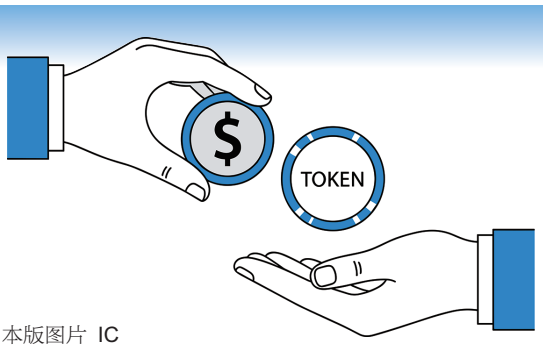
你向AI提问,背后都在“烧Token”



本版图片 IC

话题主持:本报记者 **郜阳 易蓉**

近来,AI在具体场景中落地的消息层出不穷;与此同时,开源AI智能体项目OpenClaw继续席卷全球,帮助不同行业的从业者实现了更复杂的Agent功能。

然而,越来越多的人开始发问:“做到这些究竟烧掉了多少Token(词元)?”

效率的跃升并非毫无代价——作为大模型处理信息的最小单位,AI在具体场景中施展拳脚,离不开海量Token的消耗。

与此同时,企业界也开始重新审视Token的稀缺性与战略价值。英伟达CEO黄仁勋近期提出了“Token经济学”的概念,他认为数据中心正在演变为生产Token的“AI工厂”,Token需要根据相应的标准进行分层定价供给。同时,Token将成为继工资、奖金、期权之后的“第四种薪酬”。

需求正在以指数级增长,
然而资源有限,专家提醒:

聪明用Token才是AI核心竞争力

Token,这个曾经位于技术文档底部的计量单位,正在跃升为AI时代的新货币。无问芯穹联合创始人兼CEO在不久前举行的2026中关村论坛年会上给出了一个惊人的数据:“从今年1月底开始,无问芯穹平台上的Token调用量每两周翻一倍,至今已经增长了十倍。”

他将其形容为“当年3G手机流量爆发的那种感觉”。

Token突然变成硬通货

要理解Token为何突然变得如此重要,首先需要回答一个问题:Token到底是什么?

通俗来说,Token是大语言模型处理信息的最小单位。一段文字被模型“理解”之前,需要被切分成无数个Token——一个英文单词可能对应一个或多个Token,一个汉字也可能被拆解。每一次你与大模型对话,每一次你用AI写代码、做分析、画图表,背后都是海量Token在“燃烧”。

但Token的意义远不止于此。上海交通大学副教授、无问芯穹联合创始人戴国浩提出了一个清晰的价值转化链条:输入价值→电能→算力→Token→生产力→输出价值。

在这个链条中,Token是连接算力与生产力的关键枢纽。“随着模型能力的提升,‘Token到生产力’的转化效率显著提高。”戴国浩指出,“每一个Token所能承载和释放的价值变得更高”。

这解释了为什么企业愿意为高质量Token支付更高的价格。当模型从“聊天机器人”进化为“能干活的生产力工具”时,Token就不再是单纯的成本项,而是直接对应产出的投资。

成本下降和价格上涨

然而,一个看似矛盾的现象正在发生:一方面,技术进步推动单位Token的计算成本持续下降;另一方面,不少云厂商却在上调API调用价格。

戴国浩认为,这并不矛盾,“在模型尚不可用的阶段,企业需要通过低价甚至补贴来推动使用。但当模型能力提升、后端价值显著提高后,整个逻辑会从‘推广驱动’转向‘市场化驱动’。”

换句话说,过去是“烧钱培养习惯”,现在是“按价值定价”。一个Token能完成的任务越复杂,它就越值钱。

无问芯穹联合创始人兼首席执行官夏立雪则从基础设施层面给出了更现实的解释:需求正在以指数级增长,而资源是有限的,“我们现在所有能够用到的资源,想要支撑起这样一个快速增长的时代是不够的。”

他将当前Token用量的爆发,比作当年手机流量从每月100兆到无限流量的跨越。而这个跨越,对底层算力、带宽、存储、调度提出了前所未有的挑战。

从“为人设计”到“为Agent设计”

无问芯穹的核心业务,正是解决这一矛盾。夏立雪指出,当前的云计算基础设施本质上是为人类工程师设计的——人类发起一个任务是分钟级别的,而Agent(智能体)的思考和响应是毫秒级别的。这套架构,正在限制Token效率的进一步释放。

“面向Agent时代,我们认为这是不够的。”夏立雪提出了“Agentic Infra”的概念——基础设施本身也应该是一个智能体,能够自我进化、自我迭代,甚至像一个CEO一样管理整个系统。

他透露,无问芯穹已经接入了几乎所有能看到的计算芯片,将国内十几种芯片和几十个不同的算力集群统一连接起来。“当资源不足的时候,最好的办法是:第一,把能用的资源都用起来;第二,让每一个算力都用在刀刃上。”

更进一步,夏立雪提出了一个更具野心的愿景——“可持续Token”。他希望将中国的能源优势和制造业成本优势,通过高效的Token工厂,转化为优质的Token输出到全球。

“我们不仅要盘活国内算力资源,更要打造中国特色Token经济学,复刻Made in China的优势,把我国的能源、成本优势,通过高效Token工厂,转化为高质量AI服务。”夏立雪表示。

“Token观”:不是越多越好

面对Token经济的热潮,戴国浩也给出了冷静的提醒。他并不完全认同“Token消耗量直接等于生产效率”的观点,“Token到生产力的转化效率,取决于模型能力、Prompt质量、任务类型等多个因素。高质量地使用Token,比单纯增加Token消耗更重要。”

他建议,更合理的评价方式应该是将“Token使用量”与“最终产出价值”进行联合评估,“我们更应该鼓励的是:用更少的Token产生更高的价值。”

这意味着,未来的核心竞争力,可能不再是“谁能调用更多Token”,而是“谁能更聪明地使用Token”。从“编码者”到“系统架构师”的能力跃迁,将成为每个AI时代从业者的必修课。

本报记者 **郜阳**

不是『新货币』,却是关键生产要素

塑产业、工作与社会格局 复旦专家解读Token如何重

2026年,AI领域的热词一定绕不开“Token”(词元)。从互联网巨头到每一个AI产品的用户,一个以Token为核心的新经济形态正显现在大家眼前。从社会经济的视角来看,Token究竟是什么?作为成为AI时代最基础的单元,它如何影响物理世界的经济形态?未来,Token经济将带来怎样的格局?复旦管院科创管理研究中心首席经济学家,复旦大学EMBA项目、科创企业家营授课教授,国家金融与发展实验室特聘高级研究员邵宇为大家解读。

问 如何理解Token?它在AI时代扮演什么角色?

答:Token是人工智能产出答案的最小颗粒。无论是生成一段文本、制作一个PPT,还是规划一次自动驾驶出行,背后都由成千上万甚至上百万个Token构成。

Token的核心意义在于,它成为了衡量AI工作量和价值的“公约数”。在过去,完成一项任务需要投入时间、金钱、人力等不同资源;而在AI时代,这些都可以被统一折算为Token消耗量。复杂任务需要更多Token,简单任务则消耗较少。因此,Token不仅是技术单位,更可能成为未来经济中衡量投入与产出的基本计量单位。

问 Token是如何产生的?它的成本由什么决定?

答:Token的生产依赖算力(硬件)、算法(软件)和数据三大核心要素,而这些要素都需要电力驱动。简单来说,Token是通过AI基础设施(如芯片、服务器、算法模型)将电力转化为智能输出的结果。其成本主要受到电力价格、技术效率、商业模式的影响。例如,中国电价较低,Token生产更具成本优势;中国许多AI服务采取免费或低价策略,进一步降低了Token的使用门槛。目前,中国的Token调用量已超过美国,主要原因正是其高性价比——既便宜,又在质量上逐步追赶。

问 Token会成为未来的“货币”吗?

答:尽管有观点认为Token可能像“数字世界的电力”一样成为新货币,但从经济学角度看,这仍属于“科技乌托邦”的设想。

因为货币需具备稳定性,而Token的价值波动剧烈(例如一年可能贬值90%),无法满足支付和计价所需的稳定性;缺乏储值功能,货币需要能长期保存价值,而Token的快速贬值使其难以胜任;主权与边界存在问题,Token由不同公司或国家提供(如英伟达、中国AI工厂),带有“科技主权”属性,难以成为全球统一的一般等价物。

历史上,货币从贝壳、贵金属演进到法定货币,核心在于共识和稳定性。Token更可能成为一种关键生产要素,而非货币本身。

问 尽管Token还不会取代货币,但它对经济的渗透已经真切发生,它会如何改变产业和工作模式?会带来何种挑战?

答:Token已成为生产要素,像水电煤一样进入日常账单。白领工作写PPT、写报告、写代码都可以外包给AI,用户支付的是Token的消耗。这份账单能不能覆盖,取决于你的产出值不值这个钱。

Token开始进入薪酬体系,带来成本结构变化。硅谷已有科技公司在新人薪酬包中附带每月Token额度,供员工在工作中使用。相当于以前发电脑、发文具,现在发Token,本质是把AI基础设施配给到个人。企业成本将新增“Token支出”,尤其对知识密集型行业(如编程、设计),Token消耗可能直接计入项目预算。

工作方式也发生了革新——企业通过“削峰填谷”(如凌晨调用算力)降低Token成本。但需注意,不合理的工作安排(如强制员工熬夜)反映的是管理问题,而非技术必然。

Token驱动的AI革命可能加剧两大社会矛盾:就业替代与收入分化,以及公平与价值观冲突。高技能岗位(如程序员)可能转向“人机协作”,但低技能智力劳动面临被替代风险;财富向AI基础设施所有者集中(如英伟达市值超多数国家GDP),普通人若失去劳动价值,可能陷入“无消费能力”困境;“全民基本收入”(UBI)等方案试图缓解分化,但可能违背“按劳分配”原则,削弱人的社会价值感。

中国政府强调“共同富裕”,可能通过政策引导避免过度替代,但实际执行需平衡创新与稳定。

问 未来的图景会是什么样?

答:短期内,Token将呈现动态供需平衡。

一方面,因技术进步和产能扩张(如全球AI投资激增),Token边际成本降低,刺激更多应用,价格将持续下降。另一方面,开源/免费Token适合日常需求,高精度任务仍需付费高质量Token,将带来质量分层。

长期看,若AI突破至“创新级”(如诞生AI爱因斯坦或牛顿),Token可能重构人类文明。但即使到那个时候,能源、伦理和社会结构仍是关键制约。

本报记者 **易蓉**