

大模型开打价格战 加速行业出清还是劣币驱逐良币?

证券时报记者 周春媚

继“百模大战”以后,大模型或迎来“百模价格战”。

最近一段时间以来,字节跳动、阿里巴巴、百度、智谱AI等多家国内大模型厂商均调整了旗下大模型产品的定价策略,打响了大模型领域的价格战,在行业内掀起了不小的波澜。

作为一个新兴技术,大模型在经历了一年的“技术狂飙”以后,算法更新与模型能力提升似乎已有所放缓,一些厂商开始选择将精力更多地放在优化用户体验、降低使用成本上,推动大模型商业化加速落地。业内人士普遍向证券时报记者表示,在大模型成本高企、商业模式尚未得到验证的情况下,价格战在加速竞争实力差的企业退出市场的同时,也可能会让一些专注于模型能力的初创公司面临更大的资金压力,需警惕劣币驱逐良币。

大模型厂商开打价格战

在移动互联网时代,不断压低价格、发放铺天盖地的补贴是企业跑马圈地、争夺用户与市场的惯常做法。这样的做法,也在向大模型领域蔓延。

5月21日,阿里云发布降价公告,对通义千问9款商业化和开源模型的价格进行大幅降价。其中,对标GPT-4的主力模型Qwen-Long输入价格降至0.0005元/千tokens(文本处理的最小单元),降幅达97%,输出价格降至0.002元/千tokens,降幅达90%。1元钱可以买200万tokens,相当于5本《新华字典》的文字量。

阿里宣布降价后,百度立即跟进,宣布文心大模型的两款主力模型ENIRE Speed、ENIRE Lite即刻起全面免费,开启了大模型API(应用程序编程接口)调用的“免费时代”。据百度有关负责人介绍,这两款模型于今年3月推出,支持128k、8k上下文长度,是目前百度文心大模型系列中服务用户最多的模型型号。

阿里与百度激进的价格战略,直接的导火索是一周以前字节跳动大幅降价的举动。5月15日,字节跳动宣布豆包大模型正式开启对外服务,豆包通用模型pro-32k版、pro-128k版在企业市场的推理输入价格仅为0.0008元/千tokens、0.005元/千tokens,较行业降幅分别达99.3%、95.8%,推动大模型从“以分计价”进入到了“以厘计价”阶段,并声称已经击穿了大模型行业的最低价。

一周过去,这个最低价被接连无情击穿,也让大模型价格战的火药味越来越浓。记者梳理发现,最近一段时间以来,降价已经成为大模型厂商的共同策略。不仅是字节跳动、阿里、百度这样的科技大厂,不少中小厂商也纷纷宣布降价。例如,大模型初创公司智谱AI将旗下入门级大模型GLM-3-Turbo(128k上下文)的价格从0.005元/千tokens降低到0.001元/千tokens,降幅达80%。私募巨头幻方量化旗下的大模型DeepSeek-V2(32k上下文)价格降至每千tokens输入0.001元、输出0.002元。

华龙证券研报表示,随着国内外大模型厂商技术角逐进一步激烈,大模型行业开启价格战信号明显。通过降低价格门槛,大模型厂商有望吸引更多广泛的企业用户群体,从而进一步平衡收入和成本。同时,更多C端用户有望免费使用基础AI应用,庞大的访问量有助于企业进一步提升模型服务能力,完成良性循环。

技术化与商业提速共同驱动

大模型领域常说的token,简单理解就是模型输入和输出的一种基本计量单位。目前,向C端消费者免费、以API调用量为标准向B端企业及开发者收费是大模型厂商的主流选择。此轮多家厂商将API调用量价格进行下调,主要面向的也是企业及开发者。

大模型算力与研发成本高昂,盈利水平亦有不足,为什么价格战却已早早打响?记者梳理业内人士意见发现,这背后主要有技术以及市场两个方面的考量。

技术层面,北京市社会科学院副

5月15日,字节跳动宣布豆包大模型正式开启对外服务,豆包通用模型pro-32k版、pro-128k版在企业市场的推理输入价格仅为**0.0008元/千tokens**、**0.005元/千tokens**。

5月21日,阿里云对通义千问9款商业化和开源模型的价格进行大幅降价。其中,对标GPT-4的主力模型Qwen-Long输入价格降至**0.0005元/千tokens**,降幅达**97%**,输出价格降至**0.002元/千tokens**,降幅达**90%**。

阿里宣布降价后,百度立即跟进,宣布文心大模型的两款主力模型ENIRE Speed、ENIRE Lite即刻起全面免费。

最近,字节跳动、阿里巴巴、百度等多家大模型厂商均调整了旗下大模型产品的定价策略,打响了大模型领域的价格战,在行业内掀起了不小的波澜。

图虫创意/供图 陈锦兴/制图

研究员王鹏在接受证券时报记者采访时表示,随着技术的不断进步,大模型厂商在算法优化、算力提升等方面取得了显著成果,能够以更低的成本提供服务,降价也是成本优化的直接体现。在谈及豆包降价策略时,字节跳动旗下火山引擎总裁谭涛曾公开表示,降价的基本逻辑是有信心用技术手段降低成本。据介绍,通过持续改进模型结构和训练方法、采用分布式推理和混合调度等手段,字节跳动实现了整体成本的优化,为降价赢得了空间。

市场层面,赛智产业研究院院长赵刚在接受证券时报记者采访时表示,大模型厂商打价格战,核心目的是抢占市场、争取用户,加速商业化进程。王鹏亦表示,通过降价或免费策略,大模型厂商能吸引大量企业和开发者对API的调用,增强企业和开发者与平台的粘性,提升品牌知名度,进一步培养和完善AI应用的生态系统。

“一个人现在创业,如果要花1万块钱调用大模型,可能会觉得有点贵。我们把价格降到1%都不到,100块钱就能用起来,他的想法就更容易尝试。”在谈及豆包降价策略时,谭涛说。公开资料显示,目前阿里的通义千问API日调用量超亿,服务9万多家企业用户;百度文心大模型API日调用量超过2亿,服务超过10万家企业用户;字节跳动以低价面向B端市场推出豆包大模型,或希望能够尽快抢占更多市场。

一名业内人士向记者分析,对于很多企业客户而言,选择了一家大模型产品进行部署后往往不会轻易更换。因此,以低价吸引更多“第一批客户”,是驱动大模型厂商加入价格战的重要驱动力。同时,扩大用户规模也有助于利用规模效应大幅降低模型推理的单位成本,并行成“数据飞轮”,加

速大模型性能的提升。

行业加速洗牌 警惕劣币驱逐良币

“我们也希望能够在爬楼梯和看风景之间取得一个平衡,但我们还是在爬楼梯上,而不仅仅只是去看风景。”在5月18日举行的一场论坛上,打造了爆款大模型产品Kimi的大模型创业公司月之暗面创始人杨植麟说。

然而,在价格战的步步紧逼下,大家都在挤着“看风景”,对于一些想更专注于提升模型能力,而且没有大量资金去“铺市场”的小厂来说,“爬楼梯”的时间可能不多了。

随着字节跳动、阿里、百度等大厂加入价格战,一场血雨腥风的淘汰赛也即将打响,行业或加速洗牌。赵刚表示,此前由于大模型技术创新迭代速度非常快,一些大模型企业研发投入不足,比较依赖开源大模型,技术实力不足。“大模型市场将重新洗牌,市场集中度将有所提高,一些产品成熟度不高、资金实力不强的企业可能在价格战中被淘汰。”赵刚说。

同时,即使拥有较强的技术实力和较成熟的产品,“卷价格”对于大厂与小厂的影响也是完全不同的。王鹏告诉记者,大厂能够利用规模效应和资源优势,通过多元化盈利模式和增值服务来平衡降价带来的影响,而初创公司则可能会面临更大的资金压力。赵刚认为,价格战将会为头部大模型企业带来更多的收费用户和应用场景,实现薄利多销的目标。从收入

上看,有竞争力的大模型企业收入可能反而会增加。

“对大模型企业而言,关键是其拥有的资金能否帮助企业支撑到盈亏平衡的拐点。因此,对大厂和融到资金的创业公司,价格战是保持市场竞争优势的重要手段;而对资金紧张的创业公司,价格战可能就是让企业淘汰出局的致命一击。”赵刚分析称。事实上,在字节豆包大模型大幅降价以后,有国内明星AI公司的内部人员就表示,已经有一些客户在询问其产品是否会跟进降价,“大家似乎都在期待着更低的价格”。

值得注意的是,无论国内还是国外,性能最先进的大模型产品往往并非直接来自科技巨头,而是由更专注于技术创新、模型打磨的创业公司打造。一名业内人士向证券时报记者表示,依靠雄厚资金与算力储备去抢占市场,做类似于互联网时代的“模式创新”,这并非创业公司的强项。若未能获得更多的融资支持,它们在陷入“价格战绞杀”之后也将变得更加被动,需要警惕劣币驱逐良币,从而拖累大模型技术进步的情形发生。

不过,在任何一场价格战中,无论商家如何“搏杀”,用户至少短期来看都能从中受益。“大模型降价降低了企业采用AI技术的门槛,能够推动企业数字化转型。同时,AI技术的应用也有助于企业提升运营效率和降低运营成本。”王鹏表示。对应用层的创业公司而言,赵刚认为价格战是一个利好,因为这意味着大模型的调用成本大幅降低,有利于创业公司加快创新步伐,开发出更多商业化的大模型应用,促进大模型技术的普及与推广。



通用大模型时代 机器人“适者生存”

证券时报记者 阮润生

在历年的东莞松山湖中国IC创新高峰论坛上,各家代表企业围绕特定主题推介自家产品后,圆桌论坛成为“例牌菜”。

作为会议发起人,中国半导体行业协会IC设计分会副理事长、芯原股份董事长兼总裁戴伟民客串起圆桌论坛讨论环节的主持人。不同于多数主持人的“照本宣科”,戴伟民像一位掌勺的“大厨”——对圆桌嘉宾作简要介绍,配上自己的观点,再把问题精准地抛给台上相关领域的专家;有时也会把问题同步抛给台下的观众,邀请大家扫码填写问卷,反馈直观体验,会后再逐一统计,全程互动感拉满。

今年的论坛步入了第十四个年头,圆桌讨论选取了时下最热门“机器人”主题,邀请相关从业者热议“智障”与“智能”机器人的边界,也暗暗戳中当下人们对“机器人替代”的担忧。

谈及人形机器人是否是服务机器人的必要形态时,神项科技(南京)有限公司董事长、CEO袁帝文感慨近年来风向调转,业内重新重视起人形机器人,推断与大模型发展密切相关。

“我们公司2020年创立,当时资本市场(投资人)见到仿生概念就跑掉,觉得不靠谱。”袁帝文说,如今很多大模型公司都很积极地和人形机器人合作,做

数据采集。

在业内人士看来,采用人形机器人除了使用者便于接纳与日常共情外,还有一个重要考虑是便于训练、收集数据。深圳市亿境虚拟现实技术有限公司总经理石庆提示了个细节——在马斯克的机器人视频里,有很多训练师戴着VR头显做动作,让机器人模仿学习。尽管有训练数据和大模型的加持,机器人还是有很多“智障”的一面。“机器人可以作诗作曲写文章写论文,甚至拍电影,把人休闲雅致的那一面都抢走了,但是帮我们干活的都做不好,倒垃圾、洗碗、刷马桶都做不了。”袁帝文称。机器人的功能也在快速迭代和完善。今年早前就展出了特斯拉“擎天柱(Optimus)”人形机器人自己叠衬衫等视频。而提升这些“微小”任务的智能化水平,嘉宾们认为仅仅是堆叠算力,将“头脑上云”是不够的。

瑞芯微副总裁陈锋举出了自然界碳基生物为反例,果蝇可以飞往任何地方,但并不需要很多的参数。“大模型的本质是Predict Secret,预测下一个将会出现的事物,然后去组合和汇总所有数据。”陈峰表示,如今语言类的大模型表现惊艳,但是去做一些碳基生物的事,却执行很差。

自然界“适者生存”法则,可能同样适用于机器人。乌镇智库的联合创始人张晓东接

引计算机从中心化到分散化部署的发展历史,指出人工智能终端算力飞跃的重要性,这意味着大模型将向“微型化”“小型化”演进,加强端侧能力部署,将赋能机器人。

成都启英泰伦科技有限公司创始人、CEO何云鹏建议,机器人要摆脱“智障”,还需要基于应用、模型结构、特点,甚至传统的判别式AI算法,来解决特定问题,实现稳定的执行状态。当然,几乎智能体的终极命题都逃不过“功耗”问题。随着大模型的端侧部署,机器人的感知系统也有望向低功耗方向演进。何云鹏建议,搭载不同计算类型的芯片相互协作组合,完成整体机器人的系统运行和工作。

鹏瞰科技(上海)有限公司产品市场副总裁王伟则给出了采用“光纤做机器人”的方案,即用芯片和光纤来取代传统的一捆电缆,因为光纤对于电池干扰是免疫的,而且光纤的带宽可以拓展,满足机器人从计算中心到各个关节的通信部署。

不过,光纤的成本并不低,最终还需要看计算效果来取舍。

小米生态链研发总监张秀云表示,当前机器人的发展还是比较简单,未来必然会更复杂,趋向于多模态。这需要机器人具备灵活的信息接收接口,期待富含想象力和充满惊喜的解决方案。

广汽集团：全力以赴发展海外市场

证券时报记者 李映泉

“如今正值国内汽车海外出口的黄金时期,广汽会抓住机遇,全力以赴。我经常开玩笑,现在要么我在海外,要么我正在办海外手续的路上,前天我刚从埃及回来。最近我们开发的几个国家都非常成功,特别是非洲国家有大量的潜力。”5月20日,广汽集团总经理冯兴亚在公司2023年年度股东大会上如是表示。

年报显示,2023年,广汽集团实现营业收入1287.57亿元,同比增长17.48%;净利润44.29亿元,同比有所下降。2023年全年实现汽车产销均超250万辆,规模居国内汽车企业集团前五。

进入2024年,汽车行业整体库存压力加大,新一轮价格战掀起。当前国内市场的竞争和“内卷”形势,也是此次股东大会上投资者与管理层交流的重点话题。

广汽集团董事长曾庆洪对此表示,今年汽车行业的竞争非常激烈,上半年价格战已经打得“头破血流”,预计下半年价格战会继续。希望国家相关部门出台相关政策,引导市场公平竞争。今年汽车行业预计全年微增3%—5%,主要增量是海外市场和新能源汽车,因此广汽集团会继续加大海外市场布局,并全力推进自主品牌、合资品牌向智能化、电动化方向转型。接下来两三年将会是公司的调整期和转型期。

“在营销方面,广汽确实在销售创新、销售方法工具等方面跟不上市场。广汽的产品、智能化等方面并不差,但现在是在卷老板、卷流量(的时代),所以现在我们要改变传统的销售方式。我们最近一直在做调研,现在上视频号、微博、抖音等,都能看到同时提升盈利能力。”曾庆洪说。

冯兴亚也表示,在销售领域,广汽和行业先进的营销团队相比,有很多需要创新、提升、改善的地方,特别是在一些新的营销手法的运用上,还需要进一步解放思想。“但是

通威股份：光伏行业第四季度会有起色

证券时报记者 唐强

5月20日,光伏龙头通威股份(600438)召开2023年年度股东大会,通威集团董事局主席、通威股份董事长刘汉元,通威股份副董事长严虎,董事会秘书严轲等公司管理层出席会议,并吸引了数十家券商、机构及多位个人股东参与现场会议。

2023年第三季度以来,光伏行业中的硅料环节等进入低谷期,相关产品价格持续下滑,导致光伏企业业绩同比回落。根据2023年年报显示,报告期内,通威股份实现营业收入1391.04亿元,同比下降2.33%;净利润135.74亿元,同比下降47.25%。

关于经营业绩,通威股份管理层在股东大会上表示,光伏行业仍将保持高速增长态势。不过,由于市场供应大量释放,产业链价格出现了显著回落。对通威来讲,光伏产品销售规模实现了大幅提升,但未完全覆盖光伏产业链价格大幅回落的冲击。

据中国光伏行业协会统计数据,2023年我国多晶硅、硅片、电池、组件产量分别达到143万吨、622GW、545GW、499GW,同比增幅分别为66.9%、67.5%、64.9%、69.3%。但据不完全统计,去年已有近10家上市公司相继发布产能延期投产或终止的公告。

在股东大会上,通威股份“掌门人”刘汉元坦言,光伏市场阶段性低迷给企业带来了挑战。不过,刘汉元认为,未来的一个大逻辑仍是全球要想办法实现碳中和,光伏是达成这一目标的关键路径。现在,以光伏为代表的清洁能源已经取得成本领先,能耗已经明显低于煤炭发电,发电成本低于火电。

对于通威股份的生产成本、技术路线等问题,也是投资者关注的

我们也有优势,比如国企的平台、合资的流程和民营的效率等,所以我们会发挥平台的优势,扬长补短。”

在国内激烈“内卷”的背景下,广汽集团也将目光投向了“不那么卷”的海外市场。数据显示,广汽集团今年1—4月在海外市场出口汽车4.5万辆,同比增长267.1%。

冯兴亚坦言,广汽在海外市场起步较晚,“但反过来讲,潜力也大,也避免了类似其他品牌因为质量和口碑太差退出市场又重新进入的弯路,我们出口的汽车质量都是很好的,如今正值国内汽车海外出口的黄金时期”。

在冯兴亚看来,发力海外市场一般有三个阶段,第一阶段是卖整车,第二阶段是在当地组装,第三阶段是完全的本地化。“我们现在发力海外市场有一个好处是可以直接从第二阶段切入,因为目前海外已经有产业基础,我们就直接切入到了本地化的生产,进行KD件组装。”

冯兴亚表示,对于海外市场,广汽正举全集团之力去发展,目标非常明确,措施也明确。“现在我们已经进入了40多个国家和地区,数量快速增长。去年广汽国际按照原本计划是会亏损5000万,但实际上盈利了,这也从侧面证明了我们的国际化策略效果明显。”

曾庆洪表示,维持和努力提升派息率是公司一直以来的坚持,上市以来,广汽历年累计分红比例都保持在30%以上,2023年高一档,达到36%,未来还会继续保持和提升。另外,公司还计划开展A股、H股的回购,总体资金规模在5亿—10亿元,其中H股回购股份资金总额在4亿元到8亿元之间。

“广汽正在积极转型、向着‘万亿广汽’目标迈进,对公司未来发展前景充满信心,H股回购后将注销,将增厚每股收益,提高净资产收益率,改善资本结构。未来公司将坚持常态化分红,同时根据市场情况,利用好股份回购等提升投资价值的工具,为股东创造更多价值。”曾庆洪称。

热点话题。刘汉元表示,中国光伏行业具有成本低、质量好的比较优势,在全球能源产业体系里具有极强的竞争力。中国光伏行业还有很大的成长空间,冲高就一定会回落,回落就有夯实的空间。如果是稳定、饱和的市场,将需要出清三分之一以上到一半的落后产能。

刘汉元进一步指出,2023年,全球光伏行业需求同比增长近八成,如果今年、明年的需求都达到30%至50%左右的增长水平,即使产能规模不变,光伏行业的供需也能重新恢复平衡。“只要产业前途和场还在,各环节技术和成本处于领先,我们明天的早餐就还在,没人能动你的奶酪。如果是最乐观的情况下,预计今年第四季度光伏行业就会有起色;以悲观的预期来讲,明年第四季度的早餐就还在,没人能动你的奶酪。如果是最乐观的情况下,预计今年第四季度光伏行业就会有起色;以悲观的预期来讲,明年第四季度的早餐就还在,没人能动你的奶酪。”

“随着技术的持续提升,通威股份成本下降的相对空间还有不少。”刘汉元认为,在研发方面,通威股份对市场主流的技术都保持了足够力度的研究。无论哪项技术,只有达到了技术方面的可行性、稳定性和经济方面的可行性目标,通威股份才会投入生产。通威股份有足够的研发力度防止自己走弯路、走老路、走捷径,对任何可能的技术替代路线不抱任何偏见。

严虎表示,目前来看,很多行业外资本不会再轻易进入光伏行业,而光伏行业的资本也正在快速收紧,但处于行业前列的公司仍然会继续扩张产能。在通威股份2024年至2026年的投资发展规划中,硅料、电池产能建设项目有望提前完成,公司的技术路线布局也已基本完整,后续竞争就是看谁的综合风险管控能力更强、谁的现金流能力更强。