

AI大模型带动算力产业规模猛增的研究

李 筱

首都在线科技股份有限公司 北京 100000

摘要：在今年4月25日，发布了算力商用部署白皮书。显示2022年末，中国算力规模实现180百亿亿次浮点运算/秒，存力规模大于1000EB，也就是一万亿GB，时延减小至20ms以内，并且，算力领域内模型、技术多样性，怎样协调匹配，让其体现出最大效能，已是行业关注的重点。本文先概述了AI大模型，介绍了算力产业面临的困境，然后研究了AI大模型带动算力产业规模猛增，最后探讨了怎样推进算力产业发展，希望能为相关人员提供参考。

关键词：AI大模型；算力产业规模；数字经济；发展建议

Research on the rapid increase in the scale of computing power industry driven by AI large models

Xiao Li

Capital Online Technology Co., Ltd. Beijing 100000

Abstract: On April 25th of this year, the commercial deployment white paper for computing power was released. It reveals that by the end of 2022, China's computing power will have achieved a scale of 18 billion trillion floating-point operations per second (FLOPS), with a storage capacity exceeding 1000 exabytes (EB), equivalent to ten trillion gigabytes (GB). The latency will be reduced to within 20 milliseconds. Moreover, within the field of computing power, the diversity of models and technologies presents a key concern in terms of how to coordinate and match them to maximize efficiency. This paper begins by providing an overview of AI large models, introduces the challenges faced by the computing power industry, examines how AI large models are driving a rapid increase in the scale of the computing power industry, and finally discusses ways to promote the development of the computing power industry. The aim is to offer insights for relevant stakeholders.

Keywords: AI Large Model; Computing Power Industry Scale; Digital Economy; Development Proposal

引言：

数字经济建设，数据为不可或缺要素。用来开发数据价值、完善算力，是促进数字经济的重要动力。按照IDC推测，今后五年全球数据爆炸式增长，算力需求将迅猛增长。但限于技术原理，半导体芯片已趋于性能极限。怎样加快先进计算开发与运用发展，已是获取竞争优势的制高点。很多国家都在算力方面布局，这也带动了算力产业。其是计算技术和产业统称，可以分成多个层面，比如基础软件、行业运用处理方案。算力产业的全方位发轫，有利于数据资源价值转换，支持数字化转型发展，持续增加数字经济涵盖面，延伸数字经济生态

位宽度。

一、AI大模型

目前，AI大模型已是AI界的当红明星，谷歌、英伟达和阿里等，均在布局该模型。它的诞生，使大多数产业人士认为此技术将转变产业格局，也就是将步入AI大模型时代。今后，人工智能流向终端，流向公司，先做到这点，便会在布局中获取优势。AI大模型涉及了两层含义，一是预训练，二是大模型，两者融合形成了人工智能模式，也就是在实现预训练之后用不着微调，或者只需微调少许的数据，便可以支撑运用。比如浪潮信息推出的“源1.0”，一个模型便可以提供写作、对话等运用^[1]。从运用层面来分析，大模型和预训练等组成了AI架构，也就是算法基础设施，基于此，能够支撑算法更好融入所用工具软件，实现算法运用价值商业变现。如今人工智能落地面对多项挑战，比如场景运用碎片化，

作者简介：李筱（1982.04.11-），男，汉族，山东青岛人，硕士，工程师，研究方向：人工智能算力中心在AI大模型、元宇宙方向的应用，人工智能算力中心。

而对于产业化运用痛点的处理，大模型是良方。一个模型能够支持各类运用，切实缓解反复建模困境。

二、算力产业面临的困境

根据外部环境来看，存在技术卡脖子的困境，难以构建计算体系。现阶段，计算生态还采取美国技术路线，特别是芯片及通信协议，该国制定了标准布局，能将产业链和用户进行绑定。中国在超算及云计算方面，早已占据领先地位，但针对芯片和基础软件比较依赖欧美技术，再加上技术阻断，导致中国算力产业链逐渐分割为若干“孤岛”，不但对产业发展格局造成了影响，还增加了我国的算力成本。

根据内部环境来看，不具备协同创新能力，不能达到产业链生态共享。我国提高了自主研发力度，现阶段主要是对中短期项目提供资助，注重单点应用以及技术创新，但是还没有顶层设计，并且产业化发展不具备持续造血能力，这造成项目完成后产业化不能大力推广，对成果协同效益产生了较大的影响，无法实现核心技术突破建立产业生态。在这一过程中，很多龙头企业的自主研发基本都构建了“以我为主”的目标，其技术路线及标准没有统一，缺乏一定的兼容性，不能建立产业链生态共享。

根据数据中心建设来看，存在各自为政的问题，且供需错配现象比较明显。不管是区域间算力建设，还是组织间算力建设，都没有构建协同机制，在规划过程中，通常将单独实现系统内所有需求当作目标，这造成建设过程中生态比较单一，存在重复投资等诸多问题^[2]。另外，各种各样的应用场景对算力能耗、交互性、数据安全策略等的需求存在差异，比如，矩阵式高发应用需求以及逻辑运算复杂需求等有着较大的差异。然而很多数据中心的算力基本是通用型，还没有构建多元化算力供给梯度，这极易导致算力过剩等问题。

没有建立公共交易平台，导致算力资源分配效率不高。因为算力之间没有高效链接和数据安全等因素，导致无法构建科学的算力公共平台。对于超算及云算中心等算力供给，无法达到全面整合，这对相关工作的开展造成了不良影响。另外，从算力产业有关要素来看，市场上没有建立完善的交易平台，导致被交易成本比较高，并且资源效率还不高。比如工业领域，还有着很多智能化需求和沉淀数据，然而因为算法模块化等的影响，没有形成信息透明的市场，不能找到合适的算法。

三、AI大模型带动算力产业规模猛增

1. AI大模型带动算力需求增长

算力产业规模正快速增长，根据不完全统计，从2016年至2021年，我国算力规模年均增长46%，如今，产业规模已超过2.6万亿，带动经济产出2.2万亿，从全球来看，计算设备算力规模615EFlops，增速44%，推测到2030年算力规模能实现56EFlops，年均增速实现

65%。虽然如此，算力领域还有着较大缺口，针对算力需求无法估计，由于目前很多新型赛道，比如智能驾驶均需算力支撑，根据数据类型分析，算力供给比较单一，不能支持非结构化数据，若以CPU年性能提高15%计算，技术更新还不能追上商用步伐。更为关键的是，技术路线还没有尘埃落定，通用计算全方位迸发的同时，智算平台也在迈进，多样性是领域的主要特点。之所以表现多样性，产品和商业诉求差异，是主要因素。对于服务器以及芯片市场，厂商均在推广产品、抢占份额。从数据层面来分析，如今HPE公司处在榜首，市场份额占比15.6%，戴尔份额占比15.4%、华为份额占比1.9%，在芯片方面，因为长时间被X86主导，英特尔份额常年占比约为80%，但伴随运用与需求多样化，华为等厂商也表现后来居上态势。

①大模型落地推动海量算力需求。AI大模型推出，且在多场景深入运用，大模型是人工智能发展的必然方向，如今已进入AI大模型时代。算力是建设大模型生态的重要基础，基于储存参数来扩展模型宽度，进而增强表现能力，所以模型参数表现指数规模。意味着需要更充足的计算资源，训练及推理期间需消耗大量算力，推动了需求猛增。

②规模显著增长，中国智能算力规模扩大较快。在2021年，算力规模实现615EFlops，同比提高约为44%。根据有关评估报告表明，中国算力规模已超过155.2EFlops，推测到2026年将实现1271.4EFlops。从通用算力规模来看，2021年已经达到47.7EFlops，推测到2026年将实现111.3EFlops。从2021年至2026年，推测复合增长率能实现52.3%，高于同期增长率18.5%。

③算力在AI大模型训练、信息处理上起到关键作用。其一，模型训练：训练模型需要很多的算力资源。训练期间，要持续开展矩阵运算等操作，进而调整参数，提升模型精度。算法可以加快训练，进而推动AI技术革新。其二，推理速度：AI大模型结束训练之后，需开展推理操作来处理数据。推理快慢和两方面有关，也就是算力及存储速率。较快的推理速度能够提升AI运用的准确性。其三，数据处理：AI需海量的数据开展训练以及推理。算力以及存储能力和处理量息息相关，若计算机性能不高，系统也许难以处理数据集，进而影响其精准性及性能。

④算力在应用端方面起到关键作用。为了达到业务增长、拥有竞争力以及获得较大市场份额，很多公司都入局AI领域，各行业的AI运用渗透度均呈现持续加深态势，运用场景也更加广泛。按照IDC预计，AI应用渗透度排名的行业分别是互联网行业、金融行业以及政府行业等，到2023年末，我国二分之一的制造业供应链环节会应用AI以提升效率。就入局人工智能领域的公司而言，算力对自身的发展尤为关键，计算资源能够提升产

品性能,支撑公司运营。其一,提升产品研发效率:企业需持续开发新的人工智能应用及产品,以达到市场需要及增强自身竞争力。高效计算资源能够带动人工智能技术的研发,提升研发效率。其二,提升产品性能:人工智能产品的性能与算力、效率息息相关。算力越高效,性能越佳,能够提供更快的服务,进而进一步达到使用者的需求。其三,支持企业运营:人工智能产品的运营中,数据处理也需海量的算力支撑。

2. 多样性演进和发展

对于市场现状,专家提出,多样性不该是产业发展阻碍,应是有益尝试,要加以考虑的是,怎样协调技术路线差异,确定评价标准,由此激发技术潜能。算力产业峰会上,有专家表示,多样性算力规模部署,核心在于:硬软件协同,激发算力潜能;制定测评体系,牵引产品迭代,引导使用者敢用、多用。具体至技术方面,专家表示开放性是核心点,要聚集部件规范体系,完善算力供应;探究架构体系,建设AI生态^[3]。不管是多样性演进,还是测评标准构建,均会转化成商用价值,因此,企业是此过程中的主力。从数据方面分析,智算中心达到80%运用水平,对其的投资,能推动产业增长3.4倍,带动有关产业增长大概42倍。国内多家机构共同成立委员会,如今,涵盖芯片IP、设备商等几十家公司加入其中。针对下一步发展方向,专家表示,在大模型时代,要强化智算攻关,注重突破基础设施核心技术,发展AI产业大生态,加强先进计算技术规划,将产学研协同落实到位,提高产业信心。

四、推进算力产业发展的建议

对于如何更好推进算力产业发展,本文主要从以下方面进行分析,即制定规划、建立算力网络,构建产业联盟、打造开放生态圈,强化科研人才布局、积累竞争力等,希望能为相关人员及研究提供参考。

基于算力国际竞争形势来看,应有效利用我国制度优势,把算力产业当成新基建发展任务,对重大研发项目以及区域协同发展等开展统筹规划。

构建算力产业规划以及广域算力网络。基于有效分析数字经济新业态以及算力结构性需求的同时,全面促进算力产业协同发展,实现区域、上下游及基础设施全面协同,有效落实顶层设计工作。提高对算力产业的统筹力度,科学布局各个区域下细分领域的方向,发布科学的鼓励政策,防止重复建设。为达到政府、社会及市场等方面对算力的需要,应事先建设一体化算力基础网络,为其提供安全的算力服务。可先实施网络建设试点,对区域内云计算中心等算力评估及合作进行科学统筹规划,促进高速网络设施的建设,从而促进区域一体化发展,为其提供可靠的算力保障。

构建算力产业联盟以及开放算力生态圈。促进自主

软件创新,进一步推动工具软件及核心标准体系等的建设,构建战略均衡态势。可参考美国战略联盟模式,对我国龙头企业生态联盟进行建设,对于基础软件方面,可开展供需联合创新,对技术路线进行科学布局,促进产业化应用推广。对于金融等核心领域,可实施应用示范工程,促进操作系统及中间件等的自主创新,引领大型应用软件的发展。伴随社会的持续发展,增强产品成熟度及应用兼容性等。借助开源基金会,建设版权清晰及兼容开放的原创开源社区。

强化科研及人才布局,提高竞争力。从算力方面来看,为处理卡脖子困境,应加快突破紫外光刻机等生产设备及材料的壁垒,从长期发展来看,还应加强对基础领域的研究,促进产学研合作,提高人才培育力度。比如,强化数字及新材料等方面人才的培育,促进持续创新,为其奠定人才及技术基础,事先布局量子计算等前沿领域,实现颠覆性创新^[4]。同时还应鼓励高校与龙头企业合作,建立重点实验室以及研究开发等平台。全面招引人工智能人才,培育交叉复合型人才,进而促进算力资源产业化发展。

建立开放算力市场,培育数据交易组织。根据人工智能算法还有异构数据处理需要,基于算力资源普查之上,第一时间给予社会算力资源信息,实现算力市场的优化,从企业云计算垄断变成企业协同供给以及共享。在这一过程中,还应建立基于人工智能的算法交易平台,对典型场景算法进行运用,对细分领域实施算法导流,促进市场推广运用。构建优质的公共数据集及算法集,促进数据归集,构建一致的数据格式及存储安全标准等,对数据会对利益主体产生的影响进行全面分析,参考欧盟数据治理相关办法,促进数据中介机构的建设,对各类主体在数据开发过程中的利益进行平衡,促进数据驱动系统的建设,从而实现高效的数据交易服务。

五、结论

总而言之,面对算力竞争的形势,要有效发挥中国体制优势,把算力产业当作新基建与数字经济建设的核心战略任务,对核心开发项目、产业链规划、区域协调发展以及基建投资开展统筹谋划,以更好推进算力产业发展。

参考文献:

- [1]章玉贵.算力经济发展的重要功能与战略思考[J].人民论坛·学术前沿,2023,(05):101-107.
- [2]朱民.跨越ChatGPT:大模型和AI的未来[J].新经济导刊,2023,(02):4-7.
- [3]沈阳,钟培源,杨晓彤.激发“数”活力“算”出新未来——首届“西部数谷”算力产业大会综述[J].宁夏画报,2022,(09):22-29.
- [4]刘艳,付丽丽.AI大模型:为产业智能化升级“开闸放电”[J].大众投资指南,2022,(06):39-40.