

# 以“开放计算Token谱系”破解Token焦虑

► 本报记者 张伟

当前,词元(Token)已成为大模型训练、推理及应用落地的核心生产要素。然而,算力供给碎片化、生态协同不足、应用转化周期长等问题日益凸显。如何针对性精准破题,成为各界关注的焦点。

5月28-31日,2026世界智能产业博览会(以下简称“智博会”)在天津市举办,依托国产通用计算硬件平台,由海光信息技术股份有限公司(以下简称“海光信息”)牵头成立的产业生态系统联合体——海光产业生态合作组织(以下简称“光合组织”)携生态伙伴集中亮相,系统性地展示了“开放计算Token谱系”全链赋能Token效益提升的实践成果。

“破解Token焦虑的关键在于,围绕产值转化构建开放协同体系,系统性提升全链路效率。”多位产业界人士表示。

## Token焦虑未必是坏事

何为Token焦虑?

AI(人工智能)时代,算力是核心引擎,Token是信息处理的最小运算单元,两者都是智能化生产线上的“硬通货”。然而,近年来算力新基建持续提速,Token调用量爆发式增长,真实业务场景中的价值产出却未能有效同步,业界开始从“算力不足恐惧症”转向“Token焦虑”。

面对算力投入持续加码,而Token价值难以同步提升的行业焦虑,海光信息总裁助理兼智能计算产品部总经理杜夏威认为,这未必是一件坏事。

“大家都在焦虑,说明产业要素正在快速流动。”杜夏威认为,这种焦虑可以从本质上解决。一方面,做好国产算力生态和模型生态的融合,解决算力供给的焦虑;另一方面,解决Token与传统生产型业务的深度融合,解决如何发挥Token效能的焦虑。

“Token的高效产出,始于稳定且灵活的算力供给。”杜夏威表示,

破解算力焦虑的核心在于,从算力底层解决算力高质量供应问题,用更高性能、更高安全、更高兼容的底座,托举Token要素的高价值。

“针对这一点,海光信息在过往已储备了较为完善的软件栈和生态,为上层模型算法提供支持。”杜夏威介绍说,展区现场,海光信息一字排开的CPU+DCU“双芯”底座系列产品,就能为Token全场景落地提供坚实底座。

## 开放计算Token谱系破题

2026世界智能产业博览会期间,众智FlagOS·光合组织生态联合创新启动仪式举行。18家产学研机构共同见证联合打造“算力芯片+大模型+产业生态”全链条创新平台,围绕Agent(智能体)推理、多芯调度等方向深化协同。

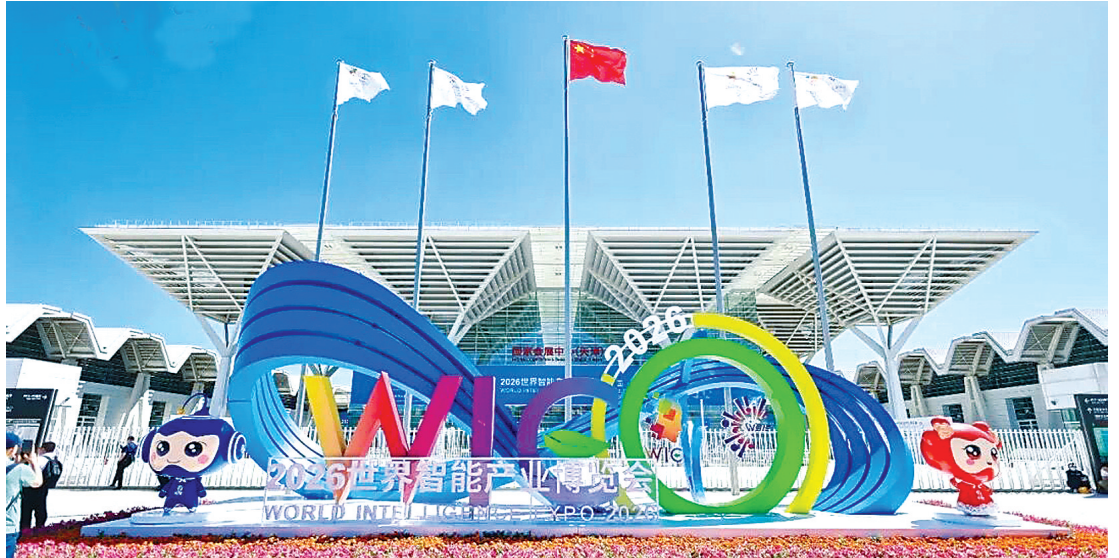
在展区现场,记者看到,光合组织联合芯片、部件、整机、平台及应用厂商全生态矩阵亮相,呈现开放计算Token谱系的三层架构,诠释提升Token效益的生态协同逻辑。

这三层架构分别为硬件层、软件层、应用层。在硬件层面,以国产芯片为底座,覆盖关键部件、服务器、超节点、超集群及终端设备,筑牢Token高效生产根基。在软件层面,整合操作系统、数据库、AI工具栈等,实现Token灵活调度与跨平台流转。在应用层面,涵盖算力运营、行业智能体、大模型等场景,形成“生产—调度—转化”闭环。

据展区工作人员介绍,目前,光合组织已联动超6000家生态伙伴,构建起免重构、低成本、广兼容的应用生态,并在本届智博会现场呈现材料计算、生物医药、气象仿真、智能制造、科学大模型等多个前沿领域落地成果。

## 接得上、跑得稳、算得快

开放计算Token谱系,带给人



上图:2026世界智能产业博览会主场馆



下图:光合组织联合芯片、部件、整机、平台及应用厂商全生态矩阵,亮相博览会

主办方供图

态伙伴最直观的感受之一是“方案落地的周期明显缩短了”,无需再为底层适配耗费大量精力。

“以前,我们做国产化方案适配,光底层调通就需要2-3周时间。而光合组织的开放计算Token谱系,把多芯调度、操作系统和上层应用拉通成为一条流水线。现在,我们从芯片到业务上线,已经可以压缩到一周以内。”湖南麒麟信安售前方案专家孙玉峰表示,Token从产生到交付,在路上“堵”的时间减少了,效益自然就体现出来了。

“以前做国产化替代,光是适配就能拖一两个月。加入光合组织之后,我们与芯片、整机、操作系统厂商在一个谱系里协同,从方案设计到安装测试,交

付周期大幅度缩短,整个国产化链条顺了,终端场景更容易跑出Token价值。”南大通用华北区总经理陈明智用3个词形容光合生态的协同感:接得上、跑得稳、算得快。

“在行业落地过程中,我们会面对大量数据。基于这些数据做应用,一次提问可能就要消耗上百万Token,这是现在面临的一个较大难题。”北京旷视科技有限公司智能体产品线负责人常豪鑫表示,加入光合组织后,该公司在数据应用方面的资源利用更加高效。“我们更希望把精力集中在面向客户的具体场景和智能体设计上。对于底层的Token性能提升则会与光合组织实现更多合作。”

“通过开放、协同的方式,把国产化软硬件全栈打通,形成一个完整的谱系,让每一最小运算单元Token从生产到价值变现的路径更短、损耗更小、价值更高。这是提升Token效益不可或缺的一环。”云宏信息科技股份有限公司华北区总经理葛玉龙认为,光合组织的开放计算Token谱系正在从好用走向高效,从高效走向划算。

“这个方向才是Token经济算力发展的主赛道。”葛玉龙说。

平凯星辰(北京)科技有限公司资深解决方案架构师毋文涛认为,国产开放算力生态未来的发展方向一定是全链路协同,光合组织开放算力最大的价值,是让数据库算力和AI负载形成更稳定的协同关系。他指出,“光合组织正在打通从芯片到应用的完整链路,数据库在其中扮演AI数据中枢角色。”

## 长安链攻克同态加密计算加速芯片关键技术

本报讯(记者 张伟)近日从国家区块链技术创新中心获悉,我国自主创新区块链技术体系长安链,在高性能隐私计算硬件领域取得重大突破:长安链团队成功研发同态加密计算加速芯片,支撑数据流通全周期高效、密态处理,效率较国内外现有标杆方案提升10倍以上。该芯片可广泛应用于端云协同计算,跨机构、跨领域数据协作等一系列关键场景。

据了解,同态加密被誉为隐私计算“皇冠上的明珠”,是一种特殊的密码算法,可实现加密数据在不解密的情况下完成计算,最终结果在解密后与明文

计算结果完全一样,可保证数据在流通、共享和计算全过程以密文开展,即使被截获也不会泄露原文,其核心价值是实现“可用不可见”,但其天然存在算力短板,密文运算速度仅为明文万分之一,高维大数据场景几乎无法落地。

长安链团队将同态加密技术硬件化,研发了同态向量运算加速芯片。该芯片从存储器中读取数据的同时即完成数据计算,利用这种“存内计算”模式实现了高维同态加密数据的“即产

即用”,大幅度降低了同态加密运算过程中海量数据读写导致的“存储墙”压力,计算效率达到了国际先进水平。

在硬件技术突破的基础上,长安链团队创新提出分布式同态数据查询计算框架,构建复杂查询语句的解析和运算架构,进一步提升密态数据的查询处理能力。

据介绍,这项成果在金融等数据流量大、安全性要求极高的场景有巨大的应用价值。当前,国内多家银行在普惠金融领

域积极探索,通过应用长安链分布式同态数据查询计算框架支持中小微企业融资授信。相关银行在数据加密状态下,就可以完成对中小微企业的信用评级,彻底规避传统模式下需解密数据分析的风险,既能充分保护各方隐私,又能精准给中小微企业“画像”,帮助企业及时获得资金支持。

长安链是我国自主可控的区块链软硬件技术体系,包括区块链底层操作系统和区块链专用芯片,性能实现了全球领先。

其中,芯片研发方面长安链96核区块链专用芯片将区块链交易性能提升了50倍,突破了超大规模区块链应用的算力瓶颈,让我国可信数字基础设施拥有了“中国芯”;底层操作系统方面,长安链创新采用了动态自适应、可适配的新型区块链架构,能够自动适配、精准构建满足复杂场景的区块链系统。国家区块链技术创新中心成立以来,长安链持续支撑贸易、金融等领域一大批国家重大数字工程建设,连续4年位居国内区块链市场占有率第一。