

IPO前夜 Anthropic抢滩企业市场



随着上市流程进入关键节点, Anthropic把用来狙击“GPT-6”的大模型直接端了出来。相比上一代, 它更擅长处理持续时间较长、步骤较多的复杂任务。更关键的变化是成本下降。Anthropic表示, 对于按Token(词元)计费的典型工作负载, Fable 5.1的预期使用成本将比Fable 5低25%。据报道, Anthropic与OpenAI均已提交不公开上市申请文件, 眼下, 持续扩大企业市场份额、将高估值转化为可见的收入增长, 对公司至关重要。

Anthropic上新Fable 5.1

当地时间9月1日, Anthropic宣布推出Claude Fable 5.1和Claude Mythos 5.1模型。该公司称这两款模型为“全球最先进的编程与知识工作模型”, 虽然采用相同的基础模型, 但有不同的安全防护级别: Fable 5.1面向所有用户开放, 而Mythos 5.1仅通过授权访问计划向经过审核的机构提供。

Anthropic强调, 除了性能升级外, 公开发布的Fable 5.1进行了多项改进, 主要针对从客户那里收到的关于价格、数据保留和安全方面的反馈, 降低Token成本, 并减少来自模型安全护栏的误报限制(指误将无害内容

识别为违规)。

价格方面, 在所有按Token计费的使用场景中, 得益于公司降低了缓存读取(即模型读取已处理并存储的输入内容)的定价, 让其大幅下降75%, 为0.25美元/百万Token。针对典型工作负载, Fable 5.1的成本预计将比Fable 5降低约25%。对于高度依赖智能体的工作任务, 节省的成本还会更多, 最高可达45%左右。此外, Fable 5.1保持了100万Token的上下文窗口和12.8万Token的最大输出长度。

这一价格结构折射出前沿模型的成本逻辑正在发生微妙变化: 输入侧的缓存降价意在降低智能体场景的调用门槛, 但模型本身更强的推理深度意味着更长的输出链, 最终单次任务的总成本不降反升。对于重度依赖

智能体 workflows 的企业用户而言, 缓存命中率将成为决定实际使用成本的关键变量。

值得注意的是, Fable 5.1在低、中等推理强度设置下, 性能已经能够接近甚至超过上一代Fable 5在最高推理强度下的表现。官方数据显示, 当Fable 5.1设置为Low或Medium推理强度时, 在多数任务上的表现与Fable 5的High设置相当或更优, 而单次任务成本显著降低。

OpenAI: 安全优先

非常凑巧的是, Fable 5.1刚刚亮相, OpenAI便刊出一篇博客文章, 称公司计划很快推出Astra模型, 但依然没有透露具体时间

表, 而是将更多表达倾斜在网络安全问题上。OpenAI CEO山姆·奥尔特曼表示, 整个夏天, 团队都全力推进安全方面的工作。如今, 能力和安全保障措施的同步提升比以往任何时候都更加重要。他同时承认, 新模型Astra的训练工作已完成一段时间, 但后续型号根据需要放慢了进度, 以确保能够充分开展安全性和适配性方面的工作。

官方称, Astra已达到准备框架下的关键网络安全能力阈值, 这意味着, 在拥有合适的工具和访问权限的情况下, 它可以发现此前未知的安全漏洞, 并开发出在多个受保护系统中利用这些漏洞的方法, 而无需人工指导每个步骤。这是OpenAI认定的首个达到此级别的模型, 因此在开发过程中和发布前需要更严格的安全保障措施。

过去几周, OpenAI推迟了Astra的部分开发和发布工作, 以便加强和测试其针对网络滥用和未经授权的模型操作的防护措施。虽然Astra并未卷入此前的Hugging Face事件, 但OpenAI表示已将该事件的经验教训融入安全策略, 并为Astra实施了更严格的安全措施, 包括训练模型更可靠地拒绝有害网络请求、增加防止滥用的保护措施以及监控以阻止潜在的未经授权的活动。

Anthropic同样在加大安全建设投入。据官方披露, 该公司正在对近期两起安全事件进行深入分析, 并计划与METR合作进行独立审查, 希望确保两项调查全面彻底, 并将在未来几周内公布更多信息。

法国人工智能安全中心(CeSIA)执行主任查贝尔·拉斐尔表示, Anthropic刚刚宣布暂停高风险强化学习环境的发展, 很可能是受近期模型安全事件影响而被迫采取的措施。在他看来, 这与OpenAI于8月18日宣布暂停前沿强化学习训练的情况类似, 而Anthropic更进一步, 他们希望尽快建立一种合法、可验证的机制。

IPO冲刺期

两家公司动作频频的背后是资本市场的狂热倒计时: OpenAI在6月秘密向美国证监会(SEC)提交了S-1上市申请, Anthropic也凭借飙升的营收尝试冲击史上最大规模的IPO。市场预计, 其10月上市时估值可达2万亿美元, 市销率约30.8倍。路透社披露, Anthropic预计到2028年年营收达到1900亿至2000亿美元。

对于Anthropic的增长, 天使投资人、人工智能专家郭涛指出, 企业不再把Claude当聊天机器人, 而是把模型嵌入真实业务流程, Token消耗量因此成倍放大。目前, 公司80%营收来自API与企业业务, 采用按量计费模式, 区别传统席位订阅, 客户业务规模扩张就能直接带动收入增长, 形成独特的增长飞轮。

而OpenAI首席财务官Sarah Friar日前在全体大会上表示, 公司计划“在2027年成为一家上市公司”, 只有在业务发展持续向好的情况下才可能提前上市。这意味着此前市场期待的OpenAI年内IPO可能性进一步降低。她同时淡化Anthropic抢先上市的影响, 称IPO“并非终点, 只是一个里程碑, 更是又一次融资的机会”。

尽管路径和现状不同, 两家公司仍面临同一个根本问题: AI大模型到底能不能成为一门利润足够丰厚的生意? 不难看出, 资本市场对AI公司的评判标准正在发生改变, 不再只比拼模型能力, 企业渗透率、增长速度、盈利能力, 权重持续抬升。

当前, 有关产业的一些逻辑已然清晰: Sarah Friar明确指出: 企业不再无脑刷Token, 而是精细计算“每单位智能的成本”。谁能提供更低时延、更高任务成功率、更便宜的算力, 谁才能留在企业账单上。

北京商报记者 赵天舒

聚焦 Focus

亚马逊遭起诉操纵广告竞价

电商巨头亚马逊被指控操纵广告竞价。当地时间8月31日, 美国联邦贸易委员会(FTC)宣布起诉亚马逊, 指控该电商巨头通过操纵其广告定价和拍卖系统, 对平台上的广告主“秘密且系统性地超收费用”。这起由22个州的检察长共同参与的诉讼指出, 自2019年变更相关规则以来, 亚马逊可能通过收取“隐藏的附加费”, 从多达120万家受影响的广告主身上榨取了超过200亿美元。

FTC已向美国华盛顿西区联邦地区法院提交诉状, 本案重点针对亚马逊在其购物平台搜索结果旁展示的赞助产品广告、品牌广告和展示广告。

此次诉讼的核心, 是亚马逊广告业务采用的竞价系统。目前, 亚马逊提供商品推广(Sponsored Products)、品牌推广(Sponsored Brands)和展示型推广(Sponsored Display)等广告服务。当消费者在亚马逊网站及移动应用上搜索商品时, 部分搜索结果和商品页面中的广告, 就会通过这一套竞价系统确定展示位置和价格。

这些广告通常采用“第二价格拍卖”(second-price auction)机制。按照这一机制, 广告商须提交自己的最高出价, 不过最终获胜者并不一定按照自己的最高出价支付, 而是支付略高于第二高出价的价格。例如, 如果某一广告位的最高出价为15美元, 第二高出价为12美元, 那么获胜广告商可能只需要支付12.01美元。这种竞价方式在数字广告行业较为常见, 其目的之一是让广告商能够按照自身对广告价值的判断参与竞价, 而不用担心最终一定支付自己的最高报价。

FTC称, 亚马逊从2018年前后开

始改变部分广告拍卖的运行方式, 在正常竞价结束后增加了一个额外的定价环节。根据FTC诉状, 亚马逊内部将这一机制称为“软保留价”(soft reserve)。在这种模式下, 广告商虽然按照正常规则参与竞价, 但亚马逊可以在拍卖结束后进一步提高广告商最终支付的价格。

FTC还指控, 亚马逊曾测试在不明显引起广告商反应的情况下, 可以将广告价格提高到何种程度。该机构援引内部文件称, 一名管理人员曾将相关定价策略描述为“希望广告商没有注意到价格变化, 并且不会因此降低出价或广告支出”。

FTC在诉状中引用了亚马逊广告高管之间的内部沟通记录并表示: “亚马逊对其拍卖做出这种隐秘的改变, 是因为它对其广告拍卖所产生的收入感到不满。”FTC提到, 在一次交流中, 公司一名高管承认使用了“虚构的拍卖参与者”来抬高价格。

因此, FTC及各州指控亚马逊的做法违反了消费者保护法, 并正在寻求民事罚款、赔偿以及其他未指明的赔偿金。

FTC主席安德鲁·弗格森(Andrew Ferguson)在声明中表示: “亚马逊有数以百万计的广告客户被误导, 支付了大幅增高的价格。而这些增加的成本大部分又转嫁给了美国消费者。”

对此, 亚马逊在一篇博客文章中回应称, FTC的诉讼属于“误导”, 该诉状“从根本上误解了广告主的运营方式”。公司强调, 本次诉讼并未提供增加的广告成本转嫁给消费者的证据。

亚马逊解释称, 公司确实在广告

竞价系统中设置了“软性底价”和“硬性底价”, 前者是实时设定的最低价格, 后者是参与竞拍的人入门槛。如果中标广告主的出价高于硬性底价和软性底价, 则他们只需支付软性底价, “类似机制在业内十分普通”。

亚马逊还指出, FTC的指控从根本上误解了广告的运营方式, 广告主会根据真实效果来调整出价, 而不是根据拍卖机制的文字说明: “退一步讲, 即便接受FTC那套‘广告主不会调整出价’的错误前提, 我们估计, 得益于亚马逊优先考虑广告相关性而非仅仅按出价高低选择广告, 广告主在2021年至2025年期间依然节省了超过80亿美元。”

亚马逊表示: “我们已经在广告主用于管理广告活动的主要工具中, 就我们的拍卖和定价向他们提供了指导, 并将对其保持更新。我们期待在法庭上阐明我们的立场。”

目前, 亚马逊已积累起全球三大数字广告业务, 仅次于谷歌和Meta。亚马逊去年获得了超过680亿美元的广告收入, 其中占绝大部分的是赞助产品广告的销售额。

此案标志着亚马逊再次成为被FTC审视的焦点。去年9月, 亚马逊与该机构达成了25亿美元的和解协议, 以解决针对其Prime会员计划的一项调查。

此外, 在FTC和美国17个州发起的另一起诉讼中, 亚马逊被控利用其“垄断权力”抬高价格、压制第三方卖家并降低购物者的服务质量, 该案预计将于明年开庭审理。亚马逊对所有指控均予以否认。

北京商报综合报道

· 图片新闻 ·

美国银行副总裁遇袭身亡



8月31日, 美国纽约, 纽约市警方在时报广场附近处理袭击案件。视觉中国/图

美国第二大银行美国银行副总裁艾琳·皮亚琴蒂(Erin Piacenti), 于当地时间2026年8月31日在纽约时报广场遭持刀袭击身亡。

据央视新闻此前引述美国方面消息, 当地时间2026年8月31日下午, 一名女子在纽约时报广场持刀刺伤两人后被警方击毙, 其中一名伤者伤重不治, 另一名腹部被刺伤者伤情稳定。纽约市警察局局长杰西卡·蒂施表示, “这似乎是一起随机且无缘无故的袭击”。

纽约市警察局局长当天晚间在记者会上说, 事件发生在下午4点后不久。初步调查显示, 一名49岁女子持刀刺伤一女一男, 其中32岁的女性伤者被送医后不治身亡, 即艾琳·皮亚琴蒂, 另一名68岁男性伤者目前情况稳定。纽约警方认为, 该事件属无端袭击, 目前尚未查到嫌疑人的过往犯罪记录, 但查到其有精神健康方面的病史记录。

公开信息显示, 美国银行为美国第二大商业银行, 业务包含个人零售银行、商业信贷、投资银行、全球市场交易以及美林品牌下的财富管理。

据英资资料显示, 艾琳·皮亚琴蒂生前担任美国银行业务遴选与利益冲突副总裁, 今年32岁, 最近她刚刚庆祝了结婚两周年纪念。

美国银行在发给媒体的声明中表示, “我们对这位同事的不幸离世感到震惊和深感悲痛。她是一位优秀的团队成员, 我们将永远怀念她。我们向她的家人和所有爱她的人致以最深切的慰问”。

纽约时报广场是纽约最著名的旅游与商业地标, 日人流巨大, 警力常年高密度部署。近年来该区域及周边曾发生过持刀或随机暴力事件, 但如此公开的持刀对峙并不多见。

北京商报综合报道