阅读整篇报告，以获取比较完整的观点与信息，不应仅仅依靠投资评级来推断结论。

本报告采用的基准指数：A 股市场基准为沪深 300 指数，香港市场基准为恒生中国企业指数（HSCEI），美国市场基准为标普 500 指数或者纳斯达克指数。

更多免费行业报告 并购家 www.ipoipo.cn