

小词元如何撬动 AI 大产业

近期,随着人工智能(AI)技术的快速迭代与产业化加速,词元(Token)作为大模型处理信息的基本单位,成为公众日益熟知、高度关注的热词。

我们每使用一次 AI 大模型,都要消耗一定数量的词元,一次复杂的智能体任务可能要消耗数万个词元。小小的词元,如何撬动 AI 大产业? 当前还存在哪些挑战、短板?“新华视点”记者进行了调查。

调用量两年间增长超千倍

“今天北京的天气如何?”“能帮我生成一个 AI 短剧吗?”“我要去演讲,帮我做一份 PPT”……如今,在工作生活中遇到问题,找大模型助手“AI 一下”越来越常见。

这些 AI 应用的背后,都离不开词元。深圳行胜数字技术有限公司 CEO 郝放表示,词元是信息时代和 AI 时代的分水岭,信息时代的度量单位是字节, AI 时代的度量单位是词元。

深圳云天励飞技术股份有限公司董事长兼 CEO 陈宁表示,词元是数字化的价值载体,可计量、定价、交易,是智能时代的价值锚点和连接技术供给与商业需求的结算单位。一个词元约等于 0.75 个英文单词,背后对应着真实的算力、电力和基础设施成本。

词元正串联起新的经济链条。如

上游 GPU(图形处理器)芯片、AI 服务器等为词元提供算力支持,中游云服务商、大模型厂商负责词元的工业化生产、分销与商业化运营,下游 AI 智能体、企业级应用等是词元消耗的最终场景。

在广东韶关,一个个数据中心日夜不停处理着大湾区的算力请求,有了这些算力,词元才能源源不断产生。而在下游的深圳华强北,玩具、眼镜、乐器等普通硬件产品,加载 AI 智能体、有了“新大脑”后,便成为词元的消耗载体。

来自国家数据局的数据表示,我国日均词元调用量从 2024 年初的 1000 亿跃升至 2026 年 3 月的 140 万亿以上,两年间增长超千倍。第三方机构预测,到 2030 年中国 AI 推理词元消耗量

中国日均词元调用量
两年增长超千倍



将增长超 300 倍。

中国(深圳)综合开发研究院通证数字经济研究中心主任李恩汉说,工厂、港口、跨境电商、金融、物流等高频场景有海量的词元调用需求。当把 AI 嵌入生产、运营和服务流程,把问答工具升级成真正能“跑业务”的智能体,词元就能变成实实在在的产值和税收。

今年初,以“龙虾”为代表的 AI 智能体爆火,推动词元调用量激增。不少海外用户调用中国大模型 API(应用程序编程接口),在中国完成推理运算,结果回传境外,“电力未出境,但价值跨境交付”。部分中国大模型收入爆发式增长,20 天的累计收入已超过 2025 年全年总收入。

相关阅读

词元和常见的数据有何不同?

“人工智能大模型的本质是一套复杂的数学运算系统,它无法像人类一样直接读懂文字、听懂声音、看懂画面。”中国移动研究院网络与 IT 技术研究所副所长唐华斌说,人工智能大模型必须先各类原始数据转化为自己能识别的“语言”,再通过运算输出结果,词元就是这套“语言”中最基础的单元。

具体来看,文字词元像“乐高积木”,一个单词或一个汉字被拆成一块块词元;音频词元像“乐谱上的音符”,一段声音被切成极短的时间片段,每一片包含音调、音量;视频词元像“拼图碎片”,每一帧画面被切成一个个小方格,并且还要考虑时间上的连续性……

“在大模型看来,三种模态的词元完全没有差别。在模型眼里,它不觉得自己在读文字或看视频,它只觉得自己是在处理一串极其复杂的数字序列。”火山引擎智能算法负责人吴迪表示。

一个个基础的词元,构成了智能经济运行的“细胞”。今年 3 月,中国人工智能大模型周词元调用量连续 3 周占据领先地位,成为全球大模型应用活跃度最高的国家之一。

“词元调用量爆发性增长,最直接的原因是模型能力的提升。每一次模型能力的提升,就会解锁更多应用场景,从而使大模型词元调用量进一步激增。”吴迪介绍,以 Seedance(豆包视频生成模型)2.0 为例,生成 1 分钟视频大约会消耗超过 100 万个词元。

同时,新的应用形态、新的商业模式,也会驱动大模型词元调用量大幅上涨。唐华斌表示,最近兴起的智能体和传统单轮问答不一样,它的运行往往意味着更长的上下文、更频繁的模型调用、更多轮次的任务拆解,以及工具执行过程中的持续反馈。

“简单任务,如果只需要单轮或几轮工具调用就能搞定,消耗词元只需要几千;如果需要几十轮工具调用,则需要至少消耗几万词元乃至更多。”吴迪说,智能体等新业态的涌现,让词元调用量又进入了一个快速增长的通道。

据新华社

仍难标准计量 治理还有难点

“词元是新的大宗商品。”在英伟达 2026 年度开发者大会上,英伟达创始人兼首席执行官黄仁勋说。

不少业界人士表示,虽然风口已至、发展空间广阔,但当前有关词元的计量、交易与管理,还有不少问题待解。

——仍难标准计量,“热闹多、闭环少”。腾讯研究院相关负责人表示,不同的词元价值不一样。百万个词元,如果是用来闲聊的,可能仅值 0.01 美元;如果是用来写代码的,可能值 200 美元;如果是用来做法律文档审阅的,可能值 1000 美元。这也导致一个可能的情况:不到 5% 的词元消耗,创造了超过 80% 的可测量价值。

在词元生产端,也存在类似情况。同时,在编程、视频生成、聊天等之外, AI 应用尚未大范围普及,不少企

业的收入增长建立在资本开支扩张和短期订单上,还未形成稳定、可复购的商业模式。

——规则尚不完善,不利于长期发展。陈宁认为,当前我国大模型正在服务全球用户,词元可直接跨境交付,不需要通过集装箱和海关;但现有的贸易统计框架、分类还无法准确描述这种新型贸易。同时,大量 AI 服务创造的价值,游离于国内生产总值(GDP)统计之外,也可能带来一定信息遗漏和盲区。

——安全隐患凸显,谨防新型骗局。国家安全部微信公众号发文提醒,词元在使用过程中存在一定的安全风险,包括泄露劫持风险、伪造篡改风险、诈骗陷阱风险等。不法分子可通过跨站脚本攻击(XSS)、公共 Wi-Fi 嗅探等方式,窃取、截获未加密的令牌。

同时,不法分子还有可能制造“虚假令牌”,诱导用户泄露身份证号、手机号等隐私信息。

“如何进一步提升词元流通的安全,这是当前亟待解决的一个问题。”郝放说。

在一些社交平台 and 私域社群里,词元被包装成新一代“暴富密码”。“现在囤词元,就像早年囤比特币、囤域名,是通往财富自由的新风口”“Token 就是数字石油! 现在不囤,以后稀缺了哭都来不及”“我有内部渠道,Token 价格打三折! 囤货转手就赚”等话术层出不穷。

深圳市人工智能行业协会执行会长郎丽艳表示,词元只是一个度量单位,真正有价值的是算力和高质量数据;要防止有人浑水摸鱼,以词元为噱头制造新型骗局。

完善治理机制 提高安全防范意识

在不少业界人士看来,要持续健全有关词元的标准、市场、公共治理等机制,促进相关产业高质量发展。

陈宁建议,加强实时算力供给与应用需求匹配能力,实现算力资源的精准匹配、动态调度与高效利用,构建与之相适应的统计制度,助力词元成为数字经济中的通用计价单位,贯穿 AI 服务、数据交易与算力租赁全场景。

合肥零次方机器人有限公司联合创始人李宜哲表示,具身智能是 AI 智能体的物理载体,可将数字世界的词元价值延伸至物理场景,实现“智能决

策+物理执行”一体化。要尽快确定标准,打通词元在人、机、物三方融合中的通用价值。

要持续推动技术突破,提升经济性、安全性。上海交通大学副教授、无问芯穹联合创始人戴国浩建议,研究打造敏感数据端侧脱敏、高智能任务上云执行的端云一体架构,通过推理优化有望降低云端成本,支撑安全、低成本、规模化的词元应用。

不少业界人士认为,要研究词元应用的能耗与碳治理,推广绿色算力认证,重视环境成本,并持续推动构建

数字经济全球治理新秩序。

郎丽艳表示,要建立可信、高效、公平的价值度量与分配体系,才能让数据、算力、算法、智能体充分释放价值。随着词元机制不断完善, AI 将全面融入产业流程,重塑生产、分配、交换、消费全链条。

有关部门提醒,面对词元热潮,我们既要理性看待其价值,又要注意信息安全、隐私安全,提高安全防范意识,做到了解、善用。如遭遇诈骗、信息泄露或发现非法活动,应及时向有关部门反映。新华社深圳 5 月 13 日电