



计算机行业研究

买入（维持评级）

行业周报

证券研究报告

计算机组

分析师：李可夫（执业 S1130525120009） 分析师：刘高畅（执业 S1130525120005）

likefu@gjzq.com.cn

liugaochang@gjzq.com.cn

Token 商业化加速

前沿模型商业化加速，Anthropic ARR 快速增长已到盈利拐点。 SemiAnalysis 数据显示，Anthropic 的 ARR 已由 2025 年底的 90 亿美元增至目前超过 600 亿美元，并有望于 2026 年三季度实现 GAAP 息税前利润转正。公司收入以用量计费的 API 业务为主，客户部署更多智能体工作流后，Token 消耗和收入可在不增加客户数量的情况下持续扩张。较高的净收入留存率进一步表明，现有客户需求仍处于快速增长阶段。OpenAI 方面预计 ARR 达到 400 亿美元，近期 Anthropic 解除对 Fable 5 和 Mythos 5 的限制，OpenAI 发布 GPT-5.6 分 Sol/Terra 和 Luna 三个档位，推出 Ultra 性能设置，进一步支撑两家模型厂商 ARR 增长。

Token 经济性逐步兑现，算力资本开支具备较强回报基础。 1) 随着模型推理效率持续提升，Anthropic 单位算力对应的收入快速增长，SemiAnalysis 预测今年晚些时候达到每 GW 600 亿美元 ARR，综合毛利率和 API 业务毛利率明显改善。按照 Epoch AI 对 1GW 人工智能数据中心建设及运营成本的测算，前沿模型每 GW 算力所能支撑的年收入显著高于其年化总成本 85 亿美元，表明头部模型厂商扩张算力投入仍具备较好的商业回报，资本开支有望继续转化为收入和利润增长。2) Meta 计划将计算基础设施容量由 2026 年的 7GW 提升至 2027 年的 14GW，并与三星合作加快自研 AI 芯片量产及先进制程供应链布局，进一步验证头部科技公司仍在通过扩大算力规模、降低单位成本来提升资本开支回报。

从 Coding 走向 Work，AI Agent 打开更广阔知识工作市场。 市场或担忧当前大模型需求主要由程序员的 Coding 场景驱动，单一领域难以长期支撑高速增长。OpenAI 发布的 Codex 研究显示，非开发者用户增速显著快于开发者，AI 编程工具的使用范围正逐步扩展至产品管理、教育、设计、营销、数据分析及财务等通用知识工作场景。国内 WorkBuddy、TRAE Work 等办公智能体产品也在加快落地。Coding 能力是智能体执行复杂电脑任务的重要基础，但更广泛的 Work 场景对应着更大的潜在用户和需求空间，当前大模型在知识工作人群中的渗透仍处于早期阶段。

投资建议

相关标的：

海外算力：东山精密、工业富联、胜宏科技、江海股份、中际旭创、中钨高新、蓝思科技、东阳光、火炬电子、三环集团、欧科亿、天孚通信、鼎泰高科、新易盛、领益智造、兆易创新、鹏鼎控股、唯科科技、天岳先进、大普微、源杰科技、麦格米特、景旺电子、英维克、京东方等；英特尔、SK 海力士、Lumentum、闪迪、铠侠、美光。

国内算力：寒武纪、海光信息、东阳光、中芯国际、华虹半导体、禾盛新材、摩尔线程、天数智芯、沐曦股份、壁仞科技、协创数据、华丰科技、杰华特、京基智农、华达科技、中科曙光、利通电子、浪潮信息、利扬芯片、胜蓝股份、华勤技术、国科微、中国长城、晶科科技、罗曼股份、盈峰环境、芯原股份、晶科科技、亿田智能、豫能控股、星环科技、鸿日达、盛视科技、神州数码、润泽科技、大位科技、润建股份、奥飞数据、瑞晟智能、科华数据、潍柴重机、欧陆通、杰创智能、奥尼电子。

大模型&云：智谱、MiniMax、阿里巴巴、腾讯控股、金山云、百度集团、优刻得、首都在线、网宿科技、云赛智联、青云科技等。

风险提示

行业竞争加剧的风险；技术研发进度不及预期的风险；AI 资本开支压力过高的风险。



内容目录

Anthropic ARR 增长提速，盈利拐点已现	3
Token 经济模型跑通：单位算力收入显著覆盖成本	4
从 Coding 走向 Work：AI Agent 打开更广阔市场	5
风险提示	9

图表目录

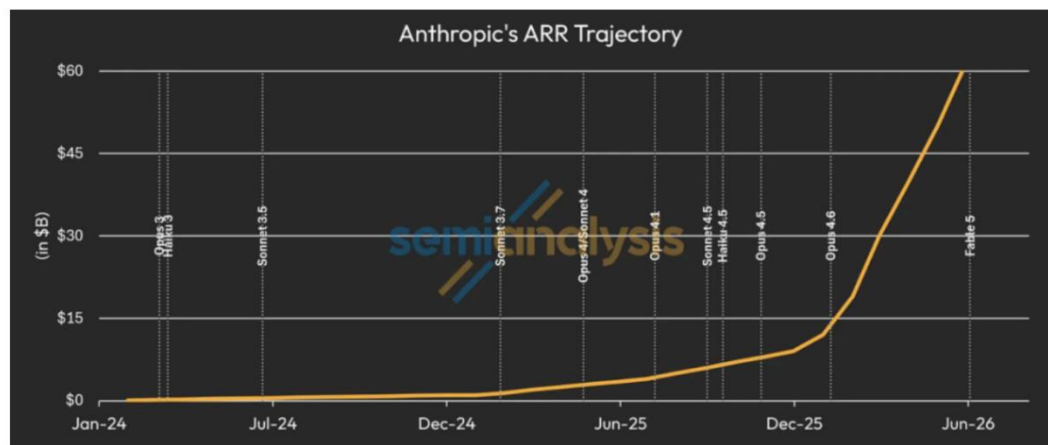
图表 1：Anthropic ARR 增长轨迹	3
图表 2：GPT-5.6 在跨专业领域的长周期智能体 workflow 表现优秀	3
图表 3：每 GW 数据中心前期资本支出	4
图表 4：每 GW 数据中心年度资本支出和运营支出	5
图表 5：Codex 的开发者和非开发者活跃用户增长，以 2025 年 8 月 1 日为基准值 1	6
图表 6：按账户类别划分 Codex 用户比例	7
图表 7：中国 PC 端 AI 原生办公智能体平台月访问量及增速	8
图表 8：NEW TRAE Work	8



Anthropic ARR 增长提速，盈利拐点已现

7月8日，华尔街见闻援引 SemiAnalysis 报告称，Anthropic 预计将于 2026 年三季度实现 10 亿美元的 GAAP 息税前利润，对应约 6% 的利润率。与此同时，其年度经常性收入(ARR) 已从 2025 年底的 90 亿美元飙升至目前逾 600 亿美元。SemiAnalysis 预测，若 Anthropic 维持每月约 150 亿美元的净新增 ARR (NNARR) 节奏，其 2027 年末 ARR 有望触及 3000 亿美元。

图表1: Anthropic ARR 增长轨迹



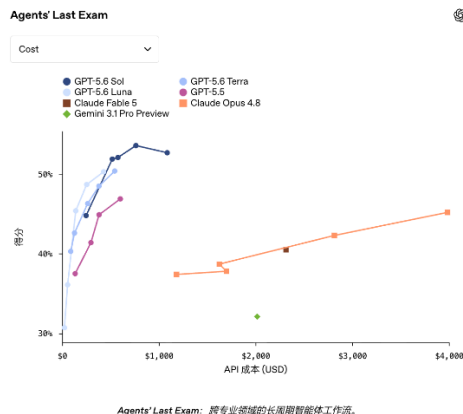
来源：华尔街见闻，SemiAnalysis，国金证券研究所

在收入结构上，Anthropic 约 75%至 85%的 ARR 来自基于用量计费的 API 业务，消费端订阅仅占总 ARR 的 5%。相比之下，OpenAI 在 2026 年一季度仍有逾 65%的收入来自订阅模式，消费端 ARR 占比约 40%。SemiAnalysis 指出，API 模式的核心优势在于无单用户收入上限——随着同一客户采用更多智能体工作流 (Agentic Workflow)，其 token 消耗量与对应收入将持续增长，无需新增客户即可实现扩张。Anthropic 首席财务官 Krishna Rao 今年 5 月在播客中披露，公司净收入留存率 (NRR) 高达 500%，即在一季度贡献 300 亿美元 ARR 的客户中，一年前这批客户仅贡献了 20 亿美元。

6月30日，Anthropic 宣布已经其前沿模型 Fable 5 和 Mythos 5 的出口管制已经解除。我们认为相关限制解除有望改善模型的全球供给与客户拓展，进一步支撑 Anthropic ARR 增长。

OpenAI 方面，SemiAnalysis 称 Anthropic 与 OpenAI 合计约有 1000 亿美元 ARR；结合 Anthropic 已超过 600 亿美元 ARR 的口径，可推算 SemiAnalysis 对 OpenAI 当前 ARR 的判断大致在 400 亿美元左右。7月9日，OpenAI 推出 GPT-5.6，包括全新旗舰模型 Sol、适用于日常工作的均衡模型 Terra，以及极具性价比的模型 Luna。

图表2: GPT-5.6 在跨专业领域的长周期智能体工作流表现优秀



来源：OpenAI，国金证券研究所



GPT-5.6 SoI 树立了智能与效率的新标杆，在编程、知识型工作、网络安全及科学领域均取得了行业前沿水平的成果，并带来了更高的单位成本对应的性能提升：在相同预算下能完成更多任务，或以更低的总成本达到同等效果。

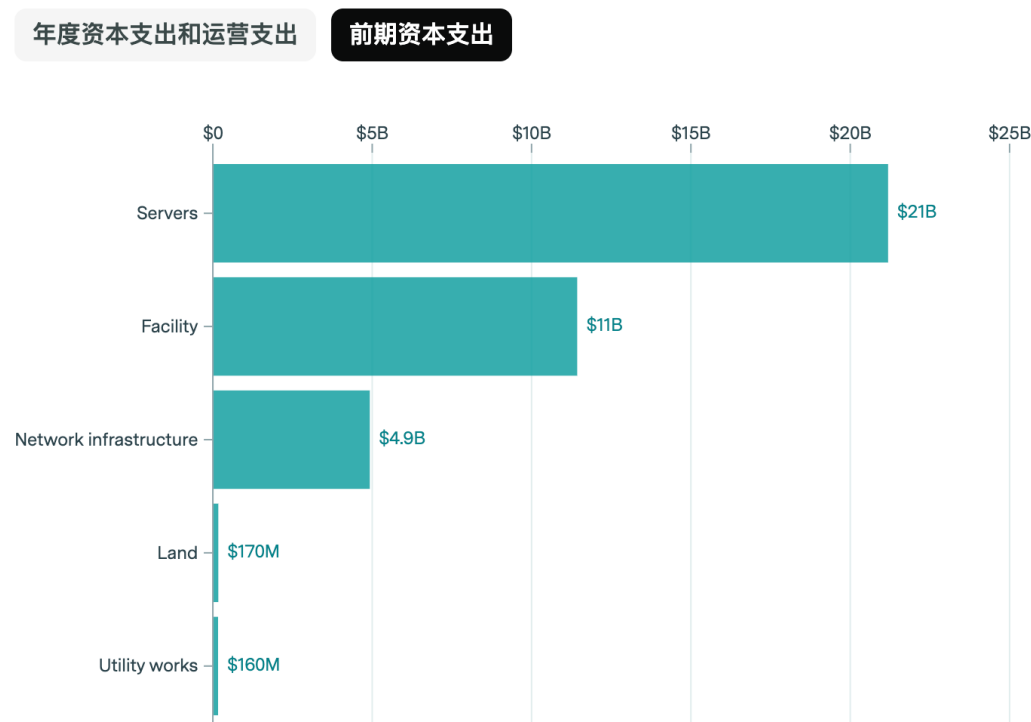
OpenAI 还引入了一种加速处理最繁重工作的全新方式：ultra 是最高性能的设置，能够跨多个并行工作流协调多个智能体，更快地完成复杂任务。更强的计算机使用能力与设计判断力，使 GPT-5.6 SoI 成为 OpenAI 目前协作体验最为出色的助手，能够自行检查、打磨并交付可直接使用的成果。

Token 经济模型跑通：单位算力收入显著覆盖成本

SemiAnalysis 估算，Anthropic 当前综合毛利率已升至 60%中段区间，而 2024 年这一数字为负 94%。其中，API 业务毛利率超过 80%。毛利率大幅改善的核心驱动力是推理效率的提升。以每兆瓦算力对应的 ARR 衡量，Anthropic 该指标将于今年晚些时候达到 6000 万美元，而九个月前仅为 1600 万美元。也就是 GW 对应 600 亿美元 ARR。

epoch AI 估算，一个典型的 1GW 人工智能数据中心需要 380 亿美元的前期资本支出 (CapEx) 和 9 亿美元的年度运营支出 (OpEx)。如果将资本支出按每项资产的使用寿命年化，则总拥有成本相当于每年 85 亿美元。服务器成本占比最高，每年达 50 亿美元，即总成本的 60%。相比之下，运营成本则相对较小：即使是运营支出中最大的能源项目，每年也仅需 6 亿美元。年度成本计算对 IT 设备的使用寿命非常敏感，Epoch AI 假设 IT 设备的使用寿命为 5 年，设施的使用寿命为 14 年。如果将 IT 设备的使用寿命缩短至 3 年，则年度总成本将增加至 120 亿美元；如果延长至 7 年，则成本将降低至 70 亿美元。

图表3：每 GW 数据中心前期资本支出

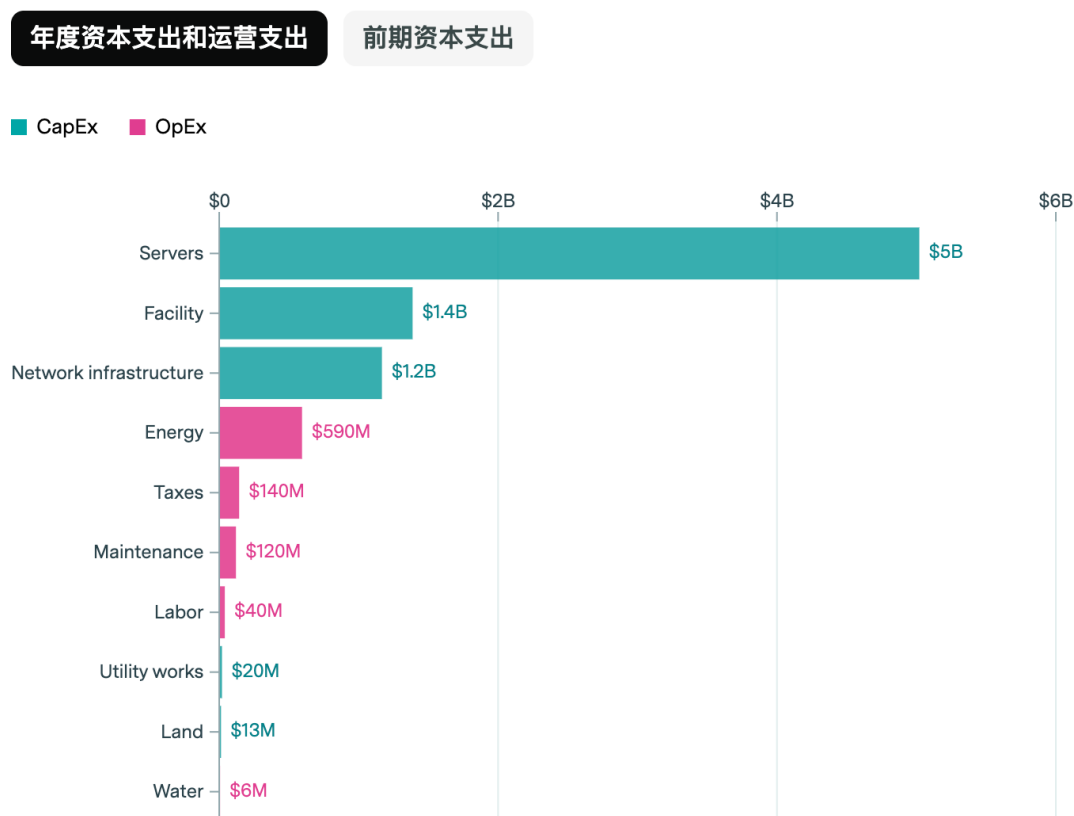


该模型基于美国超大规模人工智能数据中心的简化模型，该数据中心拥有1GW的IT设备铭牌容量。“设施”包括建筑外壳、机械系统和电气设备。

来源：epoch. ai，国金证券研究所



图表4：每 GW 数据中心年度资本支出和运营支出



基于一个美国超大规模人工智能数据中心的简化模型，该数据中心拥有1GW的IT设备额定容量。资本支出（CapEx）以资本成本作为折现率，转换为每项资产在其使用寿命周期内的年度成本。“设施”包括建筑外壳、机械系统和电气设备。

来源：epoch.ai，国金证券研究所

我们认为，据上述测算，对于 Claude 这类前沿模型，每 GW 算力对应的年收入显著高于年化成本，表明当前算力投入已具备较强的商业回报。

据财联社 7 月 10 号报道，Meta 计划将计算基础设施容量从 2026 年的 7GW 提高至 2027 年的 14GW，同时最快于 2026 年 9 月量产自研 AI 芯片 Iris。

根据韩国媒体《首尔经济日报》报道，由于台积电尖端制程产能已经被头部大客户预定一空，Meta 已计划将其后续定制化 AI 芯片的生产交由三星电子代工生产，合约总价值高达 10 万亿韩元（约合 65.64 亿美元）。根据双方的新合作协议，三星电子将利用其 2nm 制程节点，为 Meta 生产第三代至第五代的定制化 AI 芯片，首批预计将生产超过 10 万颗 AI 芯片。Meta 为了实现高达 5GW 算力数据中心的部署目标，正在维持每 6 个月推出新一代 AI 芯片的高效开发计划。

报道称，Meta 此次转单的主要原因在于，台积电的先进制程产能目前已经被其一线和二线客户全数锁定，导致没有额外的产能供给其他客户。相比之下，正在努力发展代工业务的三星电子则拥有足够的 2nm 产能供给外部客户，并且愿意从最早期设计阶段就提供一站式的共同开发解决方案，其专属芯片设计部门将与 Meta 共同开发这些未来模块的构架。这种联合模式促使 Meta 选择三星电子作为最佳合作伙伴，不仅能保障其供应链稳定，更能加速产品的推出。

我们认为 Meta 将算力容量由 7GW 提升至 14GW，进一步验证头部厂商扩张 AI 资本开支仍具备较强经济性。自研芯片加速量产并引入三星代工，也有望通过降低单位算力成本、保障先进制程供给，进一步提升算力投入回报。

从 Coding 走向 Work: AI Agent 打开更广阔市场

我们认为市场对大模型厂商 ARR 增长的可持续性存在担忧，核心原因在于当前需求主要由



程序员的 Coding 场景驱动，单一领域或难以长期支撑收入高速增长。

6 月 25 日 OpenAI 发布的论文《The Shift to Agentic AI: Evidence from Codex》显示非程序员对 Codex 的需求也非常旺盛，Codex 的采用并不局限于最初的开发者群体。开发者仍占据用户的很大比例，尤其是在个人和组织账户中，但非开发者的增长更为迅速：

图表5: Codex 的开发者和非开发者活跃用户增长，以 2025 年 8 月 1 日为基准值 1

Fig. 2: Growth in active Codex users by account type and persona

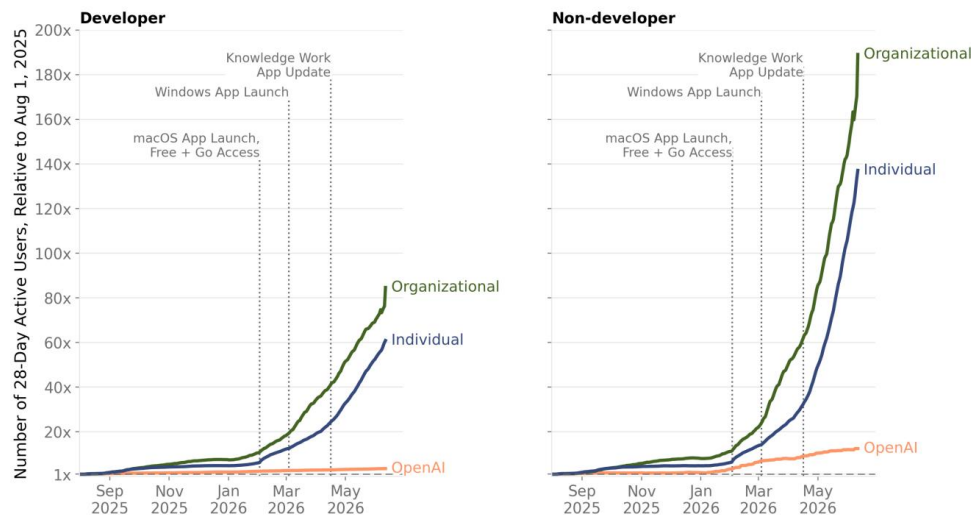


Figure shows growth in users active in the previous 28 days, by persona and account type. The Non-developer series pools users assigned to the Personal and General Knowledge Worker personas. Personas are assigned using the most common label assigned to a user's queries in the 30 days prior to June 11, 2026. Users' account types are allowed to vary over time. All series are indexed to have a value of 1 as of August 1, 2025. Estimates are based on a 3% sample of users.

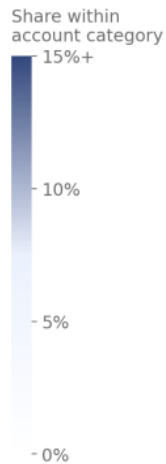
来源：《The Shift to Agentic AI: Evidence from Codex》，国金证券研究所

从 Codex 的用户类别来看，除了开发者，通用知识工作者和个人用户也占据相当高比例。Codex 的使用范围扩展至产品管理、教育、设计、营销、数据分析及财务等通用知识工作场景。



图表6：按账户类别划分 Codex 用户比例

	Organizational	Individual
Developer		
Full stack engineer	18%	21%
Other developer	5%	6%
ML/AI engineer	5%	6%
DevOps/SRE	11%	5%
Backend engineer	10%	5%
Frontend engineer	4%	5%
Mobile engineer	4%	4%
Engineering leader	4%	3%
Data engineer	3%	1%
Systems infra engineer	2%	1%
Research engineer	0.5%	1.0%
Security engineer	1%	0.8%
General Knowledge Worker		
Education/student/academic	2%	7%
Product project management	7%	4%
Design research	2%	3%
Marketing growth	2%	3%
Healthcare life sciences	1%	2%
Data analysis business ops	3%	2%
Operations admin	2%	2%
Writing content	1%	2%
Finance accounting	1%	1%
Other professional	1%	0.8%
Consulting strategy	1%	0.8%
Real estate construction	0.6%	0.9%
Sales customer success	1%	0.6%
Legal policy	0.6%	0.6%
HR/recruiting	0.6%	0.4%
Personal		
Other personal	1%	2%
Hobbies creative	0.8%	2%
Education/self-learning	0.5%	1%
Personal finance	0.5%	1%
Entertainment	0.5%	1%
Household admin	0.4%	0.7%
Relationships communication	0.2%	0.4%



来源：《The Shift to Agentic AI: Evidence from Codex》，国金证券研究所

国内，腾讯的办公 Agent 应用 WorkBuddy 近期爆火，最新发布的《中国办公智能体平台市场研究报告 2026》显示，腾讯 WorkBuddy 在国内 PC 端 AI 原生办公智能体市场中月访问量和环比增速均位列第一。



图表7：中国PC端AI原生办公智能体平台月访问量及增速

2026年3月 中国PC端AI原生办公智能体平台月访问量及增速情况			
产品名称	月访问量 (百万)	产品名称	环比增速 (%)
WorkBuddy	8.85	WorkBuddy	831%
Trae.cn	3.34	EasyClaw	449%
QClaw	2.38	CodeBuddy	360%
QoderWork	2.15	AutoClaw	226%
CodeBuddy	1.01	QClaw	123%
AutoClaw	0.76	QoderWork	82%
Coze.cn (桌面端)	0.72	Trae.cn	76%
悟空	0.53	Coze.cn (桌面端)	69%
LobsterAI	0.47	LobsterAI	68%
EasyClaw	0.38	MiniMax Agent	22%

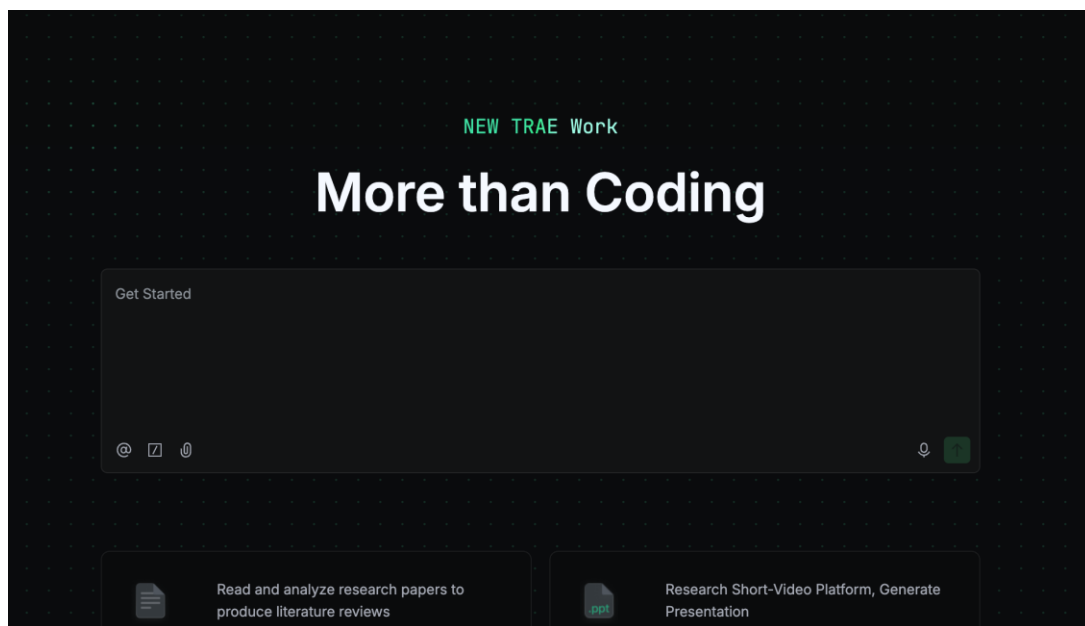
数据统计说明：“月访问量”指在统计周期内用户与PC端AI原生办公智能体平台的交互次数。部分产品为2026年3月发布正式公测版本，“环比增速”为公测与内测期间的月度环比增速。核心数据来源包括运营商数据、部分平台披露及抽样监测数据、易观用户调研数据、网络公开数据等，通过易观自有模型测算得出。仅统计通过电脑访问产品桌面端应用的用户行为数据，不包括通过Web页面/浏览器插件等浏览器端、手机/平板等移动端、办公软件平台内嵌智能体功能的数据。

© 易观分析 www.analysys.cn

来源：搜狐网，易观分析，国金证券研究所

字节的Trae也在强调Work概念,推出 New Trae Work,官网界面提出“More Than Coding”。

图表8：NEW TRAE Work



来源：trae 官网，国金证券研究所

我们认为，Coding 能力的提升，是AI Agent 执行复杂电脑操作的重要基础。相较于程序员群体，各类知识工作者对应的 Work 场景市场空间更为广阔，当前大模型在整体人群中的渗透仍处早期，增长空间并未受限于程序员需求。



风险提示

行业竞争加剧的风险：

众多新兴玩家参与到市场竞争之中，若市场竞争进一步加剧，竞争优势偏弱的企业或面临出清，某些中低端品类的毛利率或受到一定程度影响。

技术研发进度不及预期的风险：

技术开发需投入大量资源，如果相关厂商新品研发进程不及预期，表观层面将呈现出投入产出在较长时期的滞后特征。

AI 资本开支压力过高的风险：

AI 资本开支过高可能导致云厂商自由现金流持续承压，并带来折旧压力



行业投资评级的说明：

买入：预期未来 3—6 个月内该行业上涨幅度超过大盘在 15%以上；

增持：预期未来 3—6 个月内该行业上涨幅度超过大盘在 5%—15%；

中性：预期未来 3—6 个月内该行业变动幅度相对大盘在 -5%—5%；

减持：预期未来 3—6 个月内该行业下跌幅度超过大盘在 5%以上。



特别声明：

国金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。

本报告版权归“国金证券股份有限公司”（以下简称“国金证券”）所有，未经事先书面授权，任何机构和个人均不得以任何方式对本报告的任何部分制作任何形式的复制、转发、转载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券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有关的其他任何损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于 C3 级（含 C3 级）的投资者使用；本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

上海

电话：021-80234211

邮箱：researchsh@gjzq.com.cn

邮编：201204

地址：上海浦东新区芳甸路 1088 号

紫竹国际大厦 5 楼

北京

电话：010-85950438

邮箱：researchbj@gjzq.com.cn

邮编：100005

地址：北京市东城区建国内大街 26 号

新闻大厦 8 层南侧

深圳

电话：0755-86695353

邮箱：researchsz@gjzq.com.cn

邮编：518000

地址：深圳市福田区金田路 2028 号皇岗商务中心

18 楼 1806



【小程序】
国金证券研究服务



【公众号】
国金证券研究