《2026年中国人工智能计算力发展评估报告》发布

作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/42206.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

《2026年中国人工智能计算力发展评估报告》
发布
。2026年9月21日，AICC2026人工智能计算大会在京举行。会上，国际数据公司（IDC）与浪潮信息联合发布《2026年中国人工智能计算力发展评估报告》（以下简称《报告》）。

作为业内首份聚焦智能体基础设施的研究报告，《报告》指出，智能体（Agent）正成为人工智能（AI）算力需求增长的关键新变量，并以乘数效应推动词元（Token）消耗加速跃升，由此带动中国智能算力增长斜率随之大幅上调。算力跃升为与水利、电网、铁路等同等级的国家战略基础设施。

智能体成为AI算力需求增长新的驱动力

《报告》预测，全球活跃智能体数量将从2026年的7940万个增至2030年的22.16亿个，复合增长率达129.8%；同期全球Token消耗量复合增长率高达4822.6%，约为智能体数量增速的37倍。中国方面，2026年智能体部署渗透率达81.0%，高于全球领先水平的78.1%，Token消耗量复合增长率达3499.3%。

IDC研究显示，到2031年，AI预计将累计创造22.5万亿美元的全球经济价值；作为下一阶段AI投入的重要增量，智能体预计将推动企业AI采购预算进一步增长约20%。

撬动这一投入的，是智能体对Token消耗的乘数效应。《报告》指出，随着智能体承担的任务从简单的信息检索、内容生成，向需要多轮推理、工具调用和跨系统协作的复杂任务演进，单任务Token消耗已从数十Token增至数万甚至数十万Token。智能体数量、任务复杂度与Token消耗的快速增长，正带动算力需求持续攀升。2026年中国智能算力规模预计达到2576.5EFLOPS，同比增长87.9%，到2030年将进一步达到10524.3EFLOPS。

不过，智能体在各行业的落地进度并不一致。《报告》首次从智能体当前应用成熟度与长期生产力影响两个维度评估行业智能体发展，结果显示二者并不完全同步。

IDC副总裁周震刚介绍，代码开发（Coding）与办公任务（Work）凭借数字化程度高、任务边界清晰、结果可验证，已率先进入规模化应用阶段；而制造、医疗健康、能源电力等行业真正进入核心业务并形成“执行-反馈-优化”闭环的比例仍为0%，但在未来三到五年内有望逐渐落地；教育、科研与研发等知识密集型领域当前成熟度有限，但专业知识密度高、可被AI重构的环节多，长期生产力影响更为突出。

供需缺口与系统效率成新考题

《报告》预计，全球AI算力需求满足率将由2024年的79%下降至2027年71%的谷底，2030年也仅回升至77%；缺口绝对值则从2024年的389亿美元扩大到2030年的3809亿美元，供需剪刀差持续扩大。

此外，芯片、存储、服务器等关键环节的扩张周期难以与需求同步，加之能源供给约束，共同构成当前算力产业最现实的困境。

压力的另一面是机遇。《报告》指出，智能体时代AI基础设施正从规模扩张转向系统效率竞争，竞争核心从“拥有多少算力”转向“能够多高效地持续产出智能”。计算产业沿能力型智算与容量型智算双线演进：一方面以超节点为代表的Scale-up架构扩大高速互联域，支撑突破智能上限；另一方面通过多元算力协同、推理解耦以及缓存和调度优化，提高Token产出效率。Token运营也不再追求单一指标最优，而是在业务场景和服务等级约束下，找到交互性与吞吐量的最优平衡。

周震刚强调，随着算力加速成为国家战略基础设施，未来竞争的关键将不只是算力规模，而在于能否通过计算、存储、网络、调度和Token运营的系统协同，将有限算力持续、高效地转化为有效任务结果和业务价值。

作者：赵婉婷 来源：中国科学报

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发