

智能计算中心规划建设指南

国家信息中心信息化和产业发展部

二零二零年十二月

Contents

目录

	前言	
01	概念界定、主要内涵和功能定位	
	概念界定	01
	主要内涵	01
	功能定位	02
	对比分析	03
02	建设意义与应用需求	
	加速AI产业化创新发展	04
	· 应用场景1: 识别检测	04
	· 应用场景2: 语音交互	05
	· 应用场景3: AI芯片	06
	· 应用场景4: 自动驾驶	06
	· 应用场景5: 机器人	07
	驱动产业AI化转型升级	08
	· 应用场景6: 智能制造	08
	· 应用场景7: 医疗影像	09
	· 应用场景8: 无人商店	10
	· 应用场景9: 智能客服	10
	· 应用场景10: 智慧物流	11
	· 应用场景11: 智慧农林	12
	助力政府治理智能化	13
03	建设思路	
	建设原则	15
	发展定位	15
	建设要点	16

04	技术框架	
	总体架构	17
	最新人工智能理论	18
	领先人工智能计算架构	19
	智算中心作业环节	20
	· 1. 生产算力	20
	· 2. 聚合算力	21
	· 3. 调度算力	23
	· 4. 释放算力	24
	小结	25
05	预期经济社会效益	
	降低AI应用成本，提高算力效率	26
	加速AI生态对接，推动创新产业聚集	26
	促进产业转型升级，加速新旧动能转化	27
	加强政府治理，提升公共服务能力	27
	改善经济运行格局，激发经济发展活力	28
06	建设运营模式	
	智算中心“投-建-运”主体选择	29
	· 1. 投资主体	29
	· 2. 承建方主体	29
	· 3. 运营主体	30
	建设运营模式选择	30
	· 1. 基于政府独立投资的建设模式	30
	· 2. 基于特殊项目公司的建设运营（SPV）模式	30
	· 3. 基于“国家-地方-企业”共建的建设运营模式	31
07	实施路径建议	
	需求导向，明确建设思路	32
	改建结合，优化建设方式	32
	软硬兼施，提升服务能力	33
	先进适用，统一建设标准	33
	规范推进，加强项目管理	33
	丰富场景，打造产业生态	34

版权声明

本报告中所涉及的图片、表格及文字内容的版权归国家信息中心和浪潮共同所有。其中部分数据在标注有来源的情况下，版权归属原数据公司所有。本报告取得的数据来源于公开资料，如有涉及版权纠纷问题，请及时联络我们。

任何机构、个人在引用本报告数据或转载有关报告内容时，需标注来源。违反上述声明者，将追究其相关法律责任。

前言

党中央、国务院高度重视新型基础设施建设，对推进新型基础设施建设作出了战略部署，为经济社会高质量发展强基筑本。2020年4月20日，国家发展改革委首次明确新型基础设施范围，将智能计算中心（以下简称“智算中心”）作为算力基础设施的重要代表纳入信息基础设施范畴。《关于2019年国民经济和社会发展计划执行情况与2020年国民经济和社会发展计划草案的报告》指出，国家发展改革委将在2020年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重大工程，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。国务院发布的《新一代人工智能发展规划》中明确提出建设“高效能计算基础设施”。截至2020年10月，我国先后批准设立了北京、上海、合肥等13个国家新一代人工智能创新发展试验区。随着AI产业化和产业AI化的深入发展，智算中心已受到越来越多地方政府的高度关注并开展前瞻布局，已成为支撑和引领数字经济、智能产业、智慧城市、智慧社会发展的关键性信息基础设施。

2020年4月，浪潮前瞻性地提出“智算中心”概念。国家信息中心和浪潮开展联合研究，旨在明确智算中心的概念、内涵、功能定位和技术架构，初步探索智算中心建设的路线和实施路径，探讨建设智算中心的经济社会价值，为地方健康有序开展智算中心建设提供借鉴和参考。

概念界定、主要内涵和功能定位

概念界定

智算中心是基于最新人工智能理论，采用领先的人工智能计算架构，提供人工智能应用所需算力服务、数据服务和算法服务的公共算力新型基础设施，通过算力的生产、聚合、调度和释放，高效支撑数据开放共享、智能生态建设、产业创新聚集，有力促进AI产业化、产业AI化及政府治理智能化。

主要内涵

一是算力公共基础设施。智算中心面向政府、行业、企业等多用户群体提供人工智能应用所需算力服务、数据服务和算法服务，能够汇聚各行业领域数据资源、支撑各行业领域AI计算需求，智能计算中心作为公共算力基础设施，通过提供共性的算力、数据及算法服务，让算力服务更为易用，使得智慧计算像水电一样能成为基本公共服务。

二是计算架构技术领先、生态成熟。智算中心基于AI模型提供高强度的数据处理、智能计算能力，集成先进的智能软件系统和智能计算编程框架，实现云端一体化，形成技术领先、可持续发展的高性能、高可靠计算架构。智算中心核心计算单元采用先进的人工智能芯片，面向新型的人工智能场景，采用异构计算，能大幅提升对基础算力的使用效率和算法的迭代效率。同时集成生态成熟的智能软件系统和智能计算编程框架，便于不断迭代升级。

三是算力、数据和算法的融合平台。智算中心以融合架构计算系统为平台，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，使算力、数据、算法三个基本要素成为一个有机整体和融合平台。智算中心为AI算法研发提供大规模数据处理能力，也为AI产业应用提供充足的计算资源，全面支撑各类人工智能技术的应用和演进。

四是以产业创新升级为目标。围绕智算中心基础设施建设，以数据流引导技术流、业务流、资金流、人才流聚集，实现以数据驱动产业创新发展新模式是智算中心的核心目标。通过打造人工智能开放服务平台，面向AI产业、传统产业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，全面汇聚并赋能各产业领域AI应用，助力加速孵化新业态，推进数字经济与传统产业深度融合，实现AI与产业的聚合并带动形成一个多层级的AI产业生态体系，全面赋能产业创新升级。

功能定位

智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。智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。智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。智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智算中心，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。

对比分析

智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-建-运”模式等方面，与超算中心、云数据中心相比，智算中心都有所差别，对比如表1所示。

表1 智算中心与超算中心、云数据中心对比表

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

建设意义与应用需求

智算中心作为公共算力基础设施，符合中国当前社会经济发展阶段和转型需求，是促进AI产业化、产业AI化和政府治理智能化的重要引擎。智算中心将在推动国家人工智能战略实施，赋能实体经济实现新旧动能转换，提升社会治理水平，促进人工智能科研和工程技术人才培养等领域发挥重大支撑和推动作用。

加速AI产业化创新发展

据赛迪预计，到2020年中国人工智能核心产业规模将超过1600亿元，增长率达到26.2%。人工智能产业的蓬勃发展为智算中心的建设和发展提供了巨大的机遇。智算中心作为AI软硬件技术的一体化融合载体，将为AI产业的发展提供大规模数据处理和高性能智能计算支撑，加速图像识别、自然语言处理、大规模知识图谱等技术的研发、测试和应用部署进程。智能计算中心的构建将推动“平台+应用+人才”三位一体的新型AI产业发展模式，打造“算力+生态”体系，将加速AI全产业链的形成。AI产业化的主要应用场景包括，识别检测、语音交互、AI芯片、自动驾驶、机器人、视频解析、人机协同、机器翻译、精准推荐等。

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。

识别检测

面临挑战

- 识别检测技术面临的主要挑战是如何进一步提高识别的准确率，具体体现在：
1. 如何提高训练模型对未曾出现过的场景的泛化能力。特别是当训练数据规模急剧增长后，如何显著提高模型的识别能力。
 2. 如何提高算法对场景的全面理解能力，判断目标识别物中对象与对象之间的关系、层级和布局等。

市场需求	据麦姆斯咨询的数据显示，全球识别检测市场规模到2021年将达到389.2亿美元，16年-21年的复合年增长率将达到19.5%。
对智算中心的算力需求	<p>目前识别检测技术的研究热点包括三维视觉、多模态融合、图像描述、事件推理和图像与视频场景理解等，要求新的识别算法能够挖掘丰富的视频和图像数据之间的时空关系，分析事件之前的语义相关性等。</p> <p>高性能识别检测平台往往需要支持处理10万+的物体识别能力，具备分钟级别的模型生成响应速度，具备模型持续优化升级的能力，需要支持诸如深度分离卷积、残差密集网络（RDN）和隐式三维朝向学习等最新机器学习算法，同时支持各类算法的灵活配置以满足不同实际应用场景的需求。</p>

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、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。

语音交互	
面临挑战	<ol style="list-style-type: none"> 1. 语音识别。精确地理解用户复杂的、基于情感式的、语义模糊的需求进行更加深刻的分析。 2. 语义识别。真正实现人机交互和让机器理解人的“语言”，从理论和实践层面还存在巨大挑战。机器语音识别能力基于固定识别模式，无法处理分词、歧义和未知语言处理等情况。 3. 基于对用户的持续学习。具备用户引导能力，通过对话式智能交互为用户提供“建议”，还能“猜测”用户可能将要提出的需求。
市场需求	据中金企信国际咨询公布的《2020-2026年中国智能语音市场竞争策略及投资潜力研究预测报告》统计数据显示：中国对话式人工智能市场规模在2022年有望达到78亿人民币，2018-2022年之间，其市场份额将以57%的复合增速增长。
对智算中心的算力需求	将来，语音交互算法的训练语料数据量将很快突破百万小时，训练数据包含大量不同用户的口音数据、多领域歧义语料数据和具备复杂语法规则的数据。随着大规模语料样本数据的不断积累，需要更好利用智算中心具备的大数据挖掘技术、深度学习算法，构建更加高级的语音模型，为客户提供流式计算、在线计算、大规模离线计算等多种形式的服务。

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。AI加速芯片的种类十分丰富，主要包括图形处理器（graphics processing unit, GPU）、现场可编程门阵列（field-programmable gate array, FPGA）、专用集成电路（application specific integrated circuits, ASIC）和神经拟态芯片等，预计到2024年，非GPU加速芯片市场份额将超过20%，未来，AI芯片市场会继续向灵活性和多样化的方向发展。

AI芯片	
面临挑战	AI芯片设计和生产所面临的核心挑战是芯片性能、功耗和面积等。AI芯片需要更好地适应多种架构体系，以支持更多的应用系统开发。
市场需求	根据赛迪顾问预计，2019年中国AI芯片市场规模达到124.1亿元，同比增长53.6%；2020年中国AI芯片市场将保持56.1%的增长，达到193.7亿元的市场规模；2021年中国AI芯片市场规模将进一步增长至305.7亿元，同比增幅可达57.8%。
对智算中心的算力需求	在为各类用户主体提供多元化AI算力服务的同时，AI芯片产业的快速发展必将带动智能计算中心的建设部署。

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。

自动驾驶	
面临挑战	<ol style="list-style-type: none"> 1. 保障产业生态安全可控。汽车产业是典型的全球化产业，需要相关计算平台和操作系统等关键技术的自主可控，确保产业生态安全。 2. 保障自动驾驶行为安全。需要大量采用虚拟化仿真试验等技术来模拟复杂驾驶环境和加速测试过程，大幅提高实际测试效果，确保驾驶行为安全。 3. 保障自动驾驶功能安全。智能网联汽车系统逻辑功能复杂，各子系统间紧密连接，交互频繁，需要保证自动驾驶系统的功能安全。
市场需求	<p>中国工业和信息化部表示，至2020年，中国智能网联汽车的市场规模可达到1000亿元以上。根据预测，智能网联汽车将持续20年的高速发展，到2035年，将占全球25%左右的新车市场。</p>
对智算中心的算力需求	<p>自动驾驶行业具备典型的海量数据、高并发、实时处理等计算属性，呈现出高度多元主体协同、智能识别和感知、系统模拟仿真计算、决策分析和预测预警等特性。该产业的高度集成化需要智能计算中提供先进的AI技术支撑，快速推动新产品的研发、测试和应用。</p>

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。

机器人	
面临挑战	<p>机器人产业的国际竞争日益激烈，欧美等发达国家仍然掌握着机器人生产领域的关键核心技术，美国的波士顿动力（Boston Dynamics）、瑞士的ABB等科技公司依然处于绝对领导地位。我国机器人产业发展还面临着关键核心技术缺乏、创新能力不足和创新成果及产学研实际转化率较低等问题，迫切需要加强对机器人产品设计、研发和生产全过程和全流程的支持，提高研发和生产效率，降低企业AI技术测试、改进和应用部署成本，支撑机器人产业尽快突破核心关键技术，加快产学研用转换，提高产业自主创新能力。</p>
市场需求	<p>中国机器人市场规模约占全球市场三分之一，同时也是全球第一大工业机器人市场。预计到2021年将突破1200亿元。</p>
对智算中心的算力需求	<p>机器人产业的发展迫切需要智能计算中心平台提供全链条AI技术支持，包括AI芯片、视觉识别、语音识别、机械臂和导航技术等，根据不同场景的不同诉求，提供“硬件+软件+服务”的全流程快速定制化专用机器人产品，大幅缩短产业链流程。</p>

驱动产业AI化转型升级

数据洪流与智能化正以前所未有的速度重塑各行各业，人工智能深度渗透重点行业领域，AI应用场景呈现出多元化、规模化发展趋势。据相关机构预测，AI赋能全社会各行业领域发展形成的“AI+产业”有着万亿级市场规模，远大于AI产业化千亿级市场规模。智算中心作为新型基础设施成为企业低成本、高效率数字化转型的重要抓手。企业可以根据业务需要依托智算中心提供的AI模型库、AI算力调度平台等自动生成适用于实际需要的业务系统模型。同时，智算中心通过提供算力基础设施及通用软件服务，联动产业链上下游，为企业提供完整的AI服务链，帮助实现AI供给和需求的高效对接，促进产业高质量、智能化发展。产业AI化的主要应用场景包括：智能制造、医疗影像、无人商店、智能客服、智慧物流、智慧农林、无人巡检、智慧教育等。

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、自适应等功能的新型生产方式。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。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。

智能制造

面临挑战

人工智能在制造业领域的应用主要面临行业协同、技术精确性的挑战：

1. 智能制造系统构建需要建设高度智能化的工业互联网平台，开发全面覆盖的信息-物理融合系统，实现业务系统与生产现场全面贯通。此外，还需要实现全生产周期的系统联动和全合作伙伴间的信息共享，以提升企业感知、协同和运营能力，最终实现生产过程智能化。
2. 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。

市场需求	根据前瞻产业研究院的相关报告，2019年中国智能制造行业市场现状及发展前景分析，预计2024年市场规模将接近5.7万亿。IDC发布《中国AI赋能的工业质检解决方案市场分析》报告显示，2019年的工业质检市场份额达1.07亿美元，正释放出强大的AI应用潜力。
对智算中心的算力需求	赛迪研究院的研究表明，工业大数据将成为智能制造和工业互联网发展的核心、基于算法的工业智能平台将成为应用场景的重要基石、工艺装备的智能化将成为制造业转型发展的突破，这些趋势决定其对AI算力的需求是巨大的。IDC研究指出，机器制造、电子制造业将有很大的AI应用潜力，主要在于车间预测性防护、产品质检/分拣等。传统的视觉检测系统难以应对越来越复杂的检测场景以及更高的产品质量检测需求，工业质检车间的自动化和智能化越来越成为行业大势，这也为AI赋能工业质检提供了落地的土壤。由此可见，智能制造（工业质检）对于AI算力的需求是巨大的，对智算中心的应用需求也是很迫切的。

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。

医疗影像	
面临挑战	<ol style="list-style-type: none"> 1. 基础支持体系与计算能力不足。医疗数据处理的复杂度极高，对AI算法和算力的需求很强。远程医疗涉及仪器设备的精确化控制，对网络基础设施要求极高，需要确保数据传输的顺畅。 2. 智慧医疗应用的容错率问题。由于行业的特殊性，医疗行业连万分之一的错误率也无法接受。因此，目前人工智能诊断类应用屈指可数。如何让人工智能诊断类应用做到万无一失是一个巨大的难题。 3. 基于深度学习的医学影像人工智能都需要大量的标注数据进行训练，而且训练所用的标注数据本身对于训练结果的影响要大于算法，但目前严重缺乏由专业医师提供的高质量标注数据。
市场需求	据相关专业机构统计，预计到2020年，我国智慧医疗的投资规模将达到1000亿元以上，2020-2024年，年均复合增长率约为15.24%，2024年将达到1850亿元。
对智算中心的算力需求	医疗影像智能诊断应用需要基于大量的数据进行知识挖掘和深度学习训练，获取具有实际医疗应用价值的模型，这对数据的存储与计算都提出了相当高的硬件需求，因此，需要智算中心提供强大的算力支持和一体化开发环境。

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、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、更精准的识别要求，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。

无人商店	
面临挑战	<p>无人商店大规模应用落地主要面临的挑战在于高强度的计算能力支撑、更精准的智能识别技术。</p> <p>随着店铺规模不断扩张、客流量显著增大，采用机器视觉方案的无人商店识别精度难以保证。无人商店中用到的图像识别系统在小范围场景和高资本、技术投入下，准确率是可以接受的。但如果规模进一步扩大，则将面临非常大的挑战。需要更高速、精准的计算能力以及精细精准的智能识别能力。</p>
市场需求	<p>根据博思数据发布的《2018-2023年中国无人零售商店市场分析与投资前景研究报告》，2017年我国无人零售市场规模为197.0亿元，无人便利店销售规模在0.5亿元左右，预计到2025年我国无人零售市场规模将增长至3633.1亿元。</p>
对智算中心的算力需求	<p>无人商店的应用落地需要实现精准识别客户身份、商品标签、智能计算、支付等全流程智能购物，同时需要基于大量的销售数据、行为轨迹、消费习惯等大数据挖掘与深度学习分析，采用基于客户画像的精准营销算法，实现个性化精准营销，提高服务满意率。销售全流程需要以人工智能技术为核心，涵盖深度学习、计算机视觉、传感器融合、生物识别、RFID等技术，对高性能存储与计算、AI模型算法等都提出了相当高需求。智算中心强大的算力支持和人工智能技术应用是无人商店未来持续拓展的必要基础支撑。</p>

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。

智能客服	
面临挑战	<p>1. 智能客服对用户需求理解的准确度仍不高。用户的提问一般以相对口语化的方式进行，甚至夹杂地方方言，而机器系统则一般以结构化的语言去读取，例如客户的口语化提问方式、上下文智能关联等，但在客户自然语言和计算机结构化语言之间存在一定差异，目前大多数智能客服处理这类问题的能力并不强。</p> <p>2. 机器系统的自我学习能力和内容难于及时更新完善。技术上的学习主要跟系统的推荐算法相关，作为当下智能客服系统的核心算法，目前大多数智能客服系统在算法的优化更新方面的速度较为滞缓，且对投入要求也较高。</p>
市场需求	<p>据前瞻研究院预测，目前中国约有500万全职客服，客服行业市场规模约4000亿元，智能客服市场约将达到500亿-800亿元。</p>
对智算中心的算力需求	<p>智能客服实际上是集合人工智能学、计算机科学、语言学等多门学科的综合应用，需要让机器主动去认知和学习，不断强化行为模式，提高思考能力，从而更加灵活地完成各项工作任务。此外，要更加充分地利用自然语言处理技术中如语义分析、情感分析、上下文关联等技术而不单只是切词匹配，这样在应对客户多样化的提问时让机器人听懂人话，才能更加准确地判断客户需求并提供最佳答案。基于此，智算中心能够为NLP（自然语言处理）、NLU（自然语言理解）、ASR（自动语音识别技术）、TTS（语音合成技术）等核心技术提供算力支持，为实现系统的自我学习提供不可或缺的能力支撑。</p>

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	
面临挑战	<p>传统物流行业，人的工作压力大、部分工作环境苛刻，物流成本连年攀升，企业主动寻求降本增效之道。设备的海量规模和系统的柔性离散要求包括：</p> <ol style="list-style-type: none"> 1. 高可靠规划，保证最优稳定的规划方案； 2. 精准控制，精确准时的控制设备协同作业； 3. 高效能装备，保障高效安全的储存、拣选及配送。
市场需求	<p>2017年，我国智能物流市场规模达到3375亿元，同比增长21.2%。据iiMedia Research预测，到2025年我国智慧物流市场规模将超万亿。</p>
对智算中心的算力需求	<p>智慧物流产业的发展需要应用先进的AI算法、软硬一体的机器人产品与操控系统和整仓集成的完整生态。</p> <p>各类产品之间需要高效协作，实现7*24小时不间断工作，通过AI技术从图像和视频数据里获得高层次、可理解的信息，智能区分人和物，并能在50%变化率动态场景里稳健安全运行。</p> <p>因此，智慧物流产品的研发需要新型的智能计算中心助力和支撑，加快解决方案的测试和应用部署。</p>

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。

智慧农林

<p>面临挑战</p>	<p>目前，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以下几个问题：</p> <ol style="list-style-type: none"> 1. 农林渔牧数据种类繁多、数据规范化程度不高。农林渔牧副各产业生产特征不同，数据存在巨大差异。以种植业为例，涉及到的作物种类、地域分布、气候条件、基因序列等数据千差万别，没有形成数据规范。 2. 农业基础设施现状无法满足新型农业生产模式的发展。物联网是智慧农业发展的重要基础，需要大量的硬件设备投入，包括智能传感器、视频监控、通信基础设施、云服务器等，但当前的基础设施条件极不理想。 3. 新技术的应用深度不够。由于硬件投入力度和数据处理能力的限制，以及农林业生产者长期形成的传统农林业生产思维和知识、技术水平的不足，各项农林智能应用发展整体呈现杂而不精、难以推广的现象。
<p>市场需求</p>	<p>据中商产业研究院预测，到2020年智慧农业的潜在市场规模有望增长至268亿美元，年复合增长率达14.3%。智慧农业藏粮于技，通过高新技术的应用，有望将粮食产量增加70%，相当于养活近96亿人，并且能够实现农业的集约化生产，大幅度降低资源浪费，提高环境保护能力，促进生态可持续发展，还民以绿水青山。</p>
<p>对智算中心的算力需求</p>	<p>农业数据整体呈现规模庞大、类型多样、实时数据更新频繁等特征，涉及农林渔牧副各产业的多种数据。因此，为了挖掘更有价值的农林业生产数据，促进智慧农林的精准落地，针对农林渔牧副产业的不同特点，个性化定制智能解决方案变得尤为重要。需要有配套的智算中心作为支撑，实现农林感知数据的大规模存储，借助强大的算力支持农林数据的高效计算、训练和推演，最终实现快速和智能化的农林生产决策。</p>

助力政府治理智能化

当前，随着智慧城市建设的深入发展，城市基础设施智能感知网络逐步完善，以“城市大脑”为代表的城市级平台加速落地。智算中心作为“新基建”数字基础设施的主力军，其所承载的AI算力将是驱动智慧城市发展的核心动力。传统计算中心在数据处理能力、技术架构等方面均难以满足智能计算的能力需求。智算中心融合人工智能、互联网、大数据、云计算等信息技术，能迅速将线上线下各类型治理主体聚合在一起，提供数据分析、云计算平台、算法和计算能力等工具和资源，大大提升社会治理过程中的数据计算、分析、挖掘能力，推动政府治理能力现代化。

未来智算中心将有效支撑智慧交通、应急管理、防洪减灾、环境保护、地理测绘等应用场景，助力政府治理智能化。在智慧交通方面，汽车化社会带来的诸如交通阻塞、交通事故、能源消耗和环境污染等社会问题日趋恶化，交通阻塞造成的经济损失巨大。智慧交通应用场景对数据处理和算力的需求越来越高，这是由于更加复杂的交通运行管理和服务需求所决定的，随着监管车辆、人和道路基础设施的逐步数字化，

交通领域呈现出数据类型更加复杂多样（图像、语音、视频等）、数据来源更加丰富（互联网、物联网感知终端、车辆设施、基础设施）、计算模型更加复杂融合（监测预警、仿真模拟、学习推演、预测预判、评估考核、决策分析等）、计算方式更加多元（边缘计算、分布式协同计算、中心计算）等特点，这些对于数据和算力服务的需求需要智能计算中心提供有力支撑。在应急管理方面，应对重大突发公共安全事件的处置能力是城市现代化程度的一个重要标志。现阶段，我国正进入“突发公共事件的高发期”和“社会高风险期”。2020年初爆发的新冠疫情，至今仍然肆虐全球，引发了历史上前所未有的全球危机。如何快速和有效地应对这些突发事情，减少其对人类社会的损害，成为各国政府的一项重要工作内容。通过深度学习、视觉计算、知识图谱等技术手段将智能图像分析算法应用到风险识别监测和应急指挥管理过程中，通过海量视频图像数据汇聚全量的视频大数据，对人、车、物的视频分析，达到合规检测、预警预测预判的目标。

建设思路

建设原则

总体规划、政企协同。准确把握政府和市场的分工，强化政府在智算中心规划制定、统筹协调、环境营造等方面的主导作用，加大政策保障和投入支持，健全市场化机制，探索开放共赢的建设运营模式，推动形成多元化参与的政企协同机制。

需求牵引、先进适用。把握全球新一轮信息技术变革趋势，应用人工智能技术，深化供给侧结构性改革，加强供需对接，强化要素支撑，广泛开展典型示范和场景应用，加速AI产业化和产业AI化的进程。

夯实基础、培育生态。充分发挥智算中心集约化、智能化优势，夯实城市经济社会发展基础支撑能力。充分释放智算中心建设对数字经济产业、传统产业数字化的放大效应、乘数效应，培育产业生态，促进产业繁荣。

保障安全、创新发展。完善智算中心标准规范，从管理和技术两方面充分保障网络安全和数据安全，全面提升新型基础设施体系安全水平。鼓励技术创新、产品创新、模式创新、服务创新，创造新供给，激发新需求，培育新动能。

发展定位

带动智能产业——智算中心通过搭建新一代高性能人工智能开源框架、公共计算、数据开放等平台，深入推进应用场景平台建设，凝聚汇聚人工智能产业链上下游企业，打造一批人工智能创新平台，孵化一批人工智能创业企业，营造先行先试、开放包容、创新引领的产业发展环境。

打造智能生态——智算中心通过将AI技术赋能各个行业领域，衍生出全新的产业应用场景，形成涵盖“科技研发、产业孵化、创投资本、教育培训、配套政策环境”的智能产业生态圈，加速区域经济发展，拓展数字经济发展空间，提升数字经济发展成效。

支撑智能应用——围绕城市信息化建设基础与需求，将智算中心建成具有影响力的城市数字底座。同时，结合各城市特色和资源禀赋，打造创新应用支撑与服务体系，有效激发经济社会发展内生动力，全面支撑政府治理、产业升级、社会事业发展迈上新台阶。

建设要点

以全面提升AI算力为核心，以促进数据开放共享为基础，以培育区域智能生态圈为根本，在明确本地算力需求的基础上，围绕数据、算法、算力三大核心要素搭建集“智能”“开放”“融合”三大特征于一体的技术框架，通过数据深度加工、算力算法服务创新应用，着力推动AI产业资源要素聚集，带动重点产业发展，打造满足城市创新智慧发展的智能生态。

——**以全面提升AI算力为核心**。算力已成为衡量一个企业、地区甚至是一个国家智能经济发展水平的重要指标。智算中心在建设运营过程中，围绕生产算力、聚合算力、调度算力、释放算力等四大AI算力关键环节持续创新，实现AI计算力的全流程、一体化的高效交付，并通过AI计算生产方式的变革促进AI应用的效率变革。

——**以促进数据开放共享为基础**。AI技术的发展依赖于海量数据的积累，城市推进数字化、智慧化建设过程中产生沉淀了大量数据，但城市业务条块分割带来的数据孤岛将严重制约AI算力效率的提升和AI算法模型输出能力的发挥。加快推进数据开放共享，是提升智算中心AI算力及增强AI算法模型准确性的重要基础。数据开放共享程度越高，智算中心汇聚归集的数据规模越大、数据质量越高，AI算力损耗越低，基于此输出的AI算法模型可用性越高。

——**以培育区域智能生态圈为根本**。生态构建是AI算力与行业应用深度融合的关键。围绕计算生态、开放生态、应用生态，基于智算中心构建融合的技术体系和开放创新的计算生态，为用户提供可精确按需扩展、满足多样化应用场景的智能计算服务。构建服务全产业链的行业应用生态，支撑各行业领域与人工智能加速融合落地。

——**着力推动AI产业创新聚集**。以智算中心为重要载体，依托智能、开放、融合技术架构，融合大数据、云计算、人工智能等多项技术，积极推动AI技术流、资金流、价值流和人才流深度融合，围绕底层基础支撑、核心技术创新、行业领域应用三个维度打造AI产业链。加速AI算法模型在各行业领域的深度应用，通过数据深度加工、算力算法服务创新应用，实现相关产业资源要素聚集，延长AI产业链，催生出AI生态发展新模式、新业态。

技术框架

智算中心是对外提供AI算力服务的公共基础设施，如何构建一个高性能、可扩展的技术架构，是智算中心非常重要的技术支撑，所以要在智算中心完成算力的生产、聚合、调度和释放，从而实现构建平台，提供服务。智算中心的总体架构如下图所示：

总体架构

图1 智能计算中心总体架构图



整个架构可以分为四大部分，分别对应的是基础、支撑、功能和目标。基础部分是最新的人工智能理论和领先的人工智能计算架构，这是智算中心区别于其它数据中心的核点。智算中心作业环节是智算中心的支撑部分，智算中心通过作业环节实现了算力的生产、聚合、调度和释放，是区别于其它数据中心的具体体现。功能部分是四大平台和三大服务，四大平台分别是算力生产供应平台、数据开放共享平台、智能生态建设平台和产业创新聚集平台；三大服务分别是数据服务、算力服务和算法服务。目标部分是促进AI产业化、产业AI化及政府治理智能化。下面分别来讲述架构的各个部分。

最新人工智能理论

人工智能经历了三次浪潮，本次人工智能浪潮以深度学习为核心，通过创新深度学习算法和模型训练方法，在大数据和算力的驱动下，近年来在图像、语音、自然语言处理等领域取得了长足的发展和进步。

深度学习

深度学习算法的核心是深度神经网络模型。深度神经网络通过构建类似于人脑神经元的多层的、相互连接的网络结构，可以很好的学习图像、语音等复杂数据中的多层次特征。当前卷积神经网络（CNN）、循环神经网络（RNN）等各种神经网络结构的发展极大得扩展和丰富了人工智能的学习能力。

自监督学习

近年来在模型训练方法上也有很多创新和进步，其中自监督学习不需要对数据进行标注，可以直接使用互联网上的海量数据进行模型训练。比如以BERT为代表的自然语言处理模型就可以广泛的使用互联网上的文本信息，包括中英文书籍、维基百科数据、新闻数据等进行学习。

强化学习

强化学习也是一类被广泛使用的模型训练方法。强化学习是一类通过不断地尝试，从错误中学习，最后找到规律，学会达到目的的方法。此方法在复杂决策场景，比如自动驾驶、工业机器人等场景取得了很好的效果。

自动机器学习

自动机器学习是一种自动构建深度神经网络的方法。通过引入自动机器学习，可以节省人工智能算法设计人员的工作，延展了模型架构设计的广度。

跨媒体、多模态

跨媒体、多模态的建模方法可以从图像、视频、语音和文本等多个领域获取信息，实现信息转换和融合，从而提升模型性能。人们生活在一个多领域相互交融的环境中，听到的声音、看到的实物和闻到的味道等都是各领域的模态形式。通过给机器赋予学习和融合这些多领域信号的能力，可以使深度学习算法更加全面和高效地了解周围的世界。

领先人工智能计算架构

人工智能的模型训练和推理过程需要强大的计算力支撑，当前的人工智能计算以AI加速芯片为核心，构建了以CPU+AI芯片为主体的AI服务器架构，并在此基础上引入高速互连网络组建AI计算集群。

AI芯片

深度学习计算包含大量的矩阵乘加运算，AI加速芯片针对这一计算特点，通过引入针对矩阵运算的加速设计，可以实现和传统通用处理器相比10~100倍的加速。常见的加速芯片包括GPU、FPGA和ASIC等。

AI服务器

AI服务器是以CPU+AI加速芯片为主体的服务器架构。深度学习模型训练的计算需求经常会超出单个AI加速芯片的计算能力，因此需要多个AI加速芯片的协同处理，通过建立AI加速芯片之间的高速互连通信，可以有效提升AI服务器作为一个整体的计算能力。这方面的设计包括NVIDIA针对GPU设计的NVLink协议以及支持多种AI加速芯片的OAM(OCP Accelerator Module) 协议互连通信等。

高速互连

AI服务器之间可以采用高带宽、低延迟的网络实现高速通信，来应对更大规模模型训练任务。这些网络包括RDMA和RoCE等。对于深度学习常用的全局梯度归约通信操作，可以采用针对性的网络优化设计，比如采用当前效率最优的全局环状通信链路设计，可以最大化的发挥集群中任意两个计算芯片之间的通信带宽。

深度学习框架

深度学习框架的发展大大简化了深度学习模型的训练和部署流程。深度学习框架提供一系列的标准组件，比如张量、卷积、池化、自动微分等。通过调用这些组件可以较为容易的设计模型，实现模型的训练过程，以及部署模型的推理服务。目前较为常用的深度学习框架包括TensorFlow、PyTorch、Caffe、飞桨(PaddlePaddle)等。

资源调度

深度学习模型的训练和部署需要灵活的资源调度方式。基于容器的资源调度策略可以对计算资源的池化，以单个容器作为最小作业单元，并通过对计算资源的细粒度切分来提高资源的利用率。另外，基于容器的资源调度策略也可以给一个任务分配多个容器来满足大规模计算任务的需求，并能在计算任务完成后实时释放计算资源。

智算中心作业环节

1. 生产算力

生产算力是基于领先的AI服务器为算力机组，支持先进多样的AI芯片，支持成熟丰富的软件生态，形成高性能、高吞吐的计算系统，为AI训练和AI推理生产输出强大、高效、易用的计算力。生产算力主要包括四个要素：算力机组、算力芯片、算力生态和算力输出。

算力机组

AI服务器，是智算中心的算力机组，是生产算力的核心与基础，是以CPU+AI加速芯片为主体的异构架构，是面向AI应用的计算平台，是承载智算中心AI计算的核心基础设施。AI服务器最显著的特征是超强的计算性能，以应对AI应用对算力的巨大需求。与传统服务器仅支持CPU一种芯片不同，AI服务器还需要支持多颗AI加速芯片。这样对系统结构、拓扑架构、散热、噪音、能源效率、延迟等设计带来巨大挑战，需要深度优化拓扑结构，通过精准DOE SI仿真，确保在满足信号完整性的同时优化性能。需要极致优化的散热设计，才能确保高功耗下保持高稳定性，高效利用冷热风流，实现低功耗、高散热性能完美组合。

AI服务器按功能分类，AI算力机组主要由两种类型AI服务器组成，分别是AI训练服务器和AI推理服务器。

AI模型训练是消耗算力最大的部分，虽然近年来，新增的算法模型数量在逐年减少，但是单个模型的参数量和复杂程度都呈现指数级的增长趋势，具有海量参数的模型训练如果没有强大的算力支撑，很难发挥其价值，常规的服务器难以满足算力需求，需要专门的AI训练服务器。

模型训练完成后，接下来要是部署与推理。未来几年，随着推理工作负载在各个行业应用中不断增加，推理需求将呈现指数级增长。为支撑大规模、高并发的推理计算需求，需要专门的AI推理服务器。

算力芯片

智算中心需要支持各类AI加速芯片，从而覆盖从训练到推理，从边缘到数据中心的各类相关的AI应用场景。

CPU是通用处理器，适用于更好地响应人机交互的应用和处理复杂的条件和分支，以及任务之间的同步协调。深度学习计算量巨大，CPU架构被证明不满足需要处理大量并行计算的深度学习算法，需要更适合并行计算的芯片，GPU、FPGA、ASIC等各种芯片应用而生。因此CPU+AI加速芯片的架构使得CPU与AI加速芯片各司其职，分别应对大量交互响应和高并行计算。

GPU应用开发周期短，成本相对低，技术体系成熟，目前全球各大主流公司均采用GPU进行AI计算。GPU目前在深度学习训练阶段是使用范围最广的。与CPU相比，GPU在深度学习领域的性能具备绝对优势。深度学习在神经网络训练中，需要很高的内在并行度、大量的浮点计算能力以及矩阵运算，而GPU可以提升这些能力，并且在相同的精度下，相对传统CPU的方式，拥有更快的处理速度、更少的服务器投入和更低的功耗。

FPGA (Field-Programmable Gate Array)，即现场可编程门阵列，它是在PAL、CPLD等可编程器件的基础上进一步发展的产物。FPGA核心优势是在推演阶段中算法性能高、功耗和延迟低。适用于压缩/解压缩、图片加速、网络加速、金融加速等应用场景。

ASIC是一种专用芯片，与传统的通用芯片有一定的差异。是为了某种特定的需求而专门定制的芯片。ASIC芯片的计算能力和计算效率都可以根据算法需要进行定制，所以ASIC与通用芯片相比，具有以下几个方面的优越性：体积小、功耗低、计算性能高、计算效率高、芯片出货量越大成本越低。近年涌现的类似TPU、NPU、VPU、BPU等各种芯片，都属于ASIC。

尽管AI服务器可以采用多种异构形式，GPU服务器仍然占据主流地位。

算力生态

随着AI发展的不断深入，对于AI产业化和产业AI化而言，技术架构的相对稳定和编程技能的通用性是非常必要的。目前基于GPU的AI计算的软件生态已经成熟完善。在面向深度学习训练、推理计算加速等方向都有完善的软件支撑。

当前AI应用日益复杂化、多样化，需要统一的基准来衡量AI芯片/AI服务器的计算性能、单位能耗力等水平，目前国际权威的AI测试基准有MLPerf和SPEC ML两个组织。MLPerf每年组织全球AI训练和AI推理性能测试并发榜。SPEC ML成立了Machine Learning技术委员会，建立和推行一套基于ML统一可信的行业测试基准。

算力输出

基于强大的AI算力机组，支持先进多样的AI加速芯片，依托于成熟丰富的软件生态，智算中心就可以为大规模AI应用输出强大、高效、易用的AI算力，主要有数据处理算力、训练算力和推理算力。

2. 聚合算力

聚合算力，是基于智能网络 and 智能存储技术，针对多任务、大规模、高并发、高吞吐的AI应用特点，为算力机组集群构建高带宽、低延迟的通信系统和数据平台，提供弹性、可伸缩扩展的算力聚合能力。

算力集群

聚合算力的对象是智算中心的算力集群，包含AI服务器、数据存储、高速网络及配套基础设施。

算力集群面临的问题主要体现在：（1）现有数据中心网络的主要问题包括弹性和可扩展性差、功能灵活性不高，无法支撑大规模计算集群。在新基建的驱动下，智能网络将成为提升数据中心生产力水平的重要组成部分，是支撑智算中心可持续发展的核心技术；（2）智算中心数据处理面临的挑战包括数据海量、数据多元化、数据快速响应、数据快速导入。为解决上述问题，需要为智算中心配置高性能智能存储。智能存储将采用软件定义存储方式，实现文件、对象、块、大数据四种存储服务一体化设计。同时，智能存储具备高IOPS处理能力、高速缓存技术和数据加速处理芯片。

智能网络

为了解决数据中心的网络问题，采用智能网络技术，聚合AI算力集群和智能存储集群，构建弹性的、可伸缩扩展的数据中心。智能网络技术融合业务需求和专用加速方案，结合人工智能技术，提供高带宽、低延迟、高并发的计算通信系统。目前，数据中心的网络互联速率已经从10G逐步过渡到25G，并且积极探索100G和400G网络。不断提高的网络互联带宽要求智能网络提供软件定义的加速能力，优化多任务带来的性能损耗，提高网络的利用率，提供高速的数据传输通道。软件定义的网络加速能力广泛应用在人工智能计算集群中，通过虚拟化技术，提供多任务并行处理能力，通过智能网络专用加速产品，支持可编程的处理器芯片，实现智算中心所需的多种网络卸载功能，同时可以支持数据预处理的计算能力，降低主机侧的性能损耗。通过先进的网络技术和设计思想，实现网络的控制和数据转发分离，提高网络的并发能力和低延迟，也支持弹性裸金属服务器、自定义业务功能等新特性。

智能存储

为解决智算中心数据处理面临的挑战，将海量、多元化的数据快速聚合，需要为智算中心配置高性能智能存储。智能存储应具备软件定义存储功能，即存储设备使用一套软件栈，便可支持不同存储协议，文件、对象、块、大数据四种存储服务一体化设计，实现同一基础架构上不同应用之间的数据业务应用融合。

智能存储具备按需扩展功能，可以根据业务需求，灵活添加或者移除存储节点，从而将数据和应用程序从硬件抽象出来。智能存储还可实现节点动态扩展，随着节点数的增加，存储容量和计算能力线性增加，提供呈线性递增的吞吐及并发能力。智能存储扩展后，系统将通过容量均衡功能自动在后端完成数据均衡操作，实现数据全局均衡存放。

智能存储为实现高IOPS处理能力，将采用基于聚合的小文件优化技术，用户在指定的聚合目录下，对小于一定大小的文件以紧密聚合对齐的方式写入聚合文件中，从而减少数据写入硬盘次数，提高小文件读写性能，具备千亿级别小文件的并发处理能力。

在管理功能方面，智能存储还将提供文件分级功能，可以将数据按照“指定的规则”存储在“适当的存储类型”上，以实现数据的分类智能管理。分级功能将根据系统中文件的大小、类型、存放时间等属性，将满足不同分级策略的文件分别迁移到不同性能存储介质上的存储池中，实现从沸点数据到冰数据，整个数据全生命周期的一个完整的管理。为提高管理性能，未来的智能存储还将配置专用数据加速处理芯片，用户可以把项目重删、压缩、加解密等工作交给这些专用处理芯片去做。存储主控芯片专门响应客户业务需求，使智能存储发挥极高效率。

算力聚合

采用最新的网络和存储技术，提升网络传输带宽和存储性能，降低损耗，提供高效的数据传输能力，融合网络虚拟化、软件定义网络、软件定义存储等技术，提高集群聚合的灵活性，助力大规模、多元化数据处理能力。通过智能网络和智能存储的有机结合，提供弹性、可伸缩扩展的聚合能力，实现算力高效聚合。

3. 调度算力

调度算力，是基于人工智能应用对算力的需求特点，通过虚拟化、容器化等技术，将算力资源池化为标准算力单元，通过适应性策略及敏捷框架对算力进行精准调度配给，保障AI开发和AI业务的高效运行。

调度能力是智算中心连接上层应用与底层计算设备的核心能力，将聚合的CPU、GPU、FPGA、ASIC等算力资源进行标准化和细粒度切分，满足上层不同类型智能应用对算力的多样化需求，让上层应用更高效、更便捷地对算力资源进行利用。算力调度分为算力池化、算力调度、算力服务三个过程。

算力池化

随着计算机硬件技术的发展，各类异构加速芯片、设备不断涌现，满足各类上层应用对计算资源、计算能力的多样化需求。智算中心通过计算资源池化，简化算力调用过程，方便用户对大规模集群内的计算资源进行有效利用，用户无需关注计算设备的种类，将计算资源以标准算力模式提供给用户。算力池化主要包括算力虚拟化和应用容器化。

算力虚拟化的核心，是向用户直接提供计算能力，避免用户在申请、使用计算能力的过程中对集群内设备的分布、类别、性能等产生关注。通过算力虚拟化，用户避开在大规模计算设备集群中进行设备选择、设备适配的繁杂工作。

应用容器化，基于容器化的应用部署策略能够显著降低管理应用、支持业务的计算开销，具有较高的灵活性，便于快速在边缘端、计算中心端等计算场景中灵活部署和迁移。

同时，计算资源的细粒度切分能力也是算力调度过程中的关键环节。计算资源的细粒度切分会渗透至异构算力组合策略之中，协调CPU、GPU、FPGA、ASIC等各类型计算芯片，将上层应用的计算需求细粒度切分后进行定向定量分发，最大化利用算力。

算力调度

智算中心通过有效算力调度，提高算力设备的利用率，降低设备闲置率，更好地管理智算中心算力的使用情况；对用户行为进行分析和监管，优化算力设备布局规划，提升业务部署效能，使应用具有更好的稳定性和扩展性。

算力调度涉及配额策略、共享超分、负载均衡等策略。

算力配额策略，智算中心会将池化的计算资源提供给用户，为用户的算力配给进行了有效的划分。

算力共享超分策略，在保证业务正常申请需求与扩容需求的同时，减少算力资源闲置，提升计算设备利用率，保证训练和推理业务顺利进行。

负载均衡，以人工智能的推理类业务为代表，对算力的使用、申请、接入等需求进行均衡分配，对应用的稳定性、服务能力、响应速度达到最优的配置。

算力服务

训练服务，依托容器化技术，将运行环境、框架适配过程标准化、模块化，让开发者能够便捷高效地构建运行环境，提交训练任务，保证智算中心内能够部署多样化的开发环境和训练环境，且彼此隔离、不会互相影响；

推理服务，简化应用部署流程和交互模式，方便用户在智算中心中进行应用部署，保障推理业务的算力分配随推理业务量变化实现弹性伸缩，保障业务稳定可靠。

4. 释放算力

释放算力，是基于主流人工智能理论算法，采用全流程软件工具，针对不同场景应用需求，通过机器学习自动化的先进方法，产出高质量的AI模型或AI服务，提升AI应用生产效率，促进算力高效释放转化为生产力。释放算力环节通过聚合算力、数据、算法三种生产要素，建立人工智能应用服务，主要可以划分为四个层次：AI场景、AI算法、AI工具、AI服务。

AI场景

随着人工智能技术的快速发展，人工智能已逐渐应用于各个领域，例如安防、交通、互联网等，人工智能与各个行业间关系愈发密切。人工智能算法主要应用于图像、视频、语音、自然语言处理等场景。图像应用场景主要包括医学影像识别、OCR识别、图像检索等。视频应用场景主要包括工业视觉检测、视频监控等。语音应用场景主要包括智能家居、机器人等。自然语言处理应用场景主要包括情感分析、机器翻译、自动问答等。人工智能技术在上述多个场景的实际应用体现了其行业应用价值，通过人工智能技术与实际场景的深入结合，能够催生大量人工智能服务，从而构建面向不同场景的人工智能应用。

AI算法

深度学习引领了第三次人工智能浪潮的飞速发展。通过构建以卷积神经网络、循环神经网络、图神经网络、生成式对抗网络为代表的深度学习模型结构，以及反向传播优化算法、面向不同学习任务的算法及模型优化算法，实现人工智能算法的不断突破创新。人工智能算法是人工智能产业中最重要的一环，通过人工智能算法的不断升级，结合海量数据，实现算力的有效释放。

AI工具

AI工具是人工智能技术的重要基础设施之一，是数据及算法加速融合的重要驱动力。根据人工智能建模过程特点，AI工具可具体划分为数据处理、模型训练、模型推理三个组成部分。数据处理库包括NumPy（数值计算）、OpenCV(图像处理)、Hadoop(大数据处理)等数据处理及计算库，提供高效快捷数据处理方法。模型训练工具主要包括PyTorch、TensorFlow等深度学习框架，利用上述工具能够实现高性能、分布式的模型训练方案。模型推理工具包括两部分：高效模型优化方法及高吞吐推理引擎：模型量化、剪枝、蒸馏等方法可减少模型复杂度及存储需求，高吞吐推理引擎如TensorRT为深度学习应用提供低延迟、高吞吐率的部署推理，实现快速与高效的模型推理方案。

全流程的软件工具有效提升了算力释放的效率，但其中仍需要专业人员如算法人员、业务人员的人力工作。软件工具自动化、智能化能够有效摆脱人力束缚，加快应用生产效率。如自动化深度学习通过利用智能优化算法，能够面向不同任务进行自动模型搜索、模型超参数自动调节、模型自动压缩，实现AI算法及工具的自动化，提升AI应用生产效率，促进算力高效释放转化为生产力。

AI服务

面向不同AI场景应用需求，利用高精度AI算法及全流程AI工具，最终产出高质量AI服务。AI服务存在形式主要包括模型文件、开发接口、在线服务三种形式。通过构建多种形式的AI服务，可满足生产中不同阶段、不同需求、不同场景需要。AI服务是人工智能应用的最终表现形式，通过高质量及多形式AI服务，实现AI技术在各行业的渗透落地。

小结

智算中心作为承载人工智能应用需求的算力中心，以海量异构数据为资源，基于深度学习、自监督学习、强化学习、自动机器学习、跨媒体多模态等最新的人工智能理论，采用技术领先、生态成熟的AI芯片、AI服务器、高速互联、深度学习框架、资源调度等人工智能计算架构，重点围绕生产算力、聚合算力、调度算力、释放算力四大关键环节提升AI算力。通过打造算力生产供应、数据开放共享、智能生态建设和产业创新聚集平台，面向政府、企业及科研机构等多用户群体提供源源不断的算力服务、数据服务和算法服务，汇聚并赋能行业AI应用，助力行业智慧应用高效化开发，支撑和引领数字经济、智能产业、智慧城市和智慧社会应用与生态健康发展。

预期经济社会效益

降低AI应用成本，提高算力效率

随着数据增长和算法模型复杂化，AI训练过程中对数据的传输和计算要求也越来越高，当前算力供给已难以满足智能化社会构建。AI计算作为全球主流计算形态已成定势，据OpenAI统计，自2012年以来随着深度学习模型的演进，模型计算所需计算量已经增长30万倍，AI模型构建所需算力呈现出阶跃式增长特点。智算中心搭载大量AI服务器，在人工智能数据训练方面有着绝对的优势，在人工智能大数据训练方面有着绝对的优势，将显著提升数据算力效率。IDC在《2020H1中国加速计算市场调查报告》指出，预计到2024年AI加速计算市场规模将达到2019年的2.6倍。

智算中心作为算力生产供应平台，较之一般数据中心，其在构建过程中以融合开放的架构计算系统为平台，以数据为资源，在低成本提供很多开源的人工智能算法及其代码的同时，能够以强大算力驱动AI模型来对数据进行深度加工，源源不断产生各种智慧计算服务。同时，智算中心还将通过平台开放接口的方式将行业领军企业、城市运营中心等算法能力、数据资源及运营服务等输出给IT基础薄弱的企业，从而使得全社会AI应用成本得到大幅降低。

加速AI生态对接，推动创新产业聚集

智算中心未来将承担各种大规模AI算法计算、机器学习、图像处理、科学计算和工程计算任务，并以强大的数据处理和存储能力为全社会提供AI算力服务。算法创新、数据和可以用于训练的算力被认为是推动人工智能发展的三个关键要素，未来随着智算中心大规模投入使用，以“基础设施+框架+算法”为底层基础、以AI模型生成和多场景应用的AI技术生态链将逐渐形成。智算中心通过搭载高性能AI服务器，优化计算平台、框架和底层算法，能实现从模型开发、训练、部署、测试、发布的“流水式生产”与一站式交付，加速算力产业链的形成。

基于AI技术生态链的AI生态将全面加速对接。智算中心算力提升及AI模型构建输出需要大量训练数据做支撑，这些数据的存储处理离不开大数据技术的支持，进而带动一批新兴的专业化大数据企业崛起，数据清洗加工、数据交易、数据分析即服务等新业态新模式将不断涌现。随着AI模型输出能力的不断提升，识别检测、语音交互、机器人等典型人工智能应用也将加速普及，必然驱动一批创新型企业加速成长，部分龙头企业加快布局，为产业资金、人才、技术等要素聚集提供技术条件和基础支撑。同时，智算中心的建设还将带动新能源、新材料等高新技术产业发展，进而构建起以AI算力为核心的人工智能产业新生态体系。

促进产业转型升级，加速新旧动能转化

数字经济、智能产业、智慧城市等的发展离不开数据中心提供必要支撑，但现有数据中心算力服务能力已经难以有效满足实践需求。基于智能计算的IT基础设施，具有高性能计算、高可靠性、高能效和开放生态的优势，适合为大数据、分布式存储、原生应用、高性能计算和数据库等应用高效加速，可以满足多样性计算、绿色计算需求，将在各行业数字化转型过程中发挥重要作用。从技术层面来看，智算中心凭借其强大的AI算力能够为不同用户提供算法产品与服务，并将这些算法产品及服务应用到需求分析、业务流程优化、应用场景原型验证等方面，解决用户面临的业务痛点，真正实现用AI为传统行业用户赋能。

与此同时，智算中心通过提供共性的算力服务、数据服务和模型服务，让算力易用、可用并大幅降低使用成本，使得智慧计算像水电一样成为城市的基本公共服务，进而帮助城市中小企业、创新型企业与传统企业降低企业AI技术研发、应用和部署成本，增强企业创新和转型发展能力。制造行业被认为是未来一段时期内算力需求最迫切、投资规模最大的传统行业，国内制造企业转型过程中需要更多集约化、规模化、普惠型的公共基础设施提供技术支持。智算中心作为未来算力核心生产供应者，通过整合PLM、MES、ERP、3D打印等相关技术、产品和服务，可为制造业企业直接提供支持业务流程的各项服务，帮助制造业企业提升研发设计效率、生产制造能力、企业治理水平和运营效率。

加强政府治理，提升公共服务能力

算力就是生产力，算力是智慧城市发展的新生产力，未来人类所处的每个场景都离不开算力。智算中心作为新基建数字基础设施的主力军，其所承载的AI算力将是驱动城市智慧化发展的核心动力。智算中心以服务形式高效、普惠地提供智慧计算服务，能够有力支撑城市共性技术服务体系，形成以“人工智能为核心，支撑各行业关键共性技术研发”的发展布局。政府作为现代政府治理主体，能够在智算中心强大算力支撑下开展精细化、智能化政府治理。智慧城市丰富的应用场景为智算中心提供了大量AI训练数据，海量训练数据对于数据中心算力提出更高要求。智算中心作为未来城市AI算力生产供应中心，其算力能够充分满足未来智慧城市数据训练要求。

依托于智算中心强大的算力可以模拟建立更加高效精准的AI算法模型，这些算法模型在智慧交通、应急管理、防洪减灾、环境保护、地理测绘等政府治理和公共服务场景中有着极大应用空间。政府治理和公共服务场景中有着极大应用空间。

改善经济运行格局，激发经济发展活力

人工智能是新一轮科技革命和产业变革的重要驱动力量，人工智能计算需求未来将占据80%以上的计算需求，智能计算正成为未来经济的主要增长点之一。根据埃森哲对中国以及全球12个发达经济体的研究，到2035年，人工智能将帮助各国显著扭转经济增速近年来的下滑趋势。通过有效应用智能计算，中国经济增长率有望上升至7.9%，增长额高达7.1万亿美元。根据麦肯锡的预测，人工智能将每年为中国经济增长贡献0.8至1.4个百分点。融合了人工智能算力、硬件架构、云服务的智能计算产业作为数字经济的重要载体，将作为一支新兴力量，带动全球数字经济走向繁荣。

建设运营模式

当前，新型基础设施的建设运营模式主要有政府投资建设、企业建设运营、政府购买服务、政府和社会资本合作等模式。智算中心作为面向人工智能的新型基础设施之一，具有较强公益属性，能够提供AI算力、AI模型等共性技术服务，并有利于企业孵化和社会精细化治理。因此，建议各地在实际建设中，融合政府机构与社会力量，采用政企合作建设运营的框架，高效整合政企资源，在政府的统一规划和指导下开展项目的投资、建设、运营和维护等工作。注重激发市场创新活力、提高建设运营效率。确保发挥出智算中心作为创新载体的公益属性，深度推进产学研用协同攻关，促进成果转化应用，带动新兴产业发展与产业链上下游高效协同，扩大社会效益。

智算中心“投-建-运”主体选择

1. 投资主体

智算中心建设应坚持政府主导模式，政府作为投资主体加快推进智算中心落地，以智算中心为牵引打造智能产业生态圈，带动城市产业结构优化升级，增强城市创新服务能力。一是智算中心具有公共属性。智算中心属于城市新型信息基础设施，与传统城市建设中的煤水电气等公共基础设施具有相同属性，一般由政府承担投资建设职责。二是智算中心具有战略投资性。智算中心作为城市建设公共基础设施，在实际应用中会连接计算、数据和应用等多个方面，直接关系到城市安全发展，要坚持政府主导，避免私人资本、海外资本形成垄断而可能带来的潜在风险。

2. 承建方主体

企业作为最具市场活力的主体，在技术创新、产业培育等方面具有得天独厚优势。智算中心建设应选择政府主导下的政企合作模式，由企业具体承建智能计算中心。考虑到智算中心公共物品属性、技术安全、投资规模巨大等特殊要求，承建企业应为行业头部企业，在技术实力方面处于行业领先水平，技术应用方面有能力强做出前瞻性研判。同时，企业应属于国有资本控股企业，能较好规避智算中心建设和后期运营过程中被私人资本、外资等垄断所带来的不利影响。

3. 运营主体

运营主体为具体负责智算中心投入建设使用后的运营服务机构。与投资主体相比，智算中心运营主体类型应更加多元，运营模式也更加灵活，各类主体通过积极探索差异化个性化运营服务模式保障智算中心高效稳定运行。考虑到智算中心项目的公共安全、高技术门槛和资金门槛，目前更宜采取“投-建-运”一体化模式，即承建主体在项目建设验收后按照合同进行运营服务。此外，经投资主体同意，承建方可以选择将运营服务委托外包给政府出资控股或国有资本控股的专业运营公司。

建设运营模式选择

1. 基于政府独立投资的建设模式

政府直接投资和管理，需要对项目的建设周期、存在风险、成本收益等有很高的把控能力。项目建设运营资金主要来自政府预算资金。该模式的优势在于政府直接主导智算中心建设，可以起到示范效应，建设完成后智算中心所有权归政府所有。

2. 基于特殊项目公司的建设运营（SPV）模式

政府与企业共同出资成立智算中心建设运营项目公司，政府与企业按合作框架协议下按比例出资建设智算中心，政府应对项目公司具有控股权。该模式的投资主体为政府和企业，建设运营主体为项目公司及通过项目协议确定的行业领域优势企业。该模式下，政府既可以直接投资参与项目建设，也可以通过出资控股的国有混合股份有限公司参与项目建设，建设运营模式更加灵活。项目公司由政府授权，按照公司化方式独立运作，负责设计、融资、建造和运营等工作，向政府、企业提供服务或产品并收取费用。政府对项目公司运行管理及建设运营全过程实施监督和指导。该模式的优势在于能够节约政府管理部门建设的资源与成本，有效地筹集资金和转移公共部门的风险。启用企业专业化的运营团队，项目管理方式灵活多样，在项目设计、建设和运营中效率较高，能够提供较高质量的服务，增加收益与回报。

3. 基于“国家-地方-企业”共建的建设运营模式

地方政府会同相关企业共同申请与智算中心相关的国家级试点示范项目工程，以项目工程为载体，集国家、地方和企业三方力量，由国家和地方按比例出资建设智算中心。国家给予资金扶持，地方负责具体建设，企业参与运营管理，在缓解地方财政资金压力的同时对地方产业发展形成带动示范效应。国家层面可以通过给予先期引导资金或专项资金补贴等方式鼓励地方先行先试。地方层面，地方政府可通过直接投资参与项目建设或出资控股的国有混合股份有限公司参与项目建设等方式实现对项目的所有权。企业层面负责项目数据、算法等技术运维服务，进而依托智算中心项目汇聚孵化人工智能企业，促进人工智能产业发展打造“科技研发、产业孵化、创投资本、教育培训、配套政策环境”的智能生态圈。

实施路径建议

作为赋能城市经济、优化政府治理、引领高质量发展的新型基础设施，智算中心的建设部署需要在政府和企业两端发力，以应用需求为牵引，强化顶层设计，通过新建与优化升级相结合、硬件与软件相协调，进一步创新运营机制，促进人工智能与实体经济、智慧城市、社会民生深度融合。

需求导向，明确建设思路

智算中心建设规模普遍较大，建设期的一次性投入成本较高，建设思路须提前明确。首先，需明确建设需求，包括应用场景、辐射市场范围、投资收益方式等，避免盲目投资建设，除满足政府智能化场景应用需求外，还应面向产业智能化发展需求，为中小企业提供多层次、多样化的人工智能算法和模型测试服务。其次，要根据需求制定设计方案、确定承建方与建设模式。一般来说，以下两类城市对智算中心的需求会更加强烈：一类是具备一定人口规模和经济实力的中等以上城市，这类城市有必要发挥区域辐射带动作用，通过智算中心的建立或者数据中心的智慧化升级，在促进本地数字化转型的同时，实现对周边地区的数字化支撑，助推腹地中小城市数字化发展；另一类是重点发展人工智能产业的城市，这类城市具备人工智能产业发展的人才、资金、技术等基础，建设智算中心可以进一步推动各创新要素集聚，产生聚合效应促进经济发展。

改建结合，优化建设方式

面向人工智能应用和产业发展需求，各地可结合自身实际，通过新建与改造相结合模式，建设智算中心，提供智能计算服务平台。一方面，建设专用智算中心。对于产业智能化发展需求迫切、传统数据中心规模较小的地区，可新建围绕人工智能产业需求设计、为人工智能提供专门服务的智算中心，通过搭载新型人工智能芯片和新型人工智能通用算法平台，提供智能算力基础设施及通用软件服务，满足产业智能化发展需要。另一方面，开展数据中心智能化改造。对于规模较大、能耗水平较高的数据中心，可通过智能化改造使其满足人工智能对于算力和算法的要求，实现传统数据中心AI化。硬件方面，可以采用CPU+GPU芯片的方案来满足人工智能计算需要。软件方面，建立面向人工智能应用的通用算法和模型平台，为智慧城市和行业应用提供智能化解决方案。如部分地方基于已有数据中心，通过建立AI中台等方式，打造城市的“超级大脑”。

软硬兼施，提升服务能力

数据、算力和算法是推动人工智能技术进步和产业发展的“三驾马车”。随着新基建的加速建设，人工智能与大数据、云计算、物联网等融合进一步加快，智慧医疗、无人驾驶、智慧城市、智慧金融等应用场景，对于算力和算法提出了更高的要求。一方面，采用先进硬件芯片，强化算力支撑。随着数据总量的增长和智能应用需求的扩大，人工智能产业对算力的要求越来越高。因此，在建设智算中心时，可采用新型人工智能芯片和新型计算方法相结合提供更快的计算能力。另一方面，建立通用软件平台，加强算法保障。除了算力强外，智算中心的强大能力还体现在人工智能应用的算法和模型，通过建立面向人工智能应用需求的通用算法和模型平台，为智慧应用提供共性的算法支撑。

先进适用，统一建设标准

智算中心建设应向更加智能、算力更强、速度更快和节能安全方向发展，着力于为各行业构建全场景的智能解决方案。一方面，面向智算中心的长期应用，应坚持实用、经济的建设原则，采用先进的概念、技术和方法，采取相对成熟且具面向未来的结构、设备、工具，满足未来一段时间应用需求。另一方面，面向智算中心的多领域应用，应坚持资源共享、弹性扩展的建设原则，严格执行统一建设标准，确保系统的开放性和标准性，满足功能不断扩展的需求，保障技术和设备的协同运行能力。鼓励开展智算中心关键通用技术、关键领域技术、安全等重点标准的研发工作，围绕应用场景、产业创新和生态构建，开展领域标准研究与应用实践，引领行业发展。

规范推进，加强项目管理

通过出台管理办法、实施细则等，不断完善智算中心等大型建设项目的审批管理机制，明确规划和审批管理、建设和资金管理、监督管理等环节的具体内容。一方面，确保合规推进。智算中心建设项目应紧密衔接本区域或本地规划，科学论证可行性，按规定履行合理用能标准、节能评估、规划选址、用地预审、环境影响评价等审批程序并备案。对于政务信息化项目，履行审批程序前应依管理规定编报项目建议书、可行性研究报告、初步设计方案、投资概算等。另一方面，开展全流程项目管理。实施工程咨询、设计、监理、验收测试和成效评价等工程建设全过程的质量跟踪，加强承建方对于智算中心建设项目的进度管理工作和监理方的竣工验收工作，确保项目建设规范合理，能够尽早投入使用并发挥应有的效益。

丰富场景，打造产业生态

智算中心可以为多个领域提供算力支撑，辅助构建开放融合的人工智能生态，助推AI产业化与产业AI化转型发展。一方面，深度挖掘市场需求，丰富智算中心应用场景。智算中心可以为多个领域提供算力支撑，其核心特征是通过利用AI技术，支持和辐射智能交通、智能电网、智能制造、智能教育、智能金融、智能医疗、智能家居、智慧互联网等行业智能应用场景。以智慧城市建设为例，在智算中心支撑下，“城市大脑”应用有效提升了城市科学治理能力和政府治理水平。另一方面，产学研用结合，打造人工智能产业生态。建立智算中心，既是助力未来发展人工智能的必要举措，也是带动现有人工智能产业链加速渗透应用的有效路径。当前，人工智能正在从AI产业化向产业AI化的方向发展，需要建立开放融合的人工智能生态，从硬件到应用，全产业要紧密配合，面向多样化个性化的用户需求提供整体解决方案，实现以算力支撑为代表的AI基础设施、以机器学习、识别训练为代表的技术层以及以多场景领域AI化转型的应用层协同发展。

全文下载

