

中国模型加速崛起，应用端入口价值凸显

OpenRouter视角下的大模型与AI应用趋势

行业研究 · 专题报告

传媒 · 计算机

投资评级：优于大市（维持）

证券分析师：张衡

021-60875160

zhangheng2@guosen.com.cn

S0980517060002

证券分析师：熊莉

021-61761067

xiongli@guosen.com.cn

S0980519030002

证券分析师：陈瑶蓉

021-61761058

chenyaorong@guosen.com.cn

S0980523100001

推理需求扩张叠加模型快速轮换，有望推理需求进入高速扩张期，增长动力正由简单聊天向复杂工作流迁移。 1) OpenRouter作为连接模型厂商、推理服务商与AI应用开发者的中立型API聚合平台，其流量不能直接代表全球大模型需求，但可以作为观察开发者推理活动的重要窗口。2025年1月至2026年7月，平台月度日均Token由0.072T增长至7.995T，19个月增长约111倍，2026年7月单日峰值达到9.622T。2) 增长过程并非平滑线性，而是伴随新模型上线、免费版本放量以及开发者工作流迁移，表现出明显的阶跃式特征。随着推理、长上下文和工具调用逐渐成为主流，Token增长开始更多来自单次任务复杂度提升，而不仅是聊天请求数量增加，推理需求正在进入由Agent化任务驱动的新阶段。

模型供给进入百花齐放阶段，中国模型正在成为全球开发者渠道的重要力量。 1) OpenRouter最新目录包含337个模型和58家厂商，其中271个支持工具调用、212个支持推理参数、92个支持百万级上下文，但进入最新日度流量榜的模型仅约50个，前十大模型仍获得约三分之二的Token。模型竞争的关键已经不是能否上架，而是性价比、推理速度、服务稳定性以及能否进入高频应用工作流。2) 中国模型凭借能力快速迭代和成本优势，月度Token份额由2025年1月的6.0%提升至2026年7月的61.5%，最新前十大具体模型中有7个来自中国厂商。不过，头部模型生命周期正在明显缩短：半年后最新前十已经完成全部换血，模型层面的竞争优势由依赖单一爆款，转向持续推出新版本、控制推理成本并完成流量接力。3) 中国模型厂商优势持续展现，看好具备持续能力迭代的头部厂商。

持续提升CSP和推理平台的产业价值。 1) 当模型迭代周期短于企业硬件采购和部署周期，企业自行押注单一模型、建设专用算力的错配风险上升，而CSP能够通过共享资源池承接不同模型和应用的波动性需求，并提供多模型接入、弹性部署、负载调度、缓存复用、统一计费、数据安全及企业级运维服务。2) 海外相关云服务厂商有AWS、Microsoft Azure、Google Cloud和Oracle OCI，国内相关厂商入阿里云、腾讯云等具备算力规模、模型生态和客户入口优势的厂商以及算力租赁相关产业链；3) 需要强调的是，Token增长并不必然等同于云厂商利润增长，CSP价值兑现仍取决于AI云收入增速、GPU利用率、推理价格竞争，以及资本开支和折旧压力；具备资源规模、全栈平台能力和企业客户基础的厂商更可能受益。

成本下降释放AI应用需求弹性，重点关注Agent/Coding和高价值垂直场景。 1) 模型能力持续提升、推理价格下降及多模型API接入门槛降低，将显著降低AI应用的开发、试错与规模化部署成本，并通过需求弹性带动调用频次、上下文长度和任务复杂度持续提升；2) 结合OpenRouter数据，Agent/Coding应用仅占应用数量的19%，却贡献81.2%的Token，头部应用每请求Token约为其他应用的2.7倍，表明能够执行多步骤任务、调用外部工具并嵌入专业工作流的应用，更有可能成为推理需求的主要增量来源。3) 建议B端重点关注Coding Agent、通用任务型Agent，以及工业、医疗、金融、法律、设计等具有明确付费意愿和较高任务价值的垂直场景；C端建议关注具备IP/版权资源、内容生产场景、流量入口、算力优化及运营能力的公司，关注AI漫剧/游戏/情感陪伴/营销服务等方向。

风险提示：技术进步低于预期、政策风险、商业化低于预期等。

- [01] 推理需求带动OpenRouter流量进入高速增长期
- [02] 模型供给：供给扩张但流量仍集中于头部
- [03] 中国模型：由边缘供给转向主要流量来源
- [04] 竞争格局：爆款生命周期缩短，持续迭代成为关键
- [05] 应用生态：Agent/Coding是当前Token需求核心，集中度强于大模型
- [06] 投资观点：关注CSP与高价值AI应用

1. 从OpenRouter数据看趋势：推理流量进入加速期，增长动力转向复杂工作流

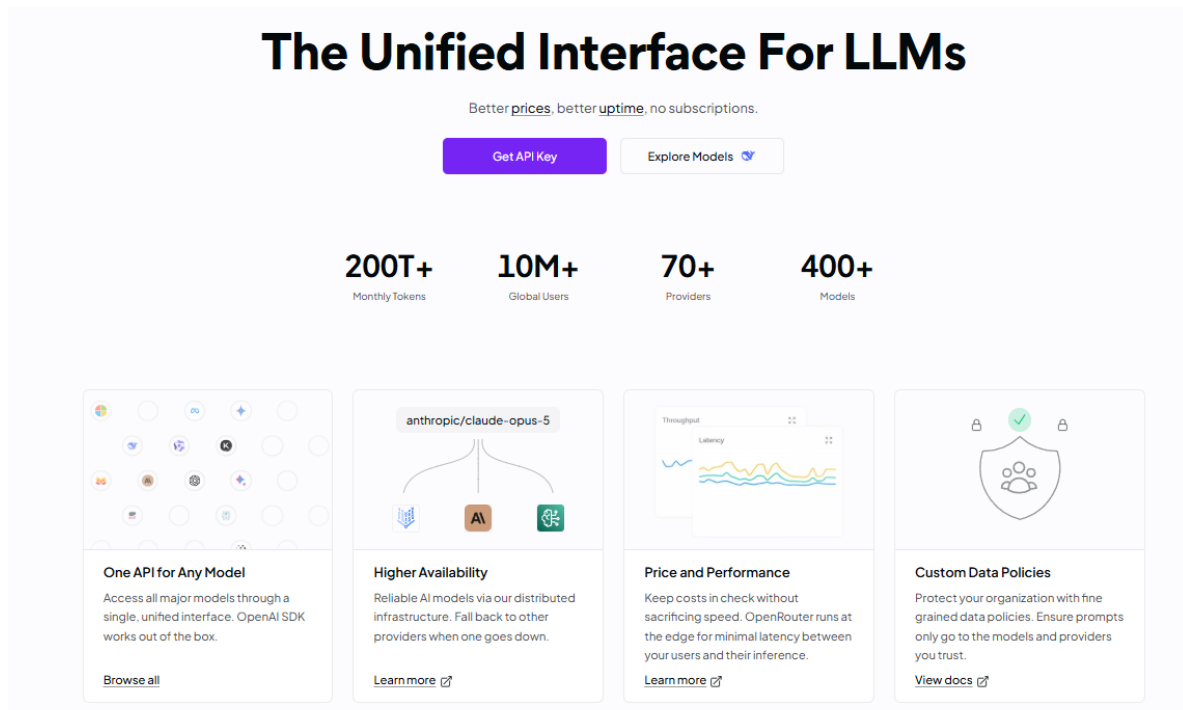
1.1 OpenRouter：连接模型供给与AI应用需求的“路由与交易层”



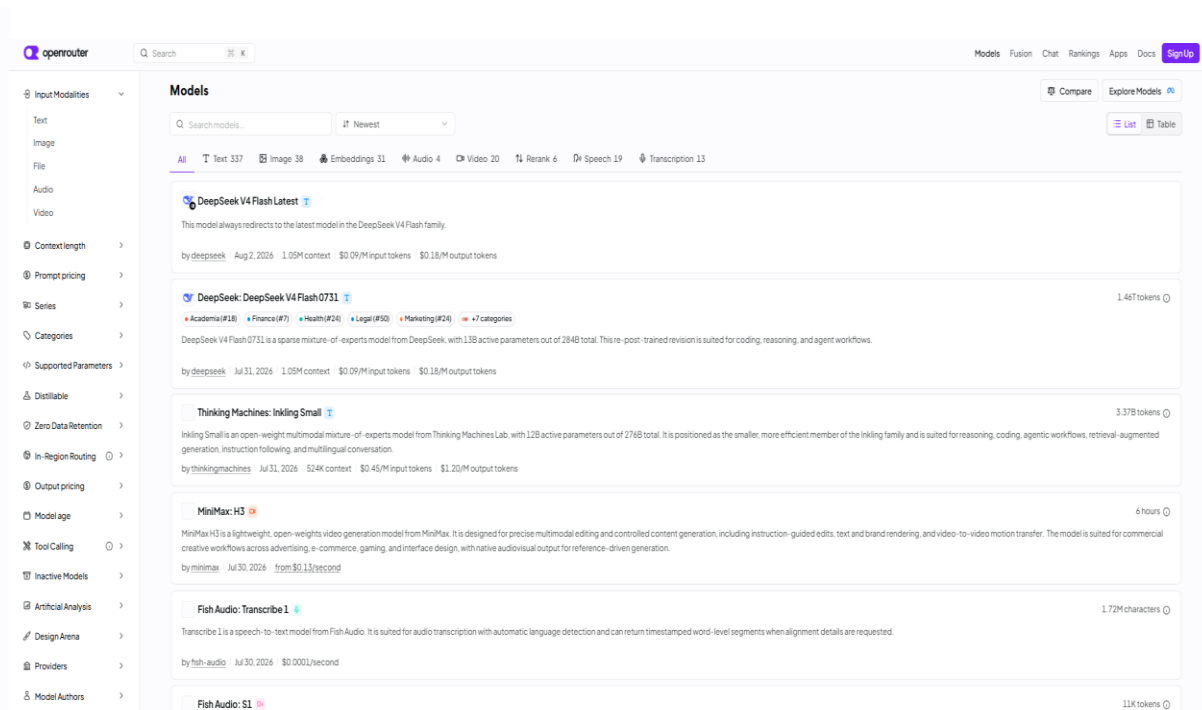
- OpenRouter汇集多个模型厂商、开发者与应用入口，其平台Token流量可以作为观察模型中立型开发者渠道活跃度的重要窗口。2025年1月至2026年7月OpenRouter月度日均Token由0.072T增长至7.995T，19个月增长约111倍；2026年7月17日单日Token进一步达到9.622T创新高；具体来看，
 - OpenRouter不是大模型厂商，而是连接模型开发商、推理服务商与AI应用开发者的中立API聚合平台。开发者只需接入一套接口，即可调用不同模型，并根据价格、时延、吞吐量、可用性和数据政策，在多个推理服务商之间进行智能路由与故障切换
 - 从商业模式看，OpenRouter本质上提供的是“模型交易基础设施”：上游模型及算力厂商借助平台获得分发渠道和增量流量，下游开发者则降低多模型接入、统一结算、供应商切换和稳定性管理成本；平台主要通过充值服务费、BYOK服务费及企业级服务实现变现。
 - 随着模型供给持续扩张、热门模型快速轮换，以及Agent应用调用量上升，OpenRouter的产业价值正由简单的API聚合，进一步升级为大模型流量的定价、分发与调度中枢。

图：OpenRouter界面

图：OpenRouter模型库



资料来源：OpenRouter，国信证券经济研究所整理



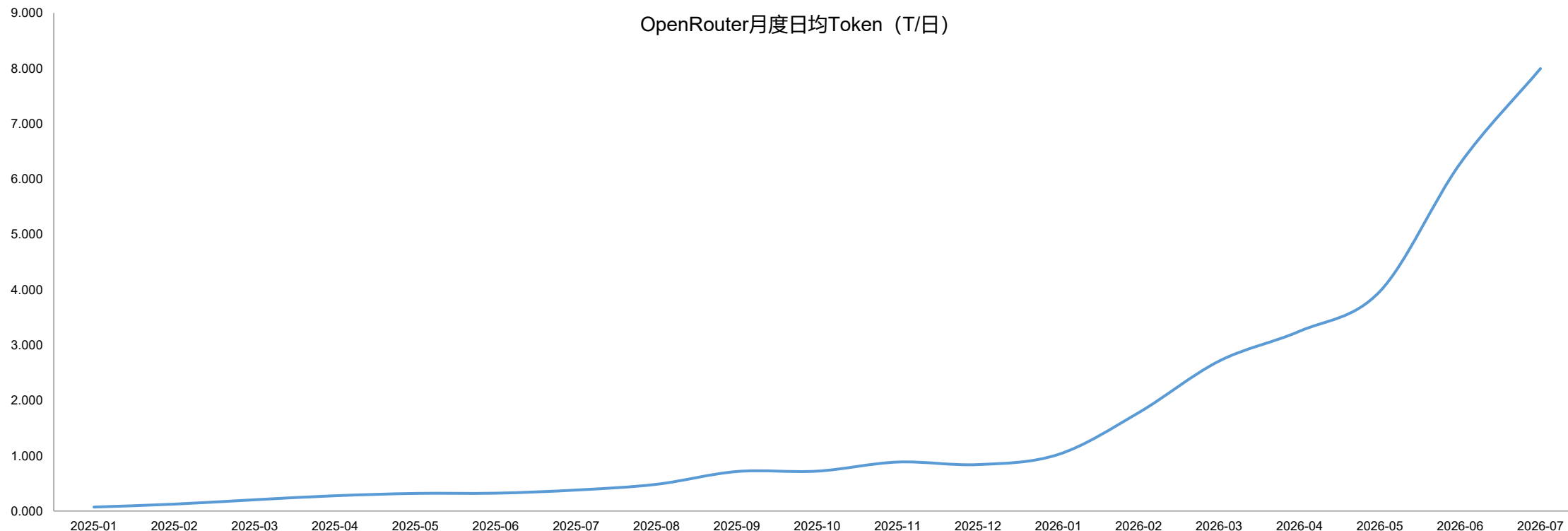
资料来源：OpenRouter，国信证券经济研究所整理

1.2 OpenRouter 日均Token在19个月内增长111倍，7月均值接近8T

➤ OpenRouter汇集多个模型厂商、开发者与应用入口，其平台Token流量可以作为观察模型中立型开发者渠道活跃度的重要窗口。2025年1月至2026年7月

OpenRouter月度日均Token由0.072T增长至7.995T，19个月增长约111倍；2026年7月17日单日Token进一步达到9.622T创新高；具体来看，

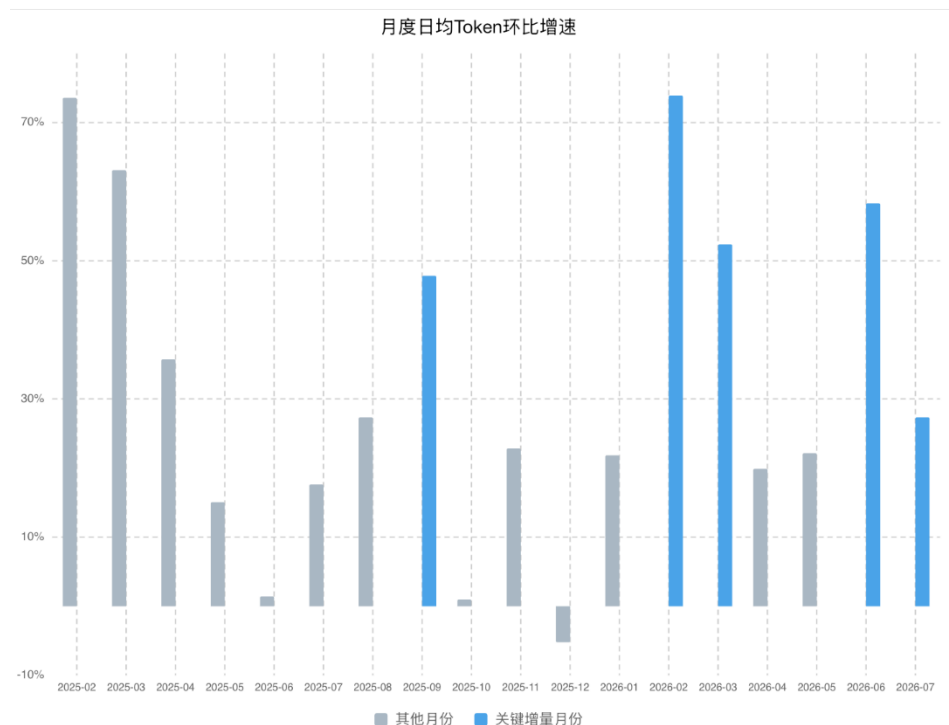
- 平台流量在2026年明显加速：日均Token由2026年1月的1.022T提升至6月的6.280T，并在7月接近8T
- 最近30天日均Token约8.03T，显示增长并非由单日异常值驱动，而是平台流量中枢整体上移
- OpenRouter正在承接更大规模的开发者推理负载，平台角色逐步从多模型调用入口扩展为AI应用工作流的重要基础设施可观察Token规模已进入万亿级日均流量阶段，模型中立型开发者渠道正在快速承接规模化推理需求



资料来源：OpenRouter，国信证券经济研究所整理

1.3 新模型上线与 workflow 迁移推动Token呈阶跃式增长

- OpenRouter Token增长并非平滑线性上升，而是表现为多个阶段性跃升。每一轮增量对应的主要模型均明显不同，说明平台增长不是由单一模型长期推动，而是由新模型上线、免费版本放量和开发者 workflow 迁移共同形成：
- 2025年9月：日均Token环比增长47.8%；主要增量来自Grok Code Fast、Grok 4 Fast免费版和DeepSeek V3.1免费版，Coding与免费模型成为这一阶段的主要流量来源。
 - 2026年2—3月：连续两个月增长73.9%和52.4%；2月主要由MiniMax M2.5、Kimi K2.5和GLM-5贡献；3月进一步由小米MiMo V2 Pro、Step 3.5 Flash免费版和Claude 4.6 Sonnet接力，新增流量呈现多厂商共同扩张。
 - 2026年6月：日均Token环比增长58.3%；MiniMax M3、小米MiMo V2.5和DeepSeek V4 Flash分别贡献约0.518T、0.464T和0.314T日均增量，形成新一轮模型集群放量。
 - 2026年7月：日均Token继续增长27.3%；腾讯HY3免费版、小米MiMo V2.5和GLM-5.2成为主要新增量来源，其中HY3免费版贡献约0.625T日均增量



资料来源： OpenRouter，国信证券经济研究所整理

关键增量阶段及主要模型贡献		
2025-09	+47.8%	Grok Code Fast +0.147T；Grok 4 Fast免费 +0.046T；DeepSeek V3.1免费 +0.024T
2026-02	+73.9%	MiniMax M2.5 +0.194T；Kimi K2.5 +0.139T；GLM-5 +0.067T
2026-03	+52.4%	MiMo V2 Pro +0.222T；Step 3.5 Flash免费 +0.139T；Claude 4.6 Sonnet +0.100T
2026-06	+58.3%	MiniMax M3 +0.518T；MiMo V2.5 +0.464T；DeepSeek V4 Flash +0.314T
2026-07	+27.3%	HY3免费 +0.625T；MiMo V2.5 +0.578T；GLM-5.2 +0.326T

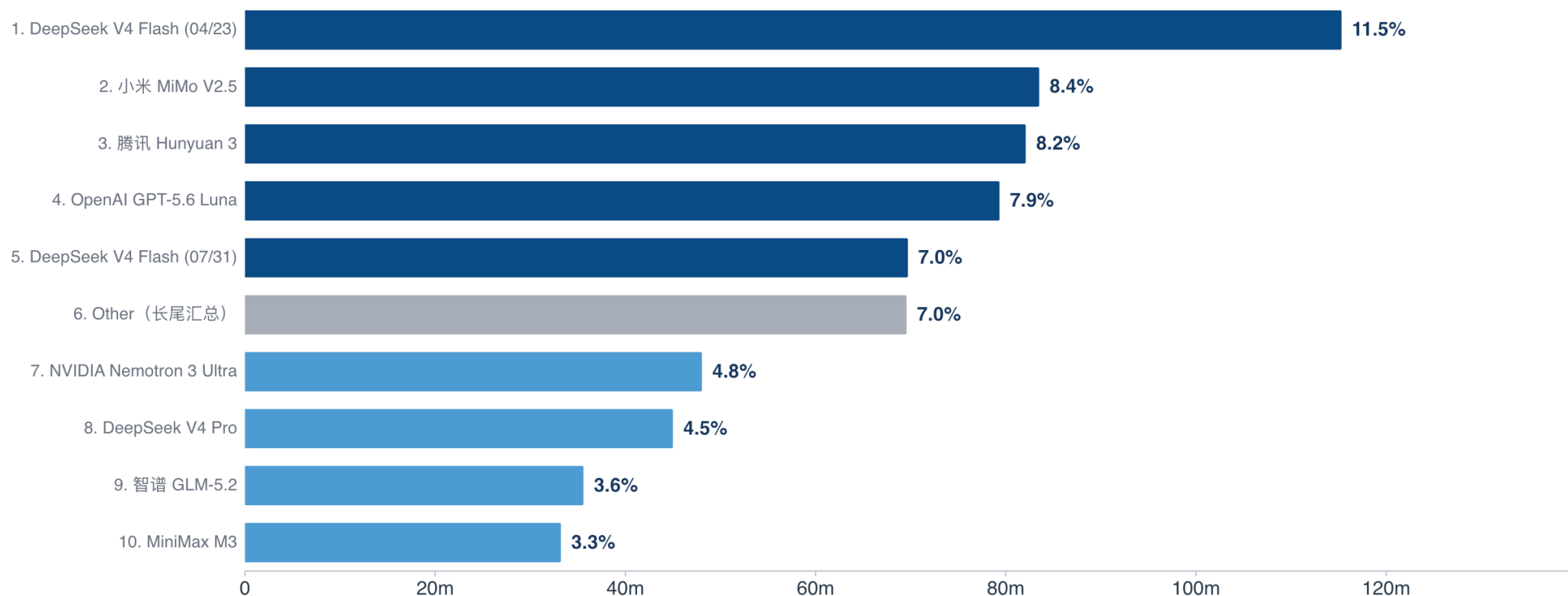
资料来源： OpenRouter，国信证券经济研究所整理

2. 模型供给：供给扩张但流量仍集中于头部

2.1 流量向少数头部选择集中：榜单前十项贡献66.1% Token

- 模型流量并未随着供给扩张而平均分散。截至最新日，第一大模型占OpenRouter当日Token的11.5%，前五大具体模型合计占43.0%，榜单前十项合计达到66.1%；其余41个榜单项仅贡献33.9%。
- 这意味着OpenRouter尚未形成单一模型垄断，但开发者流量已经明显收敛于少数头部选择，呈现出“整体高度集中、头部内部多强竞争”的格局。

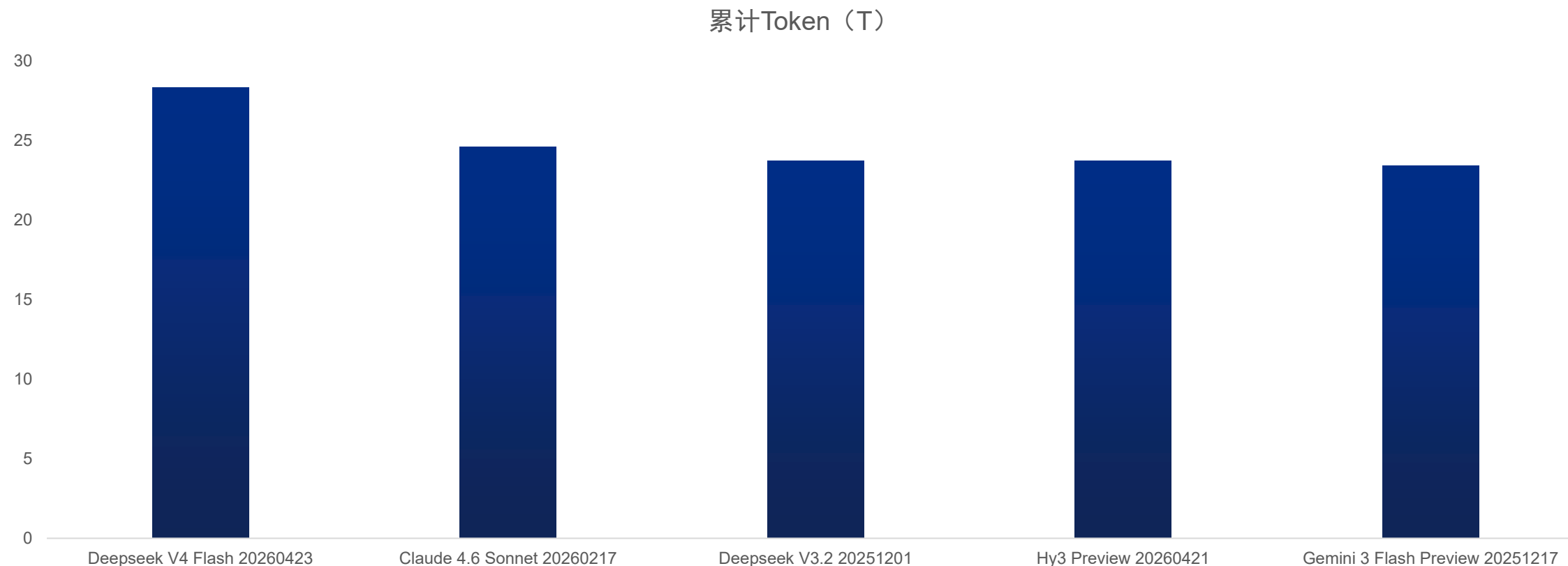
图：OpenRouter平台大模型调用量、排名及占比



资料来源：OpenRouter，国信证券经济研究所整理

2.2 整体高度集中、头部内部多强竞争

- 从累计Token看，头部模型已经不再由单一闭源系列长期垄断。剔除other后，累计Token排名靠前的模型包括Deepseek V4 Flash、Claude 4.6 Sonnet、Deepseek V3.2、Hy3 Preview和Gemini 3 Flash Preview，累计Token分别为28.3T、24.6T、23.7T、23.7T和23.4T

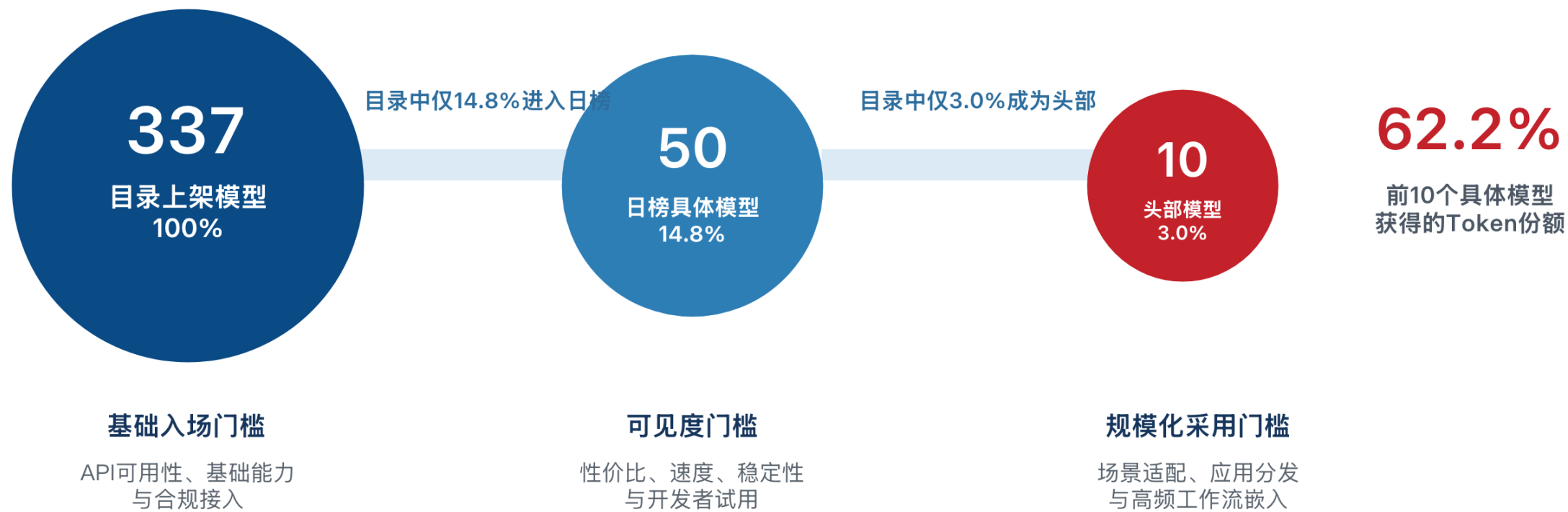


资料来源：OpenRouter，国信证券经济研究所整理

2.3 模型：前十占据62.2%的份额

- 头部模型占据较高的市场份额。截至2026年8月3日，平台共上架337个模型；但最新日榜仅列示50个具体模型，相当于目录模型的14.8%。进一步看，前十大具体模型仅占全部供给的3.0%，却获得了62.2%的Token。
- 随着工具调用、推理和长上下文等能力逐渐普及，基础能力与API接入是模型入场的基础能力；性价比、速度和稳定性决定其能否进入开发者视野，而场景适配、应用分发及高频工作流嵌入，才决定模型能否获得规模化的调用。

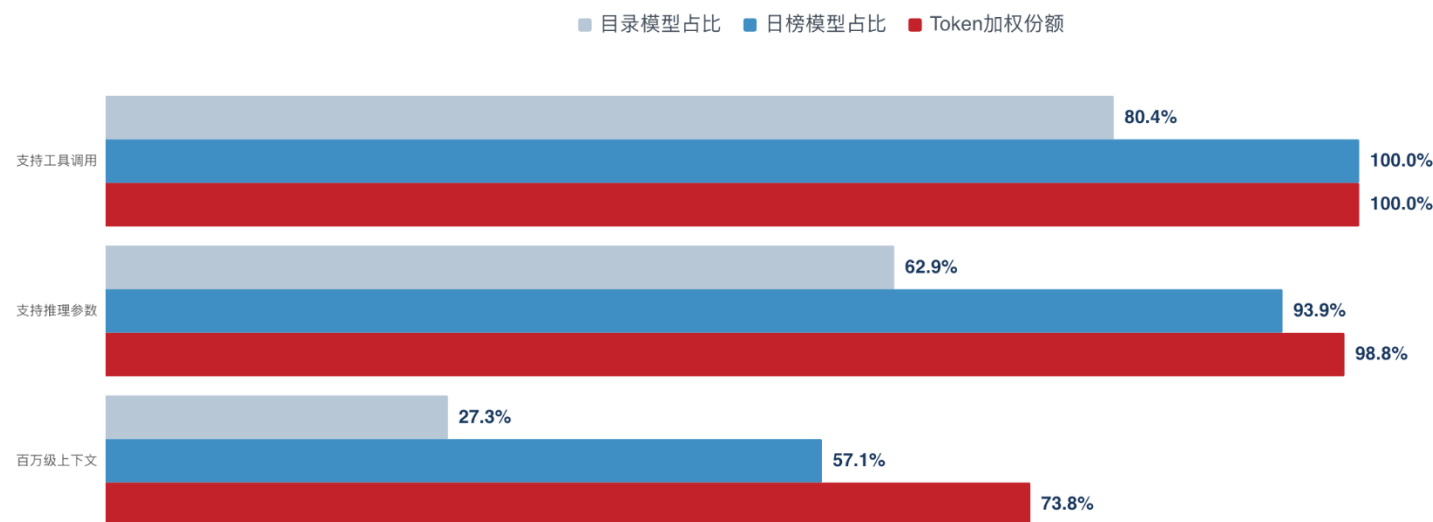
图：OpenRouter平台前十大模型调占据62.6%的份额



2.4 工具调用已成标配，推理与长上下文获得更高流量权重

- 模型能力供给与实际流量选择之间存在明显差异，Token集中到更适合Agent、编程和复杂任务的高能力模型
- 80.4%的模型支持工具调用能力（指模型API支持Function Calling，即根据任务生成结构化的外部工具调用请求，实际工具由应用程序执行），说明工具调用能力正由“加分项”变为进入主流开发者工作流的基础门槛。支持推理参数的模型占目录供给的62.9%，但在日榜模型中占93.9%，获得98.8%的Token；百万级上下文模型仅占目录的27.3%，却占日榜模型的57.1%，Token权重达到73.8%。这表明开发者并未平均消费不断扩大的模型供给，而是将更多Token集中到更适合Agent、编程和复杂任务的高能力模型
- 展望未来，随着工具调用、推理和长上下文等能力逐渐普及，基础能力与API接入只能帮助模型获得“入场资格”；性价比、速度和稳定性决定其能否进入开发者视野，而场景适配、应用分发及高频工作流嵌入，才决定模型能否获得规模化调用。

图：OpenRouter平台模型工具调用及推理、长下文状况



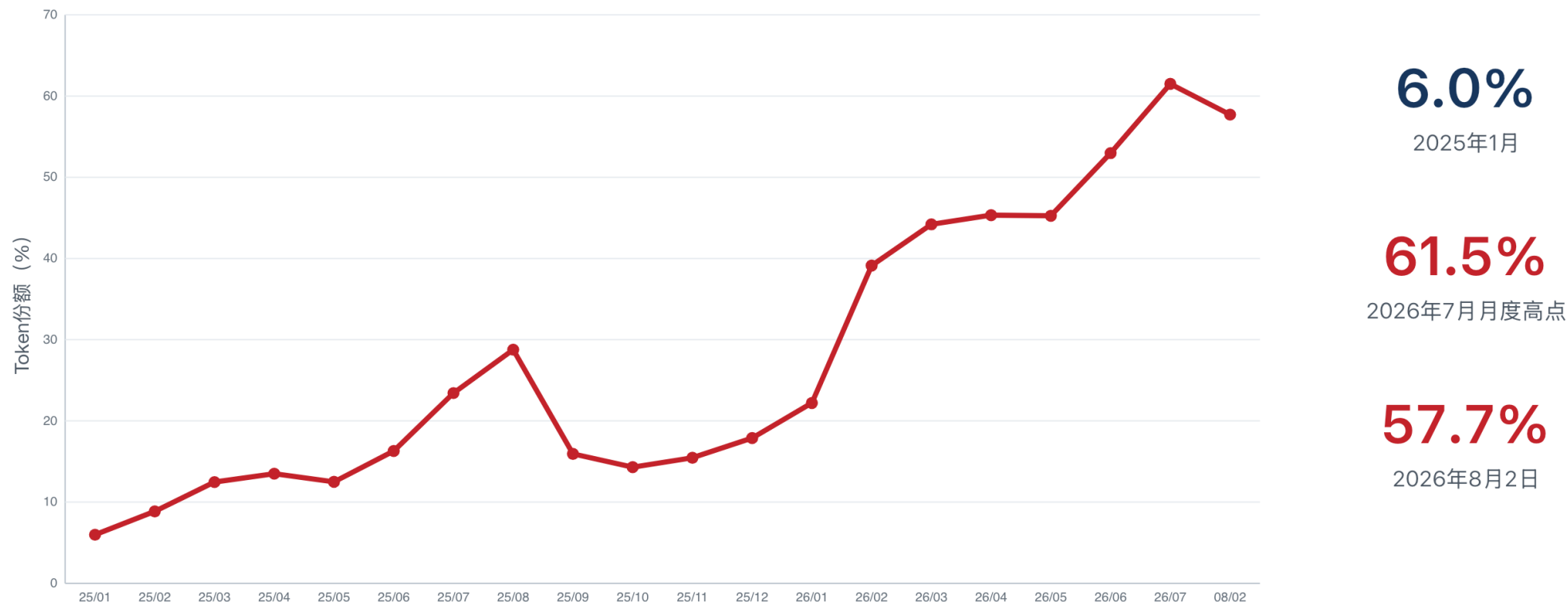
资料来源：OpenRouter，国信证券经济研究所整理

3. 中国模型正成为主导力量

3.1 中国模型已从边缘供给成为OpenRouter的主流流量来源

- 按模型开发商归属统计，中国模型在OpenRouter的Token份额由2025年1月的6.0%，上升至2025年8月的28.8%。2026年初短暂回落后，3月份升至44.2%，6月份突破53.0%，7月份进一步达到61.5%。
- 2026年8月2日，中国模型单日份额仍达到57.7%。这意味着中国模型已经由OpenRouter上的补充供给，转变为平台最主要的流量来源之一。

图：中国模型Token占比最高达到61.5%

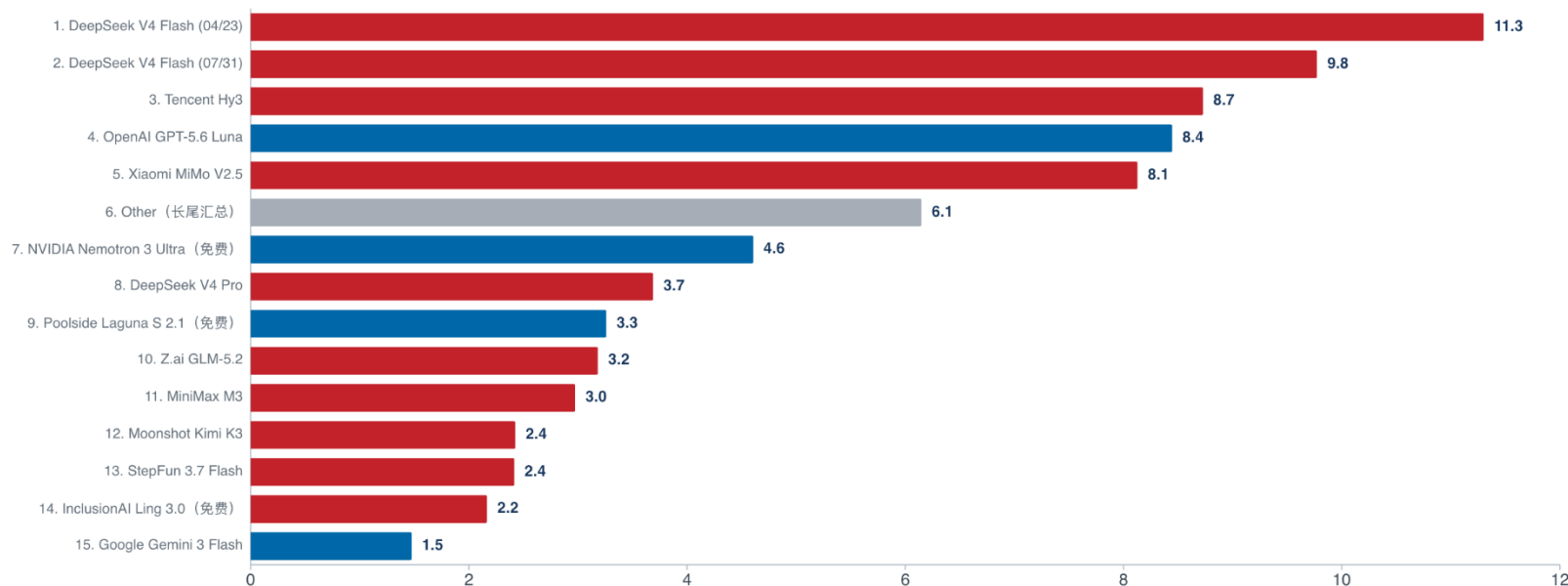


资料来源：OpenRouter，国信证券经济研究所整理

3.2 国产模型的领先已经同时体现在总体份额和头部模型席位上

- 最新数据中，国产模型占平台Token的57.7%，前五大模型中有4个来自中国厂商。
- 中国模型份额的提升并非依靠长尾模型数量堆积，而是已经进入OpenRouter流量榜最核心的位置；最新榜单中，DeepSeek包揽前两名，腾讯Hy3位居第三，小米MiMo V2.5位居第五。剔除Other长尾汇总项后，前十大具体模型中有7个来自中国厂商，此外Z.ai、MiniMax、Moonshot、StepFun等也进入前15名。这说明中国模型已经从“价格敏感型补充选项”，转变为开发者实际调用中的主流选择。

图：前5大模型中4个来自中国

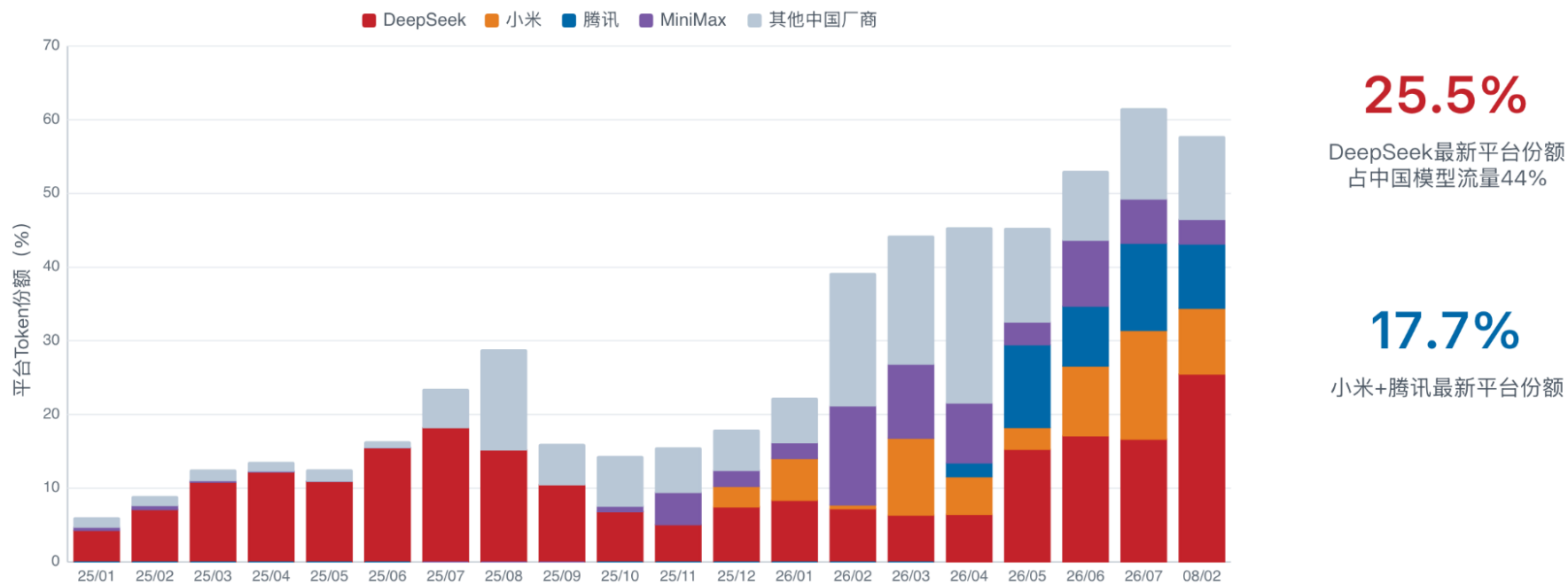


资料来源：OpenRouter，国信证券经济研究所整理

3.3 中国模型份额由多轮爆款接力推动，DeepSeek构成核心流量底盘

- DeepSeek提供基础流量，新一轮增量则由多家中国厂商共同推动
- DeepSeek仍是中国模型最重要的流量底盘。2026年8月2日，DeepSeek占OpenRouter平台Token的25.5%，约占全部中国模型流量的44%。但中国模型近期增长已经不再完全依赖DeepSeek：小米和腾讯合计占平台Token的17.7%，MiniMax、Z.ai、Moonshot、StepFun等厂商也形成了新的流量来源。
- 这种结构意味着，中国模型在OpenRouter的竞争力正在从“单一爆款模型”向“多厂商连续推出爆款”演化，份额增长的持续性有所增强。。

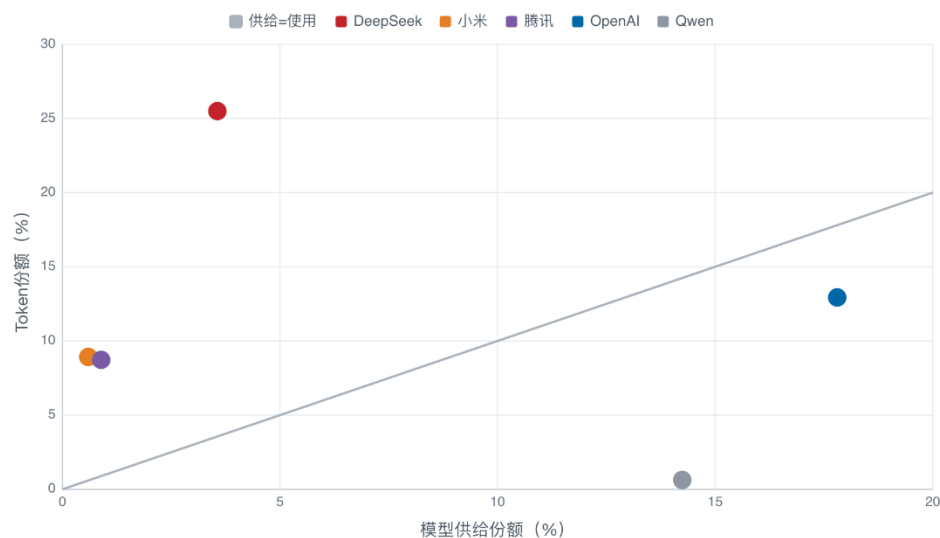
图：DeepSeek构成核心流量底盘，其他国模轮动推动整体份额提升



3.4 中国头部厂商的流量转化效率普遍更高

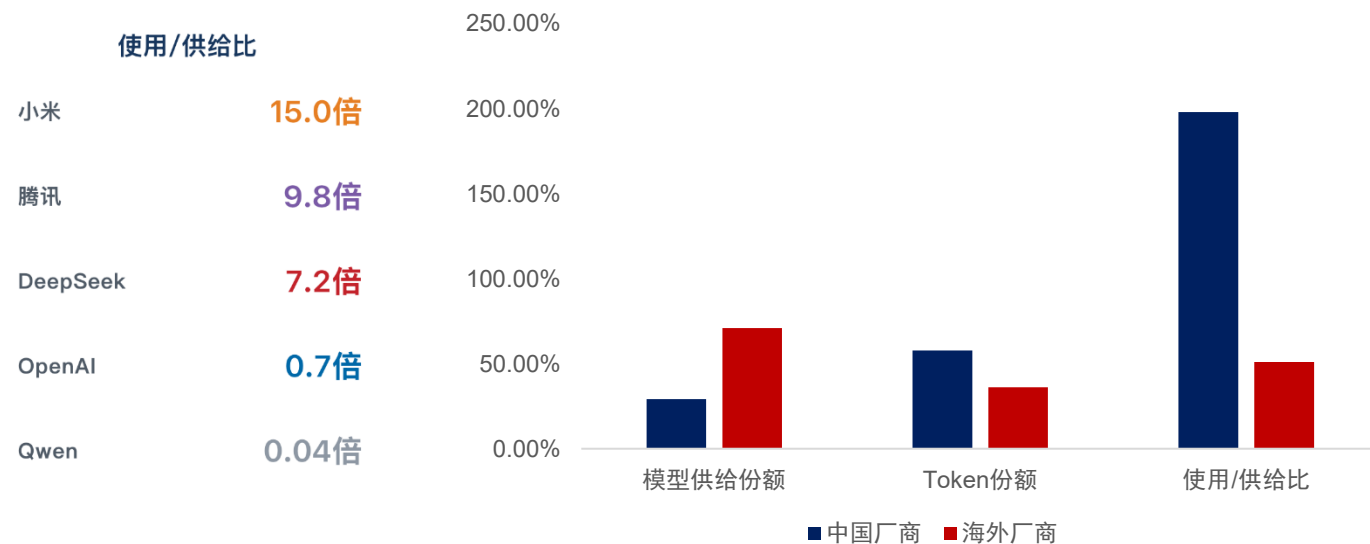
- 少量模型供给即可获得大规模Token；从转化效率上来看，中国厂商显著高于海外厂商：以29%的模型数量占比、获得58%的Token份额占比；
- 中国厂商内部来看，转化效率亦存在显著分化：
 - 小米仅占模型供给的0.6%，但获得8.9%的Token，使用/供给比达到15.0倍。
 - 腾讯占模型供给的0.9%，获得8.7%的Token，使用/供给比为9.8倍。
 - DeepSeek占模型供给的3.6%，获得25.5%的Token，使用/供给比为7.2倍。
 - OpenAI占模型供给的17.8%，Token份额为12.9%，使用/供给比为0.7倍。
 - Qwen占模型供给的14.2%，Token份额仅为0.6%，使用/供给比约为0.04倍。

图：中国厂商流量转化效率分化显著



资料来源：OpenRouter，国信证券经济研究所整理

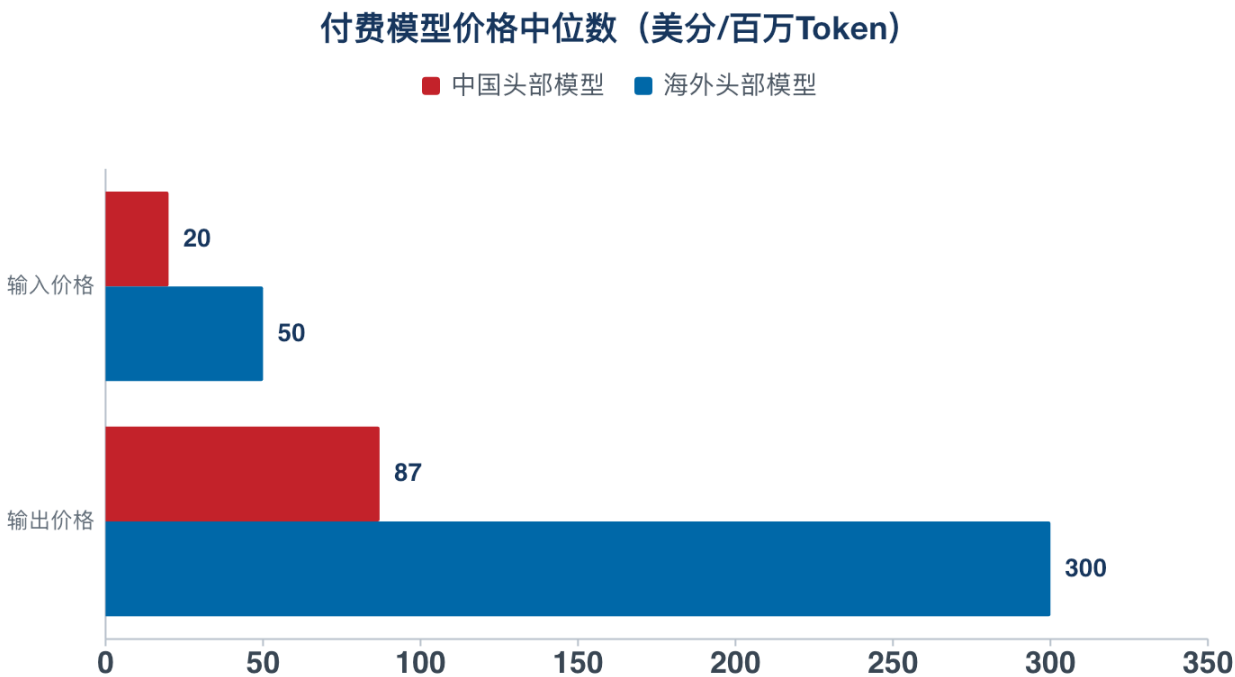
图：中国厂商以29%的模型数量占比、获得58%的Token份额占比



资料来源：OpenRouter，国信证券经济研究所整理

3.5 中国头部模型的优势主要来自性价比与能力的快速迭代

- 在最新前15大具体模型样本中，中国模型有10个，来自8家厂商；海外模型有5个，来自5家厂商。两组模型的工具调用和推理参数支持率均达到100%，百万级上下文支持率分别为70%和80%，百万级上下文支持率分别为70%和80%。
- 基础能力差距已经不是决定流量分布的唯一因素，价格差异更加明显：
 - 中国付费头部模型输入价格中位数为0.20美元/百万Token，海外模型为0.50美元，低约60%
 - 中国付费头部模型输出价格中位数为0.87美元/百万Token，海外模型为3.00美元，低约71%



资料来源：OpenRouter，国信证券经济研究所整理

能力与供给对比

	中国	海外
工具调用支持率	100%	100%
推理参数支持率	100%	100%
百万级上下文	70%	80%
样本中厂商数	8家	5家
免费端点数	1个	2个

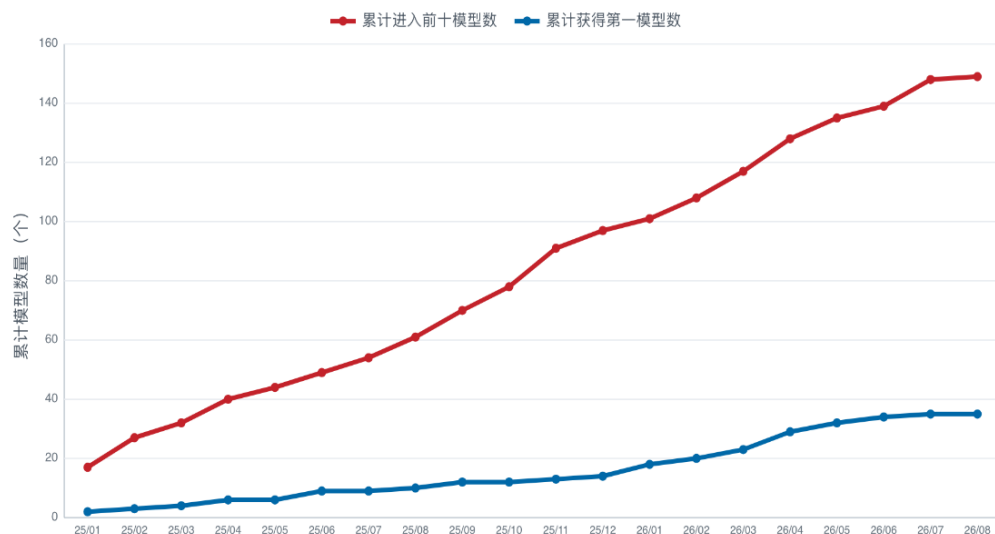
资料来源：OpenRouter，国信证券经济研究所整理

4.模型格局高度动态，爆款生命周期较短、持续迭代成为关键

4.1 头部流量池模型排序持续轮动，CSP价值或将持续提升

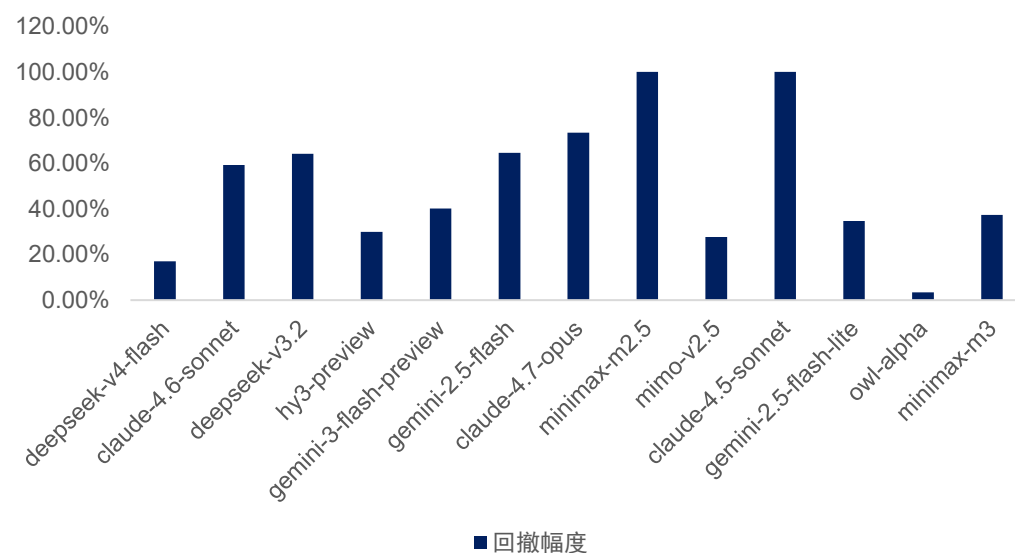
- 头部流量池持续吸收新模型，榜首位置呈现高频轮动。2025年1月1日至2026年8月2日，OpenRouter共有149个不同的具体模型进入过每日Token前十，其中35个模型获得过第一。进入前十和获得第一的累计模型数量几乎每个月都在增加，说明OpenRouter头部流量池始终保持开放。即使模型已经获得领先位置，也很难长期阻止新模型进入和替代。
- 进入日度Top10的模型中，中位达峰时间仅44天，Top10留存中位数为16天，近期Top模型的中位回撤幅度达到71.2%。这意味着模型热度可能在单个硬件采购、部署或合同周期内完成轮动，单一模型专用部署的资产结构会面临更高的错配风险：
- 未来推理基础设施的战略价值不在于押注某一个模型长期领先，而在于能否快速适配不同模型、动态路由不同任务，并把分散的模型流量转化为稳定的算力利用率。对CSP和推理平台来说，多模型接入、统一计费、负载调度、缓存复用、模型热切换和多租户资源池的重要性会持续上升。

图：榜首位置高频轮动



资料来源：OpenRouter，国信证券经济研究所整理

图：峰值回撤显著

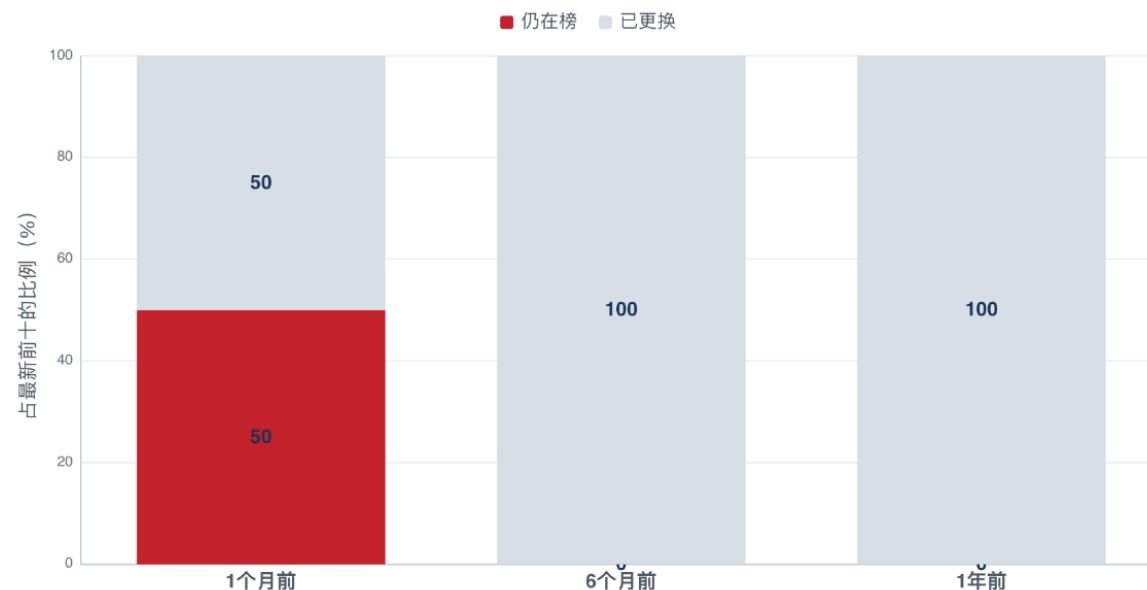


资料来源：OpenRouter，国信证券经济研究所整理

4.2 头部模型生命周期持续性较短

- 整体上来看，爆款模型短期热度可以保持，但很难跨越半年维持头部位置
- 以2026年8月2日前十大具体模型为基准：
 - 与一个月前相比，有5个模型继续留在前十，留存率为50%
 - 与六个月前相比，没有相同的具体模型继续留在前十
 - 与一年前相比，同样没有相同的具体模型继续留在前十
- 一个月后仍在榜的模型包括DeepSeek V4 Flash、DeepSeek V4 Pro、Xiaomi MiMo V2.5、Z.ai GLM-5.2和MiniMax M3。但把观察期拉长至半年，最新前十已经完成全部换血

图：头部位置难以维持超过半年



5/10

一个月后仍在榜

保留下来的5个模型

DeepSeek V4 Flash
DeepSeek V4 Pro
Xiaomi MiMo V2.5
Z.ai GLM-5.2
MiniMax M3

0/10

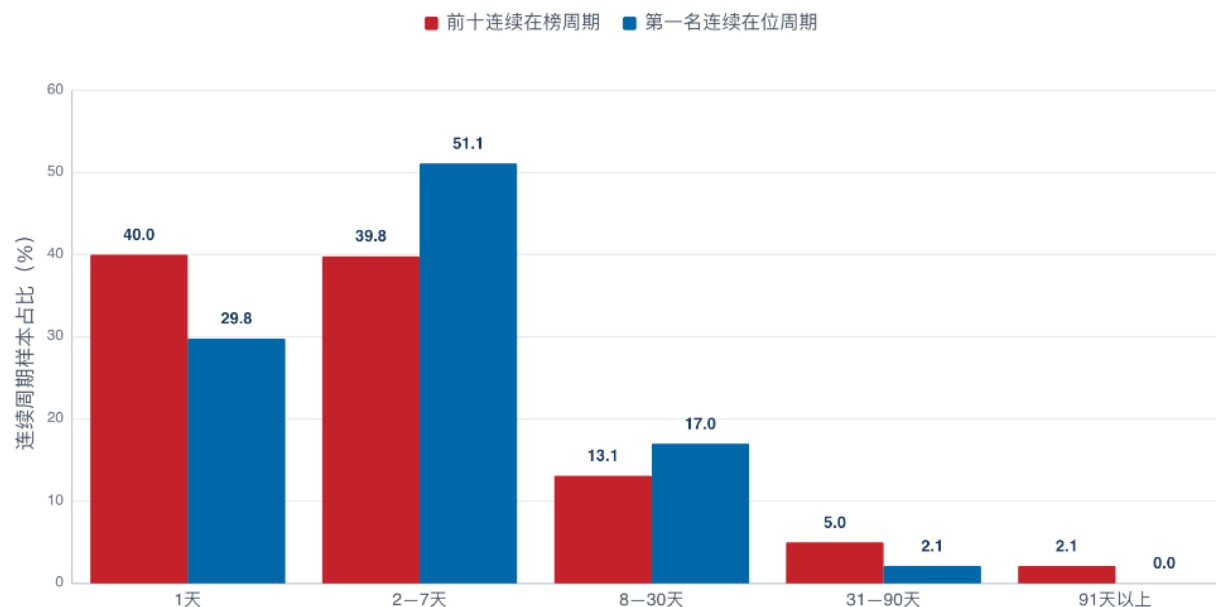
半年后仍为同一具体模型

4.3 爆款效应强，但连续流量窗口非常短

➤ 按照“连续自然日位于前十”定义单轮在榜周期，并排除观察期首尾没有完整结束的边界样本，共获得520段完整前十周期：

- 连续在榜周期中位数仅为2天。
- 40.0%的在榜周期只持续1天。
- 79.8%的在榜周期不超过7天。
- 92.9%的在榜周期不超过30天。
- 仅7.3%的连续周期达到30天以上

图：前十连续在榜中位数仅2天，八成单轮热度不超过一周



2天

前十连续在榜中位数

2天

第一名连续在位中位数

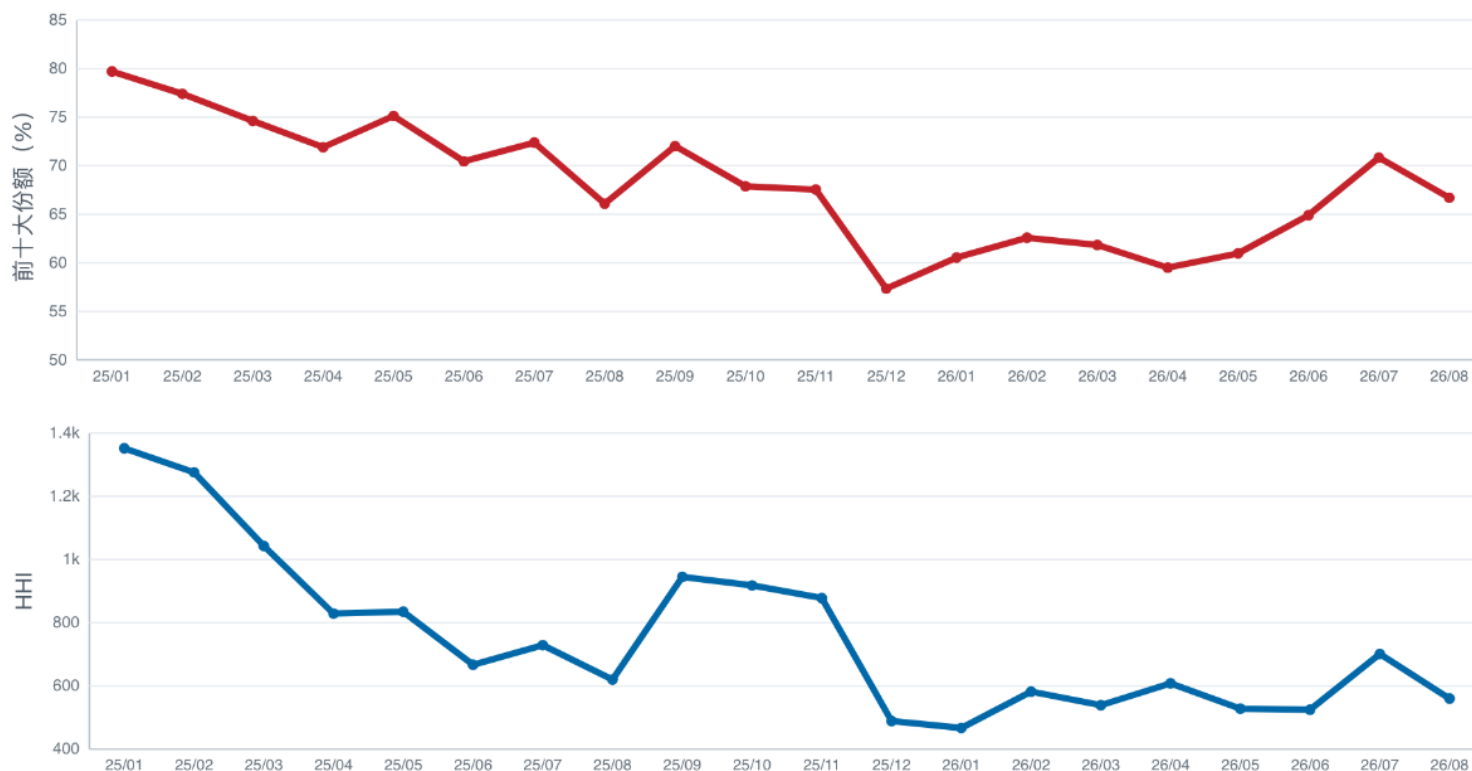
79.8%

前十单轮周期不超过7天

4.4 流量集中度持续下降，前十大模型Token份额下降明显

- 开发者流量仍快速聚集于头部，但单一模型长期控制流量的难度明显上升
- 前十大模型Token份额由76.6%下降至67.3%。HHI由1,036下降至574。集中度下降并非平滑过程。新爆款出现时，前十大份额和HHI会阶段性回升；随着更多模型进入头部，集中度又会回落。最新前十大模型仍获得约三分之二的Token，说明开发者流量依然集中于少数模型。但HHI明显下降、头部模型持续更换，说明这种集中不再稳定地归属于某一个模型

图：前十Token消耗占比67%，头部集中但是内部持续轮换



前十大份额
76.6% → 67.3%

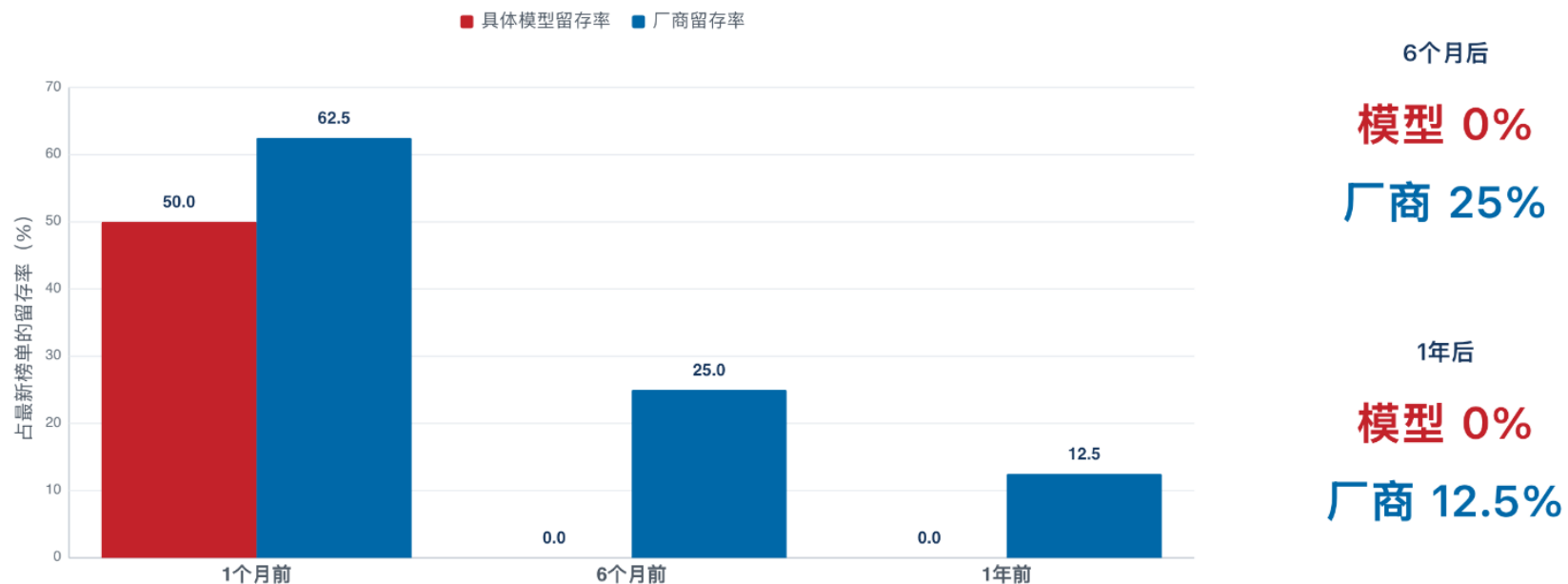
HHI
1,036 → 574

首日 → 最新日

4.5 厂商竞争关键在于持续迭代能力

- 模型层面的留存明显低于厂商层面，渠道竞争正转向持续推出爆款的产品节奏
- 2026年8月2日前十大具体模型来自8家厂商。分别从具体模型和厂商两个层面比较榜单留存来看，半年后最新前十已经没有相同的具体模型，但仍有DeepSeek和MiniMax两家厂商同时出现在两期榜单中；一年后没有具体模型留存，但DeepSeek仍然保持厂商层面的头部位置。
- 这意味着厂商无需依赖单一模型长期统治榜单，而可以通过新版本和新模型接替旧产品的流量。竞争优势正在从“拥有一个爆款”，转向“持续推出爆款并完成流量接力”

图：厂商竞争关键在于持续迭代能力



资料来源：OpenRouter，国信证券经济研究所整理

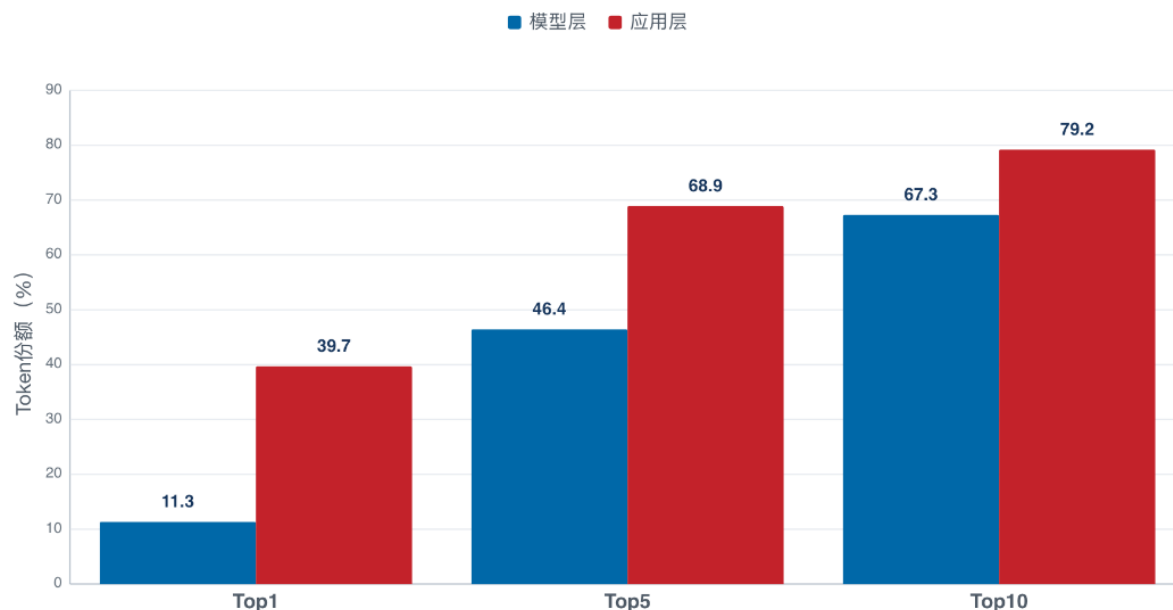
5.应用生态：Agent/Coding是Token需求核心

5.1 应用层第一名占40%流量，集中度显著高于模型层

➤ 应用层的流量集中度明显高于模型层，这或许表明模型供应商仍面临频繁轮换和多模型竞争，而应用一旦形成用户入口和稳定工作流，更容易聚集并持续获得流量：

- 应用第一名占39.7%，约为模型第一名11.3%的3.5倍。
- 应用前五名占68.9%，模型前五名占46.4%。
- 应用前十名占79.2%，模型前十名占67.3%。
- 应用层HHI达到1,850，明显高于模型层的574

图：应用层集中度显著高于模型层



39.7%

应用第一名份额

79.2%

应用前十份额

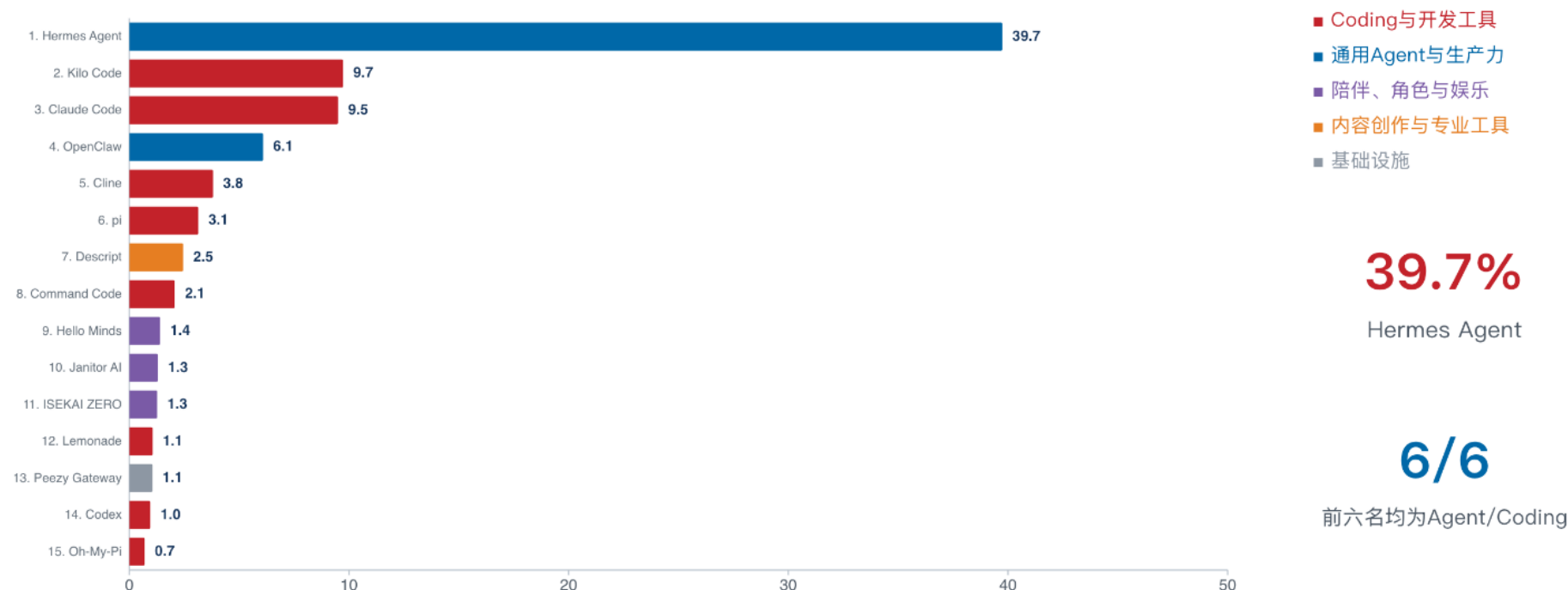
1,850 vs 574

应用层HHI / 模型层HHI

5.2 头部应用前六名全部为Agent或Coding工具

- OpenRouter的主要应用入口正在从简单聊天转向能够执行任务、调用工具和持续运行的工作流。最新应用榜前六名依次为Hermes Agent、Kilo Code、Claude Code、OpenClaw、Cline和pi，全部具有明显的Agent或Coding属性
 - Hermes Agent以39.7%的Token份额遥遥领先。
 - Kilo Code和Claude Code分别占9.7%和9.5%。
 - 前六大Agent/Coding应用合计占72.0%。
 - 传统聊天、角色扮演及内容创作类应用仍有需求，但尚未形成同等规模的头部入口

图：头部应用集中于Agent/Coding工具



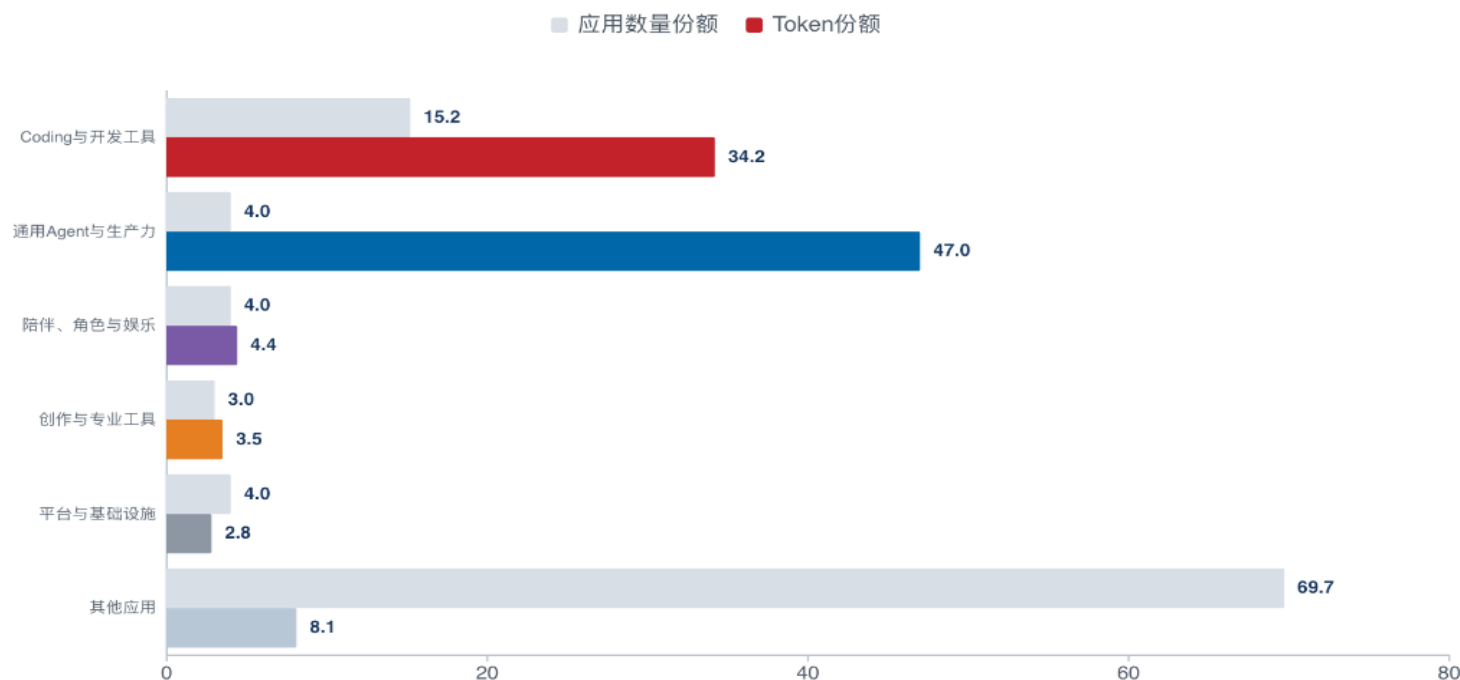
资料来源：OpenRouter，国信证券经济研究所整理

5.3 19个Agent/Coding应用贡献81%的Token

➤ 通用Agent和Coding工具仅占应用数量的19%，却构成OpenRouter应用流量中心，贡献了81.2%的Token消耗，Agent/Coding已经成为OpenRouter应用层最主要的Token需求来源

- 4个通用Agent与生产力应用贡献47.0%。
- 15个Coding与开发工具贡献34.2%。
- 其余应用合计仅贡献18.8%。
- 其中，未进入前30名的长尾应用占应用数量69.7%，但Token份额只有8.1%

图：Agent/Coding应用贡献81%的Token消耗

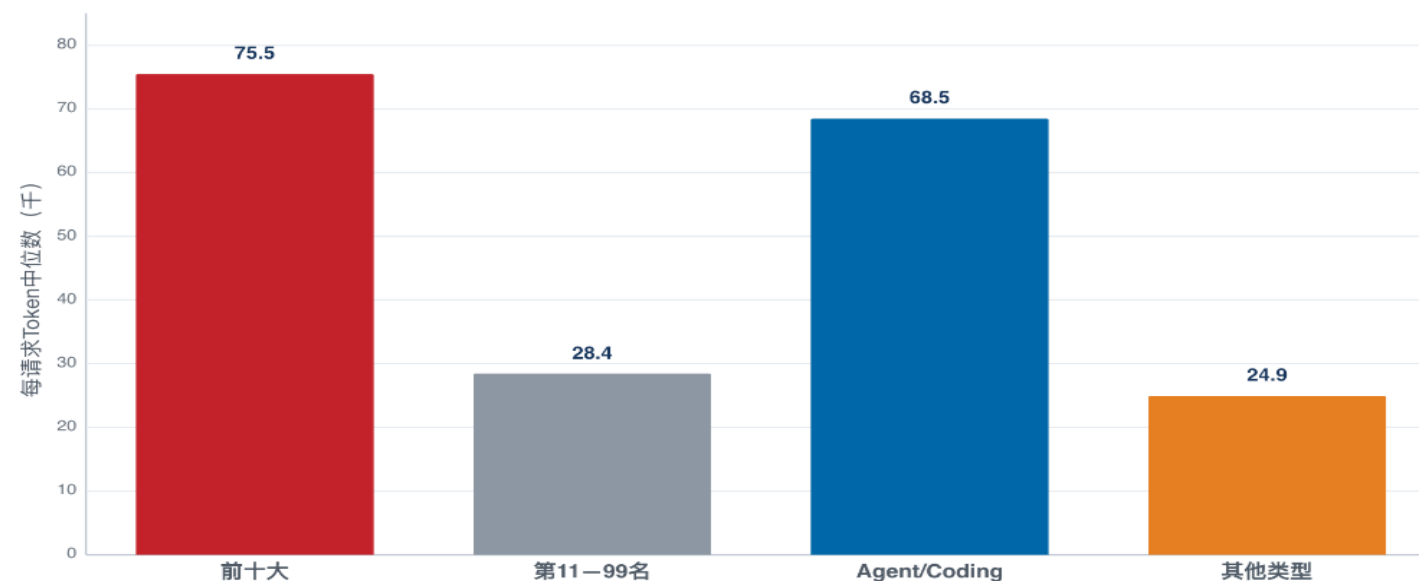


资料来源：OpenRouter，国信证券经济研究所整理

5.4 Agent/Coding更复杂的单次任务，进一步放大应用层Token需求

- 应用层的流量集中不仅来自请求数量，也来自单次任务复杂度提升；Agent和Coding任务往往包括多轮推理、代码读写、工具调用、上下文回传与结果校验，因此一次用户任务可能在后台触发多次模型调用。
 - 前十大应用每请求Token中位数为75,479。
 - 第11—99名应用中位数仅为28,430。
 - 头部应用单次请求消耗约为其他应用的2.7倍。
 - Agent/Coding应用中位数为68,483，其他类型应用为24,942，同样相差约2.7倍

图： Agent/Coding应用显著放大Token消耗

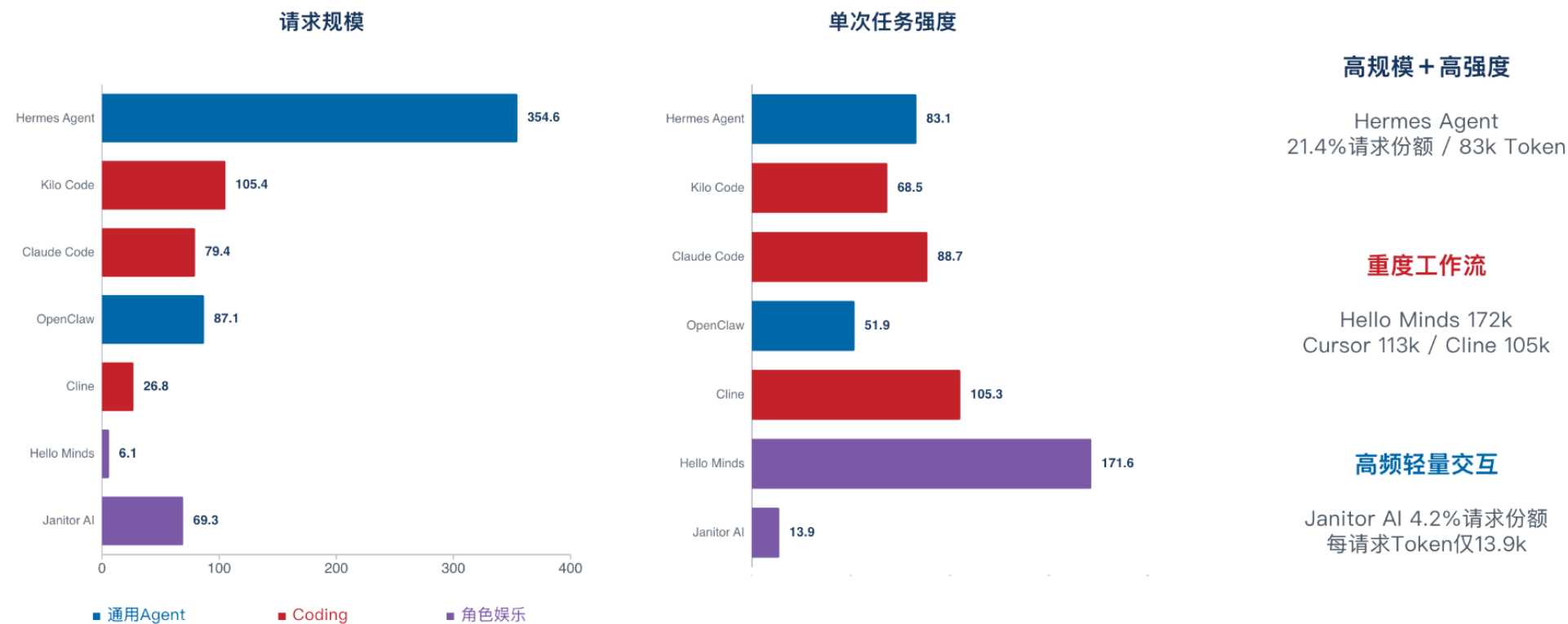


资料来源：OpenRouter，国信证券经济研究所整理

5.5 头部应用正在形成三类典型需求形态

- AI应用需求分为高频交互与重度工作流两种模式，头部应用正在形成三类典型需求形态
 - 高规模、高强度：Hermes Agent占21.4%的请求量，同时每请求消耗约8.3万Token，兼具用户规模和任务复杂度。
 - 重度工作流：Hello Minds、Cursor和Cline每请求分别消耗约17.2万、11.3万和10.5万Token，单次任务深度突出。
 - 高频轻量交互：Janitor AI占4.2%的请求量，但每请求仅约1.4万Token，主要依靠高频交互形成流量。

图：头部应用正在形成三类典型需求形态

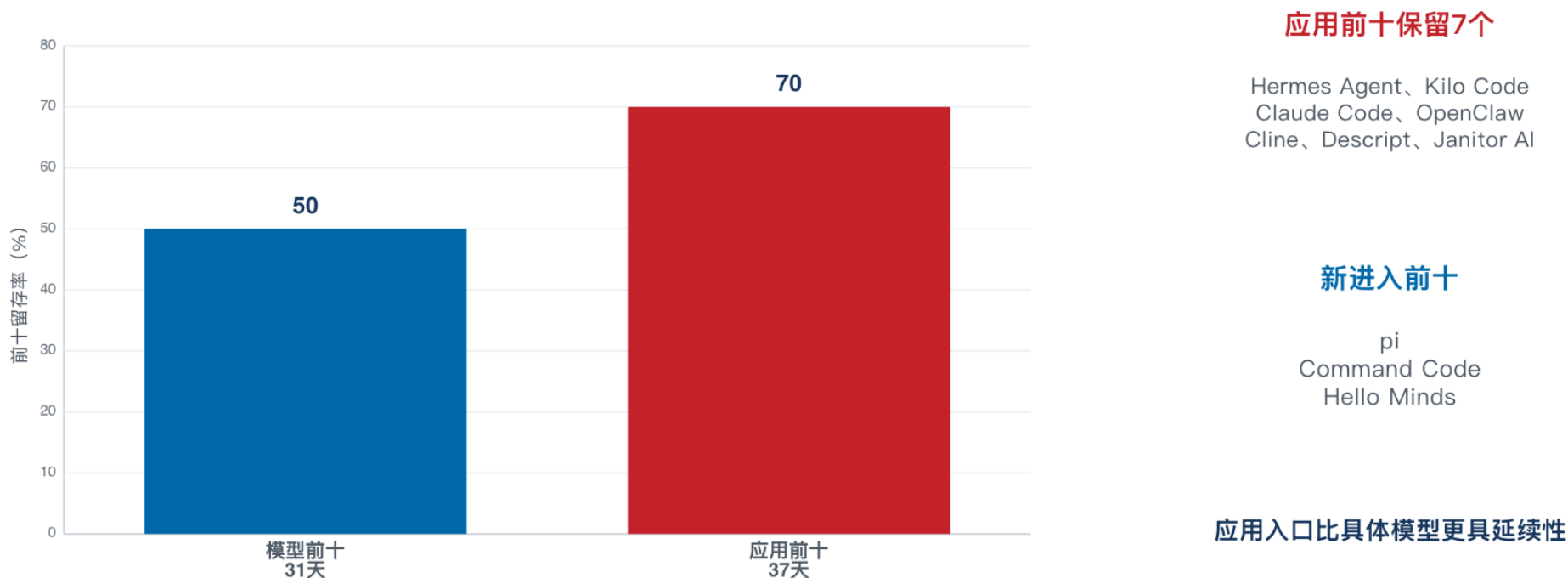


资料来源：OpenRouter，国信证券经济研究所整理

5.6 应用前十近一个月留存70%，高于模型前十的50%

- 应用入口比具体模型更具延续性，工作流和用户关系正在形成渠道价值
- 6月28日至8月3日，应用前十中有7个仍然保持在榜；作为对比，7月2日至8月2日模型前十仅有5个留存。应用前十留存者包括Hermes Agent、Kilo Code、Claude Code、OpenClaw、Cline、Descript和Janitor AI；新进入者主要为pi、Command Code和Hello Minds。这说明应用可以在底层模型变化时切换模型供应商，用户真正持续使用的是产品入口、任务流程及历史数据，而不一定是某一个具体模型。

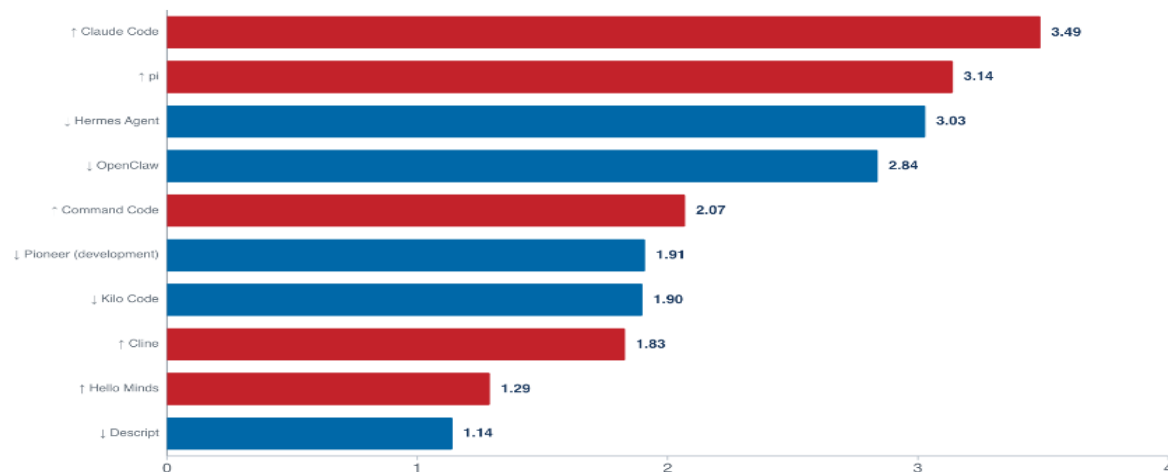
图：应用前十近一个月留存70%，高于模型前十的50%



5.7 前五大份额应用中，4个是Coding工具

- 应用头部框架相对稳定，但内部分额仍在迁移，近期增量进一步向Coding workflow倾斜。
- 近期Token份额提升最大的应用包括：
 - Claude Code：上升3.49个百分点。
 - pi：上升3.14个百分点。
 - Command Code：上升2.07个百分点。
 - Cline：上升1.83个百分点。
 - Hello Minds：上升1.29个百分点。
- 份额下降较多的包括Hermes Agent、OpenClaw、Pioneer development、Kilo Code和Descript。Hermes虽然下降3.03个百分点，但最新份额仍达到39.7%，继续稳居第一

图：Coding是主要应用场景



4/5

前五大份额增量应用为Coding

+3.49pct

Claude Code份额变化

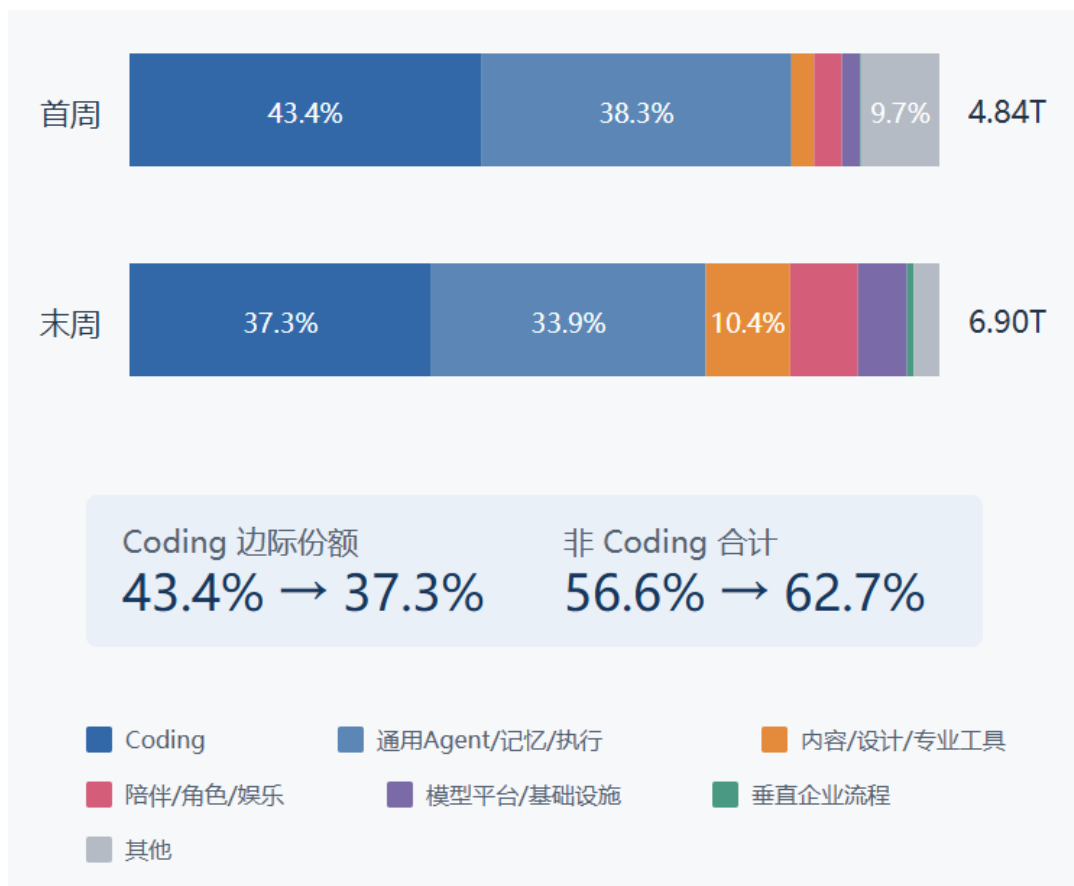
39.7%

Hermes份额回落但仍居第一

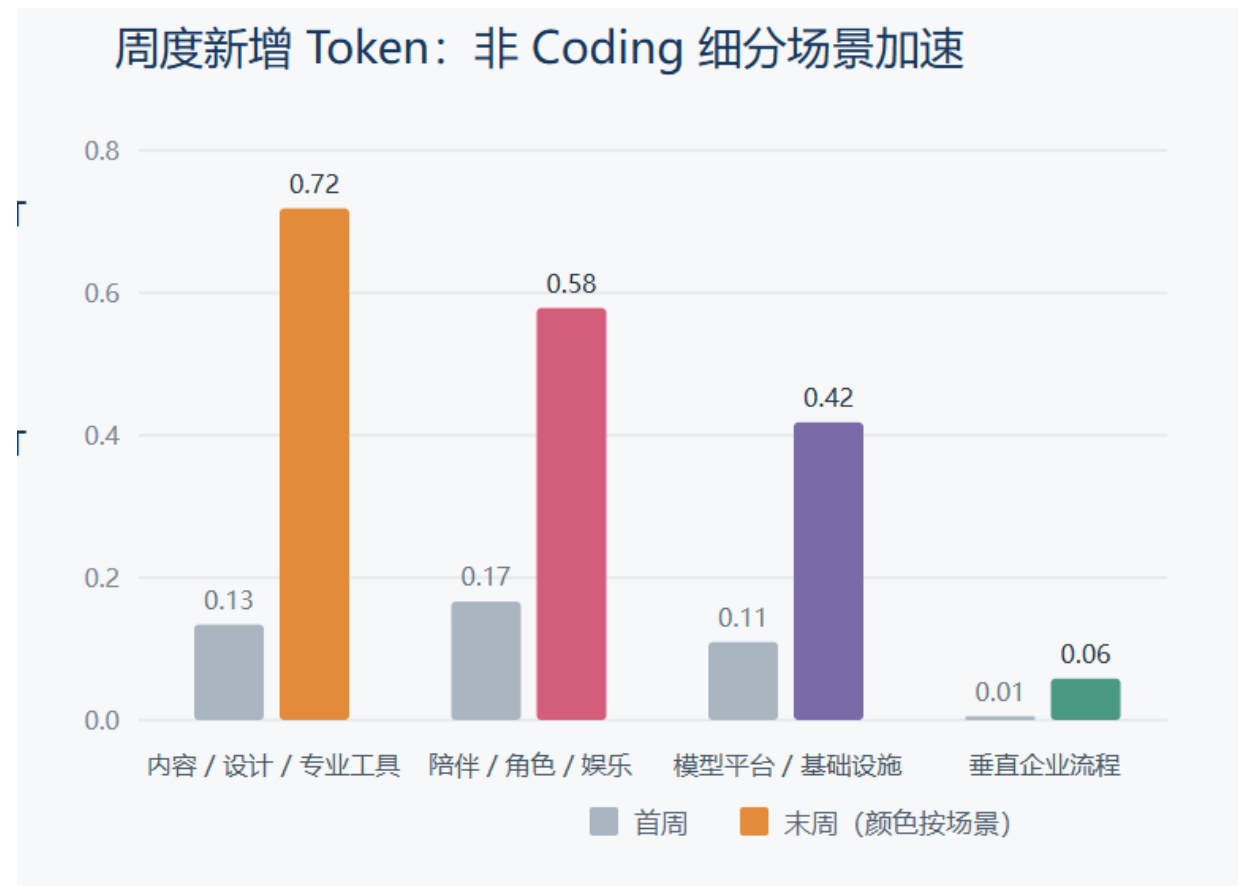
5.8 非Coding场景正在接力：新增 Token 从单一 Coding 转向多引擎

- 最近的表现来看，Coding 边际份额下降、非Coding份额提升；Token需求场景正在走向分散化，说明AI应用场景正在持续扩散；
- 非Coding场景中，内容/设计/专业工具、陪伴/娱乐等垂类B端、C端领域增长更为迅速：

图：非Coding占比提升



图：内容/设计/专业工具、陪伴/娱乐等垂类B端、C端领域增长更为迅速



资料来源：OpenRouter，国信证券经济研究所整理

资料来源：OpenRouter，国信证券经济研究所整理

6. 投资观点：持续看好AI产业趋势，关注CSP及AI应用

6.1 推理需求进入加速期，优先关注CSP，精选AI应用

- **推理需求进入高速扩张期，增长动力正由简单聊天向复杂工作流迁移。** 1) OpenRouter作为连接模型厂商、推理服务商与AI应用开发者的中立型API聚合平台，其流量不能直接代表全球大模型需求，但可以作为观察开发者推理活动的重要窗口。2025年1月至2026年7月，平台月度日均Token由0.072T增长至7.995T，19个月增长约111倍，2026年7月单日峰值达到9.622T。2) 增长过程并非平滑线性，而是伴随新模型上线、免费版本放量以及开发者工作流迁移，表现出明显的阶跃式特征。随着推理、长上下文和工具调用逐渐成为主流，Token增长开始更多来自单次任务复杂度提升，而不仅是聊天请求数量增加，推理需求正在进入由Agent化任务驱动的新阶段。
- **模型供给进入百花齐放阶段，中国模型正在成为全球开发者渠道的重要力量。** 1) OpenRouter最新包含337个模型和58家厂商，其中271个支持工具调用、212个支持推理参数、92个支持百万级上下文，但进入最新日度流量榜的模型仅约50个，前十大模型仍获得约三分之二的Token。模型竞争的关键已经不是能否上架，而是性价比、推理速度、服务稳定性以及能否进入高频应用工作流。2) 中国模型凭借能力快速迭代和成本优势，月度Token份额由2025年1月的6.0%提升至2026年7月的61.5%，最新前十大具体模型中有7个来自中国厂商。不过，头部模型生命周期正在明显缩短：半年后最新前十已经完成全部换血，模型层面的竞争优势由依赖单一爆款，转向持续推出新版本、控制推理成本并完成流量接力。3) 中国模型厂商优势持续展现，看好具备持续能力迭代的头部厂商。
- **推理需求扩张叠加模型快速轮换，有望持续提升CSP和推理平台的产业价值。** 1) 当模型迭代周期短于企业硬件采购和部署周期，企业自行押注单一模型、建设专用算力的错配风险上升，而CSP能够通过共享资源池承接不同模型和应用的波动性需求，并提供多模型接入、弹性部署、负载调度、缓存复用、统一计费、数据安全及企业级运维服务。2) 海外相关云服务厂商有AWS、Microsoft Azure、Google Cloud和Oracle OCI，国内相关厂商入阿里云、腾讯云等具备算力规模、模型生态和客户入口优势的厂商以及算力租赁相关产业链；3) 需要强调的是，Token增长并不必然等同于云厂商利润增长，CSP价值兑现仍取决于AI云收入增速、GPU利用率、推理价格竞争，以及资本开支和折旧压力；具备资源规模、全栈平台能力和企业客户基础的厂商更可能受益。
- **成本下降释放AI应用需求弹性，重点关注Agent/Coding和高价值垂直场景。** 1) 模型能力持续提升、推理价格下降及多模型API接入门槛降低，将显著降低AI应用的开发、试错与规模化部署成本，并通过需求弹性带动调用频次、上下文长度和任务复杂度持续提升；2) 结合OpenRouter数据，Agent/Coding应用仅占应用数量的19%，却贡献81.2%的Token，头部应用每请求Token约为其他应用的2.7倍，表明能够执行多步骤任务、调用外部工具并嵌入专业工作流的应用，更有可能成为推理需求的主要增量来源。3) 建议B端重点关注Coding Agent、通用任务型Agent，以及工业、医疗、金融、法律、设计等具有明确付费意愿和较高任务价值的垂直场景；C端建议关注具备IP/版权资源、内容生产场景、流量入口、算力优化及运营能力的公司，关注AI漫剧/游戏/营销服务/Token运营等方向。

6.2 附录：大模型产业链相关公司

公司名称	股票代码	上市地点	所属国家	核心大模型/产品	商业模式与业务特点
阿里巴巴	9988.HK	港股/美股	中国	通义千问（Qwen）	商业友好开源，深度融入阿里云生态，高频迭代，为多语言开发主流选择之一。
腾讯	0700.HK	港股	中国	混元大模型	整合微信、游戏、广告、办公等生态，推出AI编程工具等产品。
科大讯飞	002230.SZ	A股	中国	星火大模型	深耕教育、医疗等垂直领域
三六零	601360.SH	A股	中国	360智脑4.0	自研千亿参数通用模型，覆盖全应用场景，在税务、企业服务垂直行业落地
昆仑万维	300418.SZ	A股	中国	天工大模型、Mureka V8	开放API按Token调用计费，切入“词元经济”变现；同时布局AI音乐、漫剧等领域
智谱华章	02513.HK	港股	中国	GLM系列大模型	“模型即产品”平台路线，以MaaS（API调用）为核心，B端/G端服务为主
MiniMax	00100.HK	港股	中国	M2系列大模型、海螺AI、星野、Talkie	C端订阅+B端开放平台双轮驱动，海外收入占比过半。
字节跳动	未上市	—	中国	豆包大模型（2.1系列）、Seed 2.0/2.1系列、Seedance视频生成（2.5）、Seedream图像模型、TRAIE企业级AICoding、扣子Agent平台	B端+C端双轮驱动：①C端以豆包App为核心（日活超2亿），2026年6月上线付费订阅（68/200/500元/月三档）②B端通过火山引擎MaaS平台提供API调用，2026年MaaS营收目标150亿元（2025年约15亿元）；③视频生成模型Seedance贡献主要企业收入。国内公有云大模型Token调用份额49.5%位列第一，日均Token调用量达180万亿。
月之暗面（Kimi）	未上市（筹备赴港IPO）	—	中国	Kimi K3	2026年7月发布全球参数最大开源模型Kimi K3，在Code Arena超越Claude Fable 5和GPT-5.6 So；已调整架构，预计6个月内完成港股上市。
深度求索（DeepSeek）	未上市	—	中国	DeepSeek系列大模型	国产开源模型龙头，Token调用量稳居前列。
微软（Microsoft）	MSFT	纳斯达克	美国	Copilot、Azure AI（整合OpenAI技术）	向OpenAI投资超130亿美元，将GPT系列深度整合至全线产品；Azure云提供大模型API服务。
Alphabet（谷歌）	GOOGL / GOOG	纳斯达克	美国	Gemini系列大模型	自研全栈AI模型，云平台（GCP）提供AI基础设施；持有Anthropic约14%股权。
亚马逊（Amazon）	AMZN	纳斯达克	美国	AWS Bedrock、Titan大模型	AWS云平台提供模型托管与分发服务，持有Anthropic股权。
Meta Platforms	META	纳斯达克	美国	Llama系列大模型	自研开源大模型，深度整合至广告推荐与内容生成，为GPU综合应用大户。
Anthropic（Claude）	未上市（已保密递交IPO申请）	—	美国	Claude系列大模型	2026年6月递交IPO申请，估值达9650亿美元，最早2026年10月挂牌；Alphabet（14%）、Amazon、Microsoft参股。
OpenAI（ChatGPT）	未上市	—	美国	GPT系列、ChatGPT、Sora	全球估值最高AI初创，无公开交易股票；微软为最大战略投资者。

资料来源：Wind，公司公告，官网，国信证券经济研究所整理

6.2 附录：中美主要CSP情况



公司名称	股票代码	云业务品牌	2026年预估资本支出	业务结构
亚马逊 Amazon	AMZN.O	AWS	约2,200亿美元	综合型公有云；重点投入AWS、AI基础设施、数据中心及自研芯片。
Alphabet	GOOGL.O / GOOG.O	Google Cloud	1,950亿–2,050亿美元	TPU、Gemini、Vertex AI与公有云构成全栈体系。
微软 Microsoft	MSFT.O	Microsoft Azure	约1,750亿美元	Azure结合Microsoft 365、GitHub和Copilot；部分租赁分类影响报告口径。
Meta Platforms	META.O	自建AI基础设施	1,300亿–1,450亿美元	主要支持Llama、推荐广告系统、Meta AI及新型AI产品。
甲骨文 Oracle	ORCL.N	OCI	约600亿–700亿美元	聚焦大规模GPU集群、AI训练推理、数据库云和多云部署。
阿里巴巴 Alibaba	9988.HK / BABA.N	阿里云	约1,300亿–1,600亿元人民币	云基础设施、通义千问与AI开发平台一体化。
腾讯控股 Tencent	0700.HK	腾讯云	约1,100亿–1,300亿元人民币	结合微信、游戏、广告和企业服务，发展混元模型与GPU云。
字节跳动 ByteDance	未上市	火山引擎	约2,000亿元人民币；潜在上修至4,000亿–5,000亿元	服务抖音、TikTok、豆包及推荐系统，并向外部提供云和模型服务。
百度 Baidu	9888.HK / BIDU.O	百度智能云	约220亿–280亿元人民币	重点布局GPU云、文心模型、搜索与自动驾驶相关AI应用。
华为 Huawei	未上市	华为云	未单独披露	昇腾、盘古、华为云、存储及网络设备构成软硬件全栈体系。

资料来源：Wind，公司公告，官网，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

6.2 附录：美股B端场景垂类AI应用

公司	股票代码	市值（亿美元）	定位/产品	AI逻辑/商业化方式
Snowflake	SNOW	1098	AI Data Cloud、Cortex、Snowflake Intelligence、Cortex Code	企业数据治理、RAG和Agent运行底座；按数据存储、计算和AI用量收费，同时向业务用户延伸AI应用
Datadog	DDOG	1008	云监控、LLM/Agent可观测性、安全监控	监测模型调用、延迟、Token成本、Agent链路与生产故障；按主机、日志、事件和用量计费
Cloudflare	NET	1036	Workers AI、边缘推理、AI Gateway、网络与安全	将推理、缓存、安全和Agent运行放到边缘节点；按请求、算力、存储和安全服务收费
MongoDB	MDB	303	Atlas、向量检索、AI应用数据库	为RAG、AI应用和实时业务系统提供数据库与检索层；按云数据库用量收费
Elastic	ESTC	72	Elasticsearch、向量搜索、企业搜索、可观测性	企业知识检索、RAG、安全分析和日志智能化；订阅+云用量模式
Palantir	PLTR	3807	Foundry、AIP、Ontology、企业AI操作平台	将模型连接企业数据、权限与实际业务动作；平台许可、项目部署和持续扩容
Microsoft	MSFT	36197	Azure AI、Microsoft Foundry、Microsoft 365 Copilot、GitHub Copilot	同时覆盖云、模型、开发工具和办公入口；云用量+Copilot席位+Agent/平台调用收费
Salesforce	CRM	1581	Agentforce、Data 360、Slack Agent、AI CRM	以CRM数据和销售/客服 workflow 驱动Agent执行；订阅、数据平台及按任务/用量收费
ServiceNow	NOW	1212	Now Assist、AI Control Tower、Autonomous Workforce	在IT、HR、客服、采购等工作流中部署和治理Agent；平台订阅+AI功能包及用量收费
Adobe	ADBE	1031	Firefly、Acrobat AI Assistant、GenStudio、创意Agent	将生成式AI嵌入创意、文档与营销 workflow；订阅+生成积分/Token+企业营销产品收费
Intuit	INTU	897	财税、会计、支付和个人金融AI Agent	基于财税与经营数据辅助记账、报税、咨询和风控；订阅+交易及增值服务收费
HubSpot	HUBS	125	Breeze AI、营销/销售/客服Agent	利用CRM和营销数据自动化获客、内容、销售跟进和客服；订阅套餐+AI增值模块
Zoom	ZM	295	AI Companion、会议与协作Agent	会议摘要、知识提取、任务分配及协作自动化；基础订阅带动留存，高端功能增值
GitLab	GTLB	61	Duo、代码生成、测试、安全与DevSecOps Agent	结合代码仓库、CI/CD和安全流程提升研发效率；按开发者席位及AI用量收费
CrowdStrike	CRWD	2137	Charlotte AI、终端安全与安全Agent	利用威胁情报和终端数据自动研判、响应和处置；安全平台订阅+模块化扩容
Veeva Systems	VEEV	347	生命科学云、内容/合规与商业运营AI	基于药企合规数据和研发、商业流程提供AI能力；行业软件订阅+服务收费
Tempus AI	TEM	84	医疗数据、精准诊疗、临床AI	将临床、病理和基因组数据用于诊断与治疗决策；检测、数据服务和软件收入
C3.ai	AI	15	企业AI应用平台、工业与能源AI应用	面向能源、制造、国防等行业提供预制AI应用；订阅及项目型部署收费
SoundHound AI	SOUN	28	语音AI、车载和客服Agent	语音交互、点餐、车载助手和客服自动化；按部署、调用量或客户合同收费

资料来源：Wind，公司公告，国信证券经济研究所整理

6.2 附录：A股B端场景垂类AI应用

行业	公司	代码	市值（亿元人民币）	主要AI产品及应用	B端客户
流程工业	中控技术	688777.SH	844	TPT工业时序大模型、生产优化、预测性维护、无人化操作、能耗优化	石化、化工、电力、冶金等流程工业企业
钢铁工业	宝信软件	600845.SH	459	钢铁大模型、MES/ERP/APS智能化、设备运维、质量预测	宝武体系及外部钢铁、制造企业
建筑	广联达	002410.SZ	161	AI算量、AI组价、AI评标、AI清标、方案设计、施工安全	设计院、施工企业、造价机构、建设单位
制造业IT	赛意信息	300687.SZ	111	善谋GPT、制造智能平台、质量/设备/物流分析、企业Agent	制造业大型企业
制造业ERP	鼎捷数智	300378.SZ	99	雅典娜平台、企业AI Agent、生产排程、供应链及经营分析	制造业中大型企业
金融IT	恒生电子	600570.SH	442	LightGPT、智能投研、投顾、客服、运营、合规和知识助手	券商、基金、银行、保险、期货公司
银行IT	宇信科技	300674.SZ	119	信贷、风控、营销、客服、研发智能体	银行及其他金融机构
财税AI	税友股份	603171.SH	203	“犀友”财税大模型、财税Agent、合规税优、涉税风控、经营分析	税务机关、中小企业、财税服务机构
医疗IT	卫宁健康	300253.SZ	172	WiNEX Copilot、病历生成、临床决策、医院运营及医保应用	医院、卫健委、医保机构
法律科技	华宇软件	300271.SZ	47	智慧庭审、审判辅助、法律知识服务、仲裁及破产重组AI	法院、检察院、政法机关、高校及企业法务
城市治理	数字政通	300075.SZ	71	人和大模型、城市智能体、智能执法、12345、城市生命线	地方政府、城管及城市运营部门
网络安全	深信服	300454.SZ	552	安全GPT、威胁研判、自动处置、AI+SASE、私有化AI办公	政企、金融、医疗、教育等
教育/医疗/司法	科大讯飞	002230.SZ	1061	星火大模型、教育、医疗、司法、汽车等行业智能体	学校、医院、政府、车企、央企
企业数据与AI平台	星环科技	688031.SH	143	数据治理、多模数据库、知识图谱、LLMOps、Agent开发、异构算力管理	金融、电力、交通、政企等行业客户
企业软件平台	用友网络	600588.SH	405	企业AI、Data Agent、财务/供应链/人力等业务智能体	大中型企业、央企、制造与金融客户
企业数字化服务	汉得信息	300170.SZ	183	ERP智能化、供应链与制造业Agent、企业数据治理	制造、零售、消费等大型企业

资料来源：Wind，公司公告，国信证券经济研究所整理

6.2 附录：传媒娱乐AI应用产业链



领域	公司	代码	市值（亿元人民币）	主要AI产品及应用	落地场景
AI营销	蓝色光标	300058.SZ	565	AI创意生成、智能投放、数字人及出海营销服务	品牌广告主、电商及出海客户
AI营销/出海	易点天下	301171.SZ	241	AI营销平台、创意生成与投放优化	跨境电商、游戏、应用及AI原生企业
AI营销	因赛集团	300781.SZ	48	因赛AI引擎、营销策划/海报/商品图生成	品牌客户、营销代理及电商商家
创意软件/视频	万兴科技	300624.SZ	110	Filmora等视频创作软件的AI剪辑、生成与Agent功能	中小企业、内容机构、专业创作者；B/C混合
视觉内容/版权	视觉中国	000681.SZ	132	视觉素材版权库、AIGC内容管理、视觉内容服务	媒体、广告公司、品牌方、平台
游戏研发与运营	三七互娱	002555.SZ	448	AI用于美术、文案、客服、投放、运营及玩法探索	以内部游戏研发和发行体系为主
游戏研发工具	恺英网络	002517.SZ	393	SOON全流程AI开发平台：动画渲染、建模、脚本生成	主要内部研发团队；可关注未来工具外溢
游戏AI/内容互动	巨人网络	002558.SZ	571	自研大模型能力、多模态模型、智能体及AI产品探索	游戏研发、运营及玩家互动场景
IP内容/阅读	中文在线	300364.SZ	193	网文IP、AI辅助创作、短剧/动漫等多模态内容开发	内容平台、版权合作方、创作者生态
互动娱乐/游戏	完美世界	002624.SZ	224	AI辅助游戏研发、美术与内容生产	主要内部研发、发行及运营
AI漫剧/IP内容	中文在线	300364.SZ	193	逍遥AI、次元神笔；用于AI漫剧、AI仿真人剧、网文及短剧生产	内容平台、版权合作方、制作团队；兼具C端内容分发
AI短剧/数字阅读IP	掌阅科技	603533.SH	104	AI短剧全流程：剧本生成、视觉制作、成片交付；数字阅读IP转化	短剧平台、内容制作与发行合作方
AI影视/短剧	华策影视	300133.SZ	143	AI漫剧、AI真人剧、AI本地化翻译及国际发行	长短视频平台、海外发行渠道、品牌与内容合作方
AI短剧出海	昆仑万维	300418.SZ	571	DramaWave短剧平台、AI漫剧与AI短剧自动化制作、多语种发行	海外内容消费者、渠道及广告/分发合作方
AI影视制作	百纳千成	300291.SZ	52	AI创作、AI精准营销赋能影视、剧集和互动影视游戏	视频平台、品牌方、内容联合出品方
AI影视/IP	欢瑞世纪	000892.SZ	40	AI漫剧、AI真人剧；剧本孵化、素材生成、镜头设计与后期制作	视频平台、品牌及IP合作方

资料来源：Wind，公司公告，国信证券经济研究所整理

风险提示

技术不达预期风险；

商业化不达预期风险；

政策风险；

经营管理风险等。

免责声明



国信证券投资评级			
投资评级标准	类别	级别	说明
报告中投资建议所涉及的评级（如有）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6到12个月内的相对市场表现，也即报告发布日后的6到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作为基准。A股市场以沪深300指数（000300.SH）作为基准；新三板市场以三板成指（899001.CSI）为基准；香港市场以恒生指数（HSI.HI）作为基准；美国市场以标普500指数（SPX.GI）或纳斯达克指数（IXIC.GI）为基准。	股票投资评级	优于大市	股价表现优于市场代表性指数10%以上
		中性	股价表现介于市场代表性指数±10%之间
		弱于大市	股价表现弱于市场代表性指数10%以上
		无评级	股价与市场代表性指数相比无明确观点
	行业投资评级	优于大市	行业指数表现优于市场代表性指数10%以上
		中性	行业指数表现介于市场代表性指数±10%之间
		弱于大市	行业指数表现弱于市场代表性指数10%以上

分析师承诺

作者保证报告所采用的数据均来自合规渠道；分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求独立、客观、公正，结论不受任何第三方的授意或影响；作者在过去、现在或未来未就其研究报告所提供的具体建议或所表述的意见直接或间接收取任何报酬，特此声明。

重要声明

本报告由国信证券股份有限公司（已具备中国证监会许可的证券投资咨询业务资格）制作；报告版权归国信证券股份有限公司（以下简称“我公司”）所有。本报告仅供我公司客户使用，本公司不会因接收人收到本报告而视其为客户。未经书面许可，任何机构和个人不得以任何形式使用、复制或传播。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，一切须以我公司向客户发布的本报告完整版本为准。

本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。我公司不保证本报告所含信息及资料处于最新状态；我公司可能随时补充、更新和修订有关信息及资料，投资者应当自行关注相关更新和修订内容。我公司或关联机构可能会持有本报告中所提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或金融产品等相关服务。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意见或建议不一致的投资决策。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。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。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，我公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

证券投资咨询业务的说明

本公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询，是指从事证券投资咨询业务的机构及其投资咨询人员以下列形式为证券投资人或者客户提供证券投资分析、预测或者建议等直接或者间接有偿咨询服务的活动：接受投资人或者客户委托，提供证券投资咨询服务；举办有关证券投资咨询的讲座、报告会、分析会等；在报刊上发表证券投资咨询的文章、评论、报告，以及通过电台、电视台等公众传播媒体提供证券投资咨询服务；通过电话、传真、电脑网络等电信设备系统，提供证券投资咨询服务；中国证监会认定的其他形式。

发布证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。



国信证券
GUOSEN SECURITIES

国信证券经济研究所

深圳

深圳市福田区福华一路125号国信金融大厦36层

邮编：518046 总机：0755-82130833

上海

上海浦东民生路1199弄证大五道口广场1号楼12楼

邮编：200135

北京

北京西城区金融大街兴盛街6号国信证券9层

邮编：100032