



探寻产业发展“新引擎”

# Token即产能 大模型价值战开启

本报记者 袁传玺

近期,布局AI智能体的热潮正席卷各行业。AI智能体正以前所未有的速度融入日常的工作与生活场景中。

而这热潮的背后,是呈指数级爆发的算力需求——个人AI智能体的大规模落地带来了海量Token(计算机领域数字标识符)消耗,迅速击穿了大模型厂商原有的成本防线。

近期,北京智谱华章科技股份有限公司(以下简称“智谱”)、腾讯云等国内AI厂商密集发布AI算力产品涨价通知,部分产品涨幅甚至超过400%。这场从“烧钱换增长”到“以价补量”的战略急转,不仅标志着行业告别野蛮生长,更折射出AI智能体时代算力供需关系的深刻变局。

## 大模型定价体系再造

大模型行业的定价体系正在经历系统性再造,国内厂商也在加速上调相关产品价格。此番涨价现象,与两年前的价格战形成鲜明对比。

2024年5月份,字节跳动率先打响价格战第一枪,豆包Pro模型定价0.0008元/千Token,较行业均价低99.3%。随后阿里云通义千问主力模型降价97%,百度文心大模型两大主力模型全面免费,腾讯混元大模型价格降幅最高达87.5%。一时间,行业陷入降价潮。

“那时候的逻辑很简单:先让开发者用起来,市场份额比什么都重要。”一位从业三年的AI产品经理对《证券日报》记者透露,2024年某头部厂商内部曾定下三年内不考虑盈利的激进目标,产品定价甚至低于算力成本。

然而,低价策略的边际效应迅速递减。有行业分析师对《证券日报》记者表示,2024年至2025年的价格战加速了AI大模型的市场普及度,但也导致行业普遍陷入“高投入、低回报”的困境。当模型调用量从千亿级跃升至万亿级,算力成本呈指数级放大,单纯依靠资本输血已难以维系。2025年下半年,部分中小厂商开始悄然收缩免费额度。

“这不是简单的涨价,而是成本结构变化的必然结果。”一位头部云厂商技术负责人向《证券日报》记者分析,“以前行业是用亏损换取市场份额,2026年必须考虑可持续经营。”

## Token通胀

要理解国产大模型集体涨价,必须先理解“Token通胀”这个概念。

Token是大模型处理文本的最小单位,可以理解为AI的工作量单位。当行业谈论Token通胀时,实际上指的是AI处理任务的复杂度激增,导致同样的服务需要消耗更多的算力资源。就像以前开一盏小灯,现在要开一个工厂,电费自然水涨船高。

这股“通胀”压力首先来自海外市场需求的爆发。2026年2月份,OpenRouter平台(全球主流的大模型API分发平台)数据显示,当月全球前十AI模型的Token总消耗量突破27万亿,其中中国大模型贡献了14万亿,占比超过50%。

“这意味着国产大模型正从内需驱动转向全球输出。”广州艾媒数聚信息咨询股份有限公司CEO张毅在接受《证券日报》记者采访时表示,“海外用户的调用习惯与国内截然不同。”欧美开发者更倾向将大模型嵌入生产级工作流,单次请求往往伴随多轮工具调用,上下文检索与代码生成,“同一样API(应用程序编程接口)调用,海外场景的Token消耗可能是国内的三到五倍”。

如果说海外市场是外因,那么AI智能体的规模化落地,则是推高算力成本的内因。与早期聊天机器人时代的单次问答不同,AI智能体具备“感知—决策—执行”的闭环能力,能够自主完成复杂任务。以金融风控场景为例,一个AI智能体完成单笔信贷审批,需经用户画像检索(长上下文)、征信数据调用(工具使用)、风险评估计算(推理链)、报告生成(输出)四大环节,全程消耗Token量可达数十万级别。

2026年2月份,OpenRouter平台数据显示,当月全球前十AI模型的Token总消耗量突破27万亿,其中中国大模型贡献了14万亿,占比超过50%



图片来源视觉中国

诸多因素相互叠加,数据表现惊人。国联民生证券测算显示,中国整体日均Token消耗从2024年初的1000亿级,飙升至2026年2月份的180万亿级。且随着AI智能体向多模态、多智能体协作演进,这一数字仍在加速膨胀。

供需关系的逆转,最终传导至价格体系。2025年以来,全球AI算力基础设施面临产能荒,受HBM显存(高带宽内存,AI训练的核心部件)、先进制程GPU(图形处理器)供应紧张影响,服务器采购成本同比大幅上升。

比如,3月17日,阿里云发布公告称,因全球AI需求爆发,供应链涨价,阿里云AI算力、存储等产品最高涨价34%。

当大模型“商从‘卖水人’变成‘喝水人’”,涨价成为维持服务质量的刚性选择。智谱AI在调价公告中直言:“用户规模与调用量快速提升,公司需同步加大算力投入。”

## 商业模式重构

涨价掀开的不仅是成本缺口,更是整个行业商业逻辑的深层重构。

“价格战结束时,价值战才真正开始。”前述云厂商技术负责人判断,2026年将成为AI大规模商业化应用元年,行业竞争焦点正从单纯拥有算力,转向能否提供高效、稳定、低成本的模型服务与AI应用。

当下,大模型行业正在从“流量补贴”转向“价值筛选”。早期低价策略吸引了大量试错型用户,造成算力资源的低效占用,某厂商曾统计,其免费额度消耗中40%来自无实际业务场景的测试调用。而通过适度涨价,企业既能过滤非刚性需求,又能为优质客户保障服务稳定性。智谱、腾讯等厂商产品大幅涨价的背后,实则是将价格锚定于企业级客户的支付意愿与ROI(投资回报率)平衡点。这种“以价补量”的精细化运营,标志着中国大模型产业正从互联网式的规模扩张,迈向软件行业的价值定价。

工业和信息化部信息通信经济专家委员会委员盘和林对《证券日报》记者表示,涨价并不会抑制真实需求,反而会加速“良币驱逐劣币”的进程。企业级客户对稳定性、合规性的高要求,使其付费意愿与生命周期价值远超长尾用户,这正是大模型厂商从“流量思维”转向“价值定价”的底气所在。

这种转变,正在重塑整个产业链的利益格局。上游算力厂商(如英伟达)持续受益;中游云厂商(如阿里云、腾讯云)则在卖模型与卖算力之间寻找平衡——既想通过AI服务吸引客户,又不愿被高昂的算力成本拖累;下游应用层则出现明显分化:具备自研能力的大厂(如字节跳动、百度)可在集团内部灵活调配算力资源,来对冲涨价风险,而纯靠调用API的中小创业公司则面临成本暴增、被迫关停的困境。

企业级大模型厂商也开始关注Token经济的深层变革。滴普科技股份有限公司联合创始人、执行董事杨磊对《证券日报》记者表示:“未来,Token即产能。随着Skill-based Model(基于技能的模型)重塑软件开发、数据分析、客服外包等行业,传统的单人单天的计价方式必将被‘Token消耗’计价取代。这不仅是计量单位的变更,更是生产力范式的飞跃。”

张毅表示,站在全球竞争视角,Token通胀亦是国产模型技术跃迁的副产品。涨价不是终点,而是新一轮效率革命的起点。谁能在这场算力军备竞赛中持续优化成本结构,谁就能在AI智能体时代的全球牌桌上保留席位。

回望2024年的价格战,再看今日的集体涨价,中国大模型产业正经历痛苦的成人礼。那个依靠骨折价博眼球的时代已然落幕,依靠技术效率、客户价值与生态闭环取胜的新纪元,正在Token经济洪流中徐徐展开。

# AI算力驱动散热需求爆发 上市公司密集布局液冷赛道

本报记者 冯雨璜

AI算力需求的指数级增长,驱动散热需求爆发,液冷成为热管理领域高景气赛道。3月18日,A股液冷概念表现活跃,多只概念股涨停。投资者对相关上市公司的布局也高度关注。

在此背景下,《证券日报》记者注意到,国内液冷行业正呈现并购整合加速、上市公司定增募资火热的双重特征。

业内预测,随着算力持续放量与数据中心建设提速,液冷技术已从“可选配置”变为“强制标配”,市场将迎来前所未有的机遇。巨丰投资高级投资顾问丁臻宇对《证券日报》记者表示,液冷赛道前景非常确定,属于高景气、高确定性方向。“当前,市场已经跨过早期导入期,正处于规模化爆发的起点,头部客户开始批量落地,订单和业绩会逐步兑现,未来两三年是快速放量阶段。”

## 上市公司加码布局

今年上市公司正通过并购或定增等方式加码布局液冷赛道。记者梳理相关公告获悉,国内上市公司以定增为核心工具,集中对液冷产能建设、技术研发与算力温控等领域加大投入,自2026年初以来已有多个定增项目获证券交易所审核通过。

3月5日,依米康科技集团股份有限公司定增获深交所审核通过。本次定增拟募集资金总额不超过3.11亿元,扣除发行费用后,将全部投向算力基础设施温控产能建设项目、算力基础设施温控产品研发测试平台项目,并补充部分流动资金。

也有不少上市公司陆续披露定增预案,抢占市场先机。1月16日,深圳市澄天伟业科技股份有限公司披露定增预案,拟募集资金不超过8亿元,募集资金将用于液冷散热系统产业化项目、液冷研发中心及集团信息化建设项目、半导体封装材料扩产项目和补充流动资金。

1月29日,三河同飞制冷股份有限公司发布向特定对象发行股票预案公告,公司计划通过定增方式募集资金总额不超过12亿元,资金将重点投向液冷温控相关项目及补充流动资金,旨在进一步扩大产能规模、完善业务布局,应对液冷温控行业快速增长的市场需求。

定增火热的同时,上市公司也在积极通过并购切入液冷赛道。3月13日晚间,金富科技股份有限公司发布公告称,公司将以合计5.712亿元对价,收购佛山市卓晖金属制品有限公司、佛山市联益热能科技有限公司各51%股权,交易完成后两标的公司将纳入合并报表范围。此次交易旨在切入液冷散热高景气赛道,形成公司第二增长曲线。

## 液冷赛道步入爆发期

丁臻宇表示,今年大量上市公司密集通过并购、定增布局液冷赛道,核心驱动力就是AI算力爆发带来的散热刚需。

当前,液冷行业以冷板式、浸没式为主流技术路线,适配AI服务器、智算中心、储能、新能源车等多元场景。丁臻宇进一步表示,传统风冷已无法支撑高功耗处理器和高密度机柜,叠加数据中心PUE(数据中心总能耗与政策收紧,液冷从“可选项”变成“必选项”,行业空间被瞬间打开,企业抢的是卡位、订单和估值。

据中商产业研究院最新报告预测,2026年中国液冷服务器市场规模将达257亿元,渗透率将从2025年的20%将飙升至37%。业内认为,在这场全球算力散热竞赛中,凭借技术壁垒与订单优势的企业将率先抢占市场高地,在行业格局重塑中占据核心主导地位。

“对于布局液冷的上市公司来说,真正能走出来、具备长期竞争力的,不是靠蹭热点,而是看以下三点:即有没有核心技术和自研能力,能不能绑定头部算力厂商拿到持续订单,以及有没有规模化交付和成本控制能力。”丁臻宇直言,只有这三点都站住,企业才能在行业快速扩容中吃到红利,而不是昙花一现。

深度科技研究院院长张孝荣对记者表示,目前,液冷市场正处于导入期向规模化爆发期跨越的关键拐点。“当前以冷板式为主流,未来将向高效、智能化演进。”他谈到,行业竞争正从单一部件转向系统级整合能力的比拼,具备全栈技术能力的厂商将主导市场。

# 共封装光学板块上市公司迎发展良机

本报记者 王镜如

在AI算力需求持续爆发的背景下,共封装光学(CPO)板块再度成为市场关注焦点。

截至3月18日收盘,CPO板块多只个股表现活跃。瑞斯康达科技发展股份有限公司涨停,苏州天孚光通信股份有限公司、武汉光迅科技股份有限公司(以下简称“光迅科技”)等个股涨幅居前。

深度科技研究院院长张孝荣

对《证券日报》记者表示:“随着生成式AI和大模型应用快速发展,数据中心内部通信压力持续攀升,传统铜互联在带宽与功耗方面逐渐逼近物理极限,成为AI算力扩展的重要瓶颈。CPO技术通过将光引擎与交换处理器进行共封装,使光信号在更短路径内完成传输,从而显著降低功耗与传输损耗。”

在此背景下,国内上市公司正加快布局CPO及相关光通信技术,部分企业率先兑现业绩。例如,受

益于算力投资持续增长,高速率光模块需求显著提升。2025年成都新易盛通信科技股份有限公司预计净利润为94亿元至99亿元,预计同比增长231.24%至248.86%。

与此同时,行业内技术迭代提速。光迅科技近日披露,其3.2T硅光单模NPO模块已完成送样测试。此次送样涵盖完整NPO系统方案,包括光引擎OE、外置光源模块ELSFP及光纤管理模组FMU-Shuffle,具备面向系统集成验证的完整能力;

深圳市兆驰股份有限公司于3月份发布公告称,1.6T光模块已进入快速研发阶段,正围绕LPO、NPO、CPO等多路径同步布局下一代高速低功耗解决方案。

无锡市德科立光电技术股份有限公司则持续推进海外市场拓展,并筹划在新加坡二次上市。公司相关负责人表示,此次上市是公司国际化战略的关键举措,旨在通过成熟市场平台拓宽全球融资渠道,提升资本管理能力,增强在

光通信领域的国际竞争力。

苏商银行特约研究员高政扬对《证券日报》记者表示:“当前,全球头部科技公司受AI算力需求驱动,对高带宽、低功耗的光通信方案需求持续增加,拉动CPO市场增长。同时,国内企业凭借成熟的产品量产能力,其海外订单逐步落地,进一步印证行业的高景气度。短期资本追捧与长期产业逻辑的双重叠加,共同构成该板块持续走强的核心驱动力。”

# 中国人寿2025年十大理赔案例发布 件均赔付金额超2600万元

日前,中国人寿保险股份有限公司(以下简称“中国人寿”,股票代码:601628.SH,2628.HK)公布了2025年十大理赔案例。过去一年,该公司快速响应和受理客户理赔需求,全年处理赔付案件超6000万件,理赔总金额超1000亿元。其中,前十大理赔案件最高赔付金额4796万元,件均赔付金额超2600万元,涵盖多种重大风险场景,以真实案例印证保险业在完善社会保障体系中的重要作用。

## 案例一:疾病夺走生命,保险筑起最后一道守护

A先生具备良好的经济基础与风险保障意识。自2013年起,他先后投保了国寿鑫裕金生两全保险、国寿鑫尊宝终身寿险(万能型)(A款)等10份保单。2024年7月,A先生因转移性肺腺癌不幸身故。公司在接到其家属理赔申请后,及时兑现保障承诺,给付身故保险金4796万元。

启示:A先生的高额赔付是他用长期的保险投入在风险来临时留给家人的生活保障。

## 案例二:廿四载保障筑基,让家的温暖延续

B先生从2001年起,陆续投保国寿99鸿福两全保险、国寿鑫裕金生两全保险等11份保单。2025年6月,B先生因疾病不幸身故。2025年7月,家属提交理赔申请后,公司迅速向客户家属给付保险金共计3342万元。

启示:保险是爱与责任的延续,为家人和自己准备一份保险,就像是在风雨来临时能拥有一座坚固的避风港,让我们在风雨来临时能拥有处变不惊的勇气。

## 案例三:十载远见,以规划托举家庭安稳

C女士具备极高的风险前瞻意识。自2015年退休后,她为自己投保了国寿鑫福年年养老金保险、国寿鑫福年年年金保险、国寿鑫账户两全保险(万能型)(钻石版)3份产品。2025年11月,C女士突发剧烈腹痛,经抢救无效不幸身故,最终诊断为大面积肺栓塞。公司在收到理赔申请后迅速处理,向C女士亲属给付了身故保险金3023万元。

启示:面对无法预见的突发急症风险,保险不仅仅是一份合同,更是以金融契约的确定性,给家人提供的一份守护。

## 案例四:疾病猝不及防,保险让爱继续陪伴

D先生具备极强的风险保障意识。自2022年起为自己投保了国寿鑫尊宝终身寿险(万能型)(C款)、国寿臻享传家终身寿险等多份保单。2025年3月,D先生因胃癌身故。收到客户的理赔申请后,公司及时给付身故保险金2718万元。

启示:面对未知风险,定期体检和癌症筛查可及早发现健康问题,合理配置保险则是为自己和家庭提供一份保障。

## 案例五:四十七份保单,二十三年守护,一份沉甸甸的承诺

E先生是中国人寿的老客户,风险意识很强。2003年至2025年间,他陆续投保了康宁终身保险、国寿鑫享未来两全保险等47份保单,逐步构建起全面的保障计划。2025年4月,E先生在白血病治疗期间,因继发性出血不幸身故。家属提出理赔申请后,公司及时给付了E先生身故保险金2611万元。

启示:高额保障并非一蹴而就,它源于对风险清醒的认知,以及伴随人生阶段不断校准的长期规划。通过“增额”——适时提升既有险种的保额,以及“扩面”——配置不同类型产品以覆盖更多风险维度,方能逐步筑牢家庭经济防线。

## 案例六:未雨绸缪构筑堡垒,保险为关爱加码

F女士是中国人寿的老客户,自2013年起陆续投保了国寿鑫账户年金保险(万能型)(豪华版)、国寿福臻尊两全保险(分红型)等保单。2025年12月,F女士七岁与乳腺癌抗争,不幸离世。公司收到家属提交的理赔材料后,快速给付身故保险金2131万元。

启示:科学防治是对抗疾病的关键,合理配置保险是抵御风险的坚实后盾。

## 案例七:全面屏障布局在前,让突发风险有备无患

## 案例八:长期稳定配置保障,科学转移风险

H先生是中国人寿的客户,2006年至2015年间先后投保了国寿千禧理财两全保险(分红型)、康宁终身寿险等多份保单。2024年10月,H先生因食管恶性肿瘤身故。公司在接到家属的理赔申请后立即开展理赔服务,及时向客户家属赔付保险金1982万元。

启示:长期坚持配置保险,是一种科学的风险转移措施。它能够在极端风险发生时,为家庭提供及时、有效的经济补偿,保障家人未来的生活水平与财务稳定。

## 案例九:未雨绸缪提前规划,为至亲织就守护铠甲

知名企业家I先生自2015年起持续配置国寿鑫福年年养老金保险、国寿鑫耀东方年金保险等产品,以科学规划为家庭筑起稳健保障。2024年11月,I先生在家突发心脏骤停离

## 案例十:意外难以预料,用承诺为家人撑起晴空

J先生自2015年开始陆续投保国寿福臻尊两全保险(分红型)、国寿鑫福传家终身寿险、国寿鑫耀鸿运年金保险等多份保单。天有不测风云,2025年6月,J先生不幸溺水身故。公司收到理赔申请后,快速向其家属给付保险金1833万元,用实际行动兑现了保障承诺。

启示:高额保单从来都不只是一份合同,更是对家庭经济生命的一份郑重承诺与提前规划。它能在意外降临时,为家人未来的生活筑起一道坚实的财务屏障,守护家庭生活的稳定与尊严。

生命无常,守护常在。从重疾到意外,从壮年到银发,这些案例再次印证:保险不仅是一纸契约,更是穿越风雨的底气、延续爱与希望的桥梁。中国人寿始终坚持“以客户为中心”的服务理念,用切实的理赔服务兑现保险保障的承诺,为千万家庭筑牢安全防线,让守护时刻在场。(CIS)