Token为王，如何打赢AI时代“新大宗商品”争夺战？

证券时报记者 王小伟

无形的海量Token（词元）顺着网线，卖到全球各地，就像有形的大宗商品通过路网销往全球一样——Token正在成为AI时代的“新石油”和“新集装箱”。

随着Agent（智能体）时代的来临，尤其是OpenClaw（龙虾）应用爆火，AI的任务执行模式从人机对话升级为机器自循环，Token消耗量指数级增长，其角色从模型训练的技术副产品，一跃成为可计量、可交易的战略资产。

这种剧变重塑互联网大厂底层逻辑：商业模式从“烧钱换流量”迈向“按Token计费”，竞争模式也从“参数竞赛”转向“Token经济体系构建”。在这场围绕Token展开的竞逐中，老玩家的“二次创业重塑”与新玩家平衡“保利润还是保用户”交织而行，共同演绎着下一代AI竞争剧本。

更深远的变化也在酝酿。多位受访者认为，所有企业都值得重新评估能否被装在Token里重构，这将诞生新的产业巨头。

Token作为定义未来十年技术版图与产业话语权的关键变量，如何在这个“新大宗”全球贸易网络中寻得更多主导权，需要不同于互联网时代流量逻辑的新模式，这考验所有参与者的智慧。

新大宗商品

多数受访者都没料到，随着“龙虾”的到来，Token需求的爆发来得如此之快。

Token是大语言模型处理信息的基本单位。在AI世界，Token可以是一个词、一段代码，也可以是图像或视频中的一个像素区块。当用户向AI模型提问，模型先把用户的话“切”成Token，算完后再把结果Token拼回成句子。

百度千帆平台产品负责人张婷将Token比喻成“乐高积木”——单个Token是碎片，但大模型把成千上万个Token的“拼法”学会后，就能理解语义、生成文本、回答问题。每生成一个Token，都在调用数据中心的GPU（图形处理器）算力，并伴随着电力消耗。

过去两年，大模型竞争的核心叙事是模型能力：谁更聪明，谁就更接近AGI（通用人工智能）。参数规模、推理深度、复杂任务完成率，构成行业主要竞争指标。随着Agent时代的来临，在“自主拆解—调用模型—完成任务”新属性下，Token消耗从人机对话升级为机器自循环，消耗量级跳涨。

任职于范式智能的谷少辉说，Chatbot（聊天机器人）时代，GPU就好像餐厅服务员，一桌客人上完菜就去下一桌；Agent时代，服务员全程陪同，从点菜到结账，思考菜单的时候也不能走。粗略估算，Chatbot单轮对话消耗约1000到3000个Token，而“龙虾”跑一次深度研究要经历感知、规划、执行、反思等多个循环，稍微复杂的任务就要吃掉百万级Token。

无问芯穹联合创始人夏立雪的感受颇具代表性：从1月开始，公司Token消耗每两周就翻一番，至今已经翻了10倍。“上次见到这个速度，还是多年前3G手机流量时代。”

业内共识是，需求曲线仍处陡峭上升期。多Agent并行、长上下文推理、编程场景的爆发才刚刚开始，每一个新场景打开都意味着Token消耗量再上一个台阶。同时，AI已跨越感知与生成阶段，以智能体和物理AI为核心的执行时代，所需的Token量和计算量相比训练阶段会几何级别增加。

黄仁勋日前提出“Token经济学”概念，认为推理已成为AI最核心的工作负载，Token则是新的大宗商品——标准化、可计量、可交易。由此Token从模型训练的技术副产品，演变为驱动数字经济的核心生产要素。

3月22日，国家数据局方面表示，Token是智能时代的价值锚点，更是连接技术供给与商业需求的“结算单位”，Token有了官方翻译“词元”。

“这次定调前，圈内就在热议Token该如何翻译。”谷少辉透露，“张一鸣用计算机术语‘字节’给公司起名真是一步到位，这个基本单位的技术感、力量感并存，认知成本几乎为零”。

业内对于Token的热度也直接投射到资本市场。“Token第一股”迅策股价，已经从上市之际的40港元附近狂飙5倍。该公司收费将从订阅和交易制向Token升级，这将使收益与客户价值创造深度绑定，实现利润率扩张。

同样在港股上市的美图公司，商业模式也在从订阅模式延伸到Token消耗。公司首席产品官陈剑毅用咨询举例，“每个行业的咨询费都不一样，针对不同行业所能创造的商业价值做Token定价”。

AI时代的“集装箱”

大宗商品的特征之一是全球流动。

一个美国程序员打开电脑，调用DeepSeek的API来写代码。敲下回车后，请求数据通过太平洋海底光缆，到达中国西部的数据中心，中国的GPU集群消耗着中国电力，帮这位程序员把代码“跑”出来，结果再传回美国。整个过程不到1秒钟，算力、电力都没有离开中国，而程序员为这次服务支付了数美元。这是Token出海的一个典型场景。

上世纪50年代，全球贸易的成本一半左右是装卸费，后来集装箱被发明，货物运输被标准化，全球贸易迎来大繁荣。谷少辉将Token比作“AI时代的集装箱”——中国本土AI模型通过API接口向全球提供推理服务，按处理量计费，算力与电力实现“数字化出口”。

3月30日，据全球大模型聚合路由平台OpenRouter数据，上周全球模型调用量排名榜中，国产大模型调用量连续一个月超过海外模型。MiniMax、月之暗面等排名持续靠前。资本竞逐下，今年刚上市的智谱、MiniMax这两颗国产AI大模型“双子星”，市值直逼京东。

最新财报显示，MiniMax去年收入7903万美元，超过70%收入来自国际市场。此外，Kimi等多家平台均对证券时报记者证实，API海外调用今年以来快速增长。

中国模型足量、低价，一些企业乐意花钱补贴海外开发者，这些都放大了市场持续调用的需求，增强了中国Token服务在全球市场上的推广和使用中的优势。在谷少辉看来，这是政府工作报告中“算电协同”、“智能经济新形态”在产业层面的落地。“Token出海，或成为中国制造之后下一个出口引擎。”他说。

中美大模型性能差异不大，价格却有数倍之别，背后原因在于中国电力和算力优势。中国借助“东数西算”、数字经济等超前布局，接住了Token消耗狂飙的红利。有券商测算，国产AI模型综合推理成本仅为海外的1/10至1/6，多维优势转化为中国在全球AI服务市场的定价权。

Token成为“新大宗”，数据中心从存储设施演变为“Token工厂”——进入的是数据和电力，产出的则是词元和智能体执行能力。今年以来，二级市场资金多次围猎电力板块，背后原因就与“Token工厂”可能带来电力价值重估有关。

长期聚焦AI产业投资的基金经理杨勇举例，一度电直接出口通常在0.5元左右；做成铝锭可售1.5元；用来跑大模型推理，能产生500万Token，按照国内模型定价可以卖到10元，按照OpenAI定价可以上百元。

谷少辉认为，几乎所有产业出海都与电力优势相关。“制造业出海，可以视为电力加劳动力；精炼稀土等出海可以看作电力加资源；电力加算力下的Token出海是第三种电力出海形式。Token不是实物，而是服务贸易，这可能导致未来中美双边服务贸易差缩窄。”

“所有产业都值得被Token重构一遍，企业应该评估自己能不能装到Token里卖（服务）。”谷少辉说，“设计、咨询、教育……一旦变成可计量的智能服务，就可以规模化，这将是诞生新的垂直产业巨头的土壤。”

老玩家的新风口

与作为计量单位的Token狂飙相伴生的，是作为基础设施的算力竞赛。

新玩家智谱和MiniMax均处年度亏损状态，背后与庞大的资本开支密不可分。而老玩家方面，分别聚焦社交、搜索、电商、内容赛道的腾讯、百度、阿里和字节，在AI竞赛中，也在拼模型能力、拼算力资源，在AI基础设施及相关技术研发方面的资本开支持续加大。

从2025年财报来看，大厂中阿里资本支出最高，数额突破千亿元；腾讯和百度在百亿元级别。“大厂仍处在算力扩张期，多数都将获取最先进的训练芯片、更快迭代训练模型放在更高优先级，而并非先考虑降低推理成本。”杨勇表示。

国际巨头算力“军备竞赛”更为夸张，从2024年到2026年三年间，仅微软、亚马逊、谷歌、Meta四家公司的年度总资本支出就从约2000亿美元飙升到5000亿美元以上。不少科技巨头从高利润的轻资产模式，转向类似公用事业或制造业的重资产模式，一些公司负债率急升。

杨勇认为，量级虽不同，但中美巨头都在将海量资本投入到AI算力建设中，这是争夺下一代人工智能平台主导权的生死之战。只有赢家才能定义未来十年的技术格局。

但高额投入带来新问题。以折旧为例，英伟达的AI芯片迭代周期通常为一年左右。“今天花几千亿建的数据中心，GPU不到两年就不是最优版本，折旧应该怎么算呢？”杨勇反问。

中国云厂商们也进入全新的重资产设备周期。过去20年，云计算的叙事是“轻”——弹性伸缩、按需付费、用多少买多少。但算力需求膨胀使大厂算力部署表现激进，折旧摊销必将压在利润表上。“阿里云、火山引擎巨资投入，很像电信运营商3G/4G网络周期内的资本开支竞赛。”杨勇说。

这导致国际国内大厂更加依赖融资补血。但新问题随之出现。例如，新势力港股上市后，普遍要面对“是继续租算力还是自己买设备，是保利润率还是保用户”的艰难平衡。老玩家压力也很大。以阿里为例，一边是即时零售抢食美团的基本盘，另一边在“全栈式AI布局”下，算力烧钱凶猛，阿里将自身定义为“新的创业重塑”。

“阿里正试图从一个以电商为主业的平台，转型为一家覆盖从AI芯片、云计算基础设施、基础大模型到企业和消费者应用的全栈AI科技公司。转型结果将决定其在未来十年科技版图中的位置。”谷少辉说。

系统性新要求

Token调用量爆发带动算力需求，拉动了算力相关服务的价格。3月，腾讯云、阿里云和百度智能云，国内三大云厂商接连提高AI算力产品价格，十天之内涨价30%左右。同济大学经济与管理学院教授阮青松认为，涨价最先受益的是上游的芯片、服务器这些硬件厂商，而下游使用AI的应用和终端成本压力加大，倒逼企业要么提高效率，要么减少成本，可能会加速国产算力的替代进程，推动整个行业在技术创新上更进一步。

市场解读为大厂利好，但有接近腾讯的人士对记者介绍，扩建AI算力基础设施的资本开支动辄数百亿量级，但AI业务本身的利润率还很薄。目前卖Token的收入增速远远追不上建数据中心的花钱速度。“涨价不是云厂商贪心，而是供应链涨价的无奈之举。”在他看来，与此前低价引流、先把算力用起来的“Token价格战”相比，算力市场定价正在转向供需关系决定。

在三六零创始人周鸿祎看来，互联网流量是信息搬运，边际成本趋近于零，能走免费模式；但Token是智力产出，靠算力、芯片、电力支撑，用户越多、消耗越大、成本越高，边际成本递增，免费模式走不通。“现在行业都在赔钱做AI，就是因为免费惯性。智能体正在教育用户：软件免费、部署免费，但算力和Token需要付费，就像给数字员工发工资一样。这才是AI新势力活下去、互联网大厂进入可持续发展的健康模式。”

为了适配AI新经济体系，阿里宣布正式成立新的事业群，建立以“创造Token、输送Token、应用Token”为核心目标的新组织，由CEO吴泳铭直接负责，以Token Hub为核心主线，强化AI业务战略协同，推进AI战略落地。与阿里的全栈模式不同，腾讯的策略重心在于把握Token的流向入口。比如，用户在微信下达指令，Agent在云端执行任务，结果返回用户。由此，Token消耗被自然引导至腾讯云生态。

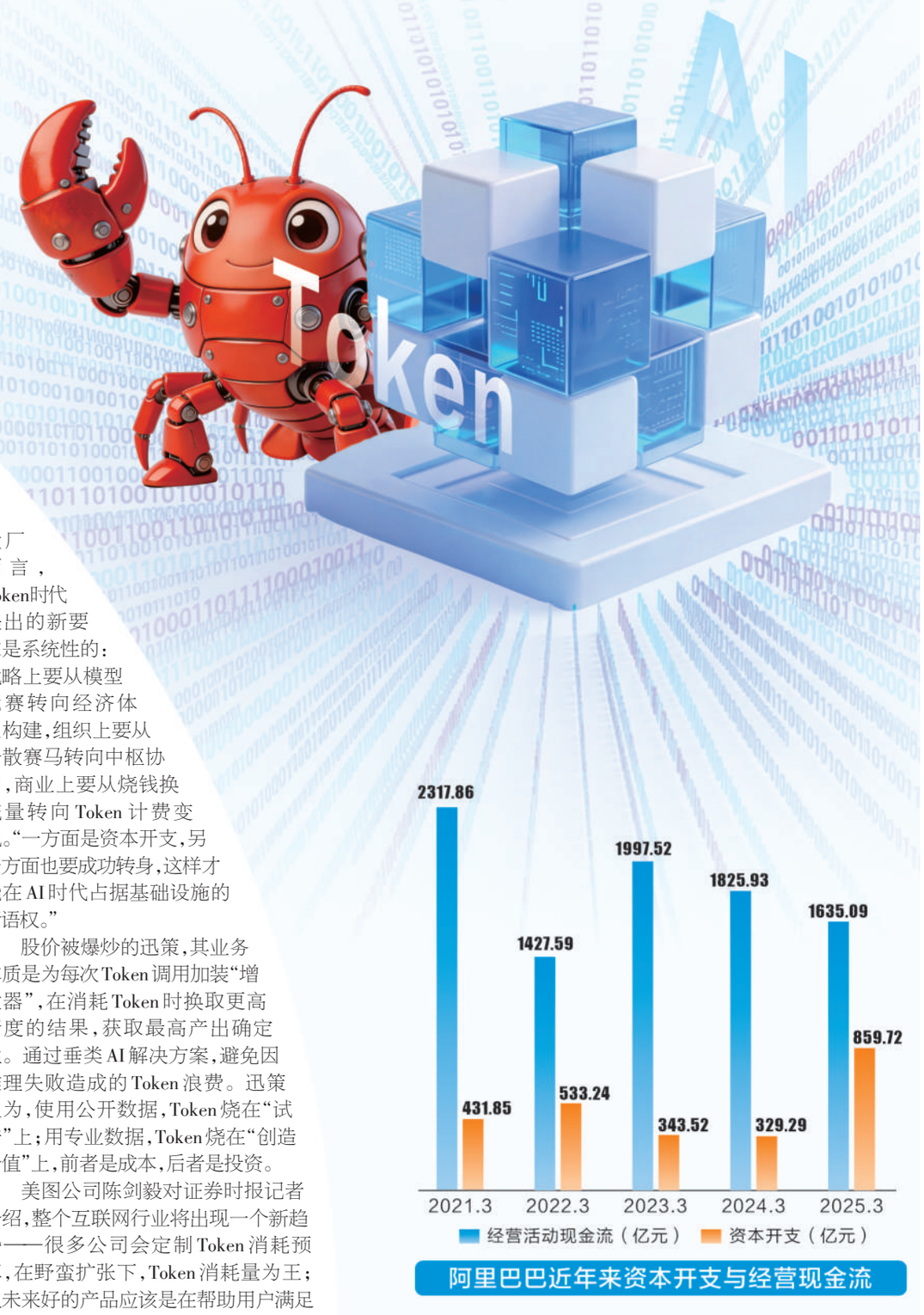
杨勇认为，对大厂而言，Token时代提出的新要求是系统性的：战略上要从模型竞赛转向经济体系构建，组织上要从分散赛马转向中枢协同，商业上要从烧钱换流量转向Token计费变现。“一方面是资本开支，另一方面也要成功转身，这样才能在AI时代占据基础设施的话语权。”

股价被爆炒的迅策，其业务本质是为每次Token调用加装“增效器”，在消耗Token时换取更高精度的结果，获取最高产出确定性。通过垂类AI解决方案，避免因推理失败造成的Token浪费。迅策认为，使用公开数据，Token烧在“试错”上；用专业数据，Token烧在“创造价值”上，前者是成本，后者是投资。

美图公司陈剑毅对证券时报记者介绍，整个互联网行业将出现一个新趋势——很多公司会定制Token消耗预算，在野蛮扩张下，Token消耗量为王；但未来好的产品应该是在帮助用户满足任务的同时尽量减少Token消耗数，可以根据所交付的商业价值给不同Token做更高的商业议价。因此，美图会关注帮用户降低Token消耗的次数，做更好的商业转化。

百度张婷畅想，五年后，Token一词可能从普通用户的视野里消失，其价值会以另一种形式存在——就像现在打滴滴不需要关心汽油消耗了多少升，用AI写一份报告不用关心消耗了多少Token，直接为成果付费。“MaaS（模型即服务）只是起点，最终希望提供的是端到端的AI能力，让企业客户不用关心底层是哪个模型、消耗了多少Token，而是关注AI解决了什么问题，带来了多少价值。”

“黄仁勋判断Token是新时代的大宗商品。但历史告诉我们，在大宗商品的全球贸易网络中，最终掌握主导权的，往往不是拥有最多原始矿藏的玩家，而是拥有高效提炼与转化技术的人。”谷少辉说，中国AI需要投入、需要效率，也需要更多不同于互联网流量时代的商业模式创新，来定义全球智能算力的贸易版图。



『Token经济』考验企业，也呼唤制度适配

证券时报记者 王小伟

Token（词元）以“新石油”的姿态登上全球经济舞台，这不仅是技术参数的跃迁，更是一场关乎产业逻辑底层重构的变革。

过去二十年，互联网商业的核心逻辑是“流量为王”——用户注意力成为最稀缺的资源，点击率、活跃用户数和停留时长构成了估值模型的基石。当AI Agent开始大规模替代人类执行复杂任务，当每一次推理、每一个决策都转化为可精确计量的Token消耗，商业价值的衡量标准正在从“注意力规模”转向“智能产出效率”。这一转变或不亚于工业革命从“马力”到“电力”的跨越，意味着企业竞争力的评判标准有望重置。

在新范式下，企业需要面对从“烧钱换流量”到“按效计费”的商业模式转型。比如，互联网行业的竞争逻辑是“先圈地后收割”——通过巨额补贴获取用户规模，再寻求变现路径。但Token是智力产出，依赖于算力、芯片和电力支撑，用户越多、消耗越大、成本越高。这意味着，AI时代的商业模型必须从一开始就建立在对价值创造的精准计量之上，而非对用户规模的粗放追逐。只有率先构建起“Token计费+价值分层”体系的企业，才能在新一轮产业竞争中占据主动。

从供给侧来看，随着AI芯片迭代周期越来越短，企业在算力布局上需有更精细的战略考量：是追求最先进芯片的绝对性能，还是构建更具成本效益的混合算力架构？是坚持全栈自研的重资产模式，还是寻求生态协同的轻量化路径？这些选择可能决定企业在“Token经济”长跑中的耐力。

需求侧的新机遇更值得重视。倘若咨询、设计、教育等知识密集型服务实现“装在Token里卖”，那些能够率先将行业know-how转化为高效Token消耗模式的企业将获得机会。正如集装箱标准化重塑了全球贸易版图一样，Token标准化也将催生新的行业冠军。

新的战略智慧呼唤制度适配。如何构建适应Token经济特点的政策框架，为数据跨境流动规则、算力资源调度机制等护航，如何引导企业从“参数竞赛”转向“价值创造”，从“规模扩张”转向“效率优化”，都是值得思考的课题。

历史经验表明，每一次价值衡量标准的变革，都会催生新的产业格局和商业文明。在这场从“流量”到“Token”的演进中，更具前瞻性的视野、更具创新性的思维，是我们定义属于自己的产业话语权的基础。

数据来源：Wind 图片来源：AI生成