

2026中国国际大数据产业博览会:

以词元为尺 度量数据新价值

8月28日至30日,2026中国国际大数据产业博览会(以下简称“数博会”)在贵州省贵阳市举行。本届数博会以“词元——数据要素价值释放新路径”为主题,吸引来自21个国家和地区的1.6万余名嘉宾注册参会、372家中外企业参展,举办各类活动89场,发布87项创新成果。

在开幕式上,人形机器人“小树”与主持人搭档主持,刚刚在第二届世界人形机器人运动会障碍赛夺得100米冠军的人形机器人“智元灵犀X2”现场演绎苗族舞蹈与武术,“词元”相关顶层设计成果发布,“五精展区”首次实体亮相……“词元”这一核心关键词贯穿整届数博会始终。词元是什么?围绕词元经济,产业界又发生了怎样的深刻变革?



2026中国国际大数据产业博览会热闹非凡

图源新华社

>> 词元新解:从计量单位到价值锚点

词元,是人工智能理解和生成信息的最小计量单位,其经营本质是提供AI服务。用户向AI输入一段话,AI将汉字拆解为词元进行运算,再将结果以词元形式返回。生成一张图片、剪辑一段视频、驱动具身智能机器人,背后均消耗词元。如果把AI大模型比作大脑,词元就是大脑读懂世界的基础语言单元,相当于大模型的识字课本。没有高质量的词元,就算有海量的数据,大模型也如同认字不全的学生,读不懂

文字背后的真正含义。

当前,词元正逐步成为人工智能服务的计量单位、结算单位与统计单位。长期以来,数据要素面临“价值难度量、收益难分配”等问题。本届数博会聚焦推动数据从资源向资产转化,共同探讨数据要素在流通、交易和价值评估中的标准化难题,为释放数据价值提供新的解决方案。

2026年3月,国家数据局正式将Token的中文名定为“词元”,定位为“智能时代的价值锚点”。5月,国家数据局召开词元经济座谈会,明确了词元经济发展的核心着力点是加强高质量数据集建设和全国一体化算力网建设。词元作为数据要素价值释放的新路径,为人工智能服务的计量、定价、交易和结算提供了重要标尺,打通数据、模型、算力之间的价值传导链路,正成为智能经济时代重要的基础计量单位与价值锚点。

>> 算力筑基:词元经济的爆发式增长

2024年初,全国日均词元调用量为1000亿次。2025年底,这一数字飙升至100万亿次。2026年3月,突破140万亿次,相当于250个图书馆的全部馆藏资源在一天内被调用、处理、生成。截至2026年6月,全国日均词元调用量已突破500万亿。

词元经济的爆发式增长,离不开算力基座的坚实支撑。截至2026年7月底,全国智算总规模达到245万PFLOPS(FP16),八大算力枢纽和三个算电协同发展区已建成。智算规模占全国比重超85%,已有145万PFLOPS智能算

力纳入国家级监测调度平台监测范围,算力资源集约化利用、一体化调度能力不断增强。国家数据发展研究院院长胡坚波表示,未来词元的服务将催生新的产业生态,AI将拓展数据产业发展新空间,数据产业规模也将保持较快增长水平。

数据产业主体活力持续释放。相关报告显示,截至2025年底,全国数据企业达48.2万家,同比增长17.9%。“和人工智能高度相关的数据产品和服务,现在已经占到数据产业总产值的七成以上,成为拉动数据产业发展的主力。”国

家数据发展研究院数据资源研究部主任王威表示,数据产业和人工智能正在相互成就、双向赋能,数智化成为产业发展新趋势。

高质量数据集的供给同步提速。截至2026年8月,全国累计建成高质量数据集超过12.6万个,数据总体量超1815PB,较2026年3月增长超过89%。国家数据局党组书记、局长刘烈宏在开幕式上表示:“当前,人工智能发展呈现鲜明的数据驱动特征。我们坚持人工智能发展到哪里,高质量数据集建设和应用就推进到哪里。”

>> 应用场景:词元赋能千行百业的落地

当算力、数据、模型、应用通过词元贯通起来,AI落地的图景正变得愈发清晰。贵州首个具身智能“数据工厂”已在贵安新区正式投用,内部搭建科研、餐饮、工业、零售、家居、办公六大类16个全真模拟应用场景,构建起“场景采集、设备校准、清洗标注”的完整数据采集闭环体系。截至2026年6月底,全国已建成启用超70家具身智能训练场,在建或规划中的训练场46家。

在工业领域,AI大模型通过海量工艺词元的模拟与优化,将电解槽分子比稳定控制在2.42的理想区间,累计节约电费超4000万元。贵州已围绕24个重点行业打造了170余个AI大模型应用场景。南方电网贵州电网公司数智运营中心相关负责人介绍,无人机能边巡检边识别,目前缺陷识别准确率超过90%,以前一个人一天的巡检工作,现在无人机只需半小时,还能规避安全风险。

在民生服务领域,天眼AI应用矩阵平台面向政务、民生、教育、消费等多元场景提

供智能化服务,同步亮相五款智能体产品覆盖智能创作、采访记录、内容安全、婚恋服务、教育升学五大应用领域。依托公共数据授权运营打造的“贵企AI”智能体,打通工商、社保、公积金等公共数据,把小微企业授信核验时长由数个工作日压缩至分钟级。城市排水管网诊断AI大模型“城脉·慧诊”,以数智手段破解管网治理痛点。

在医疗领域,“39AI医生”是国内领先的3000亿参数生成式医学大模型。其核心是将海量结构化医疗数据进行深度标注,转化为高质量词元用于模型训练,仅训练数据集体量就达4TB。相关工作人员介绍:“每一次疾病诊断、病例分析、语音问诊,本质上都是词元的调用与生成。”目前,该模型已在贵阳市29家公立医院开展试用,186家社区卫生服务中心及广西医科大学第一附属医院等三甲医院接入使用,累计使用量突破71万次,数以亿计的医疗词元正在真实诊疗场景中流转。

>> 词元工厂:从卖算力到卖词元

随着AI应用加速落地,“词元工厂”概念受到产业广泛关注。相关数据显示,今年以来,国内已有至少10家上市公司宣布探索“词元工厂”业务。王威介绍,词元工厂是在产业实践中探索形成的新路径,核心是降低词元生产成本,输出稳定可用的标准化词元服务。词元工厂的落地应用,显著降低了企业应用AI的门槛,企业不用自建算力中心,只需按词元实际消耗量付费,实现从“为算力买单”变为“为效果付费”。

在数博会现场,软通动力信息技术(集团)股份有限公司联合中国电信股份有限公司贵州分公司,共同发布了贵州壹号词元工

厂。这一落地贵安新区的新型AI基础设施,规划产能为日产万亿左右词元。展台上的演示沙盘直观还原了整套生产流程:上游接入各类行业高质量数据集,经过词元加工,形成标准化服务向下游输出,供给无人机巡检、AI康养机器人、政务智能体等各类场景。软通动力相关工作人员表示,词元工厂补齐了AI产业的关键环节,把算力资源优势转化为AI产业竞争优势。

本届数博会还展示了国内首个车载移动式词元工厂。这台采用标准化40尺集装箱的设备,整车采用电算冷、存算网一体化设计,可支持9个机柜部署,多车可快速组网形成移动算力集群,从

到场接入到完成调试运行只需6小时至8小时。南方电网数研院电算工程公司相关工程师介绍,120千瓦的供电,可以提供50P到100P的算力,理论上可以支撑两个这种国产大模型推理任务。它可以快速部署到变电站、分布式新能源电站,还有零碳园区。据测算,对比传统算力部署模式,用移动方式接入算力,整体成本降低25%左右。面向中小企业,有参展商推出词元资源生产平台,插上电就可以直接开机运用,每秒钟能够生成900多个词元,对于企业来讲,几十个员工同时使用这台设备都能满足需求。



机器人乐队表演

图源新华社

综合报道:王琳

素材来源:数博会微信公众号、央视网、央视新闻、新华网、国家数据局、中国新闻网、《人民日报》、光明网、《贵州日报》