

词元经济如何驱动智能经济发展

——访中国工业互联网研究院数据管理与应用研究所所长陈杰浩

► 本报记者 李争粉

近日,中国工业互联网研究院数据管理与应用研究所发布《Token驱动智能经济研究报告》(以下简称“报告”),首次阐释Token(词元)从技术计量单元向新型生产要素跃迁的演进逻辑,揭示其作为“能源+芯片+数据集+大模型”贯通融合产物的核心价值。

与传统数据要素相比,词元作为新型要素最本质的区别是什么?当前我国算力供需紧张、数据流通存在壁垒,词元要素化如何串联资源,破解产业链协同堵点?对此,本报记者采访了中国工业互联网研究院数据管理与应用研究所所长陈杰浩。

记者:报告首次提出词元从技术计量单元向新型生产要素跃迁。如何通俗阐释这一核心观点?与传统数据要素相比,词元的最本质的区别是什么?

陈杰浩:这里所说的“跃迁”,并不是指词元的技术定义发生了变化,而是指其在经济运行中的功能发生了变化。词元最初只是大模型切分和处理文本的技术单元,相当于机器理解语言的“最小颗粒”,随着人工智能(AI)从模型训练走向规模化推理,从单轮问答走向智能体持续执行任务,词元开始贯穿模型调用、任务拆解、推理计算、结果生成和服务结算全过程,逐步承担价值计量、成本核算、资源调度和生态协同等功能,不再只是一个后台技术指标,而是进入了真实的生产和价值创造过程。

可以作一个通俗的比喻:如果把数据看作原料,把能源和芯片看作动力与生产设备,把大模型看作生产工艺,那么词元就是这套体系加工形成的可以被调用和交付的标准化“智能服务颗粒”。企业真正需要的并不是拥有多少张显卡、多少模型参数,而是能够以多低的成本、多快的速度、多高的质量获得智能服务。词元能够将后台复杂的能源、算力、数据和模型能力,转化为可量化、可定价、可追溯的服务供给。

与传统数据要素相比,词元最本质的区别在于:数据更接近一种存量型的信息资源,词元是一种在模型运行和服务调用中形成的流量型智能服务单元。数据回答的是“掌握了什么信息和知识”,词元回答的是“这些信息和知识被转化为多少智能能力、经过多少实际调用、产生了怎样的服务价值”。数据

即使被存储起来,也不一定直接创造价值,词元则产生于数据、算力、能源和模型的协同运行过程,并且能够把每一次实际调用与资源消耗、成本归集和业务产出联系起来。

因此,词元并不是要取代数据要素,而是在数据之上进一步发挥“要素连接器”和“价值转换器”的作用:向上连接能源、芯片、数据集和大模型,向下进入研发设计、生产制造、经营管理和终端服务等场景,把多种底层要素转化为可交付的智能能力。

记者:报告提出词元是“能源+芯片+数据集+大模型”贯通融合的产物,词元要素化将如何串联这四大资源,破解产业链协同堵点?

陈杰浩:词元要素化并不是简单地给能源、芯片、高质量数据集和大模型贴上同一个标签,而是要推动4类要素协同贯通,形成连续价值转化链路:能源在芯片及相关基础设施支撑下转化为有效算力,算力经高质量数据集与大模型驱动转化为高质量词元,词元进一步注入业务流程,形成产业价值。

我国东部地区模型研发和应用需求高度集聚,高性能训练算力与低成本推理算力供给相对紧张;西部地区虽然建设了大量智算中心,但现有设施更多是按照规模化训练算力的供给逻辑布局。由此,可将西部丰富的绿色能源优势,通过芯片、数据集和大模型的协同转化为低成本、可持续交付的词元服务,并面向全国进行调度和输出,推动能源、算力、数据、模型与应用需求有效衔接,破解资源闲置、跨域调度不畅和应用需求难以承接等产业链堵点。

报告提出的词元工厂和词元调度分发体系,正是实现这一机制的产业化载体:词元工厂负责将能源、芯片、数据集和大模型整合为标准化词元供给;词元分发网络和流通服务枢纽汇聚分散产能,开展跨区域智能路由、供需匹配、统一计量和清分结算;消费端产生的调用需求、质量评价和业务结果再反向传导至生产端,指导算力扩容、数据优化和模型迭代,由此形成“需求牵引供给、调度优化配置、使用决定分润”的闭环。

因此,词元串联四大资源的关键在于,以真实调用需求作为统一需求信号,以词元作为共同计量尺度和价值接

口,改变能源、芯片、数据和模型过去各自建设、各自定价、彼此割裂的状态,打通从“纸面算力”到“有效词元”再到“业务价值”的最后一公里。当然,这一作用不会随着词元计费自然实现,还需要同步建立统一计量口径、标准接口、质量评价、数据授权、安全审计和清结算规则,才能真正把技术上的可调用转化为产业上的可流通、可协同。

记者:报告提出“以中国场景定义词元价值、以中国标准引领要素配置”。对比海外相关技术与治理路线,我国打造自主词元要素体系要避免哪些误区,标准建设优先从哪些方向突破?

陈杰浩:“以中国场景定义词元价值、以中国标准引领要素配置”,是要把我国丰富的产业场景、完整的制造体系和超大规模市场,转化为标准形成和规则输出能力。词元价值不能只由模型厂商的分词方式和API(应用程序编程接口)报价决定,更要看它在工业生产、政务治理和社会服务中解决了什么问题、创造了多少实际价值。

从海外相关实践看,当前技术路线主要由头部平台和开源生态共同推动,围绕模型接口、调用计量、智能体协作等形成一批事实性规范,治理上则更加关注数据合规、安全审计、平台责任和跨境服务等问题。相关规则仍处于持续演进之中,尚未形成覆盖词元生产、调度、消费全生命周期的统一要素体系,这也为我国立足自身产业实践、探索系统化标准提供了空间。

我国打造自主词元要素体系,需要重点避免几个误区:一是不能把词元标准简单理解为统一分词算法或者统一计费单位。二是不能把词元数量等同于词元价值。三是不能把自主可控理解为封闭隔离。四是不能在计量、质量、权属和审计规则尚未完善时,过早推动词元泛金融化。

标准建设可以按照“先立尺、再通路、能评价、算清账、守底线”的顺序推进。中国标准的竞争力不在于创造一套新的技术概念,而是能否把复杂、多元的产业需求转化为可复制的计量规则、质量标准和治理方案。可优先选择工业制造、政务服务、智能运维等边界清晰、效果可验证的场景开展试点,在实践中不断检验和完善标准,进而将我

国的场景优势和市场优势转化为智能经济的规则话语权。

记者:面向中长期智能经济发展,围绕词元要素培育,政策层面、产业层面、科研层面应当协同推进哪些关键举措?

陈杰浩:面向中长期发展,词元要素培育需要形成“政策制定规则、产业验证模式、科研提供支撑”的协同机制,并通过场景试点不断反哺标准和政策。

政策层面,重点是“立规则、建底座、守边界”。一是要进一步明确词元的产业属性和应用边界。二是要加快构建覆盖计量、接口、质量、计费、结算和安全审计的标准体系,推动不同模型、平台和算力节点互联互通。三是依托全国一体化算力网和算电协同布局,统筹建设词元生产、调度与分发基础设施,探索词元价格、质量和产业赋能指数。四是可以选择工业制造、政务服务、智能运维等边界清晰的场景开展试点,建立包容审慎、全程可追溯的监管机制,为新模式留出验证空间。

产业层面,重点是“建体系、跑场景、形成闭环”。算力运营商、模型企业、数据服务商、工业企业和平台机构需要共同打通“生产—调度—消费”链条,推动词元工厂、调度分发枢纽和应用生态协同建设。企业内部建立统一的词元台账、预算管理和质量评价体系,将消耗进一步归集到部门、工序、设备、产品和订单,建立“词元投入—任务完成—业务效果”之间的对应关系。产业发展要形成一批能够提高产品良率、减少设备停机、优化能源消耗、改善服务质量的高价值应用,以真实效果检验词元要素化的商业可行性。

科研层面,重点是“测得准、调得动、说清价值”。要加强异构算力适配、低成本推理、缓存复用、智能路由、多智能体协同和边缘部署等工程技术研究,提高单位能耗和单位成本的有效词元产出。同时,要深化词元经济学和治理研究,重点回答不同模型的词元如何比较、词元价格如何形成、数据和模型贡献如何分配、词元消耗如何影响企业绩效等问题。还要推动人工智能、经济管理、工业工程、标准化和安全治理等领域交叉研究,避免将词元研究局限于单一技术或计费问题。