



我国超算时隔九年再次排名全球第一

本报讯 据国家超级计算深圳中心发布的信息，2026年6月23日，在德国汉堡ISC2026大会上，我国全国产自主研发的“灵晟”超级计算机以2.19EFlops（每秒10¹⁸次浮点运算）持续双精度浮点性能，登顶大会发布的全球超算TOP500榜单。图灵奖得主、美国两院院士Jack Dongarra教授发表赞誉称：“中国的‘灵晟’系统让世界看到了超算通向AI4Science新型系统架构的希望之光。”

“灵晟”超算系统总设计师、国家超算深圳中心主任、中山大学教授卢宇彤在现场介绍该系统时指出，超算发展已迈入超智融合新阶段，“灵晟”系统首创Online Acceleration的全CPU架构，打破传统

CPU-GPU异构架构壁垒，内嵌AI矩阵加速单元，回归计算加速的本质，实现超算智算等多种计算模式的高效协同，全面赋能科学智能计算，完成了算力性能登顶和应用广泛落地的双重突破，为全球超算技术升级与规模应用提供了可借鉴的方案，获得了世界同行的高度评价。

据悉，“灵晟”定位为面向科学与工程智能计算的国产全栈融合基础设施。在芯片层，自研LX2 CPU引入多精度与矩阵加速等能力，实现片上超算算力与智能算力深度融合，并集成了首款国产HBM，内存带宽相比传统CPU提升10倍；在网络层，自主设计的灵启高速互连网络可支持200万个端口、10万节点的超大规模组网；在存储层，采用分

层架构，兼顾高性能作业区与容量型数据区，可扩展至E级；在系统层，构建大规模超智融合框架，自研全栈软件，将底层硬件能力以可用、可编程、可优化的方式释放给应用，提高各领域应用性能；在节能方面，首创100%全液冷散热计算机柜，能效比达51GFlops/W。

据介绍，自系统部署以来，“灵晟”已支撑大气海洋、工程仿真、材料科学、药物发现、脑科学、科学AI、大模型推理等多领域应用，支撑混合精度计算、工作流和复杂多任务并行运行，在大规模并行环境下平均扩展效率达84.4%，实现超过千万核心的全系统高效可扩展运行。

（江楠）

OpenAI和博通公司联合推出LLM优化推理芯片

本报讯 记者姬晓婷报道：6月24日晚间，记者从OpenAI官网了解到，OpenAI和博通宣布发布一款名为Jalapeno的智能处理器，这是一款基于OpenAI对LLM（大语言模型）推理未来的愿景而设计的加速器，也是两家公司共同构建的多代计算平台中的首款AI加速器，旨在使先进的AI更快、更可靠，并让更多人更容易获得。

OpenAI基于其对机器学习模型（LLM）基本原理的深刻理解，并结合其模型、内核、服务系统和产品需求的路线图，从零开始设计了这款芯片。OpenAI与合作伙伴博通和Celestica携手，通过芯片实现、电路板和机架

系统集成、高性能网络以及可扩展的生产系统，助力该平台实现产业化。Jalapeno芯片的设计兼顾灵活性，能够与所有LLM兼容。至此，Jalapeno芯片的工程样品已在实验室中以生产目标频率和功耗运行机器学习工作负载，包括GPT-5.3-Codex-Spark。

OpenAI仍在评估最终性能，但早期测试表明，Jalapeno的每瓦性能将显著优于目前最先进的技术。该架构减少了数据传输，并平衡了计算、内存和网络资源，从而使实际利用率更接近理论峰值性能。博通的芯片实现和网络技术，助力该平台实现大规模量产。

OpenAI称，Jalapeno是一个专为现代

LLM推理而设计的全新方案，并非基于早期AI工作负载改造的通用加速器。它借鉴了OpenAI日常运行的系统，包括Chat GPT、Codex、API以及未来的智能体产品，同时也面向当前及未来业界的LLM应用。其目标是将当今领先的AI加速器的性能和吞吐量与接近最快专用推理系统的延迟相结合，从而使Jalapeno非常适合大规模交互式LLM产品。

OpenAI表示，Jalapeno从最初设计到最终流片仅用了九个月时间便完成了联合开发，而其定制的AI加速器项目代表了该公司所认为的高性能先进半导体领域迄今为止最快的ASIC开发周期。

企业AI落地：从“租用智能”到“拥有智能”

本报记者 张琪玮

当前，中国企业正以截然不同的姿态涌入AI浪潮。它们或是在用Claude Code自研MES系统，或是在产线上用深度学习模型做实时质检，或是在构建自己的智算中心。这些场景看似分散，却指向同一个命题：企业如何将AI从“试用品”变成“核心能力”？

“现在遇到的核心问题是AI落地雷声大、雨点小。表面上看起来热热闹闹，最后却看不到实质价值，ROI（投资回报率）也没有。”在前不久举办的媒体沟通会上，IBM大中华区首席技术官翟峰这样看待当前中国企业普遍面临的“AI困境”。IBM将目前阶段称为AI变革的“第零日”——AI更多停留在“工具租用”和局部业务场景，尚未进入核心流程、决策机制与基础架构。从“租用智能”到“拥有智能”的质变，IBM认为还远未到来。

AI落地的三步走“必修课”

“AI时代的企业竞争优势，将取决于进化速度。对中国企业而言，数据是否实时可信、流程能否全局打通、洞察如何赋能决策，决定着核心业务的自动化、智能化水平——这正是AI从战略到落地、从试点到规模化重要标志，战略、组织和技术平台缺一不可。”IBM方面表示。

数据的实时可信，是企业落地AI能力的关键。大模型的能力越来越强，但企业最核心的“护城河”，是能把自己私域的数据整理好再喂给模型。“企业不是没有数据，而是没有高价值、可直接给模型使用的可用数据。这是目前卡住大家的第一步。”翟峰说道。

流程能否全局打通同样重要。翟峰观察到，企业里90%以上的流程是固定的，如报销、审批等，不需要AI去泛化或创造，更需要的是自动化。“只有在遇到复杂的、无法预先设计固定流程的任务时，才考虑用智能体来加速。”

洞察如何赋能决策，是企业把控方向的核心。“企业一把手今天的决策，是基于数据提供的洞察，还是仅仅基于过去的经验？”翟峰认为，在大多数时候，决策并没有基于很好地利用数据。

基于这三重观察，翟峰给出了一个判断：很多企业在做AI试点时，这里上一个小程序，那里调用一下第三方大模型——本质上是在“租用智能”，而非“拥

有智能”。智能究竟是“租用”好还是“拥有”好？翟峰认为，这并非一道选择题，而是一个逐步演进的过程。在他看来，企业要构建自己的智能化能力，需要在私域或本地环境中部署AI体系，让本地数据、核心算法模型与智能体结合，形成安全、可控的闭环。

基于许多企业落地AI的实践，翟峰提出两个概念：“+AI”是在原有工具和流程上辅助加一些AI能力，而“AI+”则是组织流程和业务建立在AI之上。他认为，大多数中国企业还走在从“+AI”到“AI+”的路上。

大部分企业还在从“+AI”到“AI+”的路上

AI风潮持续已久，但要真正实现AI落地，居高不下的成本是让不少企业头疼的难题。基于此，翟峰提出：“在企业发展的不同阶段，要选择用对的技术、最适合的技术、成本最优的技术来解决最核心的问题。”翟峰介绍了5月初举行的Think 2026大会上IBM发布的四款产品，并称其正好“对应中国企业经营的‘四大战场’”，分别是面向研发创新的IBM Engineering Lifecycle Management (ELM)，面向生产制造的IBM Maximo，面向供应链协同的IBM webMethods B2B集成方案，面向智慧经营的IBM Planning Analytics。

至于ROI的挑战，翟峰认为症结在于三个层面：“一是技术层面的‘技术债’，许多企业数字化没做好，AI拿过来没法用；二是AI战略的缺失，高层和基层方向不一致；三是组织文化层面的阻力，很多人知道AI最后会动了他的奶酪。”

同时，他也给出了解法。“把数据、流程、决策三个基本功做好，真正的智能化也就水到渠成。”此外，对于已经走过试点阶段的大企业，他认为需要从散点工具走向统一的AI平台。“平台的优势是统一标准、统一大家的协作，真正把分散在各地的员工技能和AI能力沉淀成企业可以复用的资产。”记者了解到，IBM在Think 2026上提出的“AI运营模式”蓝图——智能、行动、信任、运营四大要素，正是为这一阶段的企业提供的框架性答案。

“不能低估AI长远的价值，但也不能高估它短期内的价值。”翟峰总结道。正如他所说，AI终有一天会像电一样——“现在谁还天天谈电？当它慢慢变成水电煤的时候，就没有人谈了。大部分的企业还在从‘+AI’到‘AI+’的路上。”



稳增长 强创新

促融合 优治理 防风险

确保实现“十五五”良好开局