

解码词元经济

抢滩新赛道：多地数据集团落子Token服务

中经记者 许璐 李晖 北京报道

随着词元(Token)正成为人工智能产业发展的新度量衡,企业对AI基础设施的需求变化催生新的服务崛起。

国家数据局发布的《全国数据资源调查报告(2025年)》显示,2025年全国日均词元调用量由年初超万亿增长至年末100万亿;今

从数据运营延伸至Token服务

地方数据集团拥有属地公共数据资源,入局Token赛道具备天然禀赋。

从各地披露的信息看,地方数据集团主要依托公共数据运营、算力基础设施、数据交易平台及地方产业场景,将分散在不同运营商、算力中心和模型企业的资源,封装为可统一调用、计量和结算的人工智能服务。

7月3日,北京数据集团旗下京算公司发布京算Token工厂。

7月15日,温州市数据集团参与建设的东南Token运营中心启动,定位为面向浙南、闽北和赣东的区域服务枢纽。5月30日上线的内蒙古词元交易平台集成算力调度交易、模型调用和词元结算功能;7月23日揭牌的扬州Token区域运营中心由扬州大数据集团牵头建设,整合上游大模型与算力资源,打造“Token市集”。

温州数据集团Token业务负责人向记者表示,东南Token运营中心是由温州市政府主导,由温州数据集团控股子公司负责落地运营。作为温州市推进Token经济重要的载体,中心摒弃了全域算力、大模型自建模式,聚焦算力资源统筹聚合、算力智能调度及全周期运营服务工作,将实现统一登录认证、统一收费支付、统一宣传推广、统一政策兑付四大功能,现已兼容接入100余款

年3月,国家数据局局长刘烈宏在中国发展高层论坛披露我国词元的日均调用量突破140万亿。

今年以来,北京、贵州、内蒙古、扬州、温州等地数据集团相继建设Token工厂、词元交易平台或区域运营中心等,将原有的数据资源运营、算力调度及数字基础设施业务,延伸至模型接入、推理服务、Token计量和交易结算等环节。

主流大模型。

“应用上,企业完成一次性接入即可快速调用各类大模型能力,实现开箱即用;平台可直观展示资源消耗量、使用资费及各项惠企政策,抵扣明细,全程严守数据安全与合规管控底线。”该负责人表示。

地方数据集团进入Token服务市场,与其既有资源结构有关。《全国数据资源调查报告(2025年)》显示,地方数据集团数据流量中,66.1%为公共数据。

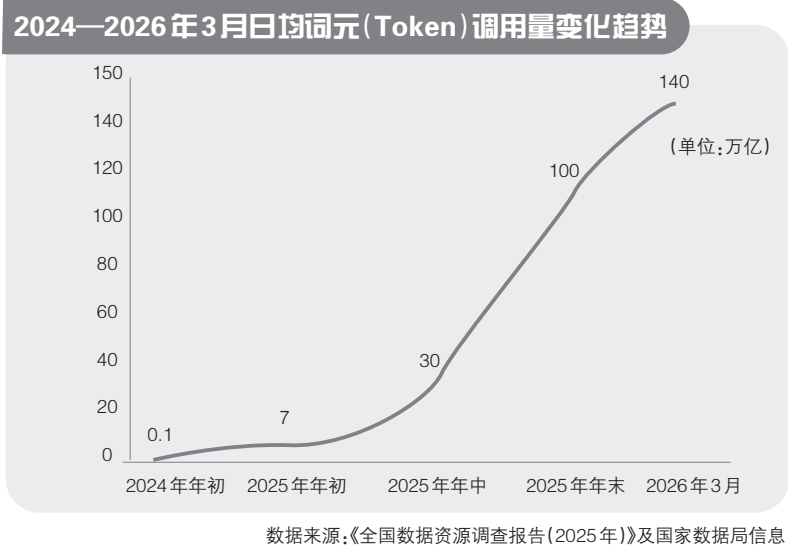
数据交易网创始人张瑶向记者表示,地方数据集团拥有属地公共数据资源、可统筹区域算力集群以及政务场景入口与国资合规信用,入局词元赛道具备天然禀赋。能源富集地区侧重算力词元供给,产业生态富集地区深耕垂直行业场景,预计后续的差异化会在具体的场景中逐步凸显。

从承担Token服务业务的主体上,多为地方数据集团通过旗下算力或数据服务子公司进行布局。2026世界人工智能大会期间,国资参股、贵州大数据集团合作伙伴——贵州数据宝发布“数据·算力·Token”三合一AI基础设施平台,由“数据要素运营商”向“数据原生AI To-

青岛数据集团首席数据官赵传启在接受《中国经营报》记者采访时表示,国内词元行业仍处于分化试点、渐进成熟的关键过渡期。目前各地数据集团聚焦两条路径,一条着眼释放数据资产价值,即数据词元化;另一条聚焦标准化供给,即算力词元化,两条路线彼此互补,共同搭建词元经济的产业骨架。

ken服务商”转型。

贵州数据宝方面向记者表示,当大模型能力逐步趋同,行业竞争的胜负手将转向底层基础设施的合规性、运转效率与场景适配能力。在产业链中,公司定位为AI基础设施服务商,把数据治理、算力调度、Token服务能力封装为标准化接口,让AI开发者实现“按量计费、即买即用”。这种转型本质上是在AI产业从“模型参数竞赛”转向“基础设施效率竞赛”背景下的必然选择。



从算力词元化到数据词元化

当前走在前列的是“算力词元化”,已完成基础工程化落地。而地方大数据集团尤其应该关注“数据词元化”。

记者从受访对象处了解到,当前市场上的Token工厂赛道模式主要指向“算力词元化”——即将异构算力、模型推理和接口服务封装为按Token调用的产品,客户购买的是模型调用能力。

在市场分工方面,云厂商和模型企业通常依托自有芯片、云平台 and 模型构建服务体系,地方数据集团则更多承担多方资源聚合、国产软硬件适配及地方场景对接。

对于地方数据集团集中进入词元服务市场这一趋势,深度科技研究院院长张孝荣认

计量定价和商业化仍在探索

目前Token服务处于从“概念验证”到“规模商用”的过渡期。

国家数据局在今年6月印发的《关于推进行业高质量数据基建建设行动的实施方案》提出,要探索词元交易等新型交易模

为,这是国资抢占AI基础设施的战略卡位,旨在填补市场信任缺口并带动产业升级。地方数据集团是责任驱动的“中立平”,整合多方资源并适配国产芯片,目的是带动区域产业和保障数据安全。

在赵传启看来,支撑产业发展的逻辑有二:其一,顶层设计鼓励词元标准化交易;其二,传统数据交易机制存在固有短板,无法适配多模态、跨主体数据融合流通场景。

赵传启向记者进一步分析

称,当前走在前列的是“算力词元化”,已完成基础工程化落地,处于成本计价向价值定价转型期。而地方大数据集团尤其应该关注“数据词元化”——即将文本、图像、基因序列等数据转换为可被模型识别和处理的词元,依托词元实现各类数据统一计量、价值核算与流通。

“数据词元化整体仍处于确权、计量、交易全链条验证探索阶段,统一词元化标准、跨场景适配规则,都是后续需要突破的核心课题。”赵传启表示。

模化商用仍存在不确定性。

张孝荣认为,目前Token服务处于从“概念验证”到“规模商用”的过渡期,深度应用有赖于Agent发展。后续增长取决于AI能否真正嵌入业务流程实现规模化落地;成本能否持续下降以驱动普惠化;统一标准能否打破利益僵局,形成规模效应。

在赵传启看来,想要规避同质化建设、实现良性发展,各地需立足自身禀赋错位布局,算力、数据、模型资源各取所长、精准发力。区域协同的核心,在于建立跨区域收益共享机制,外地调用本地特色词元资源时,加工方可获得合理分成。这种模式能倒逼各地深耕细分市场、打磨特色数据资产,最终形成互补共生的全国一体化词元流通格局。

国资Token工厂怎么建？

——访京算公司董事长颜敏

中经记者 李晖 北京报道

随着AI产业核心竞争从传统算力租赁转向“词元计量、定价、流通、应用的全生态竞争”,行业评价指标也正在从峰值算

力转向“每瓦Token数”。在此背景下,“Token工厂”这一AI产业链上的新赛道开始受到关注。今年以来,三大运营商、算力公司、地方国资等力量相继涌入……

在7月初举行的2026全球数字经济大会期间,北京数据集团旗下京算公司正式发布北京首家国资词元工厂——京算Token工厂,标志着首都智能算力服务迈入标准化、普惠化新

阶段。

国资入局Token工厂,定位是什么? 优势和挑战在哪里? 成本账怎么算? 北京数据集团副总经理、京算公司董事长颜敏在近期接受《中国经营报》记者

专访时表示:Token工厂是AI基础设施从“卖算力”向“卖服务”演进的核心产物,京算公司布局这一赛道则是市场客观条件、用户需求和公司自身发展需要的综合结果。

“整体看,这种模式在有效盘活国有资产的基础上,有利于强化客户黏性,有望推动公司从单纯的资源租赁商升级为AI能力的服务提供商。”颜敏表示。

从“卖算力”到“卖服务”

《中国经营报》:当前行业评价指标已从“峰值算力”转向“每瓦Token数”,这种转变背后的产业逻辑是什么? Token工厂与传统的智算中心本质区别在哪里?

颜敏:从“峰值算力”到“每瓦Token数”的转变,折射出竞争维度的变化——从“硬件资源比拼”转向“以词元为核心”的生态竞争。Token工厂是AI基础设施从“卖算力”向“卖服务”演进的核心产物,它的本质是把底层复杂的异构算力资源封装成标准化、可计量的Token和API服务,对算力进行一道“精加工”。

传统智算中心交付的是算力卡或服务器的租赁服务,客户拿到的是裸机,需要自己部署环境、适配模型;Token工厂交付的是标准化Token和API成品能力,按实际Token消耗量付费,客户打开浏览器就能用。从核心能力来看,传统智算中心侧重硬件采购和基础运维,Token工厂则聚焦智能调度、模型适配、推理优化和服务封装。

《中国经营报》:京算公司运营着北京最大的公共智算中心,此前主要以算力租赁为业务模式,为什么决定要下场做Token工厂? 这个业务商业回报率更高吗? 能带来哪些增量价值?

颜敏:京算公司做Token工厂的核心原因是市场需求结构的变化。2025年下半年以来,智能体应用从技术验证走向C端规模化爆

发,AI的产业关键环节也从模型训练向大规模推理转移,闲置的算力找到了用武之地,Token工厂的商业逻辑随之清晰起来。

客观层面,芯片、存储和服务器的整机能力正处在快速迭代的周期中,上一代算力卡虽然在训练场景下竞争力变弱,但面对推理需求的爆发式增长,其性价比反而凸显出来。

主观层面,国资有服务更多客户的需求。Token工厂的业务模式让算力服务从服务少数大模型公司,拓展到更多中小企业和国资客户。以前客户如果说“我只想用两个模型,你帮我部署好,把API给我”,我们可能服务不了,但是现在可以高效满足。

总结下来,这种模式一是将闲置的算力转化为可售卖的Token,能有效盘活国有资产;二是通过API绑定,客户黏性更强,形成持续的服务收入;三是从单纯的资源租赁商升级为AI能力服务提供商,更容易掌握定价权和话语权。

这种模式对客户也有很大价值——不用承担高昂的硬件采购和运维成本,按量付费,综合成本大幅降低;不用自建运维团队就能拥有接入AI能力;通过API调用的方式,还能帮国货企业规避直接使用特定型号硬件可能带来的合规风险——你只需关注调用结果,无关关心底层是哪种类型芯片,优点是安全、高效、省心。

轻资产运营、打造七大核心能力

《中国经营报》:作为北京首家国资词元工厂,国资背景带来了哪些优势和挑战? 和其他地区的国资Token工厂相比,京算公司的差异化价值是什么?

颜敏:国资背景带来的优势很具体:长期稳定的服务能力、供应链安全保障、安全合规要求、为国资客户提供定制化解决方案的能力。

在国资信用与机制灵活的平衡上,京算公司定位为国资轻资产运营平台,自身不主动持有服务器重资产,而是通过租赁模式获取算力,专注于运维、运营和商业化变现。这种模式既发挥了国资的信用与资源优势,又通过轻资产运营保持了商业上的灵活性。

和西部算力枢纽相比,我们

强健“供词元”枢纽功能

《中国经营报》:你们的核心能力是自主研发还是与生态伙伴联合攻关? 我关注到北京数据集团已联合多个合作伙伴签署Token工厂共建协议。具体如何分工?

颜敏:七大核心能力是自主研发与生态伙伴联合攻关的成果。京算公司定位为运营平台,在技术实现上采取了开放合作策略。

从分工看,北京数据集团(京算公司)作为主导方与运营方,提供Token工厂的品牌、运营

的核心优势在区位,北京能提供毫秒级以内的极低时延(如果推理时延超过50毫秒,很多客户就不用了)。同时紧邻北京庞大的国资客户和AI产业生态,这是服务首都及周边高价值推理需求的基础。

当然,国资有优势也有挑战,和创业公司相比,我们的策略还是在信用、区位和服务上的独特优势,同时积极与外部技术力量开展联合攻关,在开放中不断迭代自身能力。

《中国经营报》:京算Token工厂已具备七大核心能力,其中PD分离智能调度架构是最受关注的技术亮点。能否通俗解释一下

这项技术? 从效果看,算力吞吐量最高提升90%、大模型推理时延速度提升70%以上,是怎么实现的? 目前比较大的技术挑战是什么?

颜敏:通俗地讲,可以把大模型推理想象成一条生产“词元”的流水线。传统方式下,Prefill(预填充)和Decode(解码)两个阶段在同一条流水线上串行作业,争抢同一块芯片的计算和存储资源,效率受到很大制约。PD分离就是把流水线拆成两个专业车间:Prefill车间负责“读懂”用户输入的整个问题、生成第一个Token,这个环节计算密集,适合用高并行计算的芯片处理;Decode车间则负责在第一个Token基础上逐个“吐出”后续Token,这个

环节更看重持续吞吐能力。两个车间独立运作、并行生产,可以大幅提升效率。

实测效果也验证了这一架构的价值——由于两个阶段不再争抢同一芯片资源,芯片利用率显著提高,算力吞吐量最高可提升90%;同时,因调度链路更精简、资源竞争消除,推理时延降低70%以上。

目前技术上最大的挑战是底层硬件的多样性和软件栈的不兼容——不同厂商的芯片有不同的编程框架,我们需要构建一个能屏蔽底层硬件差异的统一调度层,让上层应用不用关心底层是什么芯片。这需要深厚的技术积累和与各芯片厂商的紧密合作。

元、社区用词元”的完整产业闭环,京算Token工厂在其中承担什么功能? 下一步有哪些规划和目标?

颜敏:京算Token工厂承担着关键的“供词元”枢纽功能。在生产侧,作为“工厂”,它将底层算力高效地转化为标准化的词元产品,是整个闭环的“生产源头”。在应用侧,它通过标准化API,为下游的“应用社区”和各类AI应用提供源源不断的Token服务。

在规模发展方面,预计2027

年至2028年的算力规模将有大幅度跃升。在服务能力方面,持续优化PD分离、异构调度等核心技术,进一步提升Token产出效率、降低成本,为更多国资及行业客户提供高质量、高性价比的AI推理服务。

我们的核心目标是构建安全、可控、标准化的词元服务体系,为首都数字经济发展提供坚实支撑。在生态建设方面,通过与更多合作伙伴共建,持续扩大Token工厂的生态圈。