

跳出流量内卷 运营商竞速Token赛道



套餐纷纷落地 算力消费告别大厂专属

南京棋盘石智能科技有限公司创始人王磊不久前订阅了电信智云套餐。公司部署混合算力中心,接入了阿里千问3.7、豆包等多款主流模型,每月花费约1万元。他估算,套餐使用后每月可节省约20%的AI算力开支。作为算力普惠的直接受益者,他的感受正是当前国内算力消费市场的真实缩影。

Token作为大模型处理信息的最小单元,正在成为智能时代可计量、可定价、可交易的全新数字资产。今年以来,三大运营商集中推出面向个人和企业的Token套餐,彻底打破了过去AI算力仅由头部大厂独享的局面,让算力像水电一样实现按需取用。

北京商报记者调查发现,目前三大运营商的Token套餐已经形成差异化落地路径。5月17日,中国电信推出试商用词元套餐,提供面向个人和开发者的多档选择,套餐最低价格为9.9元每月1000万词元;5月16日,中国联通上海分公司宣布向本地OPC(一人公司)用户提供词元服务,每位用户可免费领取3000万词元测试额度;4月份,中国移动在北京、广东、江苏等地推出词元套餐……

在北京区域,4月21日,北京移动正式推出面向个人用户的算力Token套餐。据北京

移动相关负责人王女士对北京商报记者介绍:“对于已有云电脑的客户,北京移动提供最低5.99元的算力次包,也有24.99元的算力月包,每个月可以让用户使用1000万Token的算力。对于没有云电脑的客户,我们推出了内置OpenClaw的云电脑融合套餐,无需自行部署环境,开机即用。”记者实测发现,北京移动Token套餐可自由切换多款大模型,告别多平台分开充值的繁琐。

中航证券首席市场策略分析师张郁峰表示,运营商Token套餐的推出,有助于降低中小商家使用AI的门槛,而且将对算力产业链产生实质性的催化。它将降低下游AI应用层开发成本,催生大量的轻量化AI应用,形成需求带动基建、基建支撑应用的正向循环。

战略赛道转向 运营商跳出流量增长瓶颈

三大运营商集体发力Token赛道,背后是传统通信业务增长见顶的现实压力。2025年年报数据显示,三大电信运营商营收虽均保持正增长,但增速均已降至1%以下。在传统业务增势趋缓的背景下,智算服务已经成为运营商全新的增长曲线。

“过去运营商卖的语音、短信、流量等业务基本饱和。卖Token,是运营商在AI时代选

当企业订阅算力套餐后,每月AI开支能降低20%,智能时代的全新数字资产正加速走进现实。北京商报记者日前调查发现,三大运营商正密集落地差异化Token服务,从企业侧到家庭端,一场算力普惠的产业变革正在悄然发生,推动通信行业从“经营流量”向“经营智能”完成跨越式跃迁。

的一条新路。”产业观察家许意强说。

中国联合网络通信集团有限公司副总经理邱宝华在2026(第二十五届)中国互联网大会上指出,“电力革命催生了度电,信息革命催生了流量,今天AI革命正在催生Token,作为智能时代的通用计量单位,以Token运营为核心的新的商业范式正在逐步形成”。

他进一步表示,智能时代的Token经营逻辑正在发生根本性的转变:一是流量增长已趋于饱和,2025年底我国网民日均上网时长约为4.6小时,而Token可以7×24小时持续调度;二是价值密度差距显著,处理1亿Token对比传输1GB的流量承载的智能服务的价值增长了数十倍乃至数百倍,这意味着运营商的运营模式需要从经营通道向经营智能进行跃升。

“个人与家庭产品焕新是运营商实现主营业务增长的关键突破口。”华为运营商业务总裁杨扬在2026年上海世界移动通信大会现场表示,依托庞大的用户基本盘,运营商将AI深度融入主营业务,能够有效提升用户ARPU,探索出收入增长的新路径。在家庭场景,AI赋能多模态交互与跨生态设备联动,提供个性化、自动化智慧家居服务;在个人场景,传统通话上网业务升级为个人通话助理与随身助理,依托运营商多人口优势,提供无处不在的贴身智能服务。

场景边界延伸 算力网络触达家庭客厅

很多普通用户的第一反应是:用大模型没花过钱啊,为什么要买Token套餐呢?工信部信息通信经济专家委员会委员盘和林认为,普通用户一般用量小,网页版、免费版就够了。但企业需要调用API,将AI嵌入到自己的产品和业务中,从而会产生Token费用。三大运营商更多的是面向具有Token安全诉求的企业,尤其是政府端和央企国企用户。

此外,随着赛道逐步成熟,Token的应用边界正在从企业侧快速向个人与家庭侧延伸。像北京移动于近日正式发布“一站式Token运营服务平台”,围绕Token的生产、运营到服务,构建起全链路闭环,为各类用户提供普惠、易用、安全、可信的一站式服务。

据北京移动相关负责人介绍,所谓“一站式”,在于平台依托北京移动算力基础设施,整合多元算力和模型资源,像一个大而全的“模型超市”。用户无需费时费力自建环境和自行适配镜像,只需一次接入即可调用平台上的全部模型资源。

但痛点背后,终端侧的创新探索已经在悄然发生。日前四川长虹在投资者互动平台公开回复,旗下长虹新网科技正和三大运营商深度合作,为其提供云盒、智能摄像头、云电脑、智能音箱等家庭终端硬件,同时承担配套技术运维服务。

“我们的云盒、AI算力终端等产品现在已经不是传统的终端硬件,它是算力网络伸向家庭的一个边缘节点。”长虹相关负责人在接受北京商报记者采访时透露,他们在2025年8月申请的“基于设备算力的边缘智能预处理系统及方法”专利,既能大幅降低云端算力消耗,又能让AI响应速度提升数倍。

“算力普惠从来不是把云端的价格打下来就完事了,而是让算力真正下沉到每一个家庭。”该负责人给北京商报记者演示了一个场景:普通家庭用户使用家里的云盒等产品,无需购买昂贵的高端显卡,不用复杂的环境部署,就能直接通过运营商提供的Token算力,让老人和孩子也能一键生成AI绘画、完成家庭影像智能剪辑等。

行业协同破局 标准统一筑牢长期发展

“我就简单问了两个日常问题,系统直接扣掉14万Token。”一位北京的普通用户向北京商报记者反映了自己使用Token套餐时遇到的体验问题。这并非个例,在算力普惠快速推进的过程中,终端侧的体验痛点正在逐步显现。

北京商报记者在调查中发现,赛道的快速扩张也伴随着诸多现实痛点。首先是用户认知门槛尚未完全打破,多数普通用户对“100万Token能完成多少次AI对话、生成多少张图片”没有清晰概念,远不如对1GB流量能刷多久视频的认知直观,这直接制约了C端市场的规模化渗透。其次是行业计量标准尚未统一,不同厂商对Token的定义、计费规则、权益范围存在差异,输入Token和输出Token分开计费、不同模型的百万Token价格差可达数倍等乱象,让中小企业难以精准控制算力成本。

“现在很多平台单纯比Token单价,完全不看背后的模型能力和业务场景,绘图、短视频生成的Token消耗量从来没有明确标注,行业没有统一的计量标准,中小企业根本没法精准核算算力成本。”许意强在接受采访时直指核心问题。

面对这些挑战,三大运营商已经开始协同探索破局路径。中国移动通过集约化运营实现单位Token成本下降约30%,资源占用率降低50%以上,依托统一Token量纲体系,让不同模型的Token消耗封装为标准积分,大幅降低跨模型使用门槛;中国电信将Token与连接、安全、云和大模型服务深度组合,打造“连接+算力+智能”的综合服务体系,面向办公场景推出星辰超级智能体,让企业用户实现开箱即用;中国联通构建“三个一”集约化Token运营体系,依托全国算力大网和统一调度平台,重点面向交通、物流、金融等14个行业输出高价值场景智能体,覆盖近6万家企业用户。

多位业内人士一致认为,Token赛道的长期健康发展,必须建立可持续的盈利模式,才能支撑算力网络的持续迭代升级。杨扬也特别提到,运营商要抓住AI技术爆发的战略契机,用AI焕新自身产品体系,靠升级后的算网基础设施,为Token的高效流通筑牢底座,开辟通信产业前所未有的增长新空间。

北京商报记者 卓洋

透视人形机器人运动会:竞技台前,供应链幕后

8月22日至26日,第二届世界人形机器人运动会将在北京的国家速滑馆“冰丝带”举行。六大洲16个国家的666支队伍、2056台机器人将同台竞技,队伍数量较首届增长138%,机器人数量翻了两番。参赛阵容中,既有美国、德国、日本等机器人传统强国的队伍,也涵盖了国内157家企业、200所院校和科研机构。

据大会官方介绍,作为全球首个以人形机器人为竞赛主体的国际综合性科技体育赛事,本届竞赛项目分为竞技赛、场景赛与表演类。竞技赛在田径、舞蹈、武术、体操等基础上,新增跳远、举重、拔河、乒乓球等高强度对抗项目,考验机器人脑、小脑、本体结构及核心零部件的极限工况;场景赛延伸至工业、酒店、家庭、物流等真实环境,设置装配上料、家政服务、消防救援、图书整理等赛项,瞄准“乏、脏、险、难”场景,还专门开辟灵巧手赛项,以粉末称量、镊子夹豆、开瓶撬盖等任务,比拼“看得清、拿得稳、做得准”的精细操作。

新的角逐开幕在即,赛场上的冲刺、跳跃与对抗又将刷新公众对机器人能力的想象。通用具身智能机器人公司无界动力创始人张玉峰向北京商报记者表示,与过去更多围绕运动本身相比,这一次的赛事进一步丰富了各种实际场景中的操作比拼,“不仅让机器人从研发走向落地,也给了产业一个非常好的评测和竞技平台”。他还提到,本届赛事吸引了海内外更多科研机构和创业企业加入。

历经两年沉淀,具身智能从“实体”快步走向“实业”。与此同时,一个此前被忽视的维

度也逐渐凸显:本体的竞技表现固然吸睛,但支撑其跑跳、操作与持续运行的供应链,其价值正从幕后推至台前。

隐形参赛者

如果把人形机器人比作运动员,关节模组就是它的肌肉与骨骼。跑得更快、跳得多高、能持续运动多久,赛场直观外显的指标,底层都由零部件与模型算法配合而出的性能决定。

作为关节模组供应商,卓誉科技是这场运动会背后的“隐形参赛者”。从电机起步,研发团队逐步覆盖驱动器、减速器、编码器,最终形成整套关节模组的自研能力。据称,在人形机器人行业,能够同时做这四样核心零部件的企业并不多见,公司客户群体已经覆盖了超过半数的头部具身智能公司。

“我们不造机器人,但我们是机器人相当核心的那一块。”卓誉科技董秘谭鹏这样向北京商报记者解释。在他看来,关节模组的性能直接映射到赛场表现——同样体积下输出的力矩越大、关节越轻,机器人就能跑得更快、跳得更高、续航更长;运行时温度控制得好,机器人才有可能在未来走进家庭而不烫伤人。

性能之后,寿命正在成为新的痛点。此前很少有人关注关节模组能用多久——1000小时、5000小时还是1万小时?但一台机器人如果工作1000小时就要换零件,“到底是人服务机器人,还是机器人服务人,就本末倒置了”。据悉,卓誉科技为此建了加速老化实验室,从减速器

的油脂配方到电机的温升控制,逐项打磨。公司今年推出的新一代力矩电机,运行温升比上一代低了约20℃,输出能力提升了15%到30%。

这些变化,很可能在运动会的赛场上有所体现。更轻的关节意味着更高的跳跃和更快的奔跑,更低的温升意味着更长的竞技时间,更长的寿命则意味着更多机器人从赛场走向工厂、家庭的可能性。

如何训练机器人运动员

有了强健的肌肉,机器人还需要学会如何运动,这一片是数据服务商的战场。

诺亦腾是另一家出现在运动会背后的供应链企业。与前述硬件供应商不同,诺亦腾依托十余年人体运动捕捉技术积累,为机器人提供运动数据。公司副总裁、合伙人李遥向北京商报记者介绍,诺亦腾从十多年前就专注于把人的动作数字化、数据化,这种积累在具身智能时代找到了新的应用场景。

在机器人训练中,数据分为多种类型。其一是真机数据——人穿着动作捕捉设备直接驱动机器人本体,采集其运动信息,没有本体差异,训练效果最好,但采集效率低、成本高,多用于模型训练的最后调优;另一类是“以人为本”采集的人类运动数据,虽然适配不同本体时需要做一道翻译、会损失部分信息,但可迁移、可复用,适合基础模型阶段的大规模训练。

“以人为本的数据采集到了人类既有的经验,且可复用性强,同一批数据可以用到各

种不同的机器人上。”李遥说。

而对于另一种仿真数据,李遥打了个比方:就像让机器人看费德勒打网球的视频,看10万个小时也学不会,因为视频里缺失了大量物理信息——握拍用多大力、来球时怎么卸力,这些只有真人真正操作时才能采集到。

北京商报记者了解到,在备赛周期内,诺亦腾接到了针对性的数据需求,跑步、球类竞技等运动会项目背后,都有数据服务商的身影。“现在机器人的跑步能力已经有了长足进步,但如果能有人类更多不同跑姿的数据喂给机器人的模型训练,它的协调性一定会更好”,李遥说。

但整个行业的数据积累还远远不够。李遥判断,目前行业积累的数据量距离真正需求还差两个数量级,高质量数据集的质量也“远远不够”。这也解释了运动会的价值不止于竞技——每一次奔跑、每一次抓取、每一次跌倒后的调整,都可能为企业,也为数据供应链积累素材。

更较真的赛场

如果说上届运动会的看点还在于热闹、在于看机器人“能跑能跳”,那么本届赛事的一系列规则调整,似乎正在把对赛队也是对行业的评判标准,从“能否”推向“优劣”。

北京商报记者此前从赛会了解到,本届比赛首先是赛程有所拉长。正赛从首届的3天6个单元扩展到5天9个单元,比赛时间调

整为下午和晚上。与之对应的是赛制的全面优化——取消外围赛,武术、体育舞蹈等项目纳入竞技赛;足球细化为3v3 U19、5v5大型组和5v5中型组三个组别;场景赛从6项增至21项,覆盖九大真实场景,机器人也从首届的单兵作战升级为双人甚至四人同台。

比赛程更值得关注的是规则的收紧。田径项目中,100米完赛时间从3分钟压缩到1分钟,400米从15分钟压到5分钟,1500米从40分钟压到15分钟,留给机器人的时间只有上届的三分之一。

对机器人独立自主的要求也在提高——除两项障碍赛外,所有竞技赛项均须机器人全自主完成;场景赛中全自主完成的分数权重为1,遥控完成仅为0.5,靠人在背后全程遥控的空间被大幅压缩。体操、武术、体育舞蹈等项目则加大了难度系数,障碍赛从一条赛道变为两条同场竞技。

规则升级的底气来自机器人能力的提升。今年4月的人形机器人半程马拉松上,冠军跑出50分26秒,已超越人类半马世界纪录,本届田径各赛项纪录有望全面刷新。足球赛场上,机器人已完成带球跑动、大力抽射、飞身扑救;自由搏击贯穿8个竞赛单元,三个级别竞技水平较去年大幅提升;体育舞蹈参赛队伍从29支增至108支。

如今,伴随又一届人形机器人运动会启幕,具身智能全链也进一步向落地靠拢。竞技背后,是产业技术跨越后迎接应用拐点。

北京商报记者 王天逸