

# 智能计算中心发展态势研究



## Advances of Intelligent Computing Center

焦奕硕/JIAO Yishuo, 邸绍岩/DI Shaoyan

(中国信息通信研究院, 中国 北京 100191)  
(China Academy of Information Communication Technology, Beijing  
100191, China)

DOI: 10.12142/ZTETJ.202303011

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20211018.1755.002.html>

网络出版日期: 2021-10-19

收稿日期: 2021-09-15

**摘要:** 智能计算中心是中国新型基础设施建设的重要组成部分, 是满足人工智能算力需求、支撑国家智能化转型的重要力量之一。智能计算中心已得到国家和业界高度关注, 将会迎来规模化建设和应用。智能计算中心作为智能算力的系统集成和融合创新平台, 将从需求侧拉动各类人工智能技术创新进步。此外, 智能计算中心将大大降低智能算力的使用门槛, 推动人工智能普惠化发展, 助力各行业智慧化转型升级。

**关键词:** 智能计算中心; 人工智能; 算力基础设施; 新基建

**Abstract:** The intelligent computing center is an important part of China's new infrastructure construction, and it is one of the important forces to meet the needs of artificial intelligence (AI) computing power and support the country's intelligent transformation. The intelligent computing center is highly concerned by the country and the industry and will usher in large-scale construction and application. As the systematic integration of various kinds of intelligent computing devices and techniques, the intelligent computing center will lead to the creative development of AI techniques. In addition, the intelligent computing center will greatly reduce the threshold of the use of intelligent computing power, promote the inclusive development of artificial intelligence, and help the intelligent transformation and upgrading of various industries.

**Keywords:** intelligent computing center; artificial intelligence; computing infrastructure; new infrastructure construction

**引用格式:** 焦奕硕, 邸绍岩. 智能计算中心发展态势研究 [J]. 中兴通讯技术, 2023, 29(3): 59-63. DOI:10.12142/ZTETJ.202303011

**Citation:** JIAO Y S, DI S Y. Advances of intelligent computing center [J]. ZTE technology journal, 2023, 29(3): 59-63. DOI: 10.12142/ZTETJ.202303011

随着国民经济各行业向数字化、网络化和智能化的加速转型, 人工智能技术不断发展, 应用场景不断拓展。然而, 人工智能模型训练所需的海量数据和大规模算力已逐渐超过了多数企业的能力范围。这与日益增长的人工智能应用需求产生了较大矛盾。建造智能计算中心, 能够大规模生产智能算力, 并按需为人工智能应用落地提供算法、算力和数据服务。智能计算中心越发受到地方政府关注, 有望成为支撑和引领数字经济、智慧社会发展的关键新型基础设施之一。

## 1 智能计算中心发展背景

### 1.1 智能计算中心是满足人工智能算力需求的重要途径

随着人工智能在各行各业的加速落地, 数据处理需求加速增长, 算法模型复杂化程度不断提升, 智能算力需求也急速增长。据 OpenAI 的统计数据, 自 2012 年以来, 人工智能模型所需的计算量增长了 30 万倍。目前, 暴力计算仍是人

工智能解决一些应用问题最有效的手段。神经网络的能力边界有待于进一步探索, 预计未来会需更高的算力。在全球引起轰动的 ChatGPT 参数量达到了 1 750 亿, 微软甚至专门为 OpenAI 搭建了人工智能计算中心来支撑相关模型所需算力。然而, 除了以微软为代表的顶级科技巨头, 绝大多数企业无法承担人工智能计算中心的建设和运营费用。据互联网数据中心 (IDC) 调研, 74.5% 的企业期望采用具有公共设施意义上的人工智能计算中心。那么, 通过统一建设高性能、大规模的智能计算中心, 并面向公众以服务形式提供算力, 成为解决该问题的重要途径。

### 1.2 智能计算中心是“新基建”国家战略的重要组成部分

2018 年底的中央经济工作会议提出了“新型基础设施建设” (后简称为“新基建”)。“新基建”的概念正式进入公众视野。2020 年 4 月, 国家发展和改革委员会 (后简称为“国家发展改革委”) 在新闻发布会中进一步明确了“新基建”的范围, 提出“新基建”包括信息基础设施、融合基础

设施、创新基础设施3个方面。其中,信息基础设施包括以5G、物联网、工业互联网、卫星互联网为代表的通信网络基础设施,以人工智能、云计算、区块链等为代表的新技术基础设施,以数据中心、智能计算中心为代表的算力基础设施等。2020年5月,国家发展改革委在《关于2019年国民经济和社会发展规划执行情况与2020年国民经济和社会发展规划草案的报告》中提出,制定加快新型基础设施建设和发展的意见,并实施全国一体化大数据中心建设工程,在全国布局10个左右区域级数据中心集群和智能计算中心。工业和信息化部在2021年8月发布的《新型数据中心发展三年行动计划(2021—2023年)》中提出,要加快推进边缘数据中心、智能计算中心等标准建设,支撑新技术新应用落地。2022年1月,国务院印发的《“十四五”数字经济发展规划》中提出,推动智能计算中心有序发展,打造智能算力、通用算法和开发平台一体化的新型智能基础设施。

### 1.3 智能计算中心建设开始在全国各地落地

以广州、武汉、济南等为代表的各地纷纷宣布启动智能计算中心建设。2020年4月,广州市政府出台的《广州市加快打造数字经济创新引领型城市的若干措施》提出,面向人工智能和5G应用场景,建设基于图形处理器(GPU)的人工智能、区块链算力中心。2020年10月,武汉宣布启动武汉人工智能计算中心建设,计算中心将围绕武汉创建国家新一代人工智能创新发展试验区,助力武汉市智能制造、智慧医疗、智能数字设计与建造、智能网联汽车等产业发展。2020年11月,国家信息中心联合浪潮发布了《智能计算中心规划建设指南》,明确了智能计算中心的概念和架构等。浪潮集团在济南市积极筹建智能计算中心,并通过配套建立培育产业链群的专业园区,打造“中国算谷”人工智能算力平台,支撑济南建设国家新一代人工智能创新发展试验区。

## 2 智能计算中心的内涵和主要特征

### 2.1 智能计算中心的内涵

综合对比产业界对智能计算中心的理解,智能计算中心的概念可总结为:基于领先人工智能计算架构,为公众提供人工智能应用所需的算力服务、数据服务和算法服务的新型基础设施。不同于企业自建的人工智能算力中心,智能计算中心的定位是公共设施,是中国新型基础设施的有机组成部分。智能计算中心一般包含基础支撑、核心功能和提供服务等要素。其中,基础支撑主要指运行人工智能计算的相关芯片和算法等;核心功能主要指算力的生产、聚合、调度和释

放等作业环节以及支持相关环节实现的软硬件平台;提供服务主要是指计算中心对外提供的算力、数据和算法等相关服务<sup>[1]</sup>。

### 2.2 智能计算中心的主要特征

1) 开放合作。智能计算中心的公共属性决定了其建设多由政府主导、筹划,技术密集属性决定了其具体建设运营须由相关科技企业或科研机构执行。智能计算中心的建设、运行须产学研用开放合作,协同推进。

2) 创新融合。为提供领先的算力和算法等服务,智能计算中心须采用最新的技术理念,通过硬件重构和软件定义等创新技术来实现多种资源和技术要素的协同和融合。

3) 生态协同。智能计算中心的基础设施属性决定其核心功能之一是为各行各业提供智慧化转型支撑。智能计算中心需面向行业发展需求,基于算力、算法和数据等核心资源的汇聚,开展技术研发、成果转化和落地等工作,进一步吸引业务、资金和人才等创新要素集聚,共同培育智能产业生态<sup>[2-3]</sup>。

### 2.3 智能计算中心建设的主体力量

目前,智能计算中心正处于发展的起步阶段,政府、企业和学术机构均积极响应,并以多方合作的形式在智能计算中心建设领域进行探索。

1) 地方政府。以武汉、济南等为代表的地方政府均在顶层规划时就对智能计算中心进行布局,与企业开展合作,通过建设配套产业园区和人才培养平台等方式促进智能计算中心发展。

2) 业界企业。以华为和浪潮为代表的信息通信技术(ICT)基础设施企业凭借其物理设施建设优势,通过承建智能计算中心,搭建产业合作平台,打造人工智能算力生态。以寒武纪为代表的人工智能领域企业依托专精优势,通过成立合资公司等形式参与智能计算中心的建设和运营,借助智能计算中心平台扩大自有生态优势。

3) 学术机构。以国家信息中心、中国科学技术信息研究所等为代表的学术研究机构和企业合作发布智能计算中心相关白皮书,定义智能计算中心概念框架,提供建设建议,助力智能计算中心生态建设。

## 3 智能计算中心的核心架构和技术特征

### 3.1 智能计算中心的代表性架构和关键技术

当前,智能计算中心处于探索阶段,尚未开展规模化建

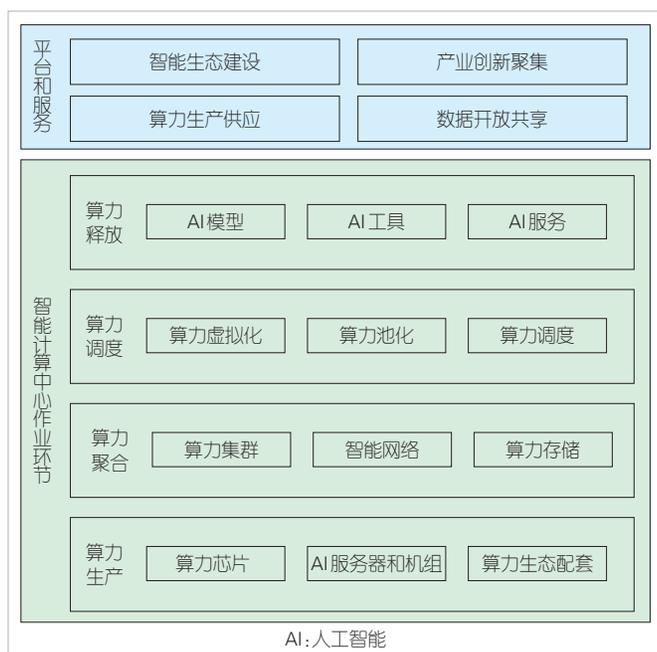
设, 因此暂未形成成熟稳定的技术指标体系。然而, 从定量层面上看, 智能计算中心以算力为主要技术指标。当前较多业界专家认为发展良好的智能计算中心所能提供的算力应在 5~10 EFlops, 与此同时, 其电源使用效率 (PUE) 大概为 1.3~1.4。从定性层面上看, 当前业界针对智能计算中心的架构和相关核心技术已达成一些共识, 其中以国家信息中心所发布的《智能计算中心规划建设指南》中所建议的架构最具有代表性<sup>[4]</sup>, 具体如图 1 所示。

算力生产环节主要包括以芯片为核心的各类硬件, 以及和其相配套的基础软件、基准性能测试等配套生态, 它们是智能计算中心对外提供服务和算力的基础支撑环节。算力生产环节的核心技术有中央处理器 (CPU)、GPU、现场可编程门阵列 (FPGA)、专用集成电路 (ASIC) 等, 以及将 GPU、FPGA 和 ASIC 各类算力进行系统集成的 AI 服务器和算力机组。算力聚合环节通过高速互联网络将各计算节点、存储节点进行连接, 针对智能计算多任务、大规模、高并发等特点, 构建高带宽和低延迟的数据汇聚交换平台, 从而将算力生产环节所产生的算力在计算中心内进行聚合。算力聚合环节的核心技术有数据中心网络和大数据混合加速等。数据中心网络又可分为节点内互联和节点外互联。寒武纪所研发的思源 MLU-Link 技术可以实现节点内 600 Gbit/s 级别的互联速率, 节点外互联以 Infiniband 为代表, 速率能够达到 200 Gbit/s。此外, 华为发布了 AI Fabric 智能无损数据中心网络方案, 通过将无损网络拥塞控制算法和网卡等硬件集成

协作, 承载远程直接数据存取 (RDMA), 形成无时延、无吞吐损失和丢包损失的开放以太网。为解决数据访问慢及数据多平台重复存储等性能瓶颈问题, 大数据混合加速主要针对存储和计算分离, 可在计算节点下构建虚拟数据湖, 在计算框架和存储框架中间增加中间缓存层, 从而实现缓存效率提升以及数据快速定位读取, 整体降低数据存储和访问成本<sup>[5]</sup>。

基于云计算领域虚拟化和容器编排调度等技术, 算力调度环节将算力进行精准调配, 保证上层平台和服务的算力供应。算力调度环节的核心技术是 GPU 虚拟化以及虚拟 GPU 调度。其中, GPU 虚拟化是指将一块物理 GPU 通过虚拟化技术虚拟成为多块虚拟 GPU (vGPU), 通过将 vGPU 纳入统一管理并在多个工作负载中统一调配, 实现算力资源池化; 虚拟 GPU 调度多基于云原生核心技术 Kubernetes, 通过容器编排调度将算力更为优化地分配到有需求的节点, 从而减少资源空置, 提升整体效能和稳定性, 实现服务能力的最优配置<sup>[5]</sup>。

算力释放环节是指在算力生产、算力聚合和算力调度的基础上, 面向各类应用场景需求, 基于最新的 AI 理论和算法, 为客户提供各类 AI 模型、服务和算力的过程。由此, 我们可以构建人工智能技术和产业生态。算力释放环节的核心技术是 AI 算法、AI 服务以及各类相关配套工具。其中, AI 算法以深度学习为核心, 主要包括卷积神经网络、图神经网络、循环神经网络等模型结构, 以及反向传播优化算法等面向不同任务的新型模型算法; AI 服务是人工智能计算中心和下游应用厂商合作的主要渠道, 主要有模型接口、开发接口和在线服务等服务提供形式; 相关配套工具是为 AI 推理和训练提供数据处理等方面配套的软件工具, 如 Hadoop 用于大数据处理, OpenCV 用于图像处理等<sup>[5]</sup>。



▲图 1 智能计算中心架构

### 3.2 智能计算中心的新技术特征

作为新一代数据中心, 智能计算中心由于服务对象和工作模式有别于当前主流的超级计算中心和云数据中心, 因此在技术特征上也和传统数据中心有所区别。一是理论基础不同: 超算中心的顶层架构是基于并行计算的算法和设计理论, 云计算中心顶层架构基于虚拟化的算法和设计理论, 而智能计算中心则基于深度学习的理论进行顶层架构设计。二是所采用的计算架构不同: 智能计算中心采用不同于传统数据中心的、较为领先的人工智能计算架构。为向客户提供优质服务, 智能计算中心须采用最新的人工智能运算架构来提升运算和相应速度。当前, 芯片间互联和开放架构的设计是业界的技术研究热点。2019 年, 在美国举办的开放计算项

目(OCP)全国峰会上,微软、Facebook和百度联合制定了OCP加速器模块(OAM)标准。该标准用于指导AI硬件加速模块和系统设计,通过定义AI加速模块、主板、互联拓补等方面的设计规范,推动不同AI加速模块和系统建立互操作性,实现多计算节点间的高速互联通信,未来有望成为智能计算中心内规模化采用的重要技术。

## 4 智能计算中心的重要作用

### 4.1 智能计算中心对人工智能产业发展的支撑作用

1) 满足大规模预训练场景的算力需求。近来,大规模预训练成为人工智能的技术热点,通过大规模数据和超大模型算法训练通用模型,再在此基础上逐步缩小场景模型,可有效降低成本。智能计算中心通过搭载大量人工智能服务器,对算力进行大规模集中生产、聚合、调度和释放,实现算力的提升和快速交付,从而提高算法效率和演进节奏。

2) 推动人工智能普惠化发展。智能计算中心既可通过服务形式为有需求的企业提供算力支撑,省去企业投资建设和运营费用,又可通过平台开放接口的方式将行业领军企业的算法、数据资源及运营服务等创新要素输送给信息技术(IT)基础相对薄弱的企业,进一步降低人工智能使用门槛,助力各行业智慧化转型升级<sup>[4]</sup>。

### 4.2 智能计算中心对人工智能技术发展的支撑作用

1) 完善供需对接机制。智能计算中心可作为供需对接平台,从需求侧拉动核心技术发展。通过在智能计算中心搭载新兴芯片和算法等核心技术,面向应用场景提供服务,可为原创技术提供一个接受市场检验的快速通道,通过行业应用反馈形成可迅速迭代的良性闭环,助力芯片和算法的发展。

2) 促进技术创新融合。智能计算中心作为新技术融合发展平台,将大大促进融合架构技术发展。智能计算中心通过集成最新的人工智能加速芯片和存储介质等,成为各新兴计算单元进行大规模融合的重要载体,可从需求侧刺激硬件重构和软件定义等融合架构技术创新发展。通过推进平台、框架和算法的协同优化,打通人工智能软硬件产业链,智能计算中心加速人工智能算力技术和产业生态形成<sup>[6]</sup>。

## 5 智能计算中心发展所面临的主要问题

中国智算中心当前建设的整体框架、建设运营、生态构建等方面仍存在一些问题和挑战,业界需要加大协同力度,形成创新发展合力,共同推动中国智能计算中心高质量

发展。

### 1) 整体框架不统一,多点建设面临碎片化风险

智能计算中心的概念框架不清晰。从各地智能计算中心规划建设情况来看,一方面,智能计算中心的核心架构、服务内容等并不完全一致,建设主体思路存在一定差异;另一方面,多点建设未形成合力。智能计算中心建设和运营过程中的业务逻辑、软硬件规范和信息安全等方面仍缺乏一个较为通用的标准体系。多点建设过程存在属地化、碎片化风险,对跨区域协同创新和生态构建造成不利影响。

### 2) 建设和运营模式不成熟,资金压力较大

智能计算中心的建设多以政府为主导、企业合作投资的形式开展。而政府和企业的投资划分、建设和运营主体职能和权责大多处于探索阶段,尚缺乏清晰统一的界定。此外,智能计算中心建设和运营成本高昂,不论是政府还是企业投资均面临较大的资金压力。且智能计算中心当前暂未经过实践验证的成熟商业运营模式,难以保证建设运营的长期运行。

### 3) 应用尚未落地,生态构建任重道远

当前智能计算中心多处于概念和规划阶段,全国各地尚不存在稳定运行并面向公众提供服务的智能计算中心。智能计算中心的服务提供模式及其所能支撑的人工智能应用仍待进一步验证。智能计算中心在应用生态构建过程中,仍面临本身能力建设、地理位置选择、合作模式探索和合作平台搭建等多方面挑战。

## 6 智能计算中心发展建议

### 1) 加大研发力度,突破核心关键技术

面向重点应用场景计算需求,鼓励芯片企业、平台软件企业、行业解决方案企业开展协同创新,突破人工智能芯片、计算架构、平台软件和模型算法等智能计算中心核心技术,提升技术关键环节自主水平。提前布局前瞻性技术,在保障信息安全的基础上加大新兴引领技术先行先试力度。引导支持新型高效节能技术在各中心内应用,推动各中心低碳化、绿色化发展。

### 2) 优化建设模式,形成高效发展路径

加大政企协作力度,强化政策保障和要素支持,积极学习借鉴其他国家相关人工智能算力中心建设和运营经验,探索开放共赢的建设运营模式和多方协同的合作机制。支持建设和运营主体围绕数据脱敏开放、多主体收益分成等主题,积极探索新型商业模式,推动智能计算中心高质量、可持续发展。

### 3) 加快应用落地,引领塑造产业生态

强化需求驱动，组织行业用户梳理热点、痛点需求，支持有关行业组织加大供需对接力度，为智能计算中心提供更多应用场景。针对重点行业的特色应用开展应用示范，形成一批可推广的典型应用创新模式。引导有智能计算需求的企业积极接入智能计算中心，使用智能计算中心服务，加速企业集聚和数据共享。

#### 4) 完善标准制定，引导产业有序发展

加大企业、协会、高校和科研院所等行业主体的合作力度，加快制定架构体系、数据接口、信息安全、软硬件规范等方面的标准体系。引导各地新建智能计算中心间建立对接机制，形成各中心互联互通，避免出现信息孤岛。支持各智库、行业组织等第三方机构推进评估评测等方面的研究，找准各智能计算中心发展优劣势和各发展环节“堵点”，推动各行业、各地域间协调配合，平衡产业供需，优化地域布局，形成优势互补、错位发展的良性发展格局。

## 7 结束语

随着以大模型为代表的人工智能技术飞速进步，应用场景加速拓展，经济社会发展对于智能算力的需求仍将持续走高。智能计算中心作为智能计算的重要载体，预计将在经济社会发展中发挥重要的支撑作用，未来发展前景一片光明。智能计算中心在各地的加速落地，将驱动人工智能和各行各业加速融合，推动各行业优化升级。

## 参考文献

- [1] 孙杰贤.《智能计算中心规划建设指南》解读[J].中国信息化,2020(12):17-18. 10.3969/j.issn.1672-5158.2020.12.008
- [2] 王恩东.智算中心成为新基建的基本条件与智慧时代动力源[J].中国工业和信息化,2020(4):44-50
- [3] 王恩东.智算中心是智慧时代的新基建[J].信息化建设,2020(5):56-57
- [4] 国家信息中心信息化和产业发部.智能计算中心规划建设指南[R].2020
- [5] 杨明川,刘倩,赵继壮.人工智能数据中心研究[J].信息通信技术与政策,2021,47(4):1-7. DOI:10.12267/j.issn.2096-5931.2021.04.001
- [6] 张金颖,郑子亨.夯实人工智能“新基建”发展经济新动能[J].网络安全和信息化,2020(6):53-54

## 作者简介



**焦奕硕**，中国信息通信研究院信息化与工业化融合研究所高级工程师；主要从事先进计算、5G通信、智慧城市等方面的研究工作，主持并参与多个课题、规划编制的相关研究工作。



**邸绍岩**，中国信息通信研究院信息化与工业化融合研究所高级工程师；主要从事集成电路、5G通信、数字经济等领域研究工作，主持并参与多个课题、规划编制的相关研究工作。