

解码词元经济

数据价值释放之年：词元催化行业全链条“焕新”

中经记者 许璐 郑瑜 贵阳报道

随着全国数据市场逐步实现一体化,进一步增强数据、算力、模型三者的耦合性成为行业共识。

在近期举行的 2026 中国国际大数据产业博览会(以下简称“数

场景激活数据新供给

数据供给的重点正从扩大规模转向满足特定场景和任务的需求。

《全国数据资源调查报告(2025 年)》显示,2025 年我国用于人工智能训练和推理的数据总量达到 199.48 艾字节,同比增长 42.86%。其中,推理数据量为 101.34 艾字节,首次超过训练数据量的 98.14 艾字节。具身智能领域的数据生产量同比增长 477.78%。

这显示,人工智能进入产业应用阶段后,数据供给的重点正从扩大规模转向满足特定场景和任务的需求。

记者注意到,该变化在具身智能领域表现较为明显。例如,机器人在制造、康养等现实环境中执行任务,不仅需要识别物体,还要理解

智能技术嵌入开发链条

人工智能正与模型行业实现深度融合。

场景激活数据新需求的同时,如何将分散的数据加工为模型可用的高质量数据集,并在跨主体使用过程中保障安全与权益,成为数据供给和模型应用的中间环节。

在 2026 数博会开幕式上,刘烈宏披露,截至 2026 年 8 月,全国累计建成高质量数据集超过 12.6 万个,数据总体量超过 1815PB,较 3 月增长超过 89%。

从产业实践看,高质量数据集建设并非简单汇集原始数据,而是需要经过采集、清洗、标注、质量评测等环节。与此同时,人工智能技

词元计价催生新服务模式

随着企业采购和使用 Token,新的服务需求也在形成。

数据经过加工进入模型后,如何计量模型调用和智能服务,成为数据、算力、模型与应用形成价值循环的又一环节。

2025 年,我国全年词元调用量约 21100 万亿,日均调用量由年初超过 1 万亿增长至年末 100 万亿。词元调用量增长的同时,围绕模型接入、算力调度、调用管理和统一结算的服务需求开始出现。

刘烈宏在 2026 数博会“词元——数据要素价值释放新路径”交流活动中指出,词元为人工智能

博会”)上,国家发展改革委党组成员,国家数据局党组书记、局长刘烈宏提出“三个全新”:一是数据赋能人工智能发展迈入全新阶段;二是人工智能技术成为激活数据要素潜能、开发利用数据资源的全新利器;三是词元正开辟数据要素价

物理规律、动作顺序以及人与环境之间的关系。与语言模型主要依托文本、图片等互联网数据不同,具身智能训练需要大量带有物理属性和动作反馈的数据,但真实场景采集往往面临成本高、周期长和极端样本不足、场景有限等问题。

对此,跨维智能贵阳区域总经理陈泓向记者表示,生产企业一条产线可能价值两三千万元,让机器人直接在产线上进行两至三个月的实训并不现实。相比之下,可以在虚拟环境中搭建产线,让机器人以仿真形式完成序列任务,经过上千次乃至上万次训练后,再将相关

术也开始进入数据处理环节,用于自动识别、预标注等环节,减少部分重复性人工操作。

此外,人工智能正与模型行业实现深度融合。

“我们做了一个工业质检的行业大模型,用 DeepSeek 微调,因为要把 15G 的数据送到大模型微调,通过互联网大概要 4 小时,但通过确定性网络只需 22 秒。大模型训练完以后再传回企业,互联网要 2 小时,通过确定性网络只需 2 秒。”中国工程院院士、紫金山实验室首席科学家刘韵洁在数博会上介绍。

服务的计量、定价、交易和结算提供了重要标尺,打通了数据、模型、算力之间的价值传导链路。

北京交通大学信息管理理论与技术国际研究中心特聘教授张向宏提出,数据资源开发利用面临三个瓶颈:一是制度瓶颈,二是技术瓶颈,三是基础设施瓶颈。词元为破解上述三个瓶颈提供了一条新路径。

国家数据发展研究院院长胡坚波认为,词元产业分为生产端、流通端、消费端,核心在于生产与

值释放的全新路径……并在此基础上形成“场景牵引数据、数据驱动模型、模型赋能应用、应用创造价值”的良性循环。

《中国经营报》记者在采访中发现,围绕上述循环,产业探索正在转向全链条协同。即具身智能

数据用于机器人本体训练。

人工智能进入产业应用阶段后,数据供给的重点正从扩大规模转向满足特定场景和任务的需求。

上海理工大学机器智能研究院执行院长、上海卓益得机器人有限公司创始人李清都将人工智能分为物理人工智能与社会人工智能两类。他表示,当前具身智能企业积累的数据主要集中于机器人执行具体任务中,而机器人与人交互过程中产生的眼神、表情和肢体动作等社交数据大量缺失。

一家具身智能企业高管告诉记者,行业正面临四个瓶颈:一是缺乏高

记者注意到,高质量数据来自政府、企业和个人等不同主体,部分数据包含个人隐私、商业秘密等信息,难以通过直接复制和交付的方式流通。可信数据空间由此成为数据进入加工、训练和应用环节的基础设施之一。

零数科技创始人兼 CEO 林乐向记者表示,在“数据—模型—Token—应用”链条中,可信数据空间承担数据流通利用枢纽的作用,数据供需双方可依托可信数据空间进行多方接入、权限管控、合规确权 and 存证审计。“没有可信数据空

间提供的可信流通底座,原始数据就无法安全、合规地输送给数据集加工环节,后续模型训练、词元生产、业务应用就变成无源之水。”

相较于传统的数据流通链条,安全要求正在由防止数据泄露进一步扩展至保护数据使用过程。

安恒信息(688023.SH)CTO 刘博表示,可信数据空间 and 高质量数据集涉及较多隐私数据,即使不存在攻击者的恶意行为,数据持有方也可能不愿让其他主体直接看到原始数据,因此需要实现数据“可用不可见”和“可用不可得”。

值得注意的是,模型进一步进入医疗、工业等专业领域,市场关注点也将由通用 Token 的价格和供给逐步延伸至应用效果。汤寒林认为,下一阶段的竞争将转向场景 Token,其形成取决于企业能否获得合规的高价值数据,并用于模型二次开发和解决具体问题。

通用 Token 的价格、稳定性和供给能力。企业员工分散购买 Token 后,还会产生用途难以掌握、报销管理成本增加等问题,由此带来统一采购、调用留痕和集中结算等需求。

值得注意的是,模型进一步进入医疗、工业等专业领域,市场关注点也将由通用 Token 的价格和供给逐步延伸至应用效果。汤寒林认为,下一阶段的竞争将转向场景 Token,其形成取决于企业能否获得合规的高价值数据,并用于模型二次开发和解决具体问题。

数据流通

零数科技 CEO 林乐：AI 生产走向“标准化流水线”

数据宝董事长汤寒林：词元竞争走向“高场景价值”

中经记者 许璐 李晖 贵阳报道

经过长期实践发现,可信数据空间是数据流通利用的核心枢纽,为原始数据到词元产出、应用落地的全流程提供安全与规则保障。没有可信数据空间提供的可信流通底座,原始数据就无法安全、合规地进入数据集加工环节,后续的模型训练、词元生产及业务应用也就成为“无源之水”。

对此,在数据要素价值释放体系中,可信数据空间是数据流通利用的核心枢纽,数据供需双方依托可信数据空间完成多方接入、权限管控、合规确权、存证审计等。

目前,地方政府、数据集团、交易机构和科技企业都在建设可信数据空间。零数科技作为数据流通服务商,主要支撑城市

及其他新场景牵引数据供给,可信数据空间 and 高质量数据集建设提升数据开发利用效率,词元计量、计价与交易推动模型调用能力转化为可交易、可结算的服务。数据价值释放之年,供给方式、开发机制和商业模式正在同步“焕新”。

质量数据,二是难以跨场景泛化,三是决策“大脑”与动作控制“小脑”之间存在割裂,四是模型能否自主进化。其中缺乏高质量数据排在第一位。

不同应用场景需要的数据类型不同,也使数据采集和加工从相对通用的流程转向以任务为导向。

山东数据集团副总经理张帆向记者表示,同一组数据用于金融征信与用于医疗、消费分析、旅游场景,清洗方式和最终形成的数据集并不相同。因此,企业应从自身研发场景出发提出数据要求,再由数据加工成词元、供模型训练、模型反哺应用并产生新数据,形成闭环。

间提供的可信流通底座,原始数据就无法安全、合规地输送给数据集加工环节,后续模型训练、词元生产、业务应用就变成无源之水。”

相较于传统的数据流通链条,安全要求正在由防止数据泄露进一步扩展至保护数据使用过程。

安恒信息(688023.SH)CTO 刘博表示,可信数据空间 and 高质量数据集涉及较多隐私数据,即使不存在攻击者的恶意行为,数据持有方也可能不愿让其他主体直接看到原始数据,因此需要实现数据“可用不可见”和“可用不可得”。

值得注意的是,模型进一步进入医疗、工业等专业领域,市场关注点也将由通用 Token 的价格和供给逐步延伸至应用效果。汤寒林认为,下一阶段的竞争将转向场景 Token,其形成取决于企业能否获得合规的高价值数据,并用于模型二次开发和解决具体问题。

通用 Token 的价格、稳定性和供给能力。企业员工分散购买 Token 后,还会产生用途难以掌握、报销管理成本增加等问题,由此带来统一采购、调用留痕和集中结算等需求。

值得注意的是,模型进一步进入医疗、工业等专业领域,市场关注点也将由通用 Token 的价格和供给逐步延伸至应用效果。汤寒林认为,下一阶段的竞争将转向场景 Token,其形成取决于企业能否获得合规的高价值数据,并用于模型二次开发和解决具体问题。

通用 Token 的价格、稳定性和供给能力。企业员工分散购买 Token 后,还会产生用途难以掌握、报销管理成本增加等问题,由此带来统一采购、调用留痕和集中结算等需求。

值得注意的是,模型进一步进入医疗、工业等专业领域,市场关注点也将由通用 Token 的价格和供给逐步延伸至应用效果。汤寒林认为,下一阶段的竞争将转向场景 Token,其形成取决于企业能否获得合规的高价值数据,并用于模型二次开发和解决具体问题。

通用 Token 的价格、稳定性和供给能力。企业员工分散购买 Token 后,还会产生用途难以掌握、报销管理成本增加等问题,由此带来统一采购、调用留痕和集中结算等需求。

词元竞争

数据宝董事长汤寒林：词元竞争走向“高场景价值”

词元竞争走向“高场景价值”

中经记者 许璐 李晖 贵阳报道

现阶段市场竞争主要集中在通用 Token 的成本和供给能力。随着人工智能应用深入,客户需求将从“有 Token 可用”转向关注实际使用效果,具有专业数据支撑的场景 Token 服务将成为下一阶段行业竞争重点。

随着词元经济走热,整个 Token 行业至少会经历两个阶段。初始阶段主要竞争的是通用 Token,客户首先关注有没有 Token 可用,以及 Token 的价格和稳定性。因此,各类市场主体比拼的是谁能够生产出高性价比的通用 Token。

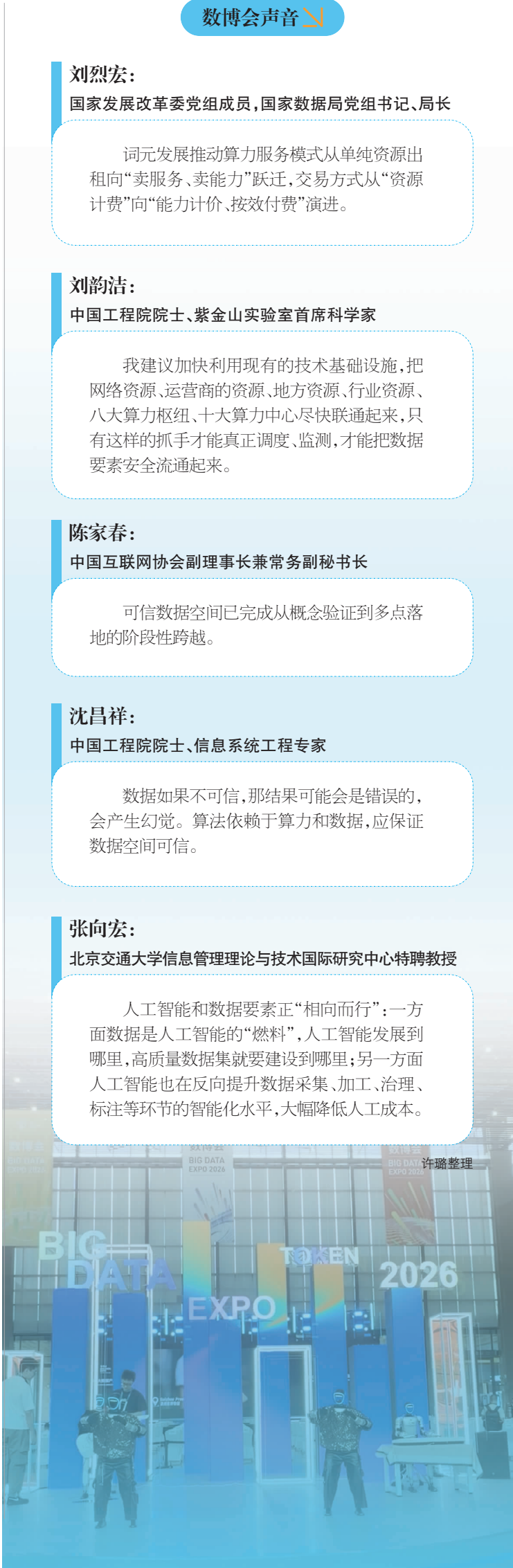
可以预见,市场很快会进入第二个阶段,即客户开始产生特定场景的 Token 需求。比如在医疗场景中,通用大模型能够回答一些医疗问题,但如果另一个

模型经过二次开发,加入了大量专业医疗知识和数据,即使其 Token 价格更高,用户也可能愿意选择。

数据宝今年从“数据要素运营商”升级为“数据原生 AI Token 服务商”,围绕战略升级,我们对需求进行分层分类,为不同需求匹配不同模型,帮助企业降低 Token 使用成本。数据宝从数据 API 走到 Token API,再延伸至算力 API,是客户需求推动我们一步步向前走。

同时,数据宝提出以“数据宝、智算宝、词元宝”覆盖数据、算力和 Token 三个环节。三项业务主要服务的是同一批 B 端客户。

数据宝现阶段主要开展 Token 服务和算力服务,我们把自身定位为“数据原生 AI Token 服务商”,而不是更宽泛的数据原生 AI 服务商。



AI 安全

安恒信息 CTO 刘博：AI 迭代快是核心安全挑战

AI 迭代快是核心安全挑战

中经记者 许璐 李晖 贵阳报道

从客户应用看,当前多数智能体仍用于个人或小规模实验场景,真正进入生产系统的应用“才刚刚开始”。随着数据依次进入模型训练、推理调用和智能体执行等环节,相关安全风险呈现新变化。

新变化分为两个方面。一方面是新环境下应对外部攻击风险的技术挑战,另一方面是隐私问题。现在以可信数据空间为代表的数

据基础设施,包括高质量数据集,涉及很多隐私相关的问题,在数据“可用不可见”“可用不可得”的要求下,正成为新的挑战。

目前,安恒信息提出打造“AI+ 动态数据安全”体系,不再依赖静态规则。AI 再加上 UEBA(用户及实体行为分析)技术,可以自动帮助我们进行数据打标、数据识别,发现敏感数据在哪里、重要数据在哪里、被哪些人使用,以及相关使用行为是

否合规。

当前,最核心的挑战在于,AI 技术本身的迭代速度远超安全防护体系的跟进周期。比如智能体,两三年前也有智能体的概念,但是当时的智能体框架和技术手段与现在差别非常大。智能体可能带来的安全风险有些是恶意操作或被投毒,有些是操作失误。还有一些任务的思维链非常长,导致智能体难以持续完成任务,出现错误操作。也有些智能体被攻击者操控,他们利用智能体进行各种恶意操作。

对于我们的挑战是,当智能体的技术框架还没有固化、沉淀下来时,要快速应对,针对新的风险快速推出新的解决方案。我们判断 2027~2029 年,智能体才会在更大规模的领域应用。我们现在需要快速迭代。在客户群体上线智能体前,要先提供一个能够解决主要问题的 AI 安全整体方案,之后再根据智能体技术和风险的变化持续迭代。

对于我们的挑战是,当智能体的技术框架还没有固化、沉淀下来时,要快速应对,针对新的风险快速推出新的解决方案。我们判断 2027~2029 年,智能体才会在更大规模的领域应用。我们现在需要快速迭代。在客户群体上线智能体前,要先提供一个能够解决主要问题的 AI 安全整体方案,之后再根据智能体技术和风险的变化持续迭代。