

大数据与决策研究

(政策与技术跟踪专题)

2021年第8期(总第51期)

广西壮族自治区信息中心
广西壮族自治区大数据研究院

2021年3月31日

编者按：人工智能基础设施的建设，既是促进产业创新能力的重要手段，也是拉动新一轮经济和社会跨越式发展的新引擎。本期介绍人工智能基础设施的重要组成部分——人工智能计算中心的相关技术情况。

本期要目

- ◆ 人工智能计算中心的概念、演进、建设现状以及发展趋势
- ◆ 人工智能计算中心的功能定位、对比分析和应用场景

人工智能计算中心的概念、演进、建设现状 以及发展趋势

一、人工智能计算中心的概念

人工智能计算中心是基于人工智能芯片构建的人工智能计算机集群为基础，涵盖了基建基础设施（机房基建）、硬件基础设施和软件基础设施的完整系统，主要应用于人工智能深度学习模型开发、模型训练和模型推理等场景，提供从底层芯片算力释放到顶层应用使能的人工智能全栈能力。



图 1 人工智能计算中心重要组成

人工智能技术的迅速发展，对人工智能计算中心提出规模化建设需求。一方面人工智能算法愈发复杂、模型规模不断提升，图片、语音、视频等非结构化数据爆炸式增长，另一方面人工智能与 5G、物联网等行业领域结合落地，使人工智能的发展对算力的需求呈现指数级增长。人工智能计算中心的建设和发展，除了满足日益增长的算力需求，同时为大规模算法和模型的基础理论研究、实时复杂的智能化计算

引擎发展、人工智能应用的商业落地、关键共性技术的研发创新等方面形成条件支撑，并将一同促进人工智能硬件、软件 and 智能云服务之间相互协同的生态链发展。

二、人工智能计算中心的演进

计算是人类能力的延伸，算力的建设和社会的发展需求紧密结合，在不同历史阶段出现了超级计算中心、云计算数据中心、人工智能计算中心等不同形态的算力基础设施。

上世纪 60 年代开始，为了对重大军事研究和科学问题进行计算模拟，超级计算机和超级计算中心诞生。2000 年互联网产业兴起，以及 2007 年大数据技术和云计算技术的兴起，带动了云计算数据中心的建设。云计算数据中心面向个人或企业提供包括虚拟机计算能力、数据储存和网络传输宽带等能力，以支撑从电子商务到电子政务等方面的云服务。随着 2012 年以来新一代人工智能技术的快速发展和突破，以深度学习¹计算模式为主的人工智能算力呈指数级增长，计算机视觉²、自然语言处理等面向人工智能的处理场景越来越多，对专用定制化人工智能算力的需求大量涌现，专门的人工智能计算中心在近年来进入人们的视野。

人工智能计算中心借鉴了超级计算中心和云计算数据中心大规模并行和数据处理的技术架构，但以人工智能专用芯片为计算算力底座，同时软件架构和业务架构也与前两者

¹ 深度学习是指多层神经网络上运用各种机器学习算法解决图像，文本等各种问题的算法集合。

² 计算机视觉指用摄影机和电脑代替人眼对目标进行识别、跟踪和测量等机器视觉，并进一步做图形处理，使电脑处理成为更适合人眼观察或传送给仪器检测的图像。

完全不同，是当前人工智能快速发展和应用所依托的新型算力基础设施。

2012 年以来，以深度学习为代表的新一代人工智能技术得到快速突破和应用，并逐渐成为最重要的计算算力资源需求之一。一方面，传统的云计算数据中心及高性能计算中心，呈现出智能化服务或智能化算力的建设趋势，一定程度上提供了人工智能发展所需的算力。另一方面，以人工智能算力为主的人工智能算力中心应运而生，能够提供人工智能计算范式所需的专用算力，配合少量的通用算力以进行数据预处理和其他任务，从而能够以较低的成本提供高效的人工智能专用算力，为计算基础设施带来了新的建设方式。人工智能计算中心基于人工智能芯片构建人工智能计算系统，主要应用于人工智能模型开发、模型训练和推理服务场景。

三、人工智能计算中心的建设现状

随着人工智能计算需求的指数级增长，人工智能算力的成本也同步高涨。MIT 计算机科学家指出：深度学习正在逼近现有芯片的算力极限；计算能力提高 10 倍相当于三年的算法改进；算力提高的硬件、软件和金钱成本将越来越高。

人工智能可以建立超越专家模型，也带来昂贵的算力成本，其所需的硬件设备和算力，背后消耗的是巨额资金。在计算机视觉领域，将 Efficientnet 训练到需求精度，按照英伟达 V100 GPU 的成本估算，将需要 172032 美元；在自然语言处理领域，将 Transformer 训练到所需精度，将需要 3840

美元，而到了 2019 年的 Bert 模型，训练到所需精度花费将达到 15360 美元。华盛顿大学的 Grover 假新闻检测模型两周的训练费用约为 25000 美元。著名人工智能非营利组织 OpenAI 花费高达 460 万美元训练其 1750 亿参数的 AI 模型——GPT-3 语言模型，而 GPT-2 语言模型每小时的训练花费达到了 256 美元。算力门槛的提高，导致很多大学、研究机构的中小团队很难获得这种算力科研条件，同样将大量中小企业挡在门外。

因此，具备训练复杂先进模型和处理海量数据能力的人工智能计算中心属于投资较大的信息基础设施，是包含了机房基建、硬件基础设施和软件基础设施的大规模的系统工程，当前的建设模式和现状主要为政府主导建设和头部企业自行建设。

四、人工智能计算中心的发展趋势

（一）全栈一体趋势：专用人工智能芯片与软硬件协同优化提升计算效率。

人工智能带来的算力需求已经远超摩尔定律。各类人工智能加速芯片适应人工智能的算法特征，进行矩阵元操作的并行化加速，或进行对特定人工智能计算任务的精简优化，发展方兴未艾。我国人工智能芯片起步较晚，但发展较快，当前华为、寒武纪等已推出商用人工智能芯片，还不断有新的人工智能芯片出现。

不同的人工智能芯片设计与实现方式不同，当前的发展

趋势是需要人工智能芯片厂商或社区开发对应的软件进行精细化匹配，以发挥硬件的最大算力。谷歌把 Tensorflow 与其人工智能专用芯片 TPU 绑定式设计协同优化；英伟达的 CUDA AI 开发框架将 GPU 与上层软件优化衔接，充分挖掘和发挥 GPU 的硬件潜力。华为推出 MindSpore AI 开发框架通过 On-device 特性充分发掘其 AI 专业芯片 腾芯片的硬件潜力。

由此可见，人工智能计算范式定义人工智能芯片，软硬件协同提升计算效率成为当前人工智能计算发展的新特征，人工智能专用算力与配套软件全栈一体建设成为必然。

（二）技术融合趋势：超级计算与人工智能融合，云与人工智能融合。

超级计算的业务需要大量的人工智能算力，超级计算中心拥有支撑人工智能的能力已经是一种趋势。人工智能正在改变传统超级计算的求解方法，将人工智能技术融入超级计算系统，可以提高准确性、加快时间并降低成本。在应用驱动下，人工智能算法在医疗诊断、天文探测、地震预测等领域快速发展。

随着云计算的发展和成熟，以云化方案构建人工智能服务，对用户统一架构、统一服务和统一 API³，向用户屏蔽复杂的人工智能技术细节，降低了人工智能服务的使用门槛。目前云提供商如华为云、AWS 是这一趋势的主要推动者，

³ 应用程序编程接口

纷纷推出云上高性能人工智能计算和人工智能使能平台服务，单个用户可以创建数千处理器规模的高性能人工智能计算资源满足高效人工智能开发。云平台同时也带来了人工智能计算中心运营模式的改变，通过云上租户粒度的安全隔离、完善的运维运营系统，人工智能计算中心可以为不同用户提供安全可靠、按需使用、弹性伸缩、有服务等级保障的自助式服务。云计算中心提供裸金属服务器、虚拟机、容器等多样化的算力资源和人工智能使能平台服务，人工智能服务与云上大数据、物联网、边缘计算等服务的相互协同，满足新型应用场景综合复杂多层次的计算需求。

随着人工智能的快速发展，超级计算、云计算和人工智能技术不断融合发展，人工智能为超级计算提供新的计算求解方法，使其可以利用长期积累的大量观测数据，云计算一方面为人工智能提供算力和新的运营和赋能方式，一方面将人工智能能力通过云与边缘计算、物联网等结合，推动人工智能在云边端全场景的应用。

（三）平台赋能趋势：人工智能计算中心赋能企业，形成算力生态。

平台是可跨情境应用的资源和能力的集合，人工智能计算中心作为集超级算力和海量数据的超级大脑，已经呈现平台化发展的趋势。具备强大软硬件能力的核心企业集聚研发能力、生产经验和产业资源，在人工智能计算中心搭建基础应用平台，并依托平台的共享输出上层应用使能能力，对平

台上的小型人工智能企业和欠缺人工智能能力的传统企业进行赋能。人工智能计算中心将成为人工智能核心企业和大量初创企业能力输出的主要方式，如通过平台开放接口的方式输出龙头企业的算法能力，资源、数据支撑、运营辅导和模式优化等。人工智能计算中心逐渐构建起人工智能的生态创新架构，助力各类架构缔结产业联盟，聚拢上下游资源，吸纳高校和科研机构，提供基础研究课题和依托平台，推动专家合作、举办生态研讨等人才培养活动，为产业发展提供人才支持，最终形成算力生态，大量生态互补者协同推进人工智能产业的开发与应用。

综上，人工智能计算中心建设之后必然会形成上层应用使能能力，溢出对周边企业进行赋能，形成科研和人才培养的算力生态，应用使能能力也成为人工智能计算中心建设的核心指标和核心竞争力之一。（中国科学技术信息研究所《人工智能计算中心发展白皮书》）

人工智能计算中心的功能定位、 对比分析和应用场景

一、人工智能计算中心的功能定位

人工智能计算中心是智能时代面向社会全域多主体的新型公共基础设施，集算力生产供应、数据开放共享、智慧生态建设和产业创新聚集四大功能于一体，为有海量数据存储、处理、分析及应用支撑需求的各类场景提供载体支撑。

一是作为算力生产供应平台。AI 计算是智能时代发展的核心动力。人工智能计算中心以数据为资源，以强大算力驱动 AI 模型对数据进行深度加工，源源不断产生各种智慧计算服务，面向全行业领域提供基于深度学习技术的人工智能算法能力、算法框架和相关接口，为政府、企业及科研院所等多方用户提供生产生活各领域智慧服务，发挥新型基础设施的社会价值，降低社会服务成本，让智慧计算服务更快的普及到每个人、每个企业。

二是作为数据开放共享平台。人工智能计算中心是新型公共基础设施，通过全量汇聚各行业领域数据资源，开放共享全面提升 AI 算法训练数据质量。同时，随着数据汇聚共享能力的提升，通过跨领域数据的多次开发利用，以数据流引领技术流、业务流、资金流、人才流等聚集，深度分析挖掘应用需求，使沉淀的数据资源在各个应用场景中实现价值

最大化。

三是作为智能生态建设平台。人工智能计算中心是集人工智能、大数据、云计算等多种技术和 AI 算力服务、数据服务和模型服务于一体的新型 IT 基础设施。其广泛应用将加速推动产业 AI 化和 AI 产业化，以智能算力生态聚合带动形成多层次产业生态体系，赋能多个产业、惠及多类主体，助推数字经济与传统产业深度融合，加速孵化新业态。

四是作为产业创新聚集平台。以 AI 算力生产供给为核心的人工智能计算中心，面向政府、企业及科研机构等多主体，围绕数据、算法和算力三大要素着力构建 AI 全产业链。其通过生产、聚合、调度和释放算力推动 AI 产业要素资源聚集，汇聚不同主体资源优势打造产业创新聚集平台，在政府主导下，科创企业、科研机构和传统企业发挥各自在 AI 方面的技术优势、研发优势和场景优势，加速 AI 应用场景落地，助力传统产业转型升级，催生经济新业态新模式，优化公共服务供给。

二、人工智能计算中心的对比分析

人工智能计算中心是伴随着数据规模指数级增长、算力需求指数级增加、业务服务能力需求不断升级、AI 技术日趋成熟并广泛落地等内外部环境变化，而逐渐兴起的一种新型计算中心。在建设目的、技术标准、具体功能、应用领域和“投一建一运”模式等方面，与超算中心、云数据中心相比，人工智能计算中心都有所差别，对比如图 2 所示。

主要指标	超算中心	云数据中心	智算中心
建设目的	面向科研人员和科学计算场景提供支撑服务	帮助用户降本增效或提升盈利水平	促进AI产业化、产业AI化、政府治理智能化
技术标准	采用并行架构，标准不一，存在多个技术路线，互联互通难度较大	标准不一、重复建设 CSP内部互联、跨CSP隔离 安全水平参差不齐	统一标准、统筹规划 开放建设、互联互通互操作 高安全标准
具体功能	以提升国家及地方自主科研创新能力为目的，重点支持各种大规模科学计算和工程计算任务	能以更低成本承载企业、政府等用户个性化、规模化业务应用需求	算力生产供应平台、数据开放共享平台、智能生态建设平台、产业创新聚集平台
应用领域	基础学科研究、工业制造、生命医疗、模拟仿真、气象环境、天文地理等	面向众多应用场景，应用领域和应用层级不断扩张，支撑构造不同类型的应用	面向AI典型应用场景，如知识图谱、自然语言处理、智能制造、自动驾驶、智慧农业、防洪减灾等
“投-建-运”模式	政府科研单位投资建设运营	行业巨头或者政府投资建设 其它用户按需付费使用；以数据服务盈利，企业自主运营	政府主导下的政企合作共建模式 政府出资指导建设，企业承建运营

图 2 人工智能计算中心与超算中心、云数据中心之间的差别

三、人工智能计算中心的应用场景

应用场景 1: 识别检测

识别检测通常包括图像分类、目标定位跟踪、对象检测和对象分割等。近年来，由于深度学习技术的突破性发展，使得识别检测的表现性能有了大幅提升，对于图像分类等识别检测任务，基于多层神经网络训练后形成的模型已经能够超越人类的识别能力。识别检测正逐渐从先进技术发展成为成熟产业，被广泛应用于政府、银行、医疗、社会福利保障和电子商务等领域。尽管取得了巨大成功，但识别检测技术仍然面临着很多挑战，需要人们不断开发更加先进的学习算法，提高模型生成和测试效率，以适应不断变化和更加复杂的应用场景对识别能力的需求。

应用场景 2: 语音交互

语音交互是一个包括了语音识别、自然语言处理和语音合成的融合性 AI 技术。语音识别技术将声音转化成文字，自然语言处理技术对文字进行解释，语音合成技术将输出信息转化成声音。目前，常用的社交软件输入法、搜索工具、语音指令控制、新闻阅读推荐和智能客服等领域都在应用语音交互技术。然而，作为极具挑战性的 AI 技术之一，语音交互技术自身面临着语音识别、语义分析和多轮对话等技术难点，语音交互的工程实现需要融合使用先进的隐马尔可夫模型⁴、深度学习和先进的语音信号处理等技术，以及超级强大的训练数据存储、计算和处理能力。

应用场景 3: AI 芯片

近年来，人工智能芯片及相关技术得到高速发展，面对不断增长的市场需求、多样的 AI 计算任务和性能要求，各类专门针对人工智能应用的设计理念和创新架构不断涌现。AI 芯片需要具备高度并行的处理能力、低内存延迟和新颖的架构，以实现计算元件和内存之间灵活而丰富的连接，除此之外，还要考虑功耗和能效管理水平。在当前人工智能各领域的算法和应用还处在高速发展和快速迭代的阶段，针对特定领域而不针对特定应用的设计，将是 AI 芯片设计的一个指导原则，具有可重构能力的 AI 芯片可以在更多应用中广泛使用。

⁴ 隐马尔可夫模型是统计模型，它用来描述一个含有隐含未知参数的马尔可夫过程（在给定现在状态时，它与过去状态是条件独立的，那么此随机过程即具有马尔可夫性质。具有马尔可夫性质的过程通常称之为马尔可夫过程）。

应用场景 4: 自动驾驶

自动驾驶作为 AI 产业化的典型代表，近年来得到广泛的关注。由自动驾驶技术的特性可知，其开发部署需要人工智能硬件架构提供高性能的 AI 算力支持，需要深度学习等高级 AI 算法和模型的快速构建来支撑“自动驾驶”技术和功能实现，需要智能计算平台与操作系统、虚拟仿真试验平台等软硬件平台紧密协作。从自动驾驶产业链发展的角度看，位于产业链中游的整车生产等关键环节还需要建设智能化网联驾驶测试与评价平台（智能汽车风洞测试实验室）和自动驾驶模拟仿真测试服务平台等关键平台来有效支撑自动驾驶产业链的完善和升级。

应用场景 5: 机器人

机器人产业更需要人工智能技术的全面赋能。机器人包括工业机器人、特种机器人和服务机器人。机器人在安全风险较大的电力、交通和能源等领域的应用十分广泛。智能电力巡检机器人集探测器、无轨化激光导航定位、红外测温、智能读表、图像识别等技术于一体，对输变电设备进行全天候巡检、数据采集、视频监控、温湿度测量和气压监测等，提高输变电站内设备的安全运行程度。

应用场景 6: 智能制造

根据《智能制造发展规划（2016—2020 年）》，智能制造是基于新一代信息通信技术与先进制造技术深度融合，贯穿于设计、生产、管理、服务到制造活动的各个环节，具有

自感知、自学习、自决策、自执行、自适应等功能新型生产方式。制造业迫切需要采用人工智能技术作为产业革新的重要抓手，其中，质检分拣作为工业最关键的环节，其 AI 化的改变一直被寄予厚望。长期以来，传统工业质检面临两大痛点，即传统的检测手段效率低下，产品质量无法保障，且因工作枯燥、人口红利消失，人工质检存在“用工难”现象。因此越来越多的工厂开始使用机器视觉技术代替人工产品质量检测。基于人工智能深度学习算法的智能质检设备是由数据驱动的、自发的机器学习，通过图像处理算法快速识别产品表面存在缺陷的不良品，并结合产线自动化，实现精准归类。目前智能机器还无法赶上人眼的视觉能力，因此全球质量检查市场的机器视觉覆盖率不到 5%，仍然存在许多未解决的问题。与传统视觉技术在识别不规则缺陷方面的不足相比，随着数据量的增加，人工智能检测的准确性将继续优化，实现对生产质量数据的完全控制，并为过程优化和过程提供关键数据支持。

应用场景 7：医疗影像

应用人工智能技术，参与疾病的筛查和预测，从医疗影像检查结果中进行判断以获取诊断信息，辅助 CT 影像识别、颈椎病症识别、癌症识别等场景应用。新冠肺炎疫情的爆发加速暴露了当前医疗体系中存在的问题。在人工智能时代，传统算力设施已无法支撑对于大量医疗影像进行机器自主学习、快速判断结果的需求。为了更好地应对公共卫生突发

事件，让检查更加便捷、迅速，从而提供更高质量的医疗服务，需要应用智能计算中心的算法算力，对多层神经网络的模型进行训练，通过测试影像数据确定真实的特征目标，实现高准确度的分类诊断，提高诊断速度与工作效率。

应用场景 8：无人商店

目前中国零售业正在呈现线上线下融合发展、新业态新模式不断涌现的发展趋势，以无人商店为代表的智慧零售新业态越来越受欢迎。无人商店主要应用人工智能卷积神经网络⁵、深度学习、机器视觉以及生物识别等人工智能领域前沿技术，可实现扫码进店、直接购物、自动结账、离店后付款等无人店全智能化操作。利用深度学习的神经网络，无人商店可以记录每个客户的消费习惯，甚至抓取用户信息，实现“人店对话”。利用机器视觉技术、生物识别技术可以识别消费者身份。通过机器视觉、射频技术、多路摄像头和传感器能够精准识别商品、捕捉消费者运动轨迹和消费行为。但是随着店铺面积增大、人流量增多，无人商店需要更强大的计算能力、更精准的认识要求，无人商店行业迫切需要在新型智能化算力基础设施的支撑下快速兑现对整个行业的变革和引领。

应用场景 9：智能客服

智能客服即通过运用智能化的语音识别技术、自然语言处理技术、知识库管理技术、语音合成技术，更好地帮助客

⁵ 卷积神经网络，是一种前馈神经网络，人工神经元可以响应周围单元，可以进行大型图像处理。卷积神经网络包括卷积层和池化层。

户分析问题，通过人声模拟或文字给予用户明确的答复或相应的建议。传统的客户服务中心以电话呼叫中心为主，随着人工智能技术的不断成熟与用人成本的逐步提高，更多政企部门加大了对智能呼叫中心的投资规模，积极拓展更加经济高效的智能客户服务渠道，如在线客服等，应用以领域知识库建设为核心，通过文本或语音等方式交互的智能客服机器人系统，有效整合了多渠道的客户服务中心，能够大幅缩减客服成本、增强用户体验，从而提升服务的质量和企业创新的品牌形象。

应用场景 10: 智慧物流

2020 年的新冠肺炎迫使许多工厂被迫停工，令无数企业意识到无人化或人机协同生产线的重要性。在新基建大背景下，智慧物流将成为我国促进经济发展内循环的有力抓手。2020 年可以定义为“AI+物流”的元年，智慧物流盛宴刚刚启幕。电商巨头、初创公司和传统企业纷纷布局，抢占物流行业巨大市场份额。如何帮助物流企业快速研发和生产智慧物流解决方案和技术产品与服务，已经成为众多企业面临的直接挑战。这需要智能计算中心提供高效算力支撑中小微企业、初创高科技公司和 AI 技术公司快速研发更加先进和智能化的物流机器人技术、产品，降低企业进入智慧物流行业的门槛，进一步促进智慧物流生态体系构建，不断优化物流工作效率。智慧物流的应用还要求高可靠规划，保证最优稳定的规划方案；精准控制，精确准时的控制设备协同作业；

高效能装备，保障高效安全的储存、拣选及配送。

应用场景 11：智慧农林

农业智慧化转型升级是实现我国经济高质量发展的重要内容，人工智能技术可以全面赋能农林业生产、助力生态防治。农业种植方面，通过深度学习算法分析农作物高清图像，可以实现大规模、低成本、自动化的病虫害识别及监测预警。林业生态防护方面，利用无人机、智能图像识别技术和高速的数据处理能力，监控、分析、处理大量实时数据，在林草火灾防治、林草有害生物防治、沙尘暴防治、野生动植物疫源疫病监测防控等领域能够实现智能监测、智能预警和智能防控。农业养殖方面，高清监控系统配合图像识别技术对畜禽进行面部和行为识别，可以自动分辨畜禽情绪、进食状态和健康情况，及时反馈给养殖户并提供养殖建议。大数据配合人工智能技术可以实现鱼群数据的精准预测，自动定位捕捞，提高捕捞准确率，并有效防范过度捕捞现象，促进海洋资源的可持续发展。然而，加速智慧农林落地，实现农林业的集约化、智能化生产主要面临农林渔牧数据种类繁多、数据规范化程度不高，农业基础设施现状无法满足新型农业生产模式的发展，新技术的应用深度不够等问题。（国家信息中心《智能计算中心规划建设指南》）

广西壮族自治区信息中心
广西壮族自治区大数据研究院

编辑部地址：南宁市体强路 18 号广西信息中心 1412 号房

联系电话：0771-6113592

电子邮箱：dsjyjs@gxi.gov.cn

网 址：<http://gxxxzx.gxzf.gov.cn/>



扫描二维码获取
更多决策参考信息