下载日志已记录，仅供内部参考

中信建投证券 CHINA SECURITIES

证券研究报告•行业动态报告

Anthropic披露招股书，AI4S/物理AI高速迭代

于海宁
yuhaining@csc.com.cn
SAC 编号:
S1440526080004

汪洁
wangjietxz@csc.com.cn
SAC 编号:
S1440523050003
SFC 编号:BXV744

曹添雨
caotianyu@csc.com.cn
SAC编号:
S1440522080001

方子箫
fangzixiao@csc.com.cn
SAC编号:
S1440524070009

辛侠平
xinxiaping@csc.com.cn
SAC 编号:
S1440524070006

发布日期：2026年10月28日

本报告由中信建投证券股份有限公司在中华人民共和国（仅为本报告目的，不包括香港、澳门、台湾）提供。在遵守适用的法律法规情况下，本报告亦可能由中信建投（国际）证券有限公司在香港提供。同时请务必阅读正文之后的免责条款和声明。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

摘要

- 核心观点：目前全球大模型已经进入商业变现周期，ARR/现金流均实现快速增长，且主要模型公司已到达盈利拐点。伴随模型能力的快速提升，新的变现场景接连打开，以Robotic控制为核心的物理AI与以AI制药为核心的AI4S的突破均意味着模型仍有较大变现提升空间，而2027年可预见的训练算力占比提升，则为后续资本开支提供了较为充分的需求保障，整体AI行业仍处于高速增长阶段。
- 全球大模型商业化进程加速，Token需求与模型厂商收入持续高速增长。2026年以来，伴随Coding Agent、自动化工作流及长程任务等应用快速渗透，全球Token调用量呈现加速增长趋势。截至2026年10月25日，OpenRouter平台周度Token调用量达到188万亿，较2025年底增长约32倍，Agent类Token消耗已成为主要增长动力。商业化层面，国内头部大模型厂商ARR合计超过115亿美元，海外OpenAI与Anthropic合计ARR约1240亿美元，较2025年底增长约3.1倍。Anthropic招股书显示，公司2026年第二季度GAAP经营利润率首次转正，2026年上半年经营活动现金流转正，标志着头部大模型厂商正在从高投入阶段迈向商业化与盈利能力同步改善的新阶段。
- Anthropic招股书披露，头部模型厂商迎来盈利与现金流拐点，算力投资的商业回报得到进一步验证。根据招股书，Anthropic 2026年上半年实现收入162.9亿美元，第二季度GAAP经营利润率首次转正至8%，上半年经营活动现金流达到39.3亿美元。公司算力费用占收入比例由2025年第一季度的251%降至2026年第二季度的48%，表明模型能力提升、推理优化及规模效应正推动单位算力经济性持续改善。与此同时，公司截至2026年6月底不可撤销采购承诺约2450亿美元，反映其对中长期算力需求的坚定预期。
- 模型智能水平持续提升，性能与价格分层推动应用市场进一步扩张。海外Claude Opus 5.5、GPT-6 Astra等旗舰模型在复杂推理、Coding及Agent等任务中持续突破，国产Kimi K3.1 Pro等模型快速追赶，与海外第一梯队差距逐步缩小。与此同时，全球大模型市场呈现高性能旗舰模型与高性价比模型并行发展的格局。海外旗舰模型依托较高任务完成率维持溢价，国产模型凭借架构创新、推理效率及成本优势抢占中低端市场。随着模型能力提升及使用成本下降，大模型有望进一步向企业级Agent、复杂Coding及自动化工作流渗透，推动Token需求与商业化空间持续增长。
- 模型能力突破正在打开物理AI与AI4S全新应用空间。GPT-6 Astra在计算机操作、长流程任务及工具调用能力方面实现明显提升，推动Agent从数字世界向物理世界延伸。机器人控制范式有望从传统CV、VLA路线进一步向Robot-Use Agents演进，通用大模型通过调用机器人控制接口，实现跨任务规划、工具调用与异常修正，提升机器人泛化能力。在AI4S领域，OpenAI与Ginkgo Bioworks合作的蛋白质合成实验实现反应材料成本下降约40%，Claude已能够自主规划复杂蛋白质设计与筛选流程。随着模型与自动化实验平台深度结合，AI有望从辅助科研工具升级为能够参与真实实验及研发决策的智能系统，为模型厂商打开新的商业化空间。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

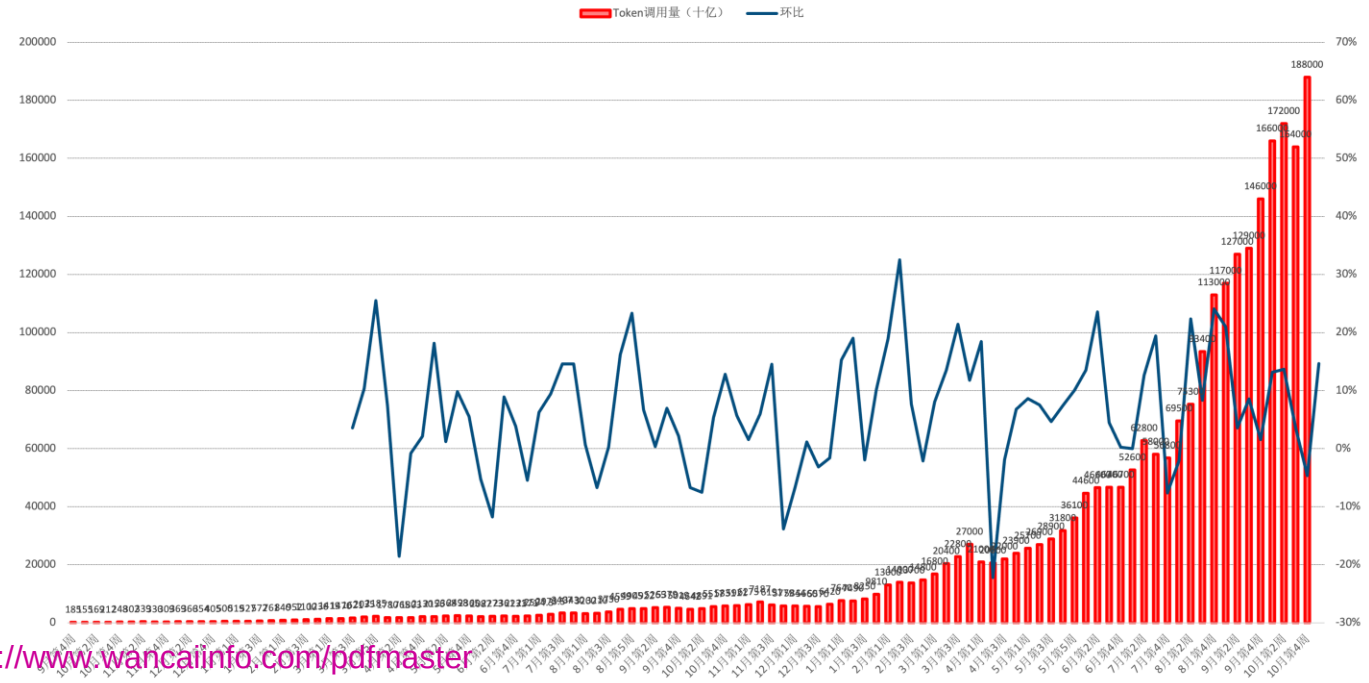
中信建投证券 CHINA SECURITIES

下载日志已记录，仅供内部参考

1.1 全球token调用量快速提升

- 2026年以来，大模型Token调用量呈现加速增长趋势。OpenRouter平台2025年底最后一周Token调用量约5.7万亿，截至2026年10月25日周度调用量已达到188万亿，**较2025年底增长约32倍**。2026年7—9月平台月度Token调用量分别达到248万亿、379万亿、576万亿，环比增速分别为32%、53%、52%，下半年增长明显提速。
- Agent及Coding应用快速渗透，成为推动Token消耗增长的核心动力。2026年2月，OpenRouter平台Agent类Token消耗首次超过人类交互；至9月最后一周，Agent类API Key的Token消耗量已达到人类交互类的约7倍。随着Coding Agent、自动化工作流和工具调用普及，单次任务由简单问答向多轮规划、执行及验证演进，推动Token消耗量持续增长。
- Token消耗结构发生显著变化，推理效率提升进一步释放Agent需求。2026年9月，OpenRouter平台Agent类调用中近90%的Token为缓存输入，较1月约60%的水平明显提升。缓存技术降低了重复上下文处理成本，推动Agent应用经济性改善。展望未来，随着长程任务、多Agent协同及推理时计算持续发展，大模型Token需求有望进一步保持高速增长。

图：OpenRouters 大模型token调用量变化（2024.9-2026.10）



资料来源：OpenRouters，Wind，中信建投

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

中信建投证券 CHINA SECURITIES

下载日志已记录，仅供内部参考

1.2 商业化开始“闭环”—ARR快速增长

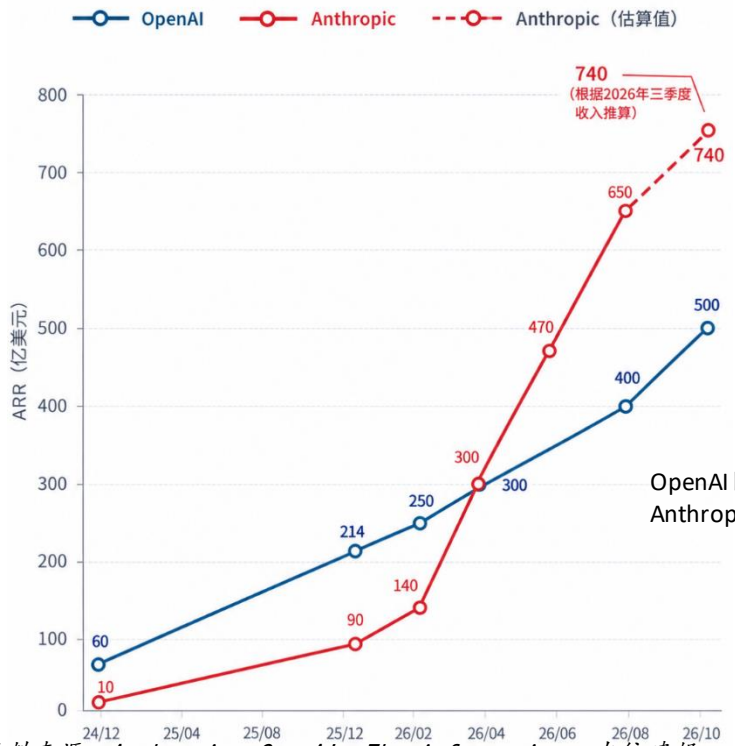
- **国内大模型商业化进程加速，头部厂商ARR合计已超过115亿美元。** 根据各公司披露及媒体报道，字节跳动、阿里巴巴、智谱、Kimi、DeepSeek、MiniMax及可灵AI等7家头部厂商ARR合计超过115亿美元。其中，字节跳动约40亿美元、阿里巴巴约24亿美元、智谱约18亿美元，Kimi与DeepSeek均达到约10亿美元。随着模型能力持续提升、API调用需求快速增长，以及Coding、Agent等应用加速落地，国产大模型商业化规模有望进一步扩大。
- **海外大模型商业化增长势头强劲，OpenAI与Anthropic合计ARR已超过1200亿美元。** 截至2026年9月底，OpenAI ARR约500亿美元，Anthropic根据三季度收入推算的ARR约740亿美元，两家合计约1240亿美元，较2025年底约304亿美元增长约3.1倍。其中，Anthropic于2026年上半年实现对OpenAI的反超，反映出Coding及企业级Agent需求快速释放，驱动大模型厂商收入持续高速增长。

图：国产模型ARR情况

公司	ARR（亿美元）	数据时点	数据来源及口径
字节跳动（豆包/Seed）	40.0	2026年7月	媒体报道，MaaS等模型服务
阿里巴巴（通义千问）	约24.0	2026年8月	管理层披露
智谱（GLM）	18.0	2026年9月	管理层披露，全业务ARR，年底指引30亿美元
月之暗面（Kimi）	10.0	2026年8月	媒体报道，年化收入
DeepSeek	10.0	2026年9月	The Information报道，投资人交流口径，主要来自API
MiniMax	超过8.0	2026年8月	CEO业绩会披露，全业务ARR
快手（可灵AI）	约5.0	2026年3月	公司业绩披露，视频生成业务

资料来源：字节，阿里，智谱，kimi，minimax，中信建投

图：海外模型ARR情况



OpenAI口径：仅计算净收入，不考虑大厂代销渠道。
Anthropic口径：全口径，9月底数据根据三季度推算

资料来源：Anthropic，OpenAI，The information，中信建投

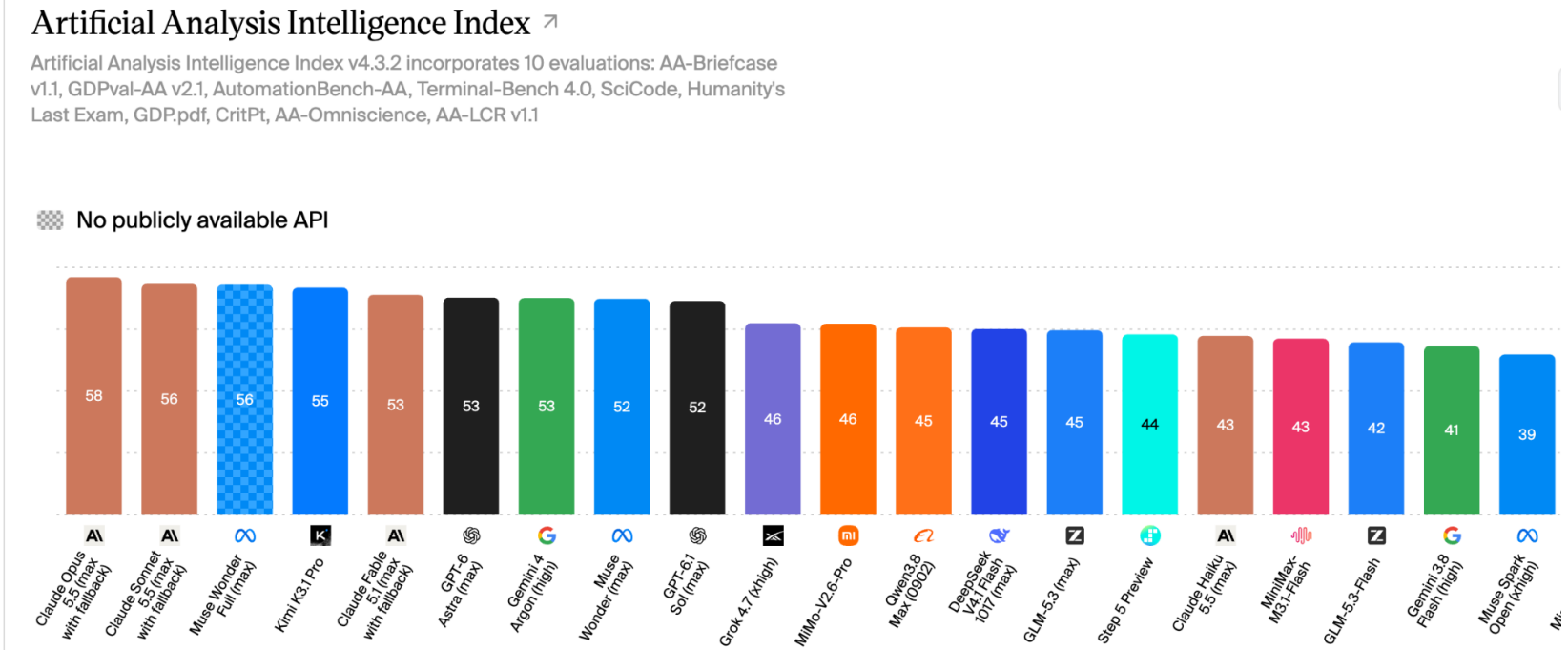
Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

1.3 模型能力：海外头部继续领跑，国产模型缩小差距

- **海外头部模型持续领跑，模型智能水平不断提升。** 根据Artificial Analysis最新Intelligence Index榜单，Anthropic Claude Opus 5.5以58分排名第一，Claude Sonnet 5.5及Meta Muse Wonder均达到56分，OpenAI GPT-6 Astra、Google Gemini 4 Argon达到53分。海外头部厂商依托持续的预训练、强化学习及推理能力优化，在复杂推理、Coding、Agent等综合任务中保持领先。
- **国产模型能力快速追赶，与海外第一梯队差距进一步缩小。** Kimi K3.1 Pro以55分进入全球第一梯队，与榜首Claude Opus 5.5仅相差3分；小米MiMo V2.6 Pro达到46分，MiniMax、DeepSeek、智谱等厂商的旗舰模型也进入全球主流模型阵营。国产模型在推理、Coding及Agent能力方面持续取得突破，部分模型已接近海外头部水平，全球大模型竞争格局呈现多元化发展趋势。

图：全球大模型能力排名



Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

资料来源：Artificial Analysis，中信建投

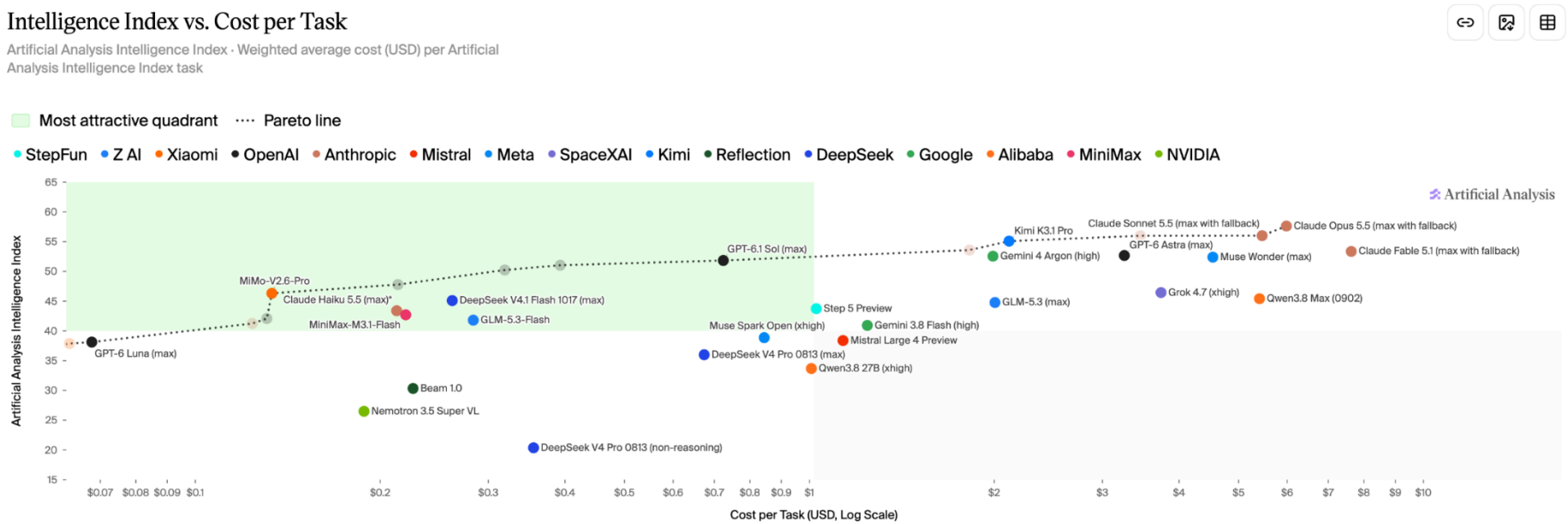
中信建投证券 CHINA SECURITIES

下载日志已记录，仅供内部参考

1.4 价格两级分化，分别抢占中低市场及高价值量场景

- **大模型价格呈现明显分层，海外旗舰模型依托性能优势维持较高溢价。** 根据Artificial Analysis Intelligence Index与Cost per Task数据，Anthropic Claude Opus 5.5、Sonnet 5.5等模型凭借领先的综合能力，单任务成本处于较高水平。其中Claude Opus 5.5单任务成本约6美元，而OpenAI GPT-6 Sol、Google Gemini 4 Argon等模型以相对更低的任务成本实现较高智能水平。随着模型能力持续提升，复杂Coding、企业级Agent等高价值场景对模型性能和任务成功率的要求不断提高，为头部旗舰模型创造差异化定价空间。
- **国产模型依托成本效率优势抢占市场，模型调用价格竞争进一步加剧。** 小米MiMo V2.6 Pro、MiniMax M3.1 Flash、DeepSeek V4.1 Flash、智谱GLM-5.3 Flash等模型，单任务成本普遍处于0.2—0.4美元区间，显著低于海外高端旗舰模型。国产厂商通过模型架构优化、推理效率提升和低价策略，加速拓展高频Coding及Agent调用场景。未来大模型竞争将逐步从单纯的Token价格竞争，转向模型能力、任务完成率和单位任务成本的综合竞争。

图：全球模型的成本&性能对比



Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

资料来源：Artificial Analysis，中信建投

中信建投证券 CHINA SECURITIES

下载日志已记录，仅供内部参考

1.5 Anthropic招股书发布，大模型已进入盈利拐点

- **2026H1首次实现经营盈利。**2025年公司实现收入45.9亿美元，2026Q1/Q2分别达到47.3/115.6亿美元，2026H1累计收入162.9亿美元，较2025全年实现大幅增长。2026Q2 GAAP经营利润率首次转正至8%，调整后经营利润率达到19%。收入结构层面，2026H1按量计费占80%，订阅收入占20%，且按量计费毛利率高于订阅收入，是核心利润来源，美国以外市场贡献28%，全球仍有较大渗透空间。
- **推理优化下算力成本快速下滑。**2025年算力与基础设施开支约73.3亿美元，占总运营费用58%；2026Q2算力费用占收入比例已由2025Q1的251%降至48%，单位算力创收持续快速增长，推理优化对公司单token毛利率提升起到了较大的效果。净亏损主要受非现金公允价值变动影响。
- **长期采购协议规模庞大且大部分不可撤销。**截至2026年6月底，公司不可撤销采购承诺约2450亿美元，且在此之外存在大量框架性订单。主要包括Google约1145亿美元、Broadcom约1612亿美元、Amazon约1100亿美元、Microsoft约449亿美元，以及SpaceX/xAI、Nscale和Lambda等协议。尽管不同协议存在部分交叉，不能简单加总，但大量的7年维度长单仍彰显了企业对未来算力需求的极强信心。

表：2024-2026H1 Anthropic 训推比测算

算力开支（百万美元）	2024	2025H1	2025	2026H1
算力与基础设施	2522	7326	2334	8945
营业成本（服务客户的工作负载）	752	2,773	1,019	2,380
销售、市场与渠道	43	568	120	3,168
销售与市场（含免费层算力）	163	911	239	3,808
训练算力消耗	1650	4210	1196	5925
训推比	65.4%	57.5%	51.2%	66.2%

资料来源：Anthropic招股书，中信建投

表：截至2026H1，Anthropic经营活动现金流已转正

现金流（百万美元）	2023	2024	2025	2026H1
经营活动现金流	-1627	-5692	-2637	+3929
资本开支	-2	-15	-131	-482
自由现金流（简算）	-1629	-5707	-2768	+3447
融资活动现金流	+2320	+7428	+19960	+110774

资料来源：Anthropic招股书，中信建投

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

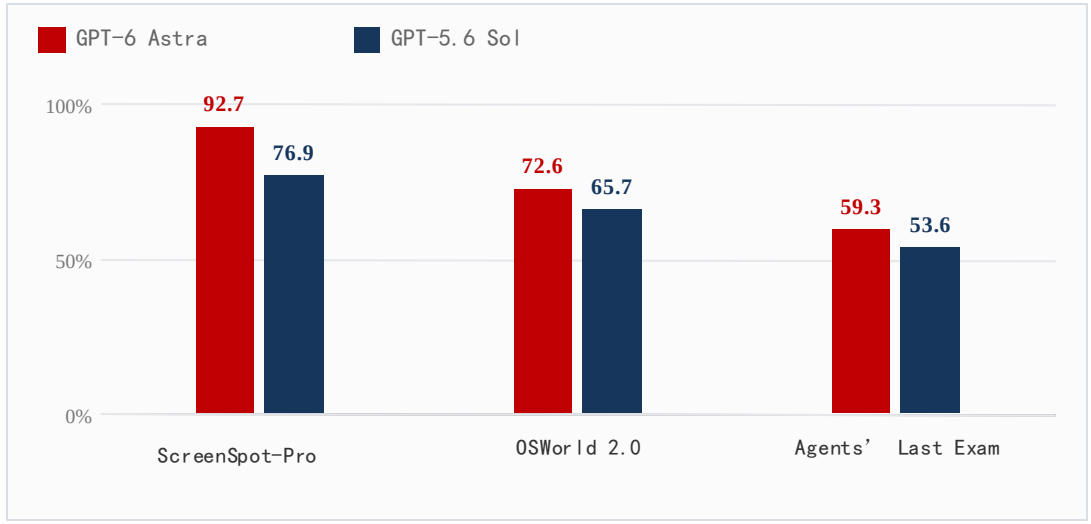
中信建投证券 CHINA SECURITIES

下载日志已记录，仅供内部参考

1.6 GPT-6 Astra发布，Computer use能力大幅提升

- **核心变化：**1）**屏幕定位更准**，ScreenSpot-Pro 从无工具条件下的 76.9% 提升至 92.7%，较 GPT-5.6 Sol 提升 15.8pct；也高于 Fable 5 的 87.3%；2）**完整任务成功率更高**，从单步操作走向端到端完成，OSWorld 2.0 从 65.7% 提升至 72.6%，高于 Opus 5 的 70.2%。提升的不只是“下一步怎么点”，而是连续执行多个步骤后，真正完成文件处理、软件设置和桌面操作等完整任务；3）**长流程执行更稳定**，Agents’ Last Exam 从 53.6% 提升至 59.3%，高于 Fable 5 的 48.7% 和 Opus 5 的 55.5%。其本质提升是：面对需要多阶段推进的任务，模型更能维持目标、组织工具调用，并根据中间结果继续执行，而不是操作几步后停滞或偏离任务
- **Astra测试结果：**ScreenSpot-Pro达**92.7%**、OSWorld 2.0达**72.6%**，较GPT-5.6 Sol分别提升**15.8/6.9pct**；Agents’ Last Exam达**59.3%**。
- **K3.1 Pro：** OSWorld V2.1达**79.6%**、Agents’ Last Exam达**43.2%（Astra 为 32.8、Opus 5.5 为 42.6）**，Agent与空间能力进入第一梯队；

图：Astra关于Computer use的benchmark一览



资料来源：OpenAI《GPT-6 Astra》（2026-09-29），中信建投

图：K3.1 Pro关于Agent与空间能力的benchmark一览



资料来源：KIMI，中信建投

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

1.7 超越CV/VLA路线，Robotics泛化边界打开

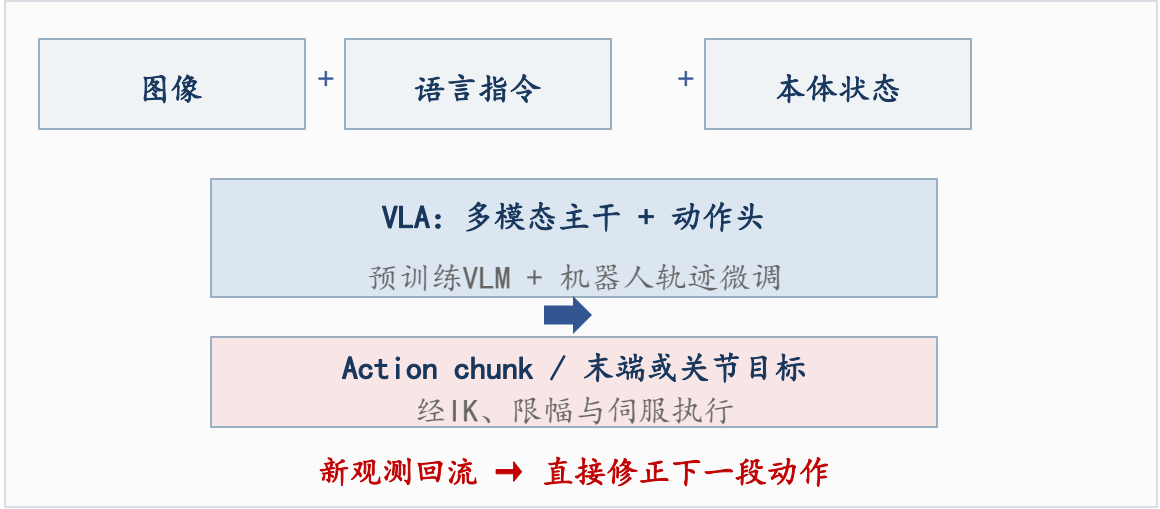
过去机器人大脑陆续经历了CV与VLA路线的变革：

- **CV路线：**图像/点云经检测、分割和位姿估计转为目标状态，再由规划器与控制器输出动作；**感知—规划—控制分离**，模块可解释，但新任务常需重做规则与标注，泛化性较差，并非由一个模型完成端到端的观察与控制。
- **VLA路线：**图像、语言指令与机器人状态输入多模态主干，动作头直接输出末端/关节动作或action chunk；能力主要由机器人轨迹训练获得。**由于从识别到动作指令是由同一个模型完成，模型反应迅速，较为适合执行连贯性动作。**
- **CV与VLA路线的本质差异：**：CV是“识别后按规则执行”，识别与执行分两阶段去做，存在一定信息丢失。VLA是“从观测到动作的统一映射”；动作后果随新观测进入下一次前向推理，**反馈隐式、连续、直达动作层，完成任务概率更高。**

图：CV路线控制机器人的示意图



图：VLA路线控制机器人的示意图



资料来源：RT-2、OpenVLA、π0公开论文，中信建投证券

资料来源：RT-2、OpenVLA、π0公开论文，中信建投证券

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

1.8 Robot-Use Agents范式下，机器人成为大模型子集

- **核心差异：**RUA将通用多模态Agent置于机器人API之上，输入观测、指令和接口文档，输出有界工具调用/控制目标，规划、计算与异常修正在线完成，不必为每个任务重训策略。**模型本身并不输出机器人的运动轨迹数据，而是通过阅读控制文档，输出对应控制代码，机器人实际运动由底层解释器翻译代码之后进行。**
- **泛化来源：**通用底座复用世界知识，输出适配本体的控制代码来屏蔽本体差异；换机器人主要换解释器适配方案而无需重新训练模型。过去的VLA每次更换机器人主体都需要做重新训练。
- **控制器要求：**本体侧需统一动作语义、状态反馈与安全约束，并承接高频伺服；Agent负责低频决策，VLA/原生控制器负责实时执行。**机器人本体所附带的解释器好坏决定了机器人实际动作执行情况。**

图：Robot-Use Agents思路示意图



资料来源：谷歌，中信建投

图：机器人控制领域各家模型仍有较大提升空间

Model	Solved	Manipulation (38 tasks)	Mobile manipulation (20 tasks)	Locomotion (11 tasks)	Driving (11 tasks)	Aerial (4 tasks)
GPT-6 Astra	16/84	16	4	0	2	1
Claude Opus 5.5	13/84	4	0	1	2	3
Kimi K3	2/84	0	1	0	1	0
DeepSeek V4.1 Flash	1/84	0	0	0	0	1
Gemini 3.8 Flash	1/84	1	0	0	0	0
Any model	21/84	9	4	1	2	3

资料来源：谷歌，中信建投

中信建投证券 CHINA SECURITIES

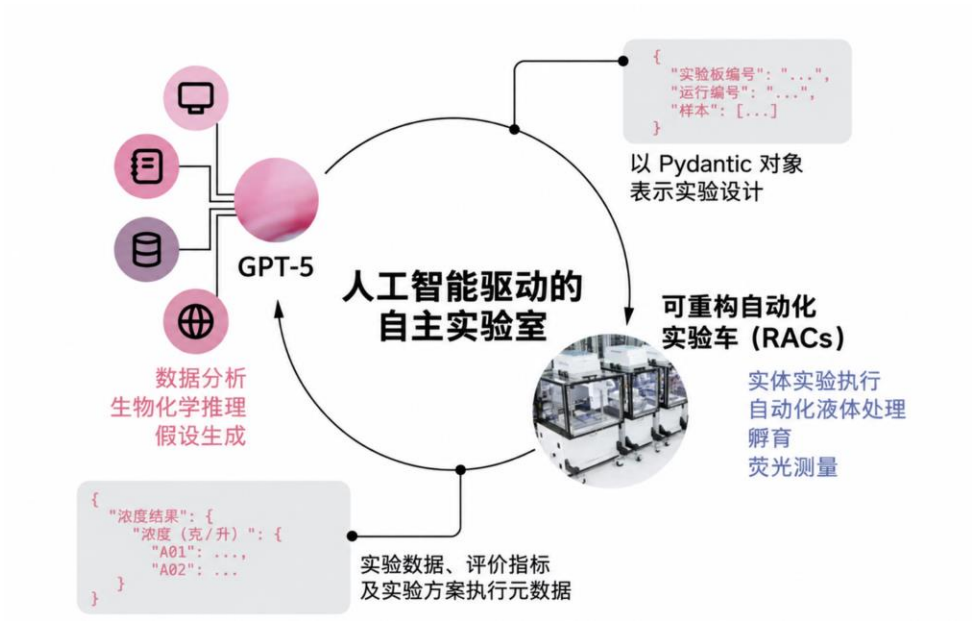
Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

1.9 GPT-5联合自动化实验室优化蛋白质合成

- OpenAI与Ginkgo Bioworks合作，将GPT-5与自动化云实验室连接，验证了大模型通过真实实验迭代优化生物生产工艺的能力。以无细胞蛋白合成为对象，由GPT-5提出反应配方、分析结果并生成下一轮假设，自动化设备负责液体处理、孵育和荧光测量，形成“设计—实验—分析—再设计”的闭环。
- 项目经过六轮迭代，累计测试约3.6万种反应配方；与此前最佳配方相比，测试蛋白的每克反应材料成本由698美元降至422美元，降低约40%，蛋白浓度由2.39克/升提高至3.04克/升，提升约27%。展示了大模型结合实验自动化，提高科研迭代效率和优化工艺的实际价值。

图：GPT-5参与蛋白质合成实验



资料来源：OpenAI，中信建投

图：蛋白质合成实验结果

Reaction composition identifier	Specific Cost ($/g)	Protein Titer (g/L)		
	2-mL tubes	2-mL tubes	96-well plates	384-well plates
RF_{opt}* (SOTA)	$698*	2.39 ± 0.28*	Not applicable*	~0.25*
1777863_77	$422	3.04 ± 0.30	0.66 ± 0.12	0.29 ± 0.01
1784943_26	$425	3.01 ± 0.27	0.86 ± 0.24	0.33 ± 0.05
1793122_35	$430	3.04 ± 0.18	1.52 ± 0.32	0.72 ± 0.38
1793665_74	$446	2.92 ± 0.19	1.31 ± 0.33	0.49 ± 0.05
1793119_36	$466	2.77 ± 0.19	1.28 ± 0.13	0.78 ± 0.58
1794089_36	$470	2.75 ± 0.27	2.00 ± 0.49	0.53 ± 0.07

资料来源：OpenAI，中信建投

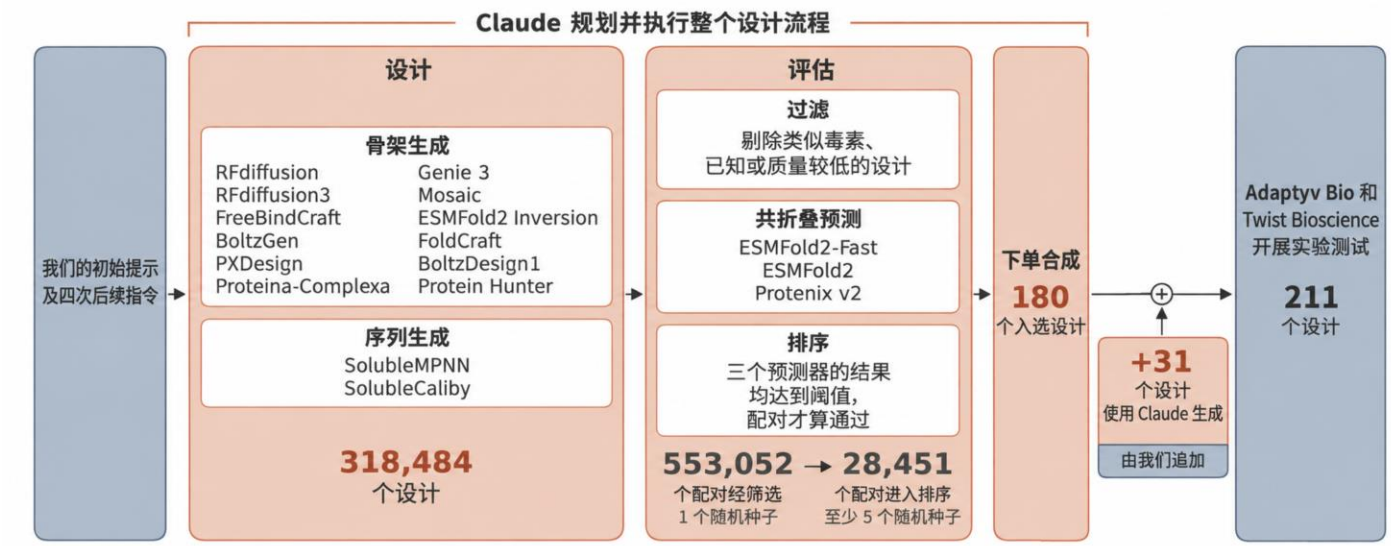
Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

1.10 Claude设计蝎毒神经毒素的结合蛋白实验流程

- **Claude设计蝎毒神经毒素的结合蛋白、同时计算筛选与实验送检。**研究人员提供初始提示及4次后续指令，由Claude规划并执行设计流程：调用专用模型生成蛋白质骨架和氨基酸序列，累计得到318484个候选设计；随后通过质量过滤、共折叠预测和排序，对553052个配对进行初筛，其中28451个配对进入使用至少5个随机种子的进一步评估，最终选出180个设计下单合成。研究人员另外追加31个使用Claude生成的设计，共计211个设计交由Adaptyv Bio和Twist Bioscience开展实验测试。
- 该流程体现了通用大模型负责规划和工具调用、专用模型完成蛋白质设计与结构计算、实验平台验证实际效果的协作方式。整个设计活动Claude发送了367条信息，使用了10000 H100 GPU小时，整个项目预算20万美元，计算花费5万美元。

图：Claude设计蝎毒神经毒素的结合蛋白实验流程



资料来源：anthropic，中信建投

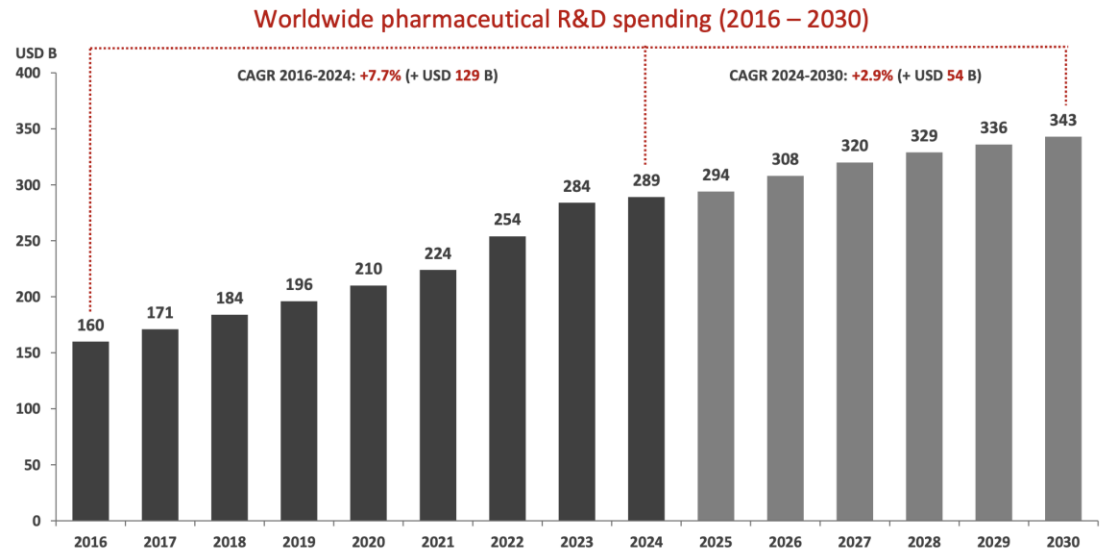
Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

1.11 AI制药有望为模型变现打开全新空间

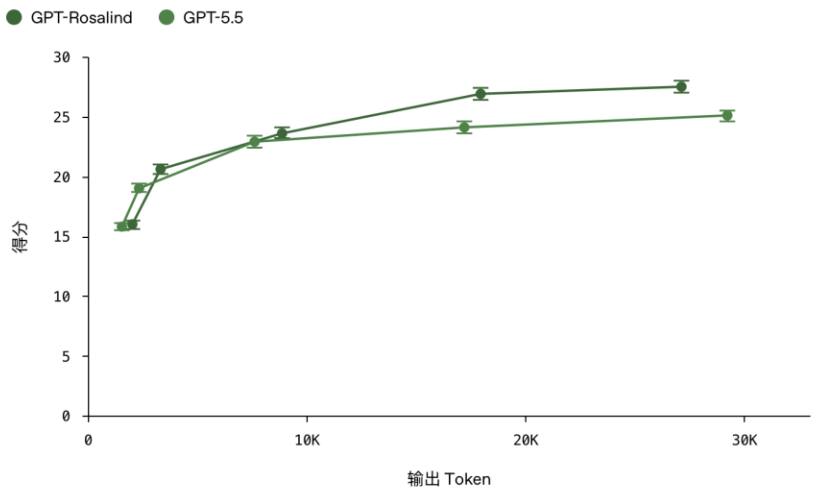
- 根据Smart Pharma Consulting《全球制药研发概览》数据，经过企业披露的研发费用汇总（剔除医疗器械）预计2026年药物研发费用达到3080亿美元。2016—2024年，全球制药研发支出由1600亿美元增至2890亿美元，年复合增长率为7.7%，2030年将进一步增至3430亿美元。
- 中国、美国、日本、德国、韩国、法国和瑞士七国制药行业的内部研发人员合计约45万人，中国约19.13万人，美国约18.30万人。
- OpenAI发布MedChemBench反映真实药物化学工作流，评估大模型的多模态化学结构理解；药物效力、毒性和吸收、分布、代谢、排泄预测；多参数先导优化决策；以及逆合成。

图：全球药物研发经费



资料来源：Smart Pharma Consulting，中信建投

图：MedChemBench测试结果



资料来源：Smart Pharma Consulting，OpenAI，中信建投

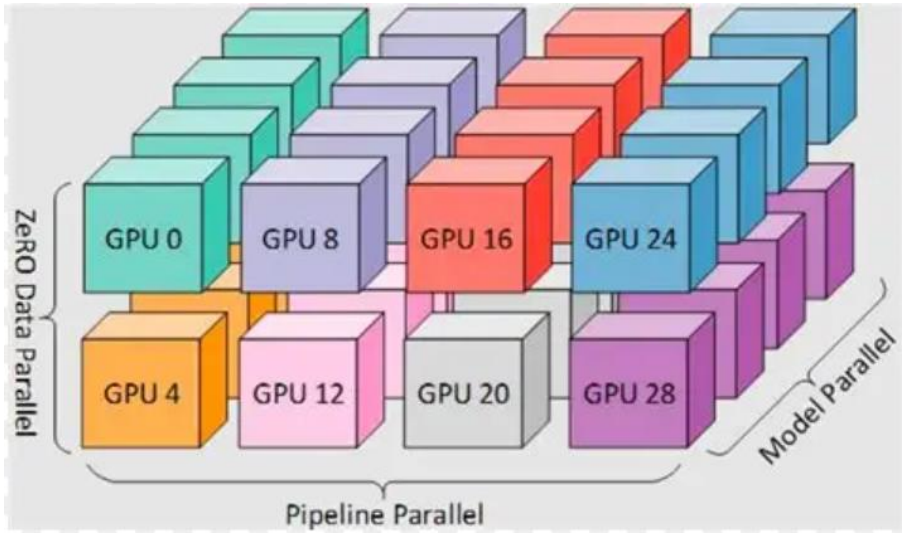
Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

1.12 下半年趋势：重新聚焦训练，核心关注RSI下的模型进展

- **产业现象：目前模型变现不足的核心原因依旧是来源于模型的智力不足。**下半年算力分配开始重新向训练聚焦（Anthropic训练算力分配比例由45%重新回到60%），以探索新的变现可能性
- **Self-envolve在北美已经初步跑通：**目前看北美已经进入每年稳定预训练参数规模翻一倍的阶段（甚至还可能加速），国内目前仍处于每年稳定提升0.5-1t阶段，海外self-envole已经初步跑通，Opus 5本身就是用anthropic内部更大的模型做的训练
- **新一代模型训练需要极为庞大的计算能力支撑：**训练本质是compute bond，依赖单卡高算力与更大的scale out规模，因此边际需求上大概率光进存退。此外，Rubin 的市占率/出货预期提升也是聚焦训练的直接后果。**较高的算力训练消耗占比为27/28年资本开支提供了较强的支撑。**

图：模型训练存在多种并行策略，可由模型自己搜索



资料来源：anthropic，中信建投

图：模型参数规模越大，训练集群规模越大，耗时越长

表格

模型	总参/激活参数	训练token量	预训练算力（6ND）
DeepSeek V4-Pro（2026/4）	1.6T / 49B	33T 百度百科	9.70×10^{24}
GLM-5 / 5.1 / 5.2（2026/2起）	744B / 40B	28.5T DeepDive · x...	6.84×10^{24}
Qwen3-235B-A22B（2025/4）	235B / 22B	36T	4.75×10^{24}
GLM-4.5（2025/7）	355B / 32B	23T DeepDive · x...	4.42×10^{24}
DeepSeek V3（2024/12）	671B / 37B	14.8T	3.29×10^{24}
Kimi K2（2025/7）	1.04T / 32.6B	15.5T	3.03×10^{24}

资料来源：anthropic，中信建投

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

风险提示

- **人工智能模型技术发展不及预期：**AI智能体模型属于先进AI算法，若后续算法更新迭代效果不及预期，则会影响人工智能模型演进及拓展，进而会影响其商业化落地等。
- **算法隐私问题：**AI智能体在进行预训练过程中，使用了大量数据，其中会涉及相关隐私问题。
- **应用不及预期：**AI智能体算法与相关应用落地之间还存在一定距离，需要注意应用落地不及预期风险。
- **算力基础设施支持不及预期：**美国制裁中国高科技企业，对中国形成芯片、算力的封锁，AI智能体模型训练过程中需要大量算力资源，需要关注中美关系带来的中国算力的压力。
- **数据数量与数据质量不及预期：**AI智能体模型需要大量的高质量数据进行训练，若数据数量和质量存在短板，则会影响模型效果。
- **伦理冲突风险：**AI智能体模型将进一步推动人工智能迈向通用型人工智能，AI智能体进一步智能化将产生人工智能欺骗、人工智能上瘾、人与人工智能之间的关系等一系列伦理问题。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster