

词元经济加速崛起 数据要素价值持续释放

■中国城市报记者 邢 灿

在近日于贵州省贵阳市落幕的2026中国国际大数据产业博览会(以下简称“数博会”)上,“词元”成为各界关注的焦点。

词元是AI处理文本时的基本单位,是大模型处理信息的最小单元。简单来说,词元是人工智能大模型为了高效处理数据,把数据进行拆分后的“最小信息载体”,可以理解为“字/词片段/符号”等。

国家发展改革委党组成员、国家数据局局长刘烈宏在数博会期间指出,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、交易和结算提供重要的标尺,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。

刘烈宏指出,当前,以词元为关键的商业逻辑持续迭代,以词元为表征的新经济形态加速培育成型。词元作为新生事物,发展机遇与挑战并存,亟须政产学研用各方协同发力,围绕以词元促进数据要素价值释放、培育多主体协同共进的产业生态,探索适配产业实际的商业模式和深化行业应用,完善词元领域制度供给促进安全有序发展等四方面,加强理论研究和交流探索。

词元市场需求 迎爆发式增长

近年来,国内词元市场需求急剧扩张,产业形态加速迭代。数据显示,国内日均词元调用量从2024年初的1000亿级,跃升至今年3月的140万亿级,两年间增幅超千倍。

这一趋势在企业的实际经营中亦得到印证。贵州数据宝网络科技有限公司董事长汤寒林向中国城市报记者介绍,今年初公司词元日调用量不足30亿次,目前已稳定在400亿至500亿次,最高单日峰值突破数千亿次,订单量从年初的数万单增至数十万单。

产业热度持续攀升的背后,是智能体应用加速兴起,叠加各行业数字化转型需求

持续释放,共同推动了词元应用场景不断落地。数博会期间发布的《中国数据产业发展报告(2026)》显示,2025年我国智能体专利授权量超过3400件,增速是2024年的两倍以上。

阿里云副总裁安筱鹏在今年数博会期间分享了他观察到的行业变化:今年2月,智能体自动调取的词元消耗首次超过人类手动调取量,到8月该数值已达到人类的5倍。

国家数据发展研究院院长胡坚波表示,未来词元服务将催生新的产业生态,AI将拓展数据产业发展新空间,AI原生等数据技术和产业创新将加速突破,数据产业规模也将保持较快增长水平。

值得关注的是,词元市场的爆发式增长,催生了一个新的基础设施形态——词元工厂。

对此,无问芯穹联合创始人、首席科学家戴国浩分析,词元需求指数增长与算力线性供给形成结构性矛盾。词元工厂通过软硬协同、跨集群异构推理技术,破解算力碎片化难题、压降推理成本。技术迭代提效将持续拉动需求增长,对词元工业化生产提出更高标准。

在业内人士看来,伴随词

元工厂的出现,企业应用AI的门槛由“建电站”降到了“插电卡”,无需再自建算力中心,只需按词元消耗量付费,从“为算力买单”转向“为效果付费”。

持续推进 高质量数据集供给

数据要素质量很大程度上决定着词元的价值成色。没有高质量数据集,词元产出的就是“劣质燃料”,无法支撑可靠的智能决策。

“数据和词元,就好比米和饭。有好米才能做出好饭。”在中央财经大学中国互联网经济研究院副院长欧阳日辉看来,只有持续供给高质量数据集,才能产出高价值的词元。

广东省政务服务和管理局党组书记、局长王天广持相似观点:“当前行业拥有较为充足的算力与较为成熟的算法能力,亟待破解的短板在于高质量数据供给不足。”

行业高质量数据集建设是提升词元价值的关键。胡坚波认为,不同场景下词元价值差别很大。同样是百万词元,通用对话中可能只值几元,在医疗金融场景中,价格

可能高出十倍甚至数百倍。词元的含金量,取决于它服务的场景。

“场景决定数据,数据赋能模型,数据是场景跟模型之间最重要的枢纽点。”在瓴泓智能科技(贵州)有限公司董事长郑科伟看来,数据质量决定了大模型性能上限。

国家数据局最新数据显示,截至2026年8月,全国累计建成高质量数据集超12.6万个,数据总体量超1815PB(存储容量单位,1PB等于1024TB),较今年一季度末大幅增长超89%。

数博会上,14家央国企发起国内首个“数据开放月”活动,面向全社会免费开放高价值数据资源,推动数据供需精准对接和高效开发利用,促进数据价值释放。

中电信人工智能科技(北京)有限公司的副总经理刘翼提醒,一段时间以来,各地纷纷建设高质量数据集,现阶段需重新审视,到底需要什么样的高质量数据集?

刘翼认为,当前地方数据集建设中主要存在两方面问题:一是碎片化。一些地方归集了大量行业数据集,但数据较为零散,难以形成规模效应。二是泛化性不强。以往地方建成不少数据集,其泛化

能力、行业应用价值,以及对模型能力的带动能力,仍有待观察评估。

场景应用是产业落地关键

目前,广东省、内蒙古自治区、贵州省、北京市、安徽省、江苏省扬州市等多地已出台词元专项政策,从不同维度展开制度探索。

内蒙古自治区依托自贸试验区呼和浩特片区,上线全国首个绿色算力全栈AI平台——词元交易平台;同时,全国规模领先的词元工厂也在和林格尔新区落地。今年7月,贵州省出台专项扶持政策明确,企业购买大模型词元调用服务可获得不超过50万元补贴,建设高质量数据集可获得不超过500万元奖励。

王天广认为,词元产业从技术概念走向产业落地,最关键的是“场景应用”,场景应用是试金石、催化剂、连接器,推动词元经济发展不仅要比拼算力,更要看场景丰富度、落地能力和实践深度。

今年以来,广东省在词元产业落地方面开展了积极探索:4月22日跑通词元出海全流程闭环,打造“数字海上丝绸之路”,形成AI玩具、AI模具、AI电商员工、开发者API调用服务等多种形态;7月30日挂牌成立广东词元交易和服务中心,解决资源汇聚、标准统一、第三方审计、政策指导等问题,让好词元卖出好价格,形成良性循环。

贵州省同样在应用场景上下足功夫,民用商用政用各方面场景应开尽开,推动词元应用与酱酒、煤矿、文旅、医药、金融等领域深度融合,加快打造推广一批大模型产品,创新做好“人工智能+”的文章,推动形成以模型应用牵引数据海量供给,以数据赋能模型迭代的良性互动格局。

在顶层设计层面,国家数据局副局长余英此前在数博会新闻发布会上表示,将深化词元应用与工业、医疗、金融等领域的深度融合,助力人工智能应用规模化落地,推动形成以模型应用牵引数据供给、以数据赋能模型迭代的良性供需互动,形成“数据飞轮”效应。



2026中国国际大数据产业博览会上,北京临一云川能源技术有限公司展区的“平流层浮空风力发电系统”吸引参观者。

中新社记者 瞿宏伦摄