

探访亚洲规模最大的智算中心 大模型时代到来 算力够用吗？

本报记者 郢阳

每秒百亿、千亿次的浮点运算能力，千万亿字节(PB)级的数据存储规模……许多人可能对这些指标没有太多概念，但要说起对答如流的大语言模型，自动驾驶中反应神速的汽车，大家应该不会感到陌生了吧？

看得见的应用背后，是看不见的算力在支撑。“大模型时代”谁能占得先机，这个问题谁说了算？在2023世界人工智能大会召开之际，记者探访了商汤科技人工智能计算中心(AIDC)。如今，基于最新人工智能理论、采用领先人工智能计算架构，提供算力服务、数据服务和算法服务的智算中心，在推进大模型能力和数字经济发展中正发挥着越来越重要的作用。



■ 商汤科技人工智能计算中心，是目前亚洲规模最大的人工智能计算中心之一
本报记者 陈灵玮 摄

AIGC时代 不可或缺的“血液”

这座位于临港新片区的商汤科技人工智能计算中心，是目前全亚洲规模最大的人工智能计算中心之一。经过一年半的运营，今天的算力规模已经达到了5000PFLPOS(1PFLPOS相当于每秒可以进行1千万亿次浮点运算)的体量，超过了当时设计的算力规模水平。

“换句话说，商汤智算中心的峰值算力可以达到每秒500亿亿次浮点运算。通过这样一个超大规模的算力，我们可以支持20个千亿参数超大模型同时进行训练，并提供完整的人工智能基础设施资源的能力。”商汤AI大装置产业赋能中心主任蒋钦告诉记者。

大模型吹响了生成式人工智能(AIGC)的号角，这必然伴随着“能源”的进一步开发。这样的“能源”，在AIGC时代，正是无数服务器中运转不息的庞大算力。打比方说，大模型若是人工智能抽象化的“身体”，那么海量的算力，则是维系身体运转的“血液”。

蒋钦解释，所谓算力并不是单一概念，它包括通用算力，就是以CPU芯片为主的算力；还包括超算，就是各种超算中心提供的高性能算力，主要服务于科学计算；还有目前非常热门的智能算力，即以GPU为典型代表的，以AI芯片为主的服务于人工智能训练和推理相关需求的计算能力。

其中，以AI芯片为主的人工智能算力近年来呈现快速增长态势。中国信通院统计显示，2016年的算力结构中，国内人工智能算力比例只占3%左右；而截至去年，整个人工智能算力比重已超过半壁江山。

各行各业 需求与日俱增

随着人工智能技术的高速发展，以及大语言模型的出现和应用，各行各业对智能算力的需求也在与日俱增。蒋钦介绍，2012年前后是一道分水岭，随着深度学习的兴起，在此之后，最先进人工智能模型的计算量每三四个月就翻一番，也就是每年增长近10倍。尤其是最近几年，在大模型取得突破、应用场景的广泛开拓与深入发展的背景下，对智能算力的需求迎来井喷。

没有人会否认，大模型江湖的混战已然愈演愈烈，“下场”的选手越来越多。由大模型为代表引发的人工智能新技术热潮，会带动整个算力需求呈现出非常旺盛的态势。

“大模型需要处理海量的数据，并且需要在短时间内完成复杂的计算任务，这就意味着需要更高的计算能力。”蒋钦指出，“大模型也需要访问大量的数据，存储大量的参数和中间结果，因此需要更大的内存容量。”与此同时，大模型需要在多个硬件设备上同时进行计算，因此需要更好的并行计算能力；而在“双碳”背景下，消耗大量能源的大模型也在呼唤更高能效比的算力。

“虽然纯CPU算力也能用来进行部分人工智能计算，但效率非常低，能耗非常高。我们不会以这样的算力形式去支撑产业的发展。”蒋钦判断，“而搭载以AI芯片为主的新型算力基础设施是未来发展的重要趋势和热潮。”

他告诉记者，在大模型的热潮下，算力服务器目前呈现紧缺状态，“大模型的发展和智能算力底座有着深刻联系，需要在大规模资源汇聚的基础上做好算力、算法、数据三要素的协同实践和联合优化，助力实体经济高质量发展。”

助力临港 打造算力新高地

作为亚洲已投入使用的、最大规模的单节点智能计算中心，商汤AIDC除了算力规模体量巨大外，与之相配套的人工智能模型开发工具，针对大模型的训练、推理、数据优化服务等，都可以高效率、低成本、规模化地支撑和服务好目前以大模型为代表的整个人工智能产业的高速发展。

今年上半年，人工智能应用向纵深发展：AIGC打造的“数字人”效果媲美真人水平；人工智能预测蛋白质结构给基础研究带来全新的研究手段；人工智能驱动的大语言模型在多轮次对话和长文本理解能力上突飞猛进……这些无不依赖算力提供支撑。

“商汤科技人工智能计算中心除了为‘日立新’大模型体系提供底层算力底座和AI工具的支撑外，也为其他行业赋能AI相关的基础能力。”蒋钦介绍，“通过云服务的方式，我们已在泛互联网、泛娱乐、自动驾驶、科研院所等多个领域里支持了十余家大模型客户的开发工作。”

商汤科技人工智能计算中心会依托临港，进一步夯实算力底座，扩大算力规模。在算力软件平台相关技术领域，持续对集群架构、集群网络、存储性能、能源管理等多方面开展技术攻关，对于大模型训练与推理定向支持等开展技术迭代更新，以支撑更大规模的算力集群落地，助力临港打造智能算力新高地。“临港本身具有非常强大的区域产业优势，集聚了集成电路、人工智能、生物医药、高端装备制造、新能源汽车等核心优势产业。这也要求我们围绕这些优势产业去打造人工智能基础设施行业解决方案与丰富的产业上下游生态，做好产业智能化升级的算力底座。”蒋钦表示。

探访张江人工智能水域应用场景 一键启航 体验“无人船”驾乘乐趣

昨天，记者提前探营本届世界人工智能大会浦东分会场，包括无人船、3D导航、AI艺术创作等应用场景将在大会期间掀起“盖头”。尤其是“无人船”首次亮相张江人工智能岛，这也是全市首个人工智能水域应用场景试验区。

自主模式 可实现“无人驾船”

昨天中午时分，酷暑炎炎，但在张江人工智能岛的智慧河上，一条白色顶篷的小船正悠然自得地融于两岸科技元素满满的科学城风光中。

这条船乍一看似寻常，但登上船便可发现这艘小船“不简单”。在这艘可以载8名左右乘客的小船上，具有自主、人工及遥控三种航行模式。原来，这

是张江AI+海洋科创中心第一个无人驾驶船落地应用项目。人们可以乘坐“无人船”，途经张江科学会堂、在建中的“张江之门”等“网红打卡点”。

开启自主模式，即可实现“无人驾船”，可以按预设游览路线自动航行、自主避障、自动靠离泊，乘客也可通过艇端控制台操控船舶体验驾驶乐趣。人在船中坐，在一个类似平板电脑的智能终端上，乘客只要按相应按钮就可一键启航。“市民现在对无人车比较熟悉，这艘无人船是L4无人驾驶级别的船，和无人车的级别是一个概念。”AI+海洋科创中心常务副主任赵辉介绍，在这个智能平板上，和我们平时开车的导航相似，根据规划路线，可以显示行驶途经的线路。原来，这艘“无人船”上配备了无轴推进器、新能源动力，

核心的自主决策的控制系统，融合了北斗卫星定位和4G定位技术的厘米级组合定位系统、激光雷达等多种先进设备和多源感知融合、人工智能识别等跨域技术，能控制推进器的转向、掉头、速度等。马上到岸了！只见“无人船”不紧不慢靠近，平平稳稳停泊，轻巧从容地陪伴乘客们度过一段美妙的午后时光。

据了解，本次世界人工智能大会期间，该测试区将推出水上清洁贪吃船、水上驰援救援机器人、水下森林探秘机器人，以及海洋科普走廊等体验活动。同时，一款已经投入商业使用的智能救生圈也将在人工智能大会期间展示。“夏天是溺水的高发期。过去发现这种情况只能靠人工救援，赶到现场还有一定时间差。智能救生圈发现呼救后，可以根据求救声音锁定来

源，迅速‘游’往溺水者附近，争分夺秒争取救援的黄金时间。即使救援人员到达后，有了智能救生圈的辅助，也能大大增加救援的成功率。”赵辉表示，未来希望“无人船”不仅服务于张江，还能拓展到全国各地。尤其浦东新区是引领区，希望参照无人车的浦东法规，拥有更大范围、更丰富的“无人船”应用场景。

浦东率先 布局未来产业“新赛道”

浦东新区作为全国首个人工智能创新应用先导区，正利用国家战略赋予的政策优势，打造上海人工智能产业发展高地，引领人工智能世界级产

业集群建设。

记者从昨天召开的浦东新区人工智能产业生态建设新闻通气会上获悉，浦东新区正在为大算力、强算法时代提供强大易用的底层支撑，推出“1+1+N”的人才政策体系和明珠计划，为人工智能产业创新汇聚全球英才，不断推动数据要素、元宇宙、智能网联汽车、机器人等前沿领域率先布局，一系列前沿科技“新成果”不断涌现，率先布局未来产业“新赛道”，成为浦东人工智能产业领域核心技术突破和制度创新新起点，浦东人工智能创新应用先导区建设再启新征程。

本报记者 宋宁华