

# 算力租赁北京有望打5折

## 市场价格的5、6折左右

就北京筹划出台“算力券”政策,北京市经信局相关负责人在接受北京商报记者采访时介绍,目前“算力券”政策正在研究中,具体细节还未敲定。不过,可以明确的是,“算力券”中算力的供应方来自北京经信局正在征集的“北京市通用人工智能产业创新伙伴计划”成员。

北京商报记者在北京市经信局官网查询发现,7月3日,北京市经信局已经发布了第二批伙伴名单,63家伙伴企业中,算力伙伴10家,分别为北京百度网讯科技有限公司、北京世纪互联宽带数据中心有限公司、中国电信股份有限公司北京分公司、中国移动通信集团北京有限公司、北京昇腾创新人工智能科技中心有限公司、京东科技信息技术有限公司、北京红山信息科技研究院有限公司、企商在线(北京)数据技术股份有限公司、北京神州数码云科信息技术有限公司、北京金山云网络技术有限公司。

此前发布的第一批伙伴成员中,共有算力伙伴2家,分别为阿里云计算有限公司和北京超级云计算中心。

关于算力券的具体价格,前述北京经信局相关负责人表示正在研究中。不过,他也指出,目前算力伙伴提供的算力价格尚未公布,但基本已经是市场价格的5、6折左右。

算力租赁的市场价格是怎样的?提供算力服务的北京昇腾创新人工智能科技中心有限公司CTO杨光告诉北京商报记者,目前市场价格约为15-20元/卡/时,1P大约是3-4张卡。

“‘卡’指单芯片,目前算力价格确实也不是很统一。但这个价格对于中小企业来说还是比较昂贵的,所以急需有普惠算力支持企业自身发展。”杨光举例,一个常规的AIGC项目租用的算力规模涉及到的“点”比较多,租用算力的费用并不太容易计算。“但是有公开数据显示,训练一个GPT-3需要花费460万美元”。

北京商报记者经查询了解到,国际算力设备龙头企业英伟达也计划开放DGX Cloud实例,企业可以“云租赁”的形式按

大模型时代,算力是底座。北京市经信局党组书记、局长姜广智在日前举办的2023全球数字经济大会人工智能高峰论坛(以下简称“人工智能论坛”)上表示,北京正筹划出台“算力券”,支持基于大模型落地应用场景的中小企业获取多元化、低成本的优质算力。眼下,大模型正推动智能算力需求爆发式增长。但智能算力很贵,大多数中小企业目前还是依赖于购买或租用算力,其成本也并不低。这成为了中小企业狂奔大模型赛道的痛点和堵点。而北京“算力券”的推出,将有望打通这一堵点,为产业打开全新的想象空间。

月租用DGX Cloud集群,价格为每实例3.7万美元/月起。DGX Cloud能为AI超级计算提供完整的解决方案,借助该集群,企业用户可扩展大型多节点训练工作负载的开发,无须等待需求量通常很大的加速计算资源。

“出台‘算力券’政策,主要是为了解决中小大模型企业算力成本问题,支持基于落地应用场景的中小企业获取多元化、低成本的优质算力。”前述北京经信局相关负责人表示,“如果企业有需要,我们很愿意提供算力供需对接的平台。”

## 需求暴增推高算力价格

本次北京经信局筹划“算力券”的背后,是呈现爆发式增长的智能算力需求。

随着ChatGPT在今天的爆发,各大科技企业都开始逐鹿细分行业大模型,包括但不限于交通、医疗、数字政务等细分领域。根据北京市经信局在人工智能论坛上披露的数据,我国目前已有80余个大型模型公开发布,其中北京约占一半。“但除了大型互联网企业、AI龙头企业具有较多的GPU算力芯片储备外,中小企业在发展AI模型、应用过程中,遭遇算力瓶颈。”



根据北京市经信局在人工智能论坛上披露的数据,我国目前已有80余个大型模型公开发布,其中北京约占一半。

一家科技企业负责人王楠告诉北京商报记者,他们公司主要通过向大厂购买算力来解决算力需求,而ChatGPT在今年突然崛起,许多大厂的GPU算力芯片价格都涨了不少,一段时间里甚至高价也很难买到。

算力价格涨了多少,从英伟达GPU芯片的交付价格即可一窥。以英伟达GPU芯片A100为例,该芯片价格从去年12月开始上涨,截至今年4月上半月,其5个月价格累计涨幅达到37.5%;同期A800价格累计涨幅达20%。同时,英伟达GPU交货周期也被拉长,之前拿货周期大约为一个月,现在基本都需要三个月或更长。甚至,部分新订单“可能要到12月才能交付”。

而算力的需求究竟有多大?资料显示,OpenAI为了训练ChatGPT,构建了由近3万张英伟达V100显卡组成的庞大算力集群,GPT-4更是达到了100万亿美元的参数规模,其对应的算力需求同比大幅增加。如今国内多家企业宣布投身大模型的公司,如果再叠加计算上围绕大模型的人工智能生态和由此产生的用于推理的算力需求,未来的算力缺口将更加惊人。

“总体上看,我国的算力需求从静态来说目前‘基本’能够实现算力供需之间的平衡,

但是从动态角度来说,未来越来越多的行业、企业要运用大模型、运用人工智能多模态,算力在未来存在非常大的缺口。”北京社科院研究员、中国人民大学智能社会治理研究中心研究员王鹏在接受北京商报记者采访时表示。

## 加大算力供给是根本

庞大的算力需求为AI算力产业带来了广阔的市场空间。来自IDC的数据与预测显示,2021年中国智能算力规模达到155.2EFLOPS(每秒一百亿亿次的浮点运算),并在之后的几年始终保持稳健增长态势,预计到2026年将突破进入每秒十万亿亿次浮点计算级别,智能算力实现1271.4EFLOPS的庞大规模,2021-2026年期间,预计年复合增长率达到52.3%。

各科技巨头在算力上的投入也十分明显。在百度2023年Q1财报中,其AI的内容贯穿始终。腾讯发布的2023年Q1财报中,在数字经济、广告和游戏三大核心业务集体发力下,腾讯重回增长曲线。值得一提的是,关于算力和人工智能的部分,同样占据了财报很大的篇幅。

然而摆在国内厂商面前的另一道难题则

是,最关键的AI芯片GPU被英伟达垄断,市场占有率达到90%以上。而一张40G显存的A100芯片,在京东平台上所标售价就达6.25万元。因此,A100芯片也一直被广大科技发烧友戏称为“小金砖”。

这也导致了自建智能算力平台成本极高。今5月,上海张江一家数据公司的相关负责人在接受媒体采访时曾表示,“一台包含8张A800的DGX超算,现在市面上的价格可能要到230万元到250万元之间,价格还在持续上涨”。

算力成本高,成为了制约中小企业的一道坎,而算力租赁或许可以成为目前解题的突破口。在杨光看来,采用租赁方式使用算力能够弹性提供算力,满足不同模型训练所需要的算力需求。

“由企业自身建设、运维算力设备所需的投入非常高,并不是一般企业所能承受的。可以参考的解决办法,首先是统一建设集中化的算力中心,统一分配,部分头部企业已经开始了这方面的布局。其次便是算力租赁,通过政府搭线,给企业提供优质的、相对廉价的算力,降低企业成本。”王鹏说道。

天使投资人、资深人工智能专家郭涛从通用算力和智能算力的细分角度进行了分析,他表示,当前部分地区已出现通用算力过剩现象,算力租赁可以有效解决算力供需对接问题。不过,当前大模型建设更需要的是智能算力,智能算力受到智能计算中心建设滞后和高端芯片难以获取等方面影响,市场供给不足,对人工智能相关产业的规模化、普惠化落地产生负面影响。

“算力租赁虽然可以在一定程度上平衡算力需求,但并不能从根本上解决算力稀缺问题。未来,随着大批智能计算中心项目建设运营,有望从根本上解决智能算力瓶颈问题。”郭涛说。

“过去几年以通用算力为主,智能时代尤其ChatGPT时代则以智能算力为主。”杨光指出,解决大模型企业的成本还是要加大智能算力基础设施建设,“北京也正在先行先试建设多个公共智算平台,除了我们位于门头沟的北京昇腾人工智能计算中心外,海淀正在规划北京人工智能公共算力平台,朝阳也在规划北京经济算力中心等。”

北京商报记者 方彬楠 陆珊珊

## X 西街观察 Xijie observation

### 高温作业,劳动者保护不能说说而已

陶凤

7月3日,网传“北京一名导游带团游览途中因中暑去世”。据媒体多方核实,此事发生在7月2日,去世导游姓蔡,年仅48岁。

夏季正值旅游旺季,这起极端事件把高温作业的危险性置于大众眼前,一些特定工种的劳动者高温作业,如何保障他们的健康安全引发关注。

这位导游所遭遇的不幸,看似在个体身上存在偶然因素,但高温对劳动者的潜在威胁,必须要引起高度重视。

一直以来,围绕高温作业都有制度上的防护、政策上的补贴和倾斜。比如,国家安监总局等部门发布的《防暑降温措施管理办法》,对高温作业提出了严格的要求。

办法要求高温下,用人单位应当根据生产特点和具体条件,采取合理安排工作时间、轮班作业、适当增加高温工作环境下劳动者的休息时间和减轻劳动强度、减少高温时段室外作业等措施。

具体到一些特殊行业,如环卫、建筑工程、快递配送等户外作业,则既需要严格遵守相关管理办法,落实到位,落实到人,还要视具体情况,调整防护措施“层层加码”。

6月以来,全国多地开启高温模式。为督促用人单位做好防暑降温工作,多部门要求加强高温作业劳动保护,防范职业性中暑事件发生,并按规定向劳动者发放高温津贴。

现实中,是否休息、高温下能否拿到足够的补贴,执行起来总会打折扣。有些行业“身不由己”,他们的收入直接关联有效劳动时间,无论是导游还是快递小哥,恶劣天气可能不会选择停工,甚至还会出于维持生计考虑,“热”在其中。

面对这种情况,一方面需要监管部门强行设置高温作业警戒线,对于高出一定温度值的特殊天气,加强干预,对应单位给予劳动者足够的健康照料,如工作状态下实时健康监测,购买配套保险,建立应急处置机制,同时提高劳动收入调节杠杆。

另一方面,则要在公共服务考虑增加有效的配套保障,这些配套措施要防患于未然,参考一些城市在街头设立补水站、爱心棚供大家避暑休息,在一些景点增配医疗救助点,应对突发情况。

而除了“身不由己”的情况,也存在个别单位违规作业、高温补贴发放不到位的情况,等于是置劳动者健康于不顾,口号一直在喊,但雷声大雨点小,甚至还会漏洞百出。

高温防护不能每一次都是说说而已,更不能年年热,还是年年难。气象预报显示,今年很有可能成为有记录以来最热的一年。高温之下,将安全保障篱笆筑牢,将防暑政策落实到位,才是对高温作业最好的尊重。

### 北京旅游行业协会:导游不带病上团

近期,北京经历持续高温,创下连续3天最高温度突破40摄氏度的历史纪录,而旅游市场也随着暑期到来而急剧升温。为了确保游客和从业人员的安全,北京旅游行业协会7月4日发布确保高温天气旅游安全提示,其中提出,作为导游员,要关注自己的身体状况,保持充足的休息和补水,备好必要的防中暑药品,不带病上团,以确保自身的健康安全。

#### 导游从业者“裸奔”

此前,有媒体报道,一名北京地接导游近日在颐和园带团游览时因中暑送医,经抢救无效而离世。这起导游中暑身亡的事件引发了人们对于旅游从业人员工作条件和安全保障的关注。

在为导游惋惜的同时,业内从业者也呼吁从制度上加强对导游权益的保障。国家金牌导游赵东勋介绍,同一个旅行团里往往旅客都上有意外险,但是导游却“裸奔”。

事实上,导游作为旅游行业的一环,他们在高温天气下为游客提供服务确实面临一定的困难和风险。

本次意外事件也让业内再次讨论起导游的职业保障问题。旅行社经营者、京骑文化负责人袁俊杰建议,给导游上人身意外保险,一方面是对旅行社有保障,另一方面也希望对导游这个群体有一个尊重。

此外,北京市应急局近期会同市相关单位对2023年7月份北京地区自然灾害风险形势进行分析后判断,7月份本市将进入自然灾害高发期。其中,极端高温热浪灾害风险不容忽视。

市应急局介绍,7月份,预计本市降水量接近常年同期,平均气温比常年同期偏高。“近期本市已出现多次35℃以上高温天气。根据气象部门预测,7月份平均温度比常年同期偏高,高温

日数略偏多,出现阶段性高温热浪的可能性大。”相关负责人介绍,高温对露天作业、老、弱、病、幼等人员都有一定影响,需重点防范高温高湿环境下有限空间作业中毒窒息风险。

#### 人才短板待补齐

根据文旅部数据,截至2022年年底,全国旅行社总数为45162家。《2023年第一季度全国旅行社统计调查报告》显示,截至第一季度末,全国旅行社总数为47069家。

据亚洲旅宿大数据研究院监测数据显示,目前四五月份新增旅行社仍处于增长势头,预计今年上半年,全国旅行社总体数量有望突破50000家。

2023年以来,我国旅游市场呈现强劲复苏势头。旅行社等旅游企业人手不足等问题也凸显出来。

过去三年,旅游行业的低迷让众多导游经历了沉寂、转行。文旅部数据显示,2019-2022年,全国旅行社从业人员由约41.6万人逐年缩减至24.3万人,签订劳动合同的导游人数由超12万人降至约8.2万人。

面对报复性的团队游需求,市场流通的导游已经供不应求。但因人才培养工作周期长、花费成本高、收入无法保障等因素,导游人才短板尚不能快速补齐。

目前,国内导游群体一直面临缺乏“归属感”的尴尬,社会性、政策性保障都是问题。袁俊杰表示,导游处于旅游产业链的下游,权益保障不到位却往往处在舆论的风口浪尖。“导游这个行业在整个旅游供应链中是属于最底层的,近十几年来一直没有形成一个稳固的体系,甚至有的导游没有基本工资、没有五险一金。”袁俊杰说。

只有从根本上加快补齐导游人才短板,才能将一系列利好政策和市场需求转化为生产力,推动旅游业高质量发展。

北京商报综合报道