

2025 云计算市场三大转向

本报记者 宋婧 陈存(实习生)

从“云计算”到“云智算”

算力是云计算“万丈高楼”的地基。随着算力需求急速攀升,如何突破算力瓶颈成为云计算产业攻关的重要课题。智能科技公司华鲲振宇方面表示,“云计算”向“云智算”转型升级,标志着算力需求从通用计算向智能计算纵深演进,对高效能、低时延、绿色集约的AI算力提出更高要求。

各行各业的数字化转型离不开强大算力支撑。为有力、有序提升算力供给,全国范围内掀起算力基础设施建设潮。中国移动在长三角打造全国首个整合通用算力、智能算力、超算、量算的“四合一”调度平台。中国电信投资20多亿元打造江西省首个省级智算枢纽——鄱阳湖智算中心。阿里云建设西部云计算中心及数据服务基地,建成后覆盖成渝两地数据业务。腾讯云中重数字孪生产业园项目被列入福建省第一批增补省重点名单。

实现算力的智能调度与优化配置成为云计算厂商共同努力的方向。华为云推出基于新型高速总线架构的CloudMatrix 384超节点集群,通过“训推共池”“灵活调度”等技术实现朝推夜训,白天进行模型推理,晚上闲时进行模型训练,算力资源利用率提升30%以上。天翼云“息壤”算力互联调度平台整合多方、跨域、异构的算力资源,可以根据实时业务需求灵活分配算力,实现资源的全局管理和弹性调度。

《中国人工智能算力发展评估报告》显示,预计到2025年中国智能算力规模将达到1037.3EFLOPS,较2024年增长43%;人工智能算力市场规模将达到259亿美元,较2024年增长36.2%。算力需求暴增的背景下,产业还呈现从“平台竞争”向“平台竞合”方向演进的趋势,通过集成多方资源形成联盟格局。

2025年5月,中国电信、中国移动、中国联通等联合启动算力互联网试验网建设,旨在打破平台壁垒,让用户得以随时、随地、随需接入算力资源。据悉,试验网首期将接入京津冀、长三角、粤港澳等八大算力枢纽节点,预计2026年实现全国TOP100数据中心互联互通。6月,华为云与无问芯穹签署协议,阿里云与并行科技达成合作,通过技术互补、生态共建实现互利共赢,推进算力基建工作。可以预见,各方的协同合作将打破原有的技术边界,催生出更多“化学反应”。

从IaaS、PaaS、SaaS到MaaS

传统云计算服务架构主要集中在IaaS(基础设施即服务)、PaaS(平台即服务)、SaaS(软件即服务)三层,大模型的升级不仅重塑了云计算架构,更催生出新的服务内容。当前,越来越多的云厂商将目光投向了MaaS(模型即服务)层,通过发布一站式落地的模

随着人工智能与算力网络深度融合,云计算已迈入智能化的新阶段,向全栈智能、开放融合的云智算升级。根据调研机构预测,未来几年,用户对国产智算云服务的需求将飞速增长,带动国产智算云服务市场规模不断攀升,预计到2029年市场规模将超4000亿元。业界已达成共识:云计算的下一个时代,已与AI深度绑定。回望2025年上半年,云厂商纷纷围绕“AI+云”展开激烈角逐,云计算市场有哪些新变化?



型服务,实现多场景的生产力解放。

一方面,云计算厂商积极推进自研大模型的建设。今年上半年,阿里云相继开源了视觉底座大模型万相2.1、推理模型通义千问QwQ-32B等诸多模型。阿里云智能集团资深副总裁、公共云事业部总裁刘伟光表示,过去一年平台上的大语言模型API调用量和接入企业数量都迎来了近百倍的增长。腾讯云上线首个混合推理MoE模型Hunyuan-A13B的模型接口,在低成本、强思维方面进一步升级。

另一方面,国产大模型与相关企业的深度合作,也为云计算行业发展注入了新动能。年初,DeepSeek推出R1版本,华为云、腾讯云、阿里云、百度智能云等国内主流云平台纷纷宣布接入,三大运营商也通过各自平台提供专属部署和算力支持,天翼云将“息壤”平台与DeepSeek-R1/V3系列大模型深度优化适配,实现全栈自主可控、模型灵活选型、异构算力融合;联通云依托DeepSeek,升级“星罗”算力调度平台,提升全域智能调度水平。

此外,各大云厂商还通过多样化的大模型策略加速大模型在产业端的普及,为企业实现业务增长和场景创新提供有力支撑。例如,百度云与DeepSeek结合,帮助山东省交通运输厅完成多模态大模型部署;腾讯云利用大模型知识引擎搭建

DeepSeek-R1联网应用,联合深圳大学、华中师范大学等推出智能问答助手。还有部分企业正在尝试以智能体(Agent)形式落地大模型技术。例如,阿里云官宣百炼3.0升级,据称能满足企业和开发者进行大模型Agent开发的“一切所需”。

浪潮云董事长肖雪指出,云的现代化、数据要素共享、智能体大爆发、动态安全防护、集成数字业务平台、智能自治运营已经成为行业数智化发展的关键趋势,同时也是企业未来布局中的一个导向性的基础考量指标,对云服务提出了更高的要求。

从“资源上云”到“深度用云”

当生成式AI被应用于银行智能风控,当人体健康问题可以被实时解答,当城市治理经由算力互联网打通“最后一公里”……2025年的云计算产业正展现出前所未有的“实用”价值。企业不再满足于资源堆砌、算力比拼,而是更关注云平台是否能真正提升业务效率,带来智能化能力的跃迁。上云的需求也从简单的“资源上云”迈入了“深度用云”。

以AI为核心的智算云,正成为云计算从资源堆砌迈向价值创造的关键引擎。华为轮值董事长徐直军表示,不是所有的应用都要追求“大”模型。企业应该根据自身不同业务

场景需求,选择最合适的模型,通过多模型组合来解决实际问题,创造价值。

可以看到,越来越多的云厂商把目光转向全栈式AI技术,既追求高效稳定的底层硬件,又强调灵活易用的上层应用开发。与此同时,各云服务厂商开始意识到,要立足场景找服务,深耕细分市场,形成自己的市场竞争力。例如,百度云在交通领域持续发力;金山云在游戏云和视频云方向长期深耕;华为云重点服务政企用户,将其细分为四类;阿里云以亚太为中心,迈向国际市场。

智算云从云端延伸到终端,辅助构建“云网边端”一体的智能系统,“云终端”或将成为2025年最火的云计算概念之一。市场调研机构数据显示,2024年中国云终端市场出货量为421.7万台,同比增长40.0%;预计2025年云终端市场将继续保持高速增长,增长率将超过16%。

近日,华为正式发布CloudDevice全系列产品,包括云手机、云PC、云Pad、云空间、云机顶盒、云游戏、云车机等产品。同时,其联合相关研究机构共同发布《云终端产业发展报告》《基于云网边端协同的云终端技术标准》。中国联通最近推出了“智家通通”语音智控设备,将云计算、人工智能、物联网等技术融入终端设备,为用户带来“无感化”的智能生活体验。

“万物上云”热潮下,智算云不仅承载着

技术创新的使命,更是提升社会治理效率、优化公共服务的重要抓手。中国移动党组书记、董事长杨杰表示,经济社会从“互联网+”“5G+”向“AI+”加速转变,AI的“规模效应”持续深化,算力的“回弹效应”加速显现,连接的“加速效应”不断拓展,呼唤基础设施新形态、释放算力服务新需求、催生信息消费新形态,为算网基础设施演进带来新机遇、提出新要求。在“由云向智”这一趋势下,算力网络的核心载体正在由传统的“云计算”向云和AI深度融合的“云智算”升级。推动云智算从理念构建走向产业实践,需要社会各方紧密协作、携手共进。

2025年中期中国电子报编辑选择

2025年中期竞争力创新产品

Dell PowerStore 4.1

作为戴尔科技集团企业级存储平台的“明星产品”,Dell PowerStore是定位关键业务负载的现代化全闪存存储,充分满足现代化数据中心解耦架构的多样化需求。

Dell PowerStore自推出以来历经八次全面的进化与迭代,最新发布的PowerStore 4.1版本在存储效能、数据保护、管理体验以及可持续性方面实现全新突破。不仅以5:1数据缩减重塑存储效能标杆,更通过全场景数据保护和现代化数据管理的提升,为企业应对数据爆炸挑战提供高效、可靠的存储底座,助力企业加速实现数智化转型。

2025年中期竞争力创新产品

海博AI云

海博思创的海博AI云聚焦人工智能和大数据技术,利用先进的物联网和边缘计算技术,实现了储能设备云边端的无缝互联,端到端的人机交互。

通过人工智能和大数据技术与储能设备深度融合,实现储能设备的全面数字化和智能化,将传统的硬件储能设备转变为智能化的资产运营服务,使电站更智能,使资产更透明,使服务更便捷。

通过“储能+X”驱动多场景能源生态重构,覆盖风电、光伏、火电等多能源场景,并围绕“储能+充电/油田/矿山/数据中心等”多负荷端应用场景,满足各差异化市场需求。

三大难题掣肘AI大模型落地

赛迪智库 王凡 赵会来 崔冰 郭雯

DeepSeek横空出世,因其成本低、性能好、生态开放等优势,掀起个人、企业乃至政府争相应用的热潮,推动开启AI大模型普及化应用的元年。

赛迪研究院分析认为,AI大模型大规模应用对于我国AI大模型快速迭代发展意义重大,但亟须同步解决好底层框架稳定性不高、跨行业融合存壁垒、生态支撑有限等三大问题。

当前,我国正值人工智能发展关键战略机遇期,亟须从技术攻关、行业适配、风险预警三方面入手,构建AI大模型应用良好发展环境,在新一轮科技革命和产业变革中下好先手棋,赢得主动权。

深度求索的DeepSeek-V3和R1大模型发布以来,以开源和高性价比的特点以及支持免费使用、任意修改和衍生开发,大大降低了使用门槛,带来大模型部署热潮,云服务厂商、科技企业、垂直领域行业企业甚至地方政府等纷纷接入DeepSeek,针对具体需求定制化开发产品与服务,推动AI应用生态的繁荣发展。

但同时,AI大模型在具体行业的推广应用仍面临可靠性与稳定性难题、深度应用的成本高企及生态支撑相对薄弱等系列瓶颈。

底层架构“黑箱”效应等难题尚未破解,适应行业深度应用的可信产品供给较为匮乏。

随着AI模型训练数据量和参数规模呈现指数级增长,算法架构也趋于复杂,导致模型缺乏可解释性和透明性,特别是深度神经网络模型的“黑箱”特性,使得在更加关注推理、决策和运行过程的具体行业应用中尤为棘手。

同时,大模型在面对新场景和处理复杂任务时也常因缺乏行业知识而输出与现实不符的“幻觉”或不可靠结论,难以确保行业应用中的可靠性与稳定性。例如,尽管大模型处理复杂数学运算能力突出,但却无法确保运算的正确率。据报道,ChatGPT-4处理简单加法和字谜时错误率甚至比部分小模型高出15%,而Meta旗下大模型LLaMA在简单两位数任务上的正确率竟未超过60%。同时,据Vectara测试团队发布的LLM大语言模型“幻觉”排行榜,ChatGPT-4、OpenAI-o1、阿里Qwen2.5-Max和DeepSeek-V3的幻觉率分别为1.8%、2.4%、2.9%和3.9%。与此同时,DeepSeek-R1以及Google PaLM-2的幻觉率均超过14%,且不管模型先进与否均无法完全避免幻觉问题。

此外,尽管AI大模型在通用领域展现

出强大能力,但具体产业场景的复杂性和专业性使其实际落地时面临显著的适配难题。特别是制造业领域数据的多模态特性要求模型具备跨模态融合能力,而现有大模型对时序数据、高精度物理参数的解析准确率还普遍偏低,且训练数据不能完全代表现实产业的复杂性,难以满足制造业产线稳定和高效运行的严格要求。由于输出结果的潜在不准确性,企业在广泛深入应用大模型时趋于谨慎,直接阻碍了AI大模型行业落地。

海量产业链上下游主体分散式接入大模型,长期“单兵作战”恐增加后期协同成本。

当前,各行业主体积极接入大模型端口,虽短期大幅提升了自身运营效率,但却忽视了长期产业链上下游协同应用的重要性,特别是厂商在基于DeepSeek-R1模型提供服务时算力成本高企,MaaS(模型即服务)价格并未降低。以满血版DeepSeek-R1每百万token(输出)收费16元计算,如果每日输出1000亿token,基于DeepSeek的服务每月的机器成本是4.5亿元,即亏损4亿元。

经某长三角新能源汽车产业集群统计,内部约38%的零部件供应商未接入Deep-



Seek系统,导致数据交互环节的传统ERP(企业资源计划)系统生成计划需延长72小时,经人工翻译适配主机厂大模型智能排产系统;成本核算环节的手工填报数据与主机厂DeepSeek动态成本模型产生15%误差,造成季度供应链金融授信额度误判1.2亿元;质量追溯环节的缺陷零件追溯时长从行业平均4.2小时激增至27小时。

由于前期大模型接入步伐不一致,主体应用间的基础能力呈现明显差距,而产业链任一环节的数字化断点都将放大整体协同成本。湖南磐云数据基于DeepSeek开发的供应链协同平台数据显示,全链条接入可使订单响应速度提升300%,异常事件处理效率提高420%。

随着AI大模型迅速渗透并赋能行业发展,后发企业的追赶成本将呈指数级增长,难以跨越基础能力鸿沟差距,进而限制AI大模型应用推广。

AI大模型推广应用的生态环境尚不健全,数据、标准、伦理等基础条件相对薄弱。高质量训练数据样本稀缺且割裂,长期将从源头掣肘我国AI大模型推广应用。例

如,医疗等垂直领域数据开放率不足,医疗影像大模型训练需10万多标注样本,但三甲医院数据因隐私保护难以共享,误诊率高居不下。麦肯锡2024年报告显示,欧盟GDPR等法规使跨境数据成本激增,企业合规支出占AI项目预算30%以上,后续难以适应复杂多元的AI大模型推广应用需求。

AI大模型应用范围和深度的持续拓展,取决于其面对复杂多变市场需求的应对能力,但当前AI大模型在处理与道德、法律、安全、隐私和社会影响等相关问题时,标准体系水平短板较为突出。根据欧盟AI监管草案,甚至全球主流十大AI模型均未达到标准,Aleph Alpha、Anthropic等产品得分低于25%,主要问题集中在版权责任不明确、能源使用报告不均衡、风险缓解披露不充分等方面。

随着我国DeepSeek大模型迭代周期缩短至45天,传统的规则制定模式难以继,政策出台大大滞后于技术创新和商业模式迭代进程,非法收集、泄露国家安全及个人隐私数据风险高企。若不能及时在上述数据、标准、伦理等领域建立保障体系,将面临推广隐患大,不推广就落后的窘境。