“满载与闲置齐飞”

到底是“算力荒”，还是“算力躺”？

▶ 本报记者 孙庆阳

11月16日，DeepSeek因算力触顶对免费用户限流。此前，Kimi K3.1 Pro仅上线48小时即因Agent（智能体）调用激增导致会员额度异常损耗，平台致歉并补偿用户。腾讯WorkBuddy作为接入DeepSeek的第三方办公产品，其爆发式推理请求亦对DeepSeek服务生态形成冲击。

而中国信息通信研究院的数据显示，2025年我国智能算力供应量约38亿卡时，同期实际使用量仅14亿卡时，算力实际使用率约36.8%，超六成建成算力处于闲置状态。

一边是算力调用频频告急，一边是算力资源闲置，“算力荒”与“算力躺”并存的局面不禁引发争议，算力到底够不够用？问题出在哪里？

不能凭一个指标判定算力利用率

“算力是否闲置，要从四个层面来看。”北京理工大学智能科技法律研究中心研究员王磊表示，“设备是否安装、系统是否开机、芯片是否执行任务、任务是否形成有效产出，分别对应‘上架’‘运行’‘GPU利用率’‘有效算力’四个指标。‘上架’是指设备被搬进机房通了电，是‘在不在’；‘运行’是指系统开机跑起来，是‘开没开’；‘GPU利用率’是指芯片有多少时间在运算，是‘忙不忙’；‘有效算力’说明任务是否真正完成、形成可交付产出，是‘有没有用出价值’。四者层层递进，统计对象和技术含义不同，不能用其中一个数字概括全部利用情况。”

国家超级计算无锡中心副主任刘钊说：“‘上架’说明这批服务器物理上到位了，代表的是部署能力，跟用没用起来是两码事。‘运行’相当于机器开机、系统跑起来了，但开机不等于在干活，就像电脑开着待机，不等于在执行办公任务。GPU利用率是最容易被误读的指标。利用率指的是一小时里GPU核心有多少时间在真正运算。利用率高不代表产出高，GPU完全可能在高强度下做无用功或低价值计算。”

“很多集群GPU利用率低，不是没有业务，而是软硬件调优不足，算力空转、功耗浪费。硬件指标好看，但实际产出效率很低。”太初（杭州）集成电路有限公司首席产品官、高级副总裁洪源介绍说，该公司在配合智能算力中心、超级计算中心落地做性能调优时，也观察到类似现象。

“有效算力决定算力价值。”刘钊总结道，“只有有效算力才是衡量有效产出的标准。这也是为什么算力行业越来越强调要从看规模、看装机转向看产出、看效能。”

结构性短缺与过剩并存

据《科技日报》2025年7月报道，部分西部千卡级智算中心服务器上架率不足50%，已上架设备实际利用率不足30%。

“实际上，紧缺的是好用、高端、有生态的算力，过剩的是同质、低端、没场景的算力。一个是结构性短缺，一个是结构性过剩，两者并存，造成了目前‘满载与闲置齐飞’的怪象。”刘钊说。

洪源介绍：“前两年全国各地发力建设智算集群，大多对标通用大模型预训练的标准模型采购，优先选择面向大规模训练、高互联带宽的硬件架构。但智算集群建成后，当地承载的业务更多集中在模型微调、多模态推理、科学计算、工业仿真、行业Agent应用等场景。这就出现了‘错配采购’，用训练型芯片集群跑推理，就像拿重型重卡跑城市快递，硬件能力释放不出来，利用率自然会被大幅压缩。”



国家超级计算无锡中心副主任刘钊在机房查看“神威·太湖之光”的运行情况。 受访者供图

刘钊认为算力行业结构性短缺与过剩并存的原因主要有三点：“一是重建设、轻运营，很多项目把主要精力花在立项、采购、装机上，缺乏专业的算力运营团队和运营能力，设备上架后没有把算力变成可交付的服务；二是重硬件、轻应用，算力是基础设施，但其价值要靠上层应用来兑现；三是重规模、轻效益，过去考核和政策导向偏向‘上架率’‘机架数’‘规模排名’这类物理指标，客观上鼓励铺摊子、争排名。”

“例如，中部某省规划建设3个智算中心、总规模1200PFLOPS（每秒千万亿次浮点运算），而当地数字经济企业不足百家、实际需求不足200PFLOPS；东部某市两个区县分别建设生物医药和工业互联网智算中心，实际入驻企业不足10家。问题不是这些产业不需要算力，而是规划规模与实际需求没有建立合适的对应关系，需求评估走在了建设后面。”王磊举例说明。

值得关注的是，这种短缺与过剩并存的现象正在好转。在2026年7月20日国新办举行的新闻发布会上，工业和信息化部发布的数据显示，截至2026年6月底，全国算力设施整体上架率达71.4%。另据《呼和浩特日报》2026年11月报道，内蒙古和林格尔集群上架率达75.8%，这表明西部算力资源具备良好利用潜力。杭州算力资源调度服务平台通过自建和集中社会闲置算力，2026年已有大集群产品出现客户排队，这表明只要运营和供需匹配，就能够释放价值。

算力建设应锚定真实需求

“十五五”规划纲要明确要统筹布局、有序建设算力设施，加强全国一体化算力监测调度；2026年9月印发的《信息通信行业发展“十五五”规划》提出，有序部署万卡、十万卡及以上集群，到2030年新建大型超大型算力设施运行PUE（电源使用效率）降至1.2以下；“算电协同”则于2026年首次被写入政府工作报告。到2026年8月底，全国一体化算力网监测调度试验验证平台已有20余家调度主体入驻、145万PFLOPS智能算力纳入监测范围。

“需求定建设、产出论英雄、全网促流通。立项端应立足需求建设。”王磊认为，新建大型项目应披露目标客户、任务类型、分期采购和全生命周期成本，以真实订单和可验证场景确定首期规模，再按利用率和服务收入扩容。财政支持也应从补设备转向补首用、补适配、补公共技术服务。

“考核上，也逐步用上架率、有效利用率、PUE和能效等指标来约束和引导企业，推动行业从拼装机转向拼应用、拼效能。”刘钊补充道，政府购买服务宜按合格任务量、Token（词元）服务量或应用效果付费，而不是按设备在线时长付费。设备开机只是提供服务的前提，不能代表实际任务产出和应用效果。

刘钊说：“算力真正的生命力，根植于真实的科研与产业需求。算力行业要走出一条从建起来、用起来再到用得好的进阶之路。”

码上读报 扫码阅读全文

窗口已至，我国颠覆性技术产业化蓄势攻坚

近日发布的《颠覆性技术产业化指数报告（2026）》（以下简称《报告》）指出，全球颠覆性技术产业化进入关键窗口期，中国首次站在未来产业范式未定的起跑线上。我国颠覆性技术产业化正处于“总量攀升、增速换挡”的攻坚跃升期，区域竞争从“比资源禀赋”转向“比转化效率”。

《报告》基于技术创新力、产业发展力、生态支撑力、战略引领力和市场变革力维度，构建了颠覆性技术产业化指数（DTII）。近六年DTII从100.00攀升至158.63，年均复合增长率约9.67%。

从细分技术领域来看，以脑机接口、6G、氢能为代表的“增长加速型”，处于产业化快速爬坡通道；以具身智能和低空经济为代表的“爬坡跨越型”，具备很强的产业化动能；以可控核聚变、量子科技、高温超导材料为代表的“长期攻坚型”，战略认可度与产业成熟度之间落差显著。

《报告》的关键发现指出，真实场景成为颠覆性技术产业化的重要牵引，AI赋能有望成为跨领域底层变量。

《科技日报》2026.11.16 沈唯

新型电网，新在哪？

当前，我国已经建成全球规模最大的电力供应体系。与此同时，我国电力体系也在发生着一系列新变化：在发电侧，新能源从边缘迈向主力，在我国电力结构中主体地位日益凸显。在用电侧，新兴产业持续壮大。

今年4月28日召开的中共中央政治局会议指出，“加强水网、新型电网、算力网、新一代通信网、城市地下管网、物流网等规划建设。”新型电网，被纳入国家顶层设计。

如何加快推进新型电网建设？国网能源研究院能源所能源战略研究室主任闫晓卿表示，要建设好新型电网，优化完善电网新架构是关键，实现主配微有机协同，为能源转型与高质量发展提供稳定电力支撑。

南方电网公司规划部规划管理处高级经理杨振纲则建议，技术上，要推动电网先进技术装备规模化应用、全环节数智化升级；标准上，要持续完善新要素、新业态的并网技术标准和服务标准，健全承载力评估机制。同时，还要充分考虑新型电力系统和新型电网的建设成本，完善全环节价格机制，实现产业链可持续发展。

《人民日报》2026.11.18 王云杉