智算产业发展白皮书

(2023 年)

中国电信研究院

天翼智库

2023 年 8 月

编制说明

主编单位：中国电信研究院

参编单位：中国电信股份有限公司安徽分公司

深圳海兰云数据中心科技有限公司

顾问专家：

中国电信研究院战略发展研究所所长：饶少阳

编委成员：

孙雪媛、陈元谋、赵静、马腾滕、熊小明、魏玥、李朔萌、谢林翰

陈锡根、王勇

联系电话：010-50902887

邮箱：sunxy11@chinatelecom.cn

目录

引言

以大模型为代表的通用人工智能不断演进，人工智能、机器学习、大数据分析等技术在金融、制造、汽车等领域持续渗透，大模型应用场景愈加广泛，正加速算力产业结构变革，智能算力将取代通用算力成为算力结构最主要构成，智算产业迎来了高速发展期。

工信部最新数据显示，我国算力总规模已位居全球第二，保持年约 30%快速增长，新增算力设施中智能算力占比过半，成为算力增长的新动能；我国算力产业创新能力持续增强，面向大模型训练、推理等高性能芯片供给持续增强，多元异构计算技术加速普及，有力支撑人工智能、区块链、元宇宙等新兴应用发展。

算力是数据中心的服务器通过对数据进行处理后实现结果输出的一种能力[1]。智算是算力的一种，指具有提供人工智能应用所需算力服务、数据服务和算法服务的智能算力，利用 CPU 与 GPU、FPGA、ASIC 等加速芯片的异构组合，实现高精度通用算力和低精度专用算力的融合供应[2,20]。智算涵盖从底层高性能芯片、服务器和网络设备，到智算中心基建、机电配套和软硬件服务平台，再到顶层人工智能应用等完整体系，产业上下链长、集聚效应显著。智算为经济增长提供数字转型、智能升级、融合创新的新动力，带动人工智能及相关产业倍速增长，成为我国数字经济发展的新引擎。

本白皮书系统分析了智算产业发展环境、产业链全景特点、最新进展及面临挑战，指出了智算产业五大发展趋势、七大技术关键词，提出了我国智算发展潜力评估体系及分省指标结果，并介绍了典型智

算中心建设场景案例。

本白皮书由中国电信研究院编制，我们希望通过此白皮书为我国智算产业市场洞察、技术创新、生态建设，高水平发展提供参考启示。

一、智算发展迎来新机遇

1、AI 大模型驱动的智算时代正加速到来

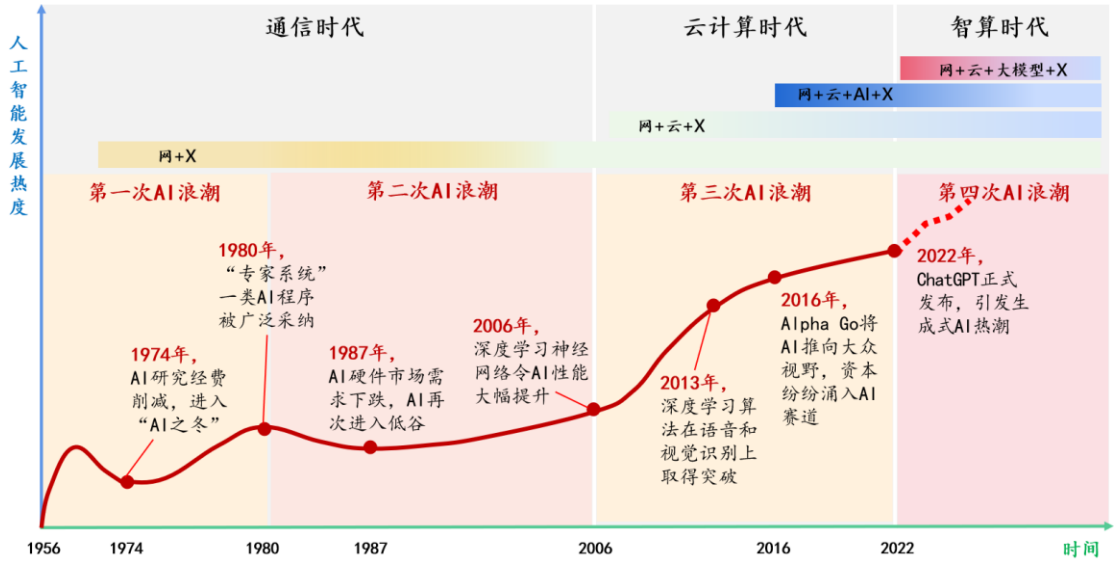

图 1 人工智能产业发展历程

人工智能自 1956 年诞生以来，历经三次发展浪潮。

第一次浪潮（1956-1970s），神经网络相关基础理论被提出，搜索式推理、自然语言等大量 AI 程序和创新研究涌现。但由于大部分 AI 程序不具备解决复杂问题的能力，造成 AI 研究经费开始大幅削减，AI 迎来第一次发展低谷。该阶段算力主要以 IBM 大型计算机为主，以集中的方式分配使用。

第二次浪潮（1980s-2000s），80 年代名为“专家系统”的 AI 程序问世，极大增强了 AI 的实用性。但“专家系统”应用仅限于特定领域，迭代升级难度及维护成本高，规模推广难度大，AI 再次进入发展低谷。90 年代小型计算机性能每 18 个月翻一番，且价格和耗电量大幅降低，算力逐渐进入分布式发展阶段[3]。

第三次浪潮（2006-2020），深度学习等算法的突破使得 AI 性能

大幅提升。移动通信技术快速发展，共享计算资源、提高算力利用率等需求催生出以云计算为中心的集中式共享算力模式。2016 年，谷歌研发的 AlphaGo 将 AI 推向大众视野，语音识别、视觉处理等 AI 应用逐渐渗透到各行各业。同年，中国电信提出“云网融合”发展方向，将云计算和网络技术有机结合，实现计算和网络资源的统一管理和优化配置，推动网络和算力一体化供给、运营和服务。

当前人工智能正迎来第四次发展热潮，加速进入大模型驱动的智算时代。2022 年 11 月，OpenAI 公司正式推出 ChatGPT，推动生成式 AI 应用进入爆发期，M6、文心一言、盘古等国内 AI 大模型层出不穷，AI 算力需求被推到“井喷”状态，开启智算时代。随着数据指数级增长，计算密度越来越高、计算节点分布越来越广，加速云网与 AI、安全等要素融合。算力逐渐由终端计算等需求驱动的“被动式”发展，转向促进 AI 大模型训练、实现通用人工智能等代表的“主动式”发展，从“技术工具”进阶为社会经济发展的“底层动力”。

2、智能算力成为数字经济发展的新引擎

人工智能产业市场前景广阔，成为推动全球经济发展的新动力。IDC 预测，全球以 AI 为中心的各类系统的软件、硬件与服务支出，2023 年将达到 1540 亿美元，到 2026 年将超过 3000 亿美元，预计 2022 年至 2026 年间复合年增长率（CAGR）为 27%[4]。2021 年中国 Al 服务器市场规模为 53.9 亿美元，预计 2025 年达到 103.4 亿美元，2021 年至 2025 年间 CAGR 达 17.7%[5]。

智算产业集群化作用显著，成为带动人工智能及相关产业快速发展的新动力。到 2035 年，人工智能的发展将给我国甚至全球经济增长带来突出贡献。预计到 2026 年，人工智能技术对于全行业的渗透率将超过 20%[6]。据信通院数据，2022 年我国算力核心产业规模达到 1.8 万亿元，其中人工智能核心产业规模达 5080 亿元，同比增长 18%[7]；2022 年我国新增算力基础设施中智能算力占比过半，智算成为算力增长新曲线，智算中心正在支撑人工智能产业的快速发展，支撑其到 2025 年达到 4000 亿，带动 5 万亿产业目标；2030 年达到 1 万亿，带动 10 万亿元产业目标[8]。

3、国家和地方密集出台政策支持智算布局

表 1. 我国部委智算中心建设相关政策

时间	部委	政策	相关内容
2021年1月	中共中央	《建设高标准市场体系行动方案》	强调要加大新型基础设施投资力度，推动第五代移动通信、物联网、工业互联网等通信网络基础设施，人工智能、云计算、区块链等新技术基础设施，数据中心、智能计算中心等算力基础设施建设。
2021年7月	工信部	《新型数据中心发展三年行动计划（2021-2023年）》	加快高性能、智能计算中心部署，推动新型数据中心与人工智能等技术协同发展，构建完善新型智能算力生态体系。
2022年1月	国务院	《“十四五”数字经济发展规划》	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。
2022年12月	国务院	《扩大内需战略规划纲要（2022-2035年）》	“加快建设信息基础设施，推动人工智能广泛、深度应用”，“发展普惠性‘上云用数赋智’”等。
2023年2月	国务院	《数字中国建设整体布局规划》	夯实数字中国建设基础，系统优化算力基础设施布局，促进东西部算力高效互补和协同联动，引导通用数据中心、超算中心、智能计算中心、边缘数据中心等合理梯次布局。
2023年4月	中共中央	政治局会议重要讲话	要重视通用人工智能发展，营造创新生态，重视防范风险。
2023年7月	科技部	国家新一代人工智能公共算力开放创新平台	正式批复9个平台建设国家新一代人工智能公共算力开放创新平台、16个平台建设国家新一代人工智能公共算力开放创新平台（筹）。

数据来源：各部委官方文件

我国高度重视智算产业发展，围绕智算中心、人工智能、大模型等先后出台系列政策文件，加快产业布局。“十四五”规划和 2035 年远景目标纲要中明确提出要“加快构建全国一体化大数据中心体系，强化算力统筹智能调度，建设若干国家枢纽节点和大数据中心集群”。工信部、国家发改委等先后出台《新型数据中心发展三年行动计划

（2021-2023 年）》、《全国一体化大数据中心协同创新体系算力枢纽实施方案》等文件，启动“东数西算”重大工程。2023 年 4 月，中共中央政治局会议中强调“要重视通用人工智能发展，营造创新生态，重视防范风险。”7 月，科技部批复 25 个平台建设国家新一代人工智能公共算力开放创新平台（含筹建）。

表 2.我国各省市智算中心建设相关政策

时间	地方	政策	相关内容
2021年5月	广东省人民政府	《广东省人民政府关于加快数字化发展的意见》	布局建设智能计算中心等新型高性能计算平台，提供人工智能算力支撑。
2021年9月	四川省人民政府	《四川省“十四五”新型基础设施建设规划》	加快建设成都鲲鹏生态基地、中科曙光先进微处理器国家工程实验室、华为成都智算中心，着力构建基于鲲鹏及昇腾、海光自主知识产权芯片及自主可控超融合异构技术等的多层次融合架构计算系统，打造国际领先的人工智能计算与赋能平台。
2021年12月	重庆市人民政府	《重庆市数字经济“十四五”发展规划（2021—2025年）》	加快建设人工智能计算中心，积极构建人工智能数据资源、模型库、算法库、标准数据集和开放平台，夯实人工智能创新发展“算法+算力+数据”基础。
2022年1月	浙江省人民政府	《建设杭州国家人工智能创新应用先导区行动计划（2022—2024年）》	支持新型智能计算架构试验验证等重大科技基础设施（装置）建设。
2022年9月	上海市人民政府	《上海打造未来产业创新高地发展壮大未来产业集群行动方案》	推动超大模型智能计算突破，培育智能计算自主框架和算法平台，发展自主智能芯片。
2023年1月	成都市政府	《成都市围绕超算智算加快算力产业发展的政策措施》	明确成都每年将发放总额不超过1000万元的算力券。
2023年5月	上海市发改委	《上海市加大力度支持民间投资发展若干政策措施》	充分发挥人工智能创新发展专项等引导作用，支持民营企业广泛参与数据、算力等人工智能基础设施建设
2023年5月	北京市人民政府	《北京市促进通用人工智能创新发展的若干措施》	高效推动新增算力基础设施建设，将新增算力建设项目纳入算力伙伴计划，加快推动海淀区、朝阳区建设北京人工智能公共算力中心、北京数字经济算力中心，形成规模化先进算力供给能力，支撑千亿级参数量的大型语言模型、大型视觉模型、多模态大模型、科学计算大模型、大规模精细神经网络模拟仿真模型、脑启发神经网络等研发。

数据来源：各省市官方政策文件

地方政府纷纷发布智算产业相关政策，开展智算中心相关基础设施建设工作，提供普惠算力服务。北京发布《北京市促进通用人工自能创新发展的若干措施》高效推动算力基础设施建设，将新增算力建设项目纳入算力合作伙伴计划，加快推动智算中心建设，形成规模化先进算力供给。上海出台《上海市助力中小微企业稳增长调结构强能力若干措施》助力中小企业数字化转型，发放“AI 算力券”，重点支持租用本市智能算力且用于核心算法创新、模型研发的企业，最高按合同费用 20%进行支持。成都印发《成都市围绕超算智算加快算力产业发展的政策措施》明确每年发放总额不超过 1000 万元的算力券，

用于支持算力中介服务机构、科技型中小微企业和创客、科研机构、高校等使用国家超算成都中心、成都智算中心算力资源。

二、智算产业全景及新进展

1、智算产业链初步形成，生态集聚效应不断增强

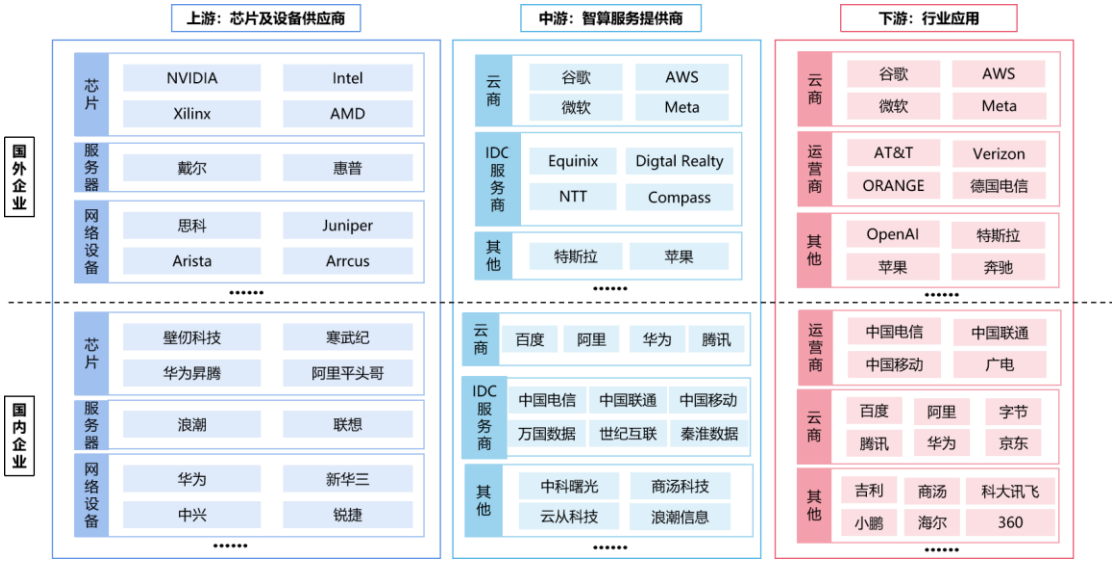

图 2 智算产业链图谱

目前，我国智算产业链已经初步形成，涵盖由芯片、软件、硬件供应商构成的上游产业，由云商、电信运营商、第三方数据中心服务商等构成的中游产业，以及由互联网、交通、金融、工业等行业等用户构成的下游产业。

（一）上游：AI 芯片领域形成多方竞争格局

GPU、FPGA 技术壁垒高，迅速形成寡头格局。英伟达（NVIDIA）凭借 NVLink、NVSwitch 等集群性能优势以及良好的 CUDA 生态，成为全球 GPU 主要供货商，其 A100 芯片占据了数据中心 GPU 市场 90%以上的份额。据 IDC 数据，预计到 2025 年 GPU 仍将占据 AI 芯片 8 成市场份额。同时赛灵思（Xilinx）和英特尔（Intel）已在 FPGA 领域形成双寡头格局，市场份额占比约 90%[9]。

TPU、NPU 逐渐兴起，呈现“百家争鸣”态势。以 TPU、NPU 为代

表的 ASIC 凭借吞吐量、功耗、算力等优势，逐渐被广泛应用于人工智能领域。国外以谷歌为首发布 TPU 芯片，国内寒武纪、华为、阿里等公司也都推出了深度神经网络加速的 ASIC 芯片，如华为昇腾 NPU、阿里平头哥 NPU。

白盒交换机以其软硬解耦、灵活可编程、高速转发等优势受到云商智算中心大规模组网青睐。Omdia 数据显示，2022 年全球数据中心以太网交换机市场份额白盒供应商占比 32%，其中 Arista 占比 18%。在北美市场，全球 TOP3 云商亚马逊、谷歌和 Meta 的白盒交换机购买规模已超市场总规模的三分之二。

InfiniBand 和 RoCE 作为智算中心高性能网络的主流方案，满足智算网络的低时延、大带宽、稳定运行、大规模以及可运维的需求。InfiniBand 网络方案及配套设备供应商主要包括英伟达、英特尔、思科，其中英伟达市场占有率超七成。支持 RoCE 的交换机厂商较多，主要以新华三、华为为主。支持 RoCE 的 NVIDIA ConnectX 系列网卡当前市场占有率比较高。

（二）中游：云商及 IDC 服务商基于自身优势提供智算服务及解决方案

云商、科技公司借助自身技术壁垒提供大模型及平台服务。主流云商一方面自建大型智算中心，如 Meta 宣布取消或暂停部分正在建设的数据中心，对其 11 个正在开发的项目进行重新设计，彻底转向人工智能数据中心的建设。另一方面加速布局 AI 大模型，如谷歌“PaLM-2”、Meta“Llama 2”等。特斯拉、苹果等科技公司基于自身

业务优势，一方面自建定制化智算中心，如特斯拉面向自动驾驶等领域建设超算中心 Dojo，拥有超过 100 万个训练节点，算力达到 1.1EFLOPS[10]。另一方面，积极布局 AI 大模型体系，巩固自身行业优势壁垒，如特斯拉 AI 机器人“擎天柱”、苹果“Apple GPT”。

IDC 服务商依托云/网资源优势，积极参与智算建设。国内运营商积极建设智算中心及平台，如中国电信推出息壤智能计算平台，提供智算、超算、通算多样化算力服务，为大模型训练、无人驾驶、生命科学等场景提供软硬一体解决方案，RDMA 吞吐可高达 1.6Tb[11]。国外 IDC 服务商仍在布局阶段，如 2023 年日本 NTT 宣布将在 5 年内投资 8 万亿日元（约合 590 亿美元）用于人工智能、数据中心和其他增长领域[12]；Equinix 的 2023 年全球科技趋势调查报告显示，人工智能应用率上升，但 IT 基础设施没有为人工智能做好充足准备。

（三）下游：车企领衔行业大模型落地应用

互联网、交通、金融、工业等行业，基于大模型带动自动驾驶、机器人、元宇宙、智慧医疗等下游产业发展。海外大模型行业应用主要在传媒游戏、机器人、办公等领域落地，如 Meta 推出 AI Sandbox 为广告生成不同的文字、Apple 推出生成式人工智能元宇宙产品 Visin Pro 头显，并计划在 siri 嵌入类 GPT 功能。哈维基于 GPT 及行业数据推出 AI 法律助手。国内大模型行业应用主要聚焦金融、医疗、传媒游戏、智能汽车等领域，如百度文心大模型助力浦发银行、泰康保险在投资决策、理赔信息检索等方面的应用。华为盘古大模型为国家电网电力巡检提供智能服务。

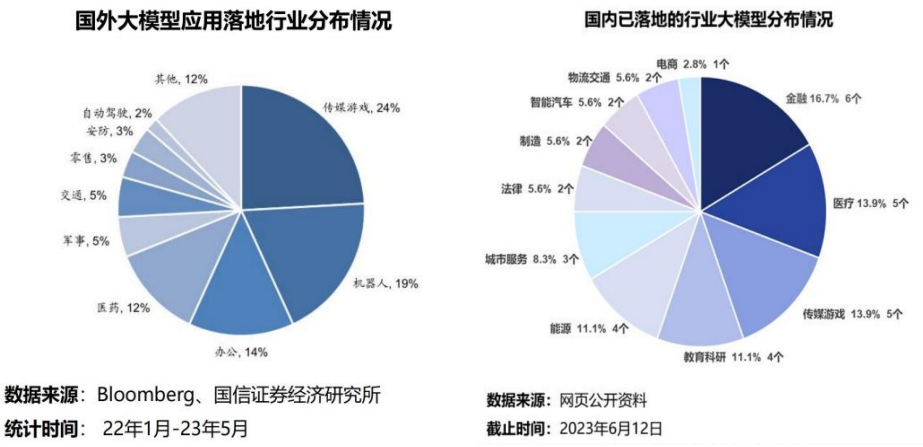

图 3 国内外大模型行业分布[13]

车企布局智算中心用于自动驾驶大模型训练。特斯拉基于 Dojo 超级计算机先后推出 BEV 大模型、端到端自动驾驶大模型，推动高阶智能驾驶落地，预计到 2024 年算力将达 100EFLOPS。吉利星睿智算中心自研汽车行业 AI 对话模型，初步完成百亿参数的大模型训练，吉利星睿智算中心（湖州）预计 2025 年算力规模将达 1.2EFLOPS[14]。小鹏汽车自动驾驶智算中心“扶摇”（乌兰察布），基于阿里飞天智算平台，算力可达 600PFLOPS，将小鹏自动驾驶核心模型的训练提速近 170 倍[15]。毫末智行智算中心“雪湖·绿洲”（山西大同），基于火山引擎智算云解决方案，算力达 670PFLOPS，模型训练效率提升 100 倍[16]。

2、国产自研 AI 芯片加速入场，短期高效供给仍受限

国产硬件厂商持续突破 AI 芯片性能，提升市场竞争力。华为推出昇腾 910，性能对标英伟达 A100，可用于智能手机、云计算、自动驾驶等领域，同时推出 AI 开源计算框架 MindSpore，支持用户进行 AI 开发。寒武纪提供云边端一体、训练推理融合等系列 AI 芯片产品及平台化基础系统软件，重点对推荐系统和大语言模型的训练推理等

场景进行优化。壁仞科技等初创公司不断与多方建立合作关系，如万国数据、浪潮、中国移动等，聚焦云端通用智能计算，重点在 AI 训练和推理、图形渲染等领域发力。

大型云商自研 AI 芯片，以摆脱对国外技术依赖。阿里面向自身电商、汽车、家电等领域需求自研 AI 芯片，基于 RISC-V 架构和自研算法推出含光 800 NPU，支持 TFlops 级别浮点运算。百度面向搜索、智能交通等领域的深度学习运算需求，推出昆仑系列 AI 芯片，用于大模型推理。腾讯依靠蓬莱实验室推出 AI 推理芯片“紫霄”，已用于腾讯会议等多个内部业务。

我国自主 AI 芯片在系统效率等方面与国际领先产品仍有差距，并存在性价比待提高、架构不够兼容、配套工具不够成熟、应用场景不够广泛等问题。**制程方面，**目前英伟达已率先到达 4 nm，而国内厂商多集中在 7 nm[17]；**算力方面，**国内厂商大多不支持双精度（FP64）计算，且仅在单精度（FP32）及定点计算（INT8）方面与国外中端产品持平；**生态方面，**与英伟达 CUDA 的成熟生态相比，国内企业多采用 OpenCL 进行自主生态建设，存在明显差距。

3、智算中心建设版图持续扩张，智算服务灵活多样

智算中心聚焦东部城市，以政府主导国产化为主。截至 2023 年 5 月，全国超 35 个城市在建或投运 44 个智算中心（在建 15 个智算中心，投运 29 个智算中心），其中明确面向 AI 大模型应用的有 11 个。**地理分布集聚一线及省会城市，与大模型研发分布强相关。**智算

中心建设以东部为主，京津冀、长三角、粤港澳共 29 个（占比近 66%），其中 9 个在建，20 个投运，面向西部枢纽节点逐渐开展布局。**东部多为政府主导建设，且国产化占比高（54%），西部以云商自建为主。**地方政府牵头主导 34 个（占比近 80%），主要满足当地 AI 产业发展，且以华为昇腾、寒武纪等为主要合作方提供国产化能力。西部以云商为主，如阿里乌兰察部智算中心、字节跳动与毫末智行合建雪湖绿洲（山西大同）智算中心。**受限于需求不清晰、高性能芯片产业生态不成熟等因素影响，智算规模普遍偏小。**智算中心规模在 100-300PFLOPS 内占比超 70%以上，超过 1EFLOPS 规模的智算中心约占 25%（超半数为云商及大型企业自建），且全部集中在京津冀、长三角和粤港澳区域。

图 4 我国智算中心及大模型分布

由于智算需求场景多样且高度定制化，相较于传统数据中心，智算中心服务模式呈现多元化特点，包括机房托管、算力租赁、智算平台、工具集及咨询等增值服务、模型即服务（MaaS）、大模型应用服务以及各种组合模式。

（一）机房托管：机房托管服务与传统数据中心服务模式类型相同，但需要面向智算提供更高层次的定制化（功耗、配电、网络等），主要面向云商、AI 公司、大型央企等客户。

（二）算力租赁：主要面向中小型科技公司、IT 公司、小参数量的模型（10B 规模）等客户，通过将闲置 GPU 资源通过云服务的形式将服务器或虚拟机租用给用户，采用按使用时间及利用率收费。

（三）大模型托管、训练、部署、订阅等从 IaaS 到 SaaS 全线服务。一是提供 GPU 主机、高性能计算、批量计算等 IaaS 产品。二是依靠智算平台提供公有云和专有云，为各类科研、公共服务和企业机构提供算力调度、数据处理、模型开发等一体化智能计算服务。三是通过 MaaS 提供模型定制、精调、部署等一站式模型服务。四是基于大模型和 MaaS 能力全面智能升级 SaaS 应用，帮助企业构建行业大模型或集成在企业应用上、以及面向公众用户提供搭载大模型应用的基于大模型的搜索引擎、数字人等服务。

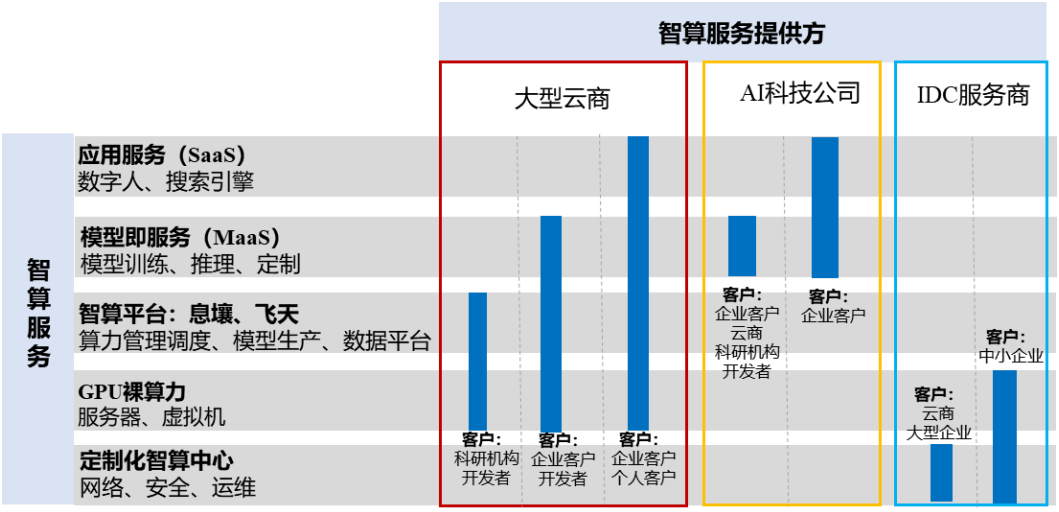

图 5 智算服务体系

未来智算服务模式将由现阶段集成 AI 大模型能力的云产品的卖方市场，逐步过渡到围绕产品提供配套衍生服务，最终形成基于标准化智算中心基于“AI 原生”生态服务的买方市场。

4、大模型呈蓬勃发展态势，助力产数业务发展

我国大模型研发快速增长，大模型研发分布以东部城市为主。从全球已发布的大模型分布来看，中国和美国大幅领先，超过全球总数的 80%，美国在大模型数量方面始终居全球最高，中国从 2020 年进入大模型快速发展期，目前与美国保持同步增长态势。据不完全统计，目前中国 10 亿参数规模以上的大模型已发布 79 个，14 个省市/地区都在开展大模型研发，与智算中心布局一致，主要集中在北京（38 个）、广东（20 个）、浙江（5 个）和上海（5 个）[18]。其中大模型开源占比过半，高校/科研机构是开源主力。清华大学的 ChatGLM-6B、复旦大学的 MOSS 以及百度的文心系列大模型开源影响力最高。

通用大模型不仅需要海量数据与雄厚算力支撑，对资金实力、人

才队伍等也提出更高要求。如 ChatGPT 单次训练成本高达数百万美金，OpenAI 核心团队 87 人，全部来自世界顶尖高等院校。**未来将呈现少数几家通用大模型，并涌现出无数更贴近产业需求的行业大模型的趋势。**

智算赋能行业应用，是产数业务发展的“加速器”。行业大模型通过对垂直细分领域的数据进行更有针对性的训练和优化，从而更好地理解行业的语义和规范，更有效地执行专业性更强的任务。如金融的风险控制和投资决策，医疗的图像识别和诊断，交通的调度和路径优化，能源的能耗预测、碳排放监测等。预计生成式 AI 能为这些行业带来 1000 亿美元到 3000 亿美元的收益。通用大模型企业基于自有通用大模型+外部行业数据的模式拓展多个行业大模型，行业公司基于开源大模型+内部行业数据赋能自身应用。截至 2023 年 8 月，国内已落地的行业大模型共 72 个，主要集中在金融（14 个，19.4%）、医疗（14 个，19.4%））、传媒游戏（8 个，11.1%）及教育科研（8 个，11.1%）。

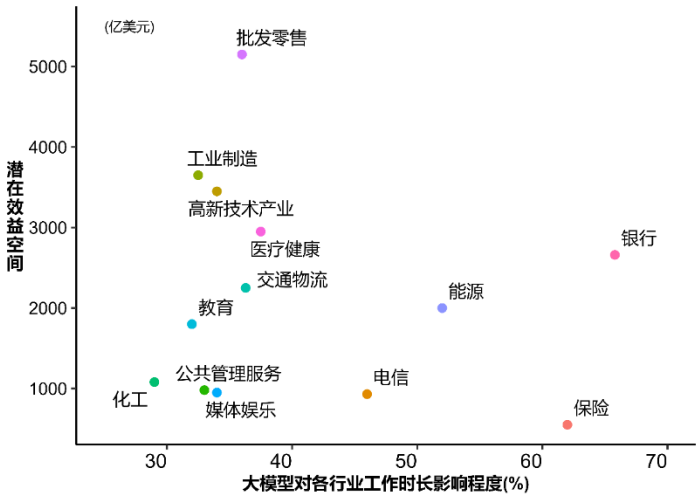

图 6 大模型潜在影响矩阵[19]

三、智算发展五大新趋势

趋势 1：国产多元异构算力融合推动智算长效发展

大模型发展推动 CPU、GPU、DPU 等“XPU”异构算力融合。一方面，模型训练、边缘推理、数值模拟等不同智能应用需要智算中心提供不同的算力，如自动驾驶、智慧医疗等场景既需要高精度通用算力也需要低精度专用算力[20]。另一方面，随着多模态大模型流量规模增长，CPU、GPU 需要拿出更多精力处理数据传输，需要利用 DPU 减负，从而更好地处理“本职工作”。中国电信自研紫金 DPU 实现服务器虚拟化零损耗，全面释放算力，同时网络 PPS 性能翻倍、存储 IOPS 性能提升两倍、网络时延降低至原来的四分之一。

高端 AI 芯片国产化能力是我国智算产业长效发展的关键。一方面，我国 AI 芯片需求增长迅猛，华为数据显示，我国对人工智能芯片的需求半年内增长了十倍以上；IDC 预测，未来 18 个月，GPU、ASIC 和 FPGA 等 AI 芯片搭载率将持续增高。另一方面，我国高端 AI 芯片性能与国际领先水平仍有差距，对美国依赖较大。随着美国对中国高端 AI 芯片的管制进一步加强，如英伟达等厂商对中国（含香港）禁运高性能 GPU，需要我国持续加强芯片技术攻关，提升 GPU 的国产化替代能力。

趋势 2：智算从单节点向区域化协同、边端部署演变

大模型驱动的智算成为东数西算的最佳实践。由于异构算力封装、

超大带宽和超低延迟传输网络技术仍未解决，以当前模型训练参数量（千亿级）为参考，大模型训练等的 AI 计算基本依靠单智算中心完成，且基本集中在同构智能算力中心。智算中心选址多位于东部地区，东部区域在传统数据中心建设方面，由于受能耗、成本等因素的影响发展放缓，但各地政府为实现大模型的创新培育与产业聚集，短期内将主导智算中心发展，形成布局一线及省会城市。长期来看（5 年以上），受成本、双碳目标以及业务模式等因素影响，集约、规模化的智算中心向全国一体化枢纽节点布局的趋势不会改变。未来随着计算机视觉、科学计算等多模态大模型的发展以及参数量的规模增长（万亿以上），将带动东数西训、东数西渲成为东数西算场景落地的最佳实践，并呈现两大趋势：一是大模型演进为多个智算中心分布式训练，且此时智算中心间可以通过全光网等方式实现 us 级时延，智算中心间交互带宽达 T 级别以上；二是业务应用调用多个专业大模型，可能形成云计算中心与智算中心间一对多的互联需求，流量规模增长。西部地区具备发展智算中心、承接东部算力需求的潜力，东西跨区域协同将更加突出。

训练-推理的集中-边缘/终端两级化布局逐步形成。现有大模型业务模式主要包括与大模型直接交互和基于大模型能力的产品改造。前者以猎奇为主，短期并发难以持续，如 ChatGPT 的访问量增长率 1 月环比增长 131.6%，5 月下降至 2.8%。后者更多实现大模型与产品、业务流程的融合，将成为主流形态，如集成了 GPT4 的 Bing 搜索引擎用户访问规模已超 ChatGPT。随着多模态大模型逐步成熟，将推动 2B

生产型和 2C 消费型流量渐成规模，以高频富媒体即时交互为主，业务应用调用多个专业大模型成为主要方式，驱动分布式推理智算中心下沉，中心（训练）-边缘（推理）将成大模型的主流部署方式。此外，随着大模型轻量化处理、终端性能的持续提升，大模型从云端到终端部署渐成发展趋势。截至 2023 年 2 月国内存量手机终端智能算力总规模是我国数据中心算力总规模 12 倍以上，相当于近一百万片英伟达 H100 芯片算力[21]，大模型的云-边-端协同应用将在未来几年快速发展。

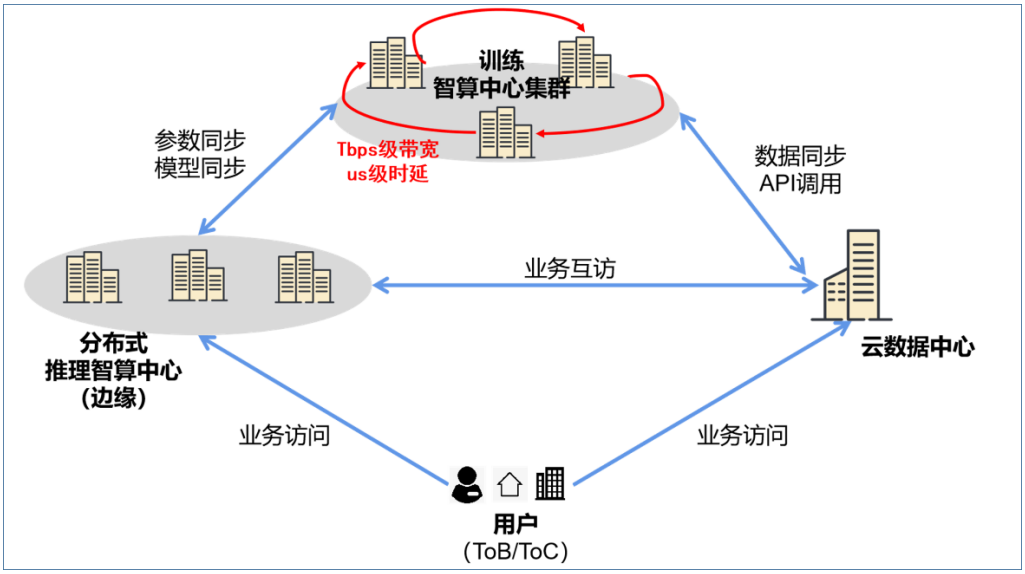

图 7 智算中心业务访问模式

趋势 3：普惠泛在的智算服务生态正逐步构建

智能算力使用具有周期性，复用难。AI 大模型以“大规模预训练 + 微调”为主，前期预训练工作量大，且需要高性能大算力 AI 芯片支撑，算力需求呈现周期性，后期推理算力对芯片计算能力要求相对较低。智算中心的算法模型、AI 架构定制化程度高，其他场景难以复用。据 IDC 调研，超过 80%的受访组织表示会考虑购买预先训练好

的人工智能模型，但未来 2-3 年私有化部署仍将是整个智算市场的主流。由于当前国内高性能芯片受限、智能算力建设及使用门槛高等原因，借助平台调度实现算力错峰使用，并整合数据集、组件、算法模型提供平台级服务，可实现全社会算力服务普适、普惠和高效利用，因此成为业界运营智算中心的趋势。

地方政府主导建设公共算力服务平台，提供普惠算力。算力服务多以场景化云服务的形式交付，用户按业务需求采购算力、存储、带宽等专业服务，实现无处不在的计算，服务模式将从“资源式”向“任务式”转变。政府以城市为单位建设公共算力服务平台，用于连接社会多方智算中心，主要面向中小型企业或科研机构提供普惠算力，同时助力当地人工智能产业孵化，如上海公共算力服务平台、北京多元智算中心等。现有智算中心的软硬件通用性和兼容性较低，需要进一步推动产业链上下游开放协同，实现不同品类、不同技术路线的芯片、算法、模型、应用等要素实现“横向”兼容、“纵向”耦合，确保各层次灵活构建，降低迁移应用门槛，共同推动行业赋能。

趋势 4：确定性、高性能网络助推大规模智算集群构建

智算中心内网络无损高速互联是关键。大模型对数据中心内网络的传输效率有着严格的要求。一是网络丢包 0.1%会导致算力损失 50%（华为实验数据），对于一个可以承载 1.6 万卡的集群而言，近 10 万个光模块平均 4 天左右就会有故障发生。二是面对千亿、万亿参数规模的大模型，训练过程中通信占比最大可达 50%，仅单次计算迭代

内梯度同步需要的通信量就达百 GB 量级。因此，无阻塞、高吞吐量成为面向大模型训练的智算中心内网络的核心诉求。

智算中心间确定性、无损网络研究，是实现跨域多元算力整合的关键。现阶段大模型的训练、推理主要在单一智算中心内进行，未来随着大模型发展以及训练任务的增多，单点算力资源无法满足训练需求，需要将物理位置上分散、归属于不同方所有的多个智算中心之间构建高性能互联网络（DCI），从而整合成一个更大规模的虚拟智算中心以期达到智算中心内部无损网络传输水平。当前中国电信已完成单波 400Gbit/s、传输容量 44Tbit/s、传输距离 1050km 的传输系统，创造了实时光传输容量距离积的新世界纪录（46.2Pbit·km/s），为部署 400G 光传输骨干先现网提供了实验验证[22]；自研算力网关在东数西渲等业务场景中，解决跨域算力调度。鹏城实验室开展深圳和广州超算 10 TB 全光网络互联研究。

趋势 5：低碳化发展格局需创新智算-电网协同模式

绿色电力不产生碳排放，助推智算中心零碳运营。中国工程院院士戴琼海表示，预计 2030 年智能计算年耗电达到 5000 亿度，占发电总量 5%。根据斯坦福人工智能研究所的研究数据，OpenAI 的 GPT-3 单次训练耗电量高达 1287 兆瓦时，相当于 120 个美国家庭 1 年的用电量、10000 辆特斯拉跑满 10 万公里消耗的电量，而这仅仅是训练 AI 模型的前期电力，占模型实际使用时所消耗电力的 40%。作为用电大户，智算中心必须因地制宜利用各种可再生能源，针对地域、时间、

天气等对绿电供给影响较大的问题，通过储能、源网储荷一体化等方法应对。

零碳是智算中心发展的长远目标。零碳是指直接或间接产生的温室气体排放总量，通过节能减排、清洁能源、碳交易等方式进行正负抵消，实现总碳排放为零。一是通过减碳，运用技术手段降低用能、提高能效、提高绿色能源使用等；二是通过碳抵消，购买绿电、绿证等来进行碳排放的消纳。谷歌宣布计划 2030 年实现零碳运营，开发并部署了碳智能计算平台，通过获得各国与地区历史、实时和未来 24 小时内每小时电力能源结构及碳强度，通过在时间或空间上转移计算任务，实现计算任务与低碳电力供应的最佳匹配。

四、智算技术发展的七大关键词

关键词 1：存算一体

存算一体作为一种新型算力，是突破 AI 算力瓶颈和大数据的关键技术。与以往的冯诺依曼架构相比，打破了由于计算单元与存储单元过于独立而导致的“存储墙”（CPU 处理数据的速度与存储器读写数据速度之间严重失衡的问题，严重影响目标应用程序的功率和性能），达到用更低功耗实现更高算力的效果。作为可 10 倍提升单位功耗算力的颠覆性技术之一，存算一体有望降低一个数量级的单位算力能耗，在 VR/AR、无人驾驶、天文数据计算、遥感影像数据分析等大规模并行计算场景中，具备高带宽、低功耗的显著优势。目前主流的实现方案包括：一是利用先进封装技术把计算逻辑芯片和存储器（如 DRAM）封装到一起；二是在传统 DRAM、SRAM、NOR Flash、NAND Flash 中实现存内计算；三是利用新型存储元件实现存算一体。当前存算一体技术仍处于早期阶段，我国存算一体芯片创新企业与海外创新企业齐头并进，在该领域的先发制人，为我国相关技术的弯道超车提供了巨大可能性。

关键词 2：一云多芯

一云多芯是指用一套云操作系统来管理不同架构的硬件服务器集群，可以支持多种类型的芯片，解决不同类型芯片共存所带来的多云管理问题，最大限度利用云上资源池的强大算力。作为 IT 产业链

承上启下的关键环节，向下纳管和兼容底层各种芯片、整机、操作系统等软硬件基础设施，向上支撑大数据、人工智能、物联网、5G 等新一代企业级应用，有效规避算力孤岛，逐步实现从算力的并存到算力的统一。一云多芯通过纳管不同指令集的芯片，包括 CPU、GPU、DPU 等，为各类应用场景提供异构多元化的算力支持，满足智算业务高性能计算和数据处理要求，助力算力平台建设标准化、统一化、服务化。中国电信云骁智算平台基于天翼云全栈自研操作系统，打造一云六芯，全面支持主流国产芯片。阿里飞天操作系统正在全面兼容 X86、ARM、RISC-V 等多种芯片架构，实现一云多芯。

关键词 3：CPO

CPO（共封装光学）是光模块未来的一种演进形式，被视为 AI 高算力下高能效方案。CPO 是指把光引擎和交换芯片共同封装在一起的光电共封装，使电信号在引擎和芯片之间更快传输，缩短光引擎和交换芯片间的距离，有效减少尺寸，降低功耗，提高效率。800G 光模块可提高服务器之间互联密度，在同等算力下计算效率倍增，高效支撑 AI 大模型 100%释放算力。随着 AIGC 发展趋势明朗，高算力需求催化更高速率的 800G/1.6T 光模块需求，LightCounting 预测，硅光模块有望在 2025 年高速光模块市场中占据 60%以上份额。多家厂商也开始大力研发用于数据中心的硅光模块，如新华三发布 51.2T 800G CPO 硅光数据中心交换机，单芯片带宽 51.2T，支持 64 个 800G 端口，支撑 3.2 万台节点单个 AIGC 集群，单位时间内 GPU 运算效率提升 25%，

硅光+液冷技术融合实现单集群 TCO 降低 30%，满足大模型智算网络高吞吐、低时延、绿色节能需求[23]。

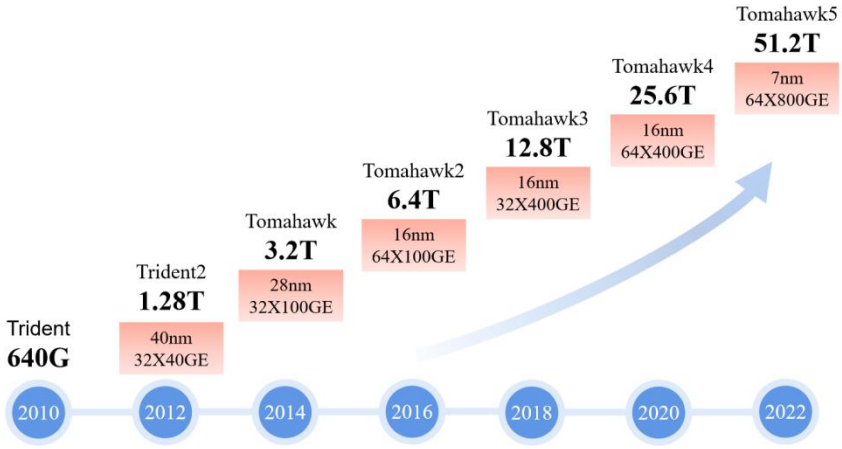

图 8 以太网转发芯片演进

关键词 4：RDMA

RDMA（Remote Direct Memory Access）是一种远程直接数据存取技术，可以有效降低多机多卡间端到端通信时延，满足智算网络的低时延、大带宽需求。当前 RDMA 技术主要采用的方案为 InfiniBand 和 RoCEv2 两种。InfiniBand 网卡在速率方面保持着快速的发展，主流 200Gbps、400Gbps 已规模商用。当前用于大模型训练的智能算力节点内部大多采用 InfiniBand 技术构建数据中心内高性能网络，提供高速连接，以及微秒级的时延、无丢包，避免 GPU 计算等待数据传输导致算力效率的下降。目前 InfiniBand 技术为英伟达独家控制，成本偏高、开放性较弱，因此业界也在考虑用 RoCEv2 等无损网络技术替代 InfiniBand 技术，但存在配置复杂、支持万卡规模网络吞吐性较弱等问题。

对比项	InfiniBand	RoCEv2
同集群端到端时延	2us	5us
网络带宽	1.6T/Server 100-400G/NIC	1.6T/Server 100-400G/NIC
网络无损	完备	待规模工程验证
建设成本	3倍+	低
网络配置	通关UFM实现零配置	手工配置
产业生态	NVIDIA独家	芯片： 博通/Marvell/NVDIA 中兴/华为/盛科 设备：Cisco/Arista/Juniper 中兴/华为/H3C/锐捷

图 9 InfiniBand 和 RoCEv2 的技术对比

关键词 5：DDC

传统 CLOS 网络架构面临多级转发导致时延高、设备低缓存、易丢包等挑战，目前业界主要围绕优化 CLOS 架构、DDC 等开展研究。

（一）云商普遍采用多轨道流量聚合优化面向大模型训练的三层 CLOS 架构，确保在大规模训练时集群的性能和加速比。在多轨道网络架构中，大部分流量都聚合在轨道内传输（只经过一级 ToR switch），小部分流量跨轨道传输（需要经过二级 switch），让任一同号卡在不同机器中的通信中的跳步数尽可能少，大幅减轻了大规模下的网络通信压力。

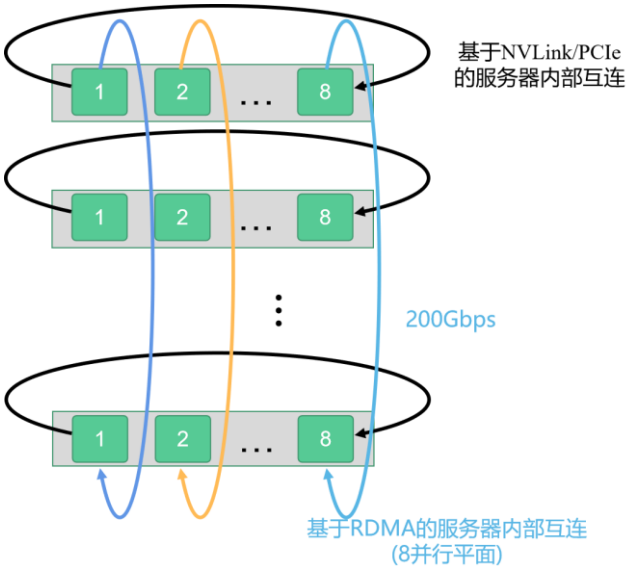

图 10 多轨道流量聚合

（二）AT&T、博通推出 DDC（Disaggregated Distributed Chassis）架构，支持 AI 超大规模集群弹性部署。DDC 将传统软硬一体的框式设备组件进行拆解，使用若干个低功耗盒式设备组成的集群替换框式设备业务线卡和网板等硬件单元，盒式设备间通过线缆互联。整个集群通过集中式或者分布式的 NOS（网络操作系统）管理，以软件化的方式灵活部署于任何一台标准服务器或多台服务器，能有效节省部署成本，提升系统冗余性和可靠性。DDC 架构简单，支持弹性扩展和功能快速迭代、更易部署、单机功耗低，可以根据 AI 集群大小来灵活选择。基于 VOQ+Cell 机制实现端到端流量调度，充分利用缓存大幅减少丢包，且解决了 ECMP 策略下流量负载不均衡的问题，能有效提升宽带利用率。但由于 DDC 硬件要求专用设备、大缓存设计增加网络成本等问题，目前可交付的 DDC 产品较少，有待进一步优化。

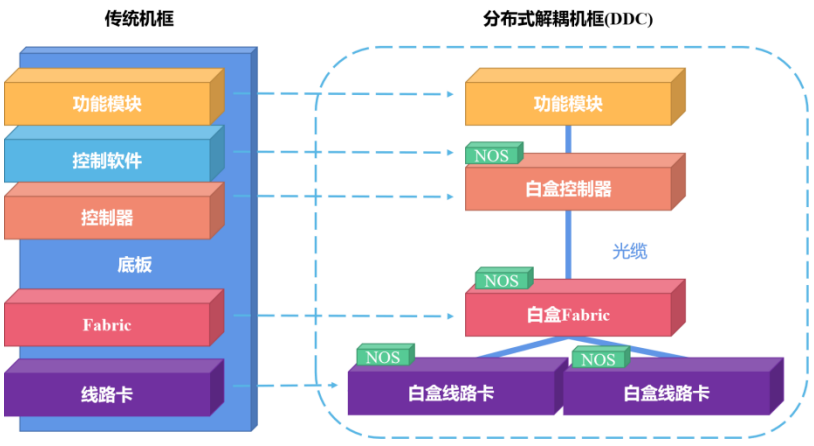

图 11 DDC 架构[24]

关键词 6：并行计算

智算在数据迁移、同步等环节，千卡以上规模的算力输出最低往往仅有 40%左右。随着大模型规模的增长，需要考虑千卡甚至万卡规模的 GPU 集群训练，在多个 GPU 上进行并行计算，将训练任务分解为多个子任务并同时训练，以提升训练速度和效率。针对大规模并行计算的特点，数据并行、模型并行、流水并行、混合专家、增量更新等一系列优化算法和技术有效提升了算法的运行效率和并发性能以及算力的资源利用率，支撑更高更复杂的训练速度和效率。当前业内普遍采用多种并行方式联合优化的策略，如在机内做张量并行，同时配合数据并行进行分组参数切分操作，在多组机器组成流水线并行，以此来承载千亿甚至万亿的模型参数。

关键词 7：液冷

AI 服务器的功率较普通服务器高 6-8 倍，通用型服务器原来只需要 2 颗 800W 服务器电源，而 AI 服务器的需求直接提升为 4 颗 1800W

高功率电源，当前商汤、阿里等高性能 AI 服务器已达到 25kw 以上，而风冷空调的极限在 25-30kw[25]。传统风冷面临散热不足、能耗严重的问题，液冷技术成为了降低数据中心 PUE 的优解，其在 15kW/柜以上时更具经济性优势。浸没式和喷淋式液冷实现了 100% 液体冷却，具有更优的节能效果，PUE 均在 1.2 以下，甚至可低至不足 1.1；浸没式液冷散热节能优势明显，在超算、高性能计算领域取得了广泛应用。在机架功率密度要求和 PUE 限制下，液冷已成为智算中心制冷必选项，预计 2025 浸没式液冷数据中心占比将达 40%[26]。

五、智算发展潜力评估

自大模型等 AI 业务爆火以来，人工智能驱动智算发展进入快速发展阶段。为了全面客观评价我国各省份智算发展水平，本章节设计了智算发展的评估方法和评估结果。该评估主要围绕各省的智算整体发展，以及智算在外部环境、基础设施、服务应用方面的发展展开评估，并依据评估结果进行了相关的分析，为全国及各省份智算发展潜力判断提供参考依据。

1、评估方法

基于全国及各省智算业务相关政策、智算发展特点、行业专家意见，并结合国内外科研机构对智能算力的评估指标研究，借助统计学、指标筛选方法等构建智算发展潜力的评估指标。我们将智算发展潜力评估简称为 ICDP-EM（Intelligent computing development potential evaluation model）。ICDP-EM 如图 1 所示，包括外部环境、基础设施、服务应用 3 个一级指标，以及相应的 8 个二级指标。

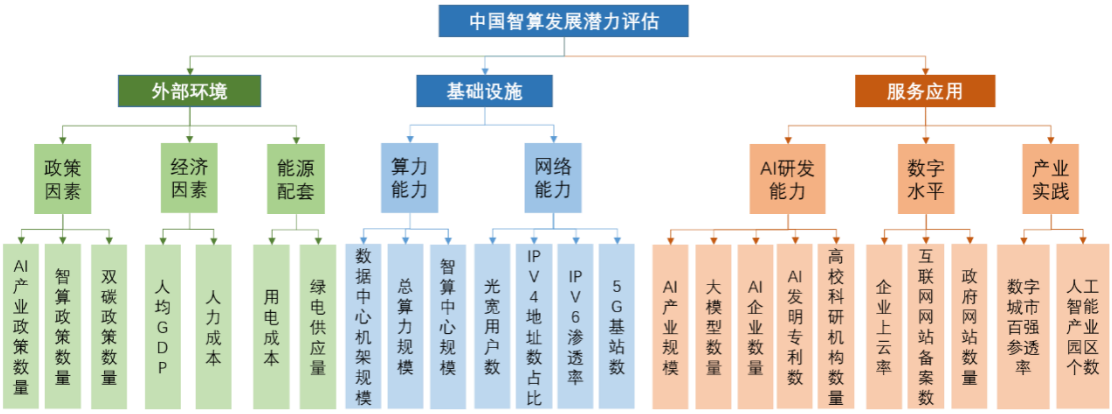

图 12 中国智算产发展潜力评估模型(ICDP-EM)

（一）模型分析

我们从外部环境、基础设施、服务应用三个方面对评估模型进行分析。

1）外部环境

AI 产业、智算中心、双碳等相关智算政策，将影响智算中心选址的具体位置。城市的商业电价、太阳能 / 风能 / 水等绿色发电能力决定了智算中心建设的总体成本，对智算中心的发展区域选择有较大影响。员工薪资、GDP 等是经济发展水平高低的体现，对智算建设能力有一定影响。

2）基础设施

网络高带宽、低延迟是提升智能算力性能的重要因素，如光宽用户数、每万人 5G 基站数、IPV6 渗透率等网络基础能力作为智算中心算力、数据互通的基础，将影响智算对大模型等 AI 业务的训练推理速度、处理能力和结果的准确性。IDC 机架规模、总算力规模影响智算中心的建设和服务能力。

3）服务应用

大模型数量、AI 企业数量、AI 发明专利数等是衡量每个区域 AI 研发能力的关键，企业上云率、互联网网站数等体现了数字化能力，将影响智算服务未来的发展潜力。数字城市百强渗透率、人工智能产业园区数促进产业实践，影响智算服务应用能力。

（二）评估方案[1]

依据 ICDP-EM 模型分析，设计评估体系的评估方案，流程如下：

[1] 详细的评估流程，见附录

1）**指标构建：**通过 ICDP-EM 模型分析，构建中国智算发展潜力评估指标体系包括一、二、三级指标，详情见附录中表 3。

2）**指标赋值：**基于省人民政府、工信部、国家统计局等官网统计三级指标对应的最新数据，为三级指标赋值提供权威、客观的依据。

3）**权重确定：**基于 AHP 和熵权法主客观结合为各指标的权重设计方案，其中一二级指标采用 AHP 方法确定权重，三级指标基于各省统计的指标赋值采用熵权法确定权重。

4）**评估指数结果：**最终根据指标的得分和权重得到各省相应的评估结果，包括综合评估指数、外部环境评估指数、基础设施评估指数、服务应用评估指数。

2、评估结果

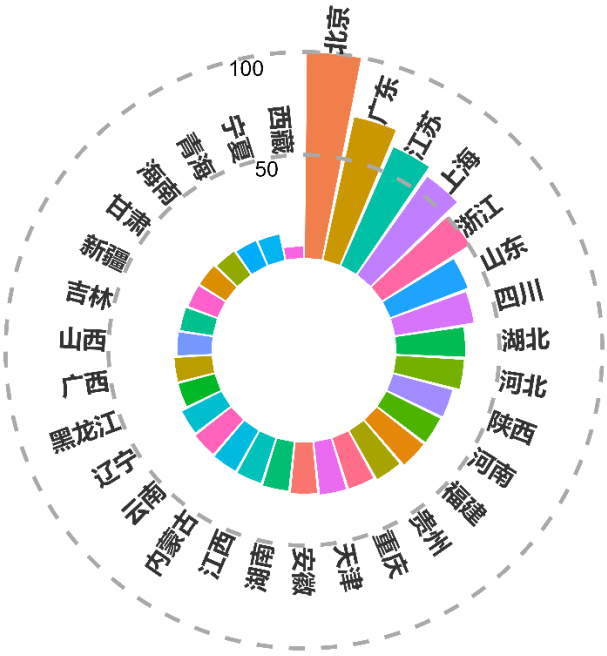

图 13　2023 中国 31 省智算发展综合评估指数

基于评估方法确定的指标、权重和评估指数，本报告从综合评估指数、发展环境评估指数、基础设施评估指数、智算服务评估指数四个方面给出了我国 31[2]省智算发展潜力排序的建议。

（一）智算发展潜力综合评估指数

京津冀、长三角地区智算发展的综合评估指数均在中上游，是具有较高智算发展潜力的城市。

由图 13、14 所示，广东、北京、江苏、上海、浙江属于智算发展的第一梯队，综合指数在 50 以上。山东、四川、湖北、河北、河南、陕西、贵州、重庆、安徽属于智算发展第二梯队，综合指数在 25 以上。如图 14 所示，以北京为代表的京津冀地区和以上海为代表的长三角地区人均 GDP 较高，拉动了智算整体的产业发展，在智算的发展建设上有更大的优势，助力大模型等 AI 业务快速发展。

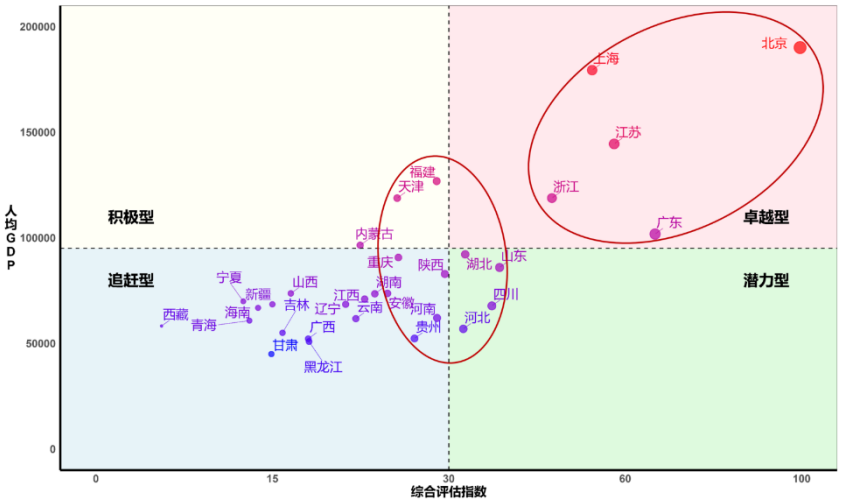

图 14 人均 GDP 与综合指数的象限图

（二）外部环境评估指数

[2] 因数据获取难度等限制，本报告只统计中国 31 省数据，不包括中国香港、中国台湾和中国澳门

中西部地区因绿电、建设成本低等特点，在智算发展的外部环境方面优势凸显。

如图 15 所示，四川、云南、湖北地区因水电等绿色能源供应量充足，内蒙古、新疆等因工业电价低，均跻身第一梯队，适合发展绿色智算相关业务。北京、上海、江苏因 GDP、高薪等因素在智算发展的外部环境方面也具有一定优势。

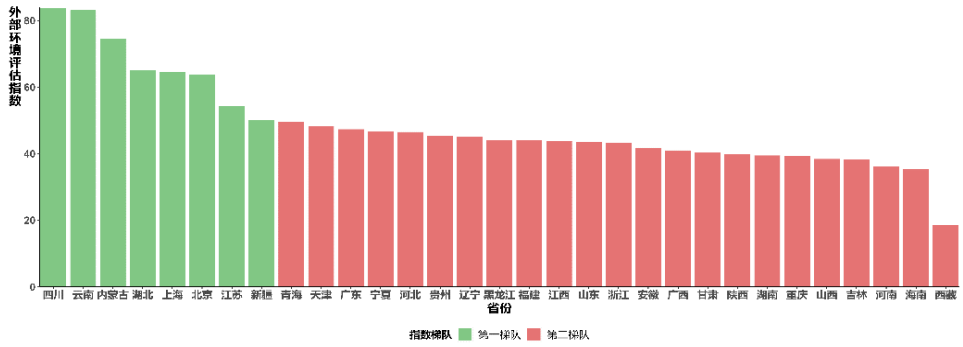

图 15 2023 中国 31 省智算发展外部环境评估指数

（三）基础设施评估指数

全国智算基础设施布局不均，北京、上海、广东为代表的京津冀、长三角等地区在基础设施建设上具有城市集群效应，远高于中西部地区。

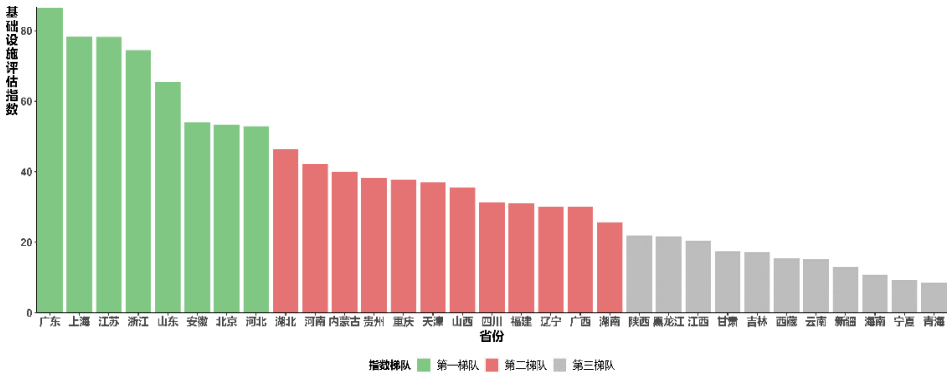

图 16 2023 中国 31 省智算发展基础设施评估指数

如图 16 所示，上海、江苏、浙江、安徽长三角地区均处于第一梯队，京津冀基础设施能力处于中上游水平，山东跻身第一梯队。西

部地区在基础设施建设上还有很大发展空间，尤其宁夏、甘肃作为八大枢纽节之二，在光纤、5G 基站、IDC 机架建设等方面可重点发力。

（四）服务应用评估指数

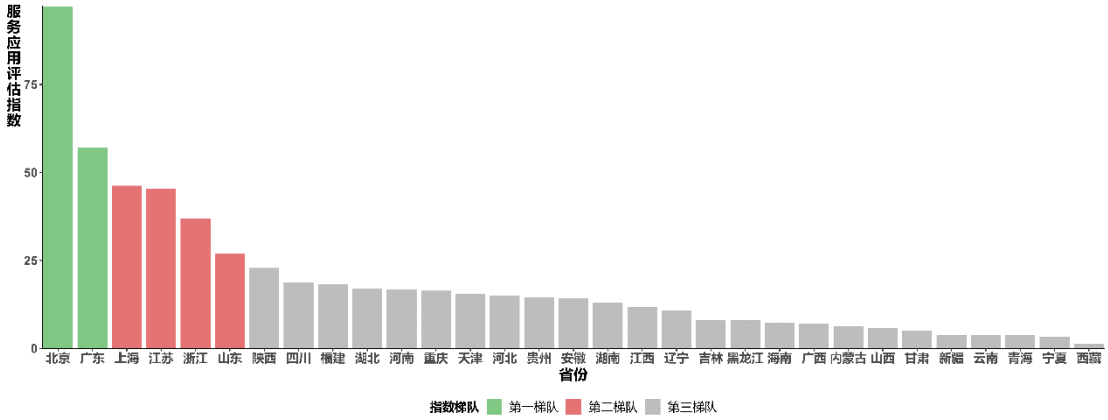

图 17 2023 中国 31 省智算发展服务应用评估指数

智算服务应用能力主要聚集在经济较发达的一、二线城市。

如图 17 所示，北京、广东处于第一梯队，尤其北京在智算服务应用方面远高于其他省份。服务应用能力受基础设施能力的影响较大，服务应用评估指数的第一梯队（北京、广东）和第二梯队（上海、江苏、浙江、山东），其均处于基础设施评估指数的第一梯队。

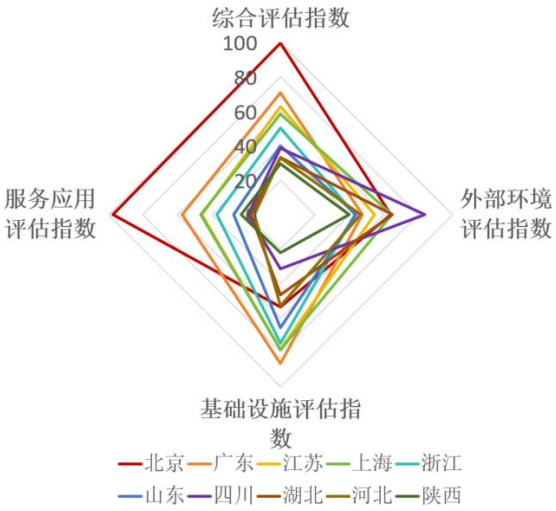

图 18 综合评估指数 Top10 省份细分评估指数对比

基于以上评估指数排序，对综合评估指数 top10 的城市进行外部

环境、基础设施、服务应用的能力分析。如图 18 所示，北京在综合能力和服务应用能力方面遥遥领先，广东、上海、江苏、浙江在基础设施能力方面占有优势，四川因出色的绿电供应（水电）使其在外部环境能力方面名列前茅。山东、湖北、河北、陕西等在各方面处于中等水平，整体能力较稳定。

六、典型案例

1、中国电信安徽智算中心

中国电信安徽智算中心位于合肥市高新区南岗科技园，园区规划占地面积 150 亩，累计投资将达 100 亿元，一期于 2021 年 12 月投产使用。该项目预计 2024 年全量完工，将成省内规模最大、标准最高、网络最快、算力最智能、绿色低碳最节能的超大型智算中心。

基础设施方面，按照国家数据中心最高 A 级标准建设，包括 6 栋数据中心，2 栋动力中心，1 栋 110KV 变电站和 2 栋产业孵化中心，建成后将具备 16000 架中高密度机柜，可容纳约 30 万台服务器，支持算力规模可达到 2.2 EFLOPS，使安徽省的整体算力规模翻番[27]。

网络方面，园区的网络出口带宽达 20 T，通过四平面全光传送网直连国家级互联网骨干节点，网络层级高度扁平化，可实现业务流量的高效疏导。同时与新建的合肥国家级互联网骨干直连点高速互通，省内互访时延将降低 90%。

算力应用方面，目前安徽智算中心已落地合肥人工智能计算中心项目，搭载 224 颗鲲鹏 920+448 颗昇腾 910 芯片，初期具备 100P 智算能力，同时还为安徽通用人工算力集群提供 3000P 算力底座能力，助力安徽算力产业高速发展。

绿色低碳方面，园区采用集中水冷式中央空调系统，充分利用自然冷源，引入液冷、光伏、多联热管空调等先进技术，辅助 AI 节能，实现智能化精确制冷，有效降低能耗，使得数据中心 PUE 降到 1.25

以下，打造长三角区域领先的绿色低碳数据中心，是安徽省唯一入选工信部 2022 年国家新型数据中心典型案例的超大型数据中心。

2、中国电信（国家）数字青海绿色大数据中心

中国电信（国家）数字青海绿色大数据中心是全国首个 100%清洁能源可溯源绿色大数据中心，以绿色、零碳、可溯源为其典型特征。2023 年 4 月，该数据中心通过权威机构“碳中和”认证，成为全国首个通过自身储备碳汇实现“碳中和”的数据中心，也是国内首个真正实现零碳排放的数据中心，年减碳量近 30 万吨。

节能技术方面，利用青海的自然条件优势，采用冷冻水+间接蒸发冷却技术，机房可以全年 314 天不开启空调压缩机，大大减少机房能耗，并且在冬天可以将机房内热量通过余热回收，满足办公室及走廊供暖需求，实现 PUE 值保持在 1.2 以下。同时配备源网荷储一体化绿电智慧供应系统，办公和基础设施用电由园区光伏发电系统供应，多余电量在园区存储备用，储电能力探顶后可向城市电网输送，数据中心从用电方变为发电者。

算力应用方面，依托青海“3+8+X”绿色算力资源布局，与云网大数据中心、青藏高原灾备中心协同为青海乃至全国提供高效算力调度和应用。目前，已为青海近 60%的各级政府部门政务云平台提供算力和存储，如青海省最大的线上教育互动平台“三个课堂”融合平台，海西文化旅游大数据平台等[28]。在民生服务、城市管理、生产制造等方面助力青海经济社会数字化转型，吸引头部互联网企业等全国客户

入驻。

3、海兰信海底数据中心

GPT 等生成式人工智能浪潮引发新一轮 AI 革命，各行业大模型训练、生成式 AI 对算力产生爆炸式需求。得算力者得天下，沿海发达城市的算力容量是其未来发展空间的关键因素。而当前受能耗指标限制，这些城市的智算中心发展受到制约。充分利用自然冷源、与可再生能源相结合是数据中心绿色低碳发展的共识和趋势。因地制宜、依海而兴，向海洋要冷源和新能源，是沿海发达城市数据中心向“零碳”发展的创新思路。

图 19 海底数据中心概念图

海底数据中心应运而生，将海洋工程、数据中心、海上新能源等多领域融合，主要由岸站基地、海底光电复合缆、分电站及数据舱四个部分组成。通过立体科技用海，实现降本增效，多产业协同；通过就地消纳海上绿电，解决数据中心能耗和高算力需求矛盾，突破沿海发达城市发展数字经济的资源限制。据海兰信测算，海底数据中心建

设成本比传统低 23%、日常运维成本比传统低 14%，绿电使用率理论可达 100%。以上海市为例，海底数据中心若能取代上海当前 11 万架陆地数据中心，将能节约地方 57.8 亿千瓦时（约 71 万吨标准煤）能耗指标。

2022 年 12 月，全球首例商用海底数据中心在海南陵水成功启用。该项目由海兰信与中国电信海南分公司合作开发。其暖通系统利用海水实现全年自然冷却，匹配重力热管技术、海水泵变频技术、空调群控技术等节能措施，舱内运营平均温度约 25 摄氏度，运行 PUE 低至 1.1，较传统数据中心节能 30%以上。由于无需蒸发散热，减少了冷却塔和冷水系统，水资源消耗为 0。此外，由于大部分设施位于海底，土地占地极少，仅有传统数据中心的十分之一。

图 20 全球首例商用海底数据中心入海瞬间

海底数据中心显著提高服务器的安全性与可靠性。相较于陆地和海面，海底环境十分平稳，系统设计能够应对百年一遇的风暴。海底数据舱内充满惰性气体，给 IT 设备提供了一个无氧、无尘、恒湿、恒压的密闭环境。这一方面使系统具备防火灾、防水涝、防极端天气

的容灾价值，另一方面对服务器及相关设备十分友好，提高了数据中心的可靠性。海南示范项目运行以来，没有一台服务器出现故障，大大降低运维工作量。此外，该项目搭建数字孪生系统，实现全链路微结点智控技术，满足海底数据中心的远程维护、少维护、甚至免维护的需求，显著降低运维成本及碳排放。

中国电信海南分公司在海底数据中心部署的业务包括媒体存储节点、CDN 节点以及海南省国资专属云资源池等。自项目首舱下水投产运行以来，各业务运行稳定，系统性能表现良好。目前海底舱内的核心路由器到省域网核心路由器平均延迟不超过 4m，网络效能达到互联网数据中心最高级别，可以承载对时延性、互通量要求最高的业务。

下一步，海底数据中心将与海上风电融合，打造算电协同新模式。该方案具有多种优势：第一，海底数据中心与海上风电可以共用海域场址、海底光电复合缆、海洋工程船，实现降本增效，相比传统陆上数据中心，建造和运维成本优势明显；第二，海上风电供绿电，大大降低海底数据中心用电成本，相较陆上市电，用电成本更低，也更加绿色低碳；第三，海底数据中心作为海上风电的有效载荷，可以原位消纳海上绿电的产业，提高海上风电场的发电效率和经济效益，助力海上风电场产业链做大做强。目前我国正在大力发展数字经济与海洋经济，海上风电等新能源产业蓬勃发展，产业融合创新进入机遇窗口期。海底数据中心与海洋绿色能源结合，将海洋电力转化为陆地算力，有助于实现绿电直供、立体用海、共建共维、产业协同的新发展格局，

打造海上新能源与数字经济融合发展新赛道。

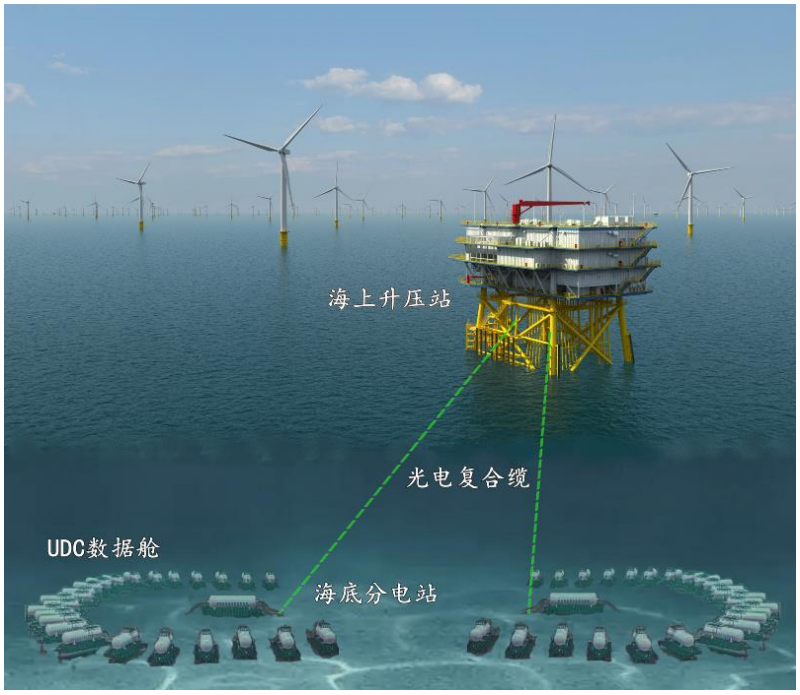

图 21 海底数据中心与海上风电融合方案

七、总结与展望

AIGC、自动驾驶、智能制造、智慧医疗、智慧城市等领域发展迅速，随之而来的超大规模 AI 模型和海量数据对算力基础设施提出更高要求，结合《“十四五”国家信息化规划》的“适度超前部署下一代智能设施体系”、《“十四五”数字经济发展规划》的“推动智能计算中心有序发展”、“东数西算”的“布局全国算力网络国家枢纽阶段”等政策背景，智算中心高质量发展正当时。与云计算中心、超算中心不同，智算中心主要为 AI 各个领域提供算力、数据、算法等服务，既能满足计算机视觉、自然语言处理等应用需求，又能用于理论研究支撑，满足新技术创新探索需求。

八、附录-智算评估实施方案

本白皮书对中国智算产业潜力发展评估的具体实施方案如下。

1、评估指标模型构建

结合模型假设的影响因素，我们编制了一级指标、二级指标以及对应的指标说明、评估单位，便于后续指标评估。

附表 1. 中国智算发展潜力评估指标体系

一级指标	二级指标	三级指标	单位
外部环境	政策因素	AI产业政策数量	个
		智算中心政策数量	个
		双碳相关政策数量	个
	经济因素	人均GDP	元
		人力成本（月薪水平）	元
	能源配套	用电成本（工业）	元/千瓦时
		太阳能、风能等绿电供应量	亿千瓦时
基础设施	网络能力	光宽用户数	万户
		5G基站数	个/万人
		IPV4地址数占比	%
		IPV6渗透率	%
	算力能力	IDC机架规模	万架
		总算力规模	EFLOPs
		智算中心规模	EFLOPs
服务应用	AI研发能力	AI产业规模	亿元
		大模型数量	个
		AI企业数量	个
		AI发明专利数	个
		高校、科研机构数量	个
	数字化水平	企业上云率	%
		互联网网站备案数	个
		政府网站数量	个
	产业实践	数字城市百强渗透率	%
		人工智能产业园区个数	个

2、评估指标赋值

基于省人民政府、工信部、国家统计局等官网统计智算相关三级评估指标的最新数据，为 31 省的三级指标赋值提供权威、客观的依据。为 31 省的 24 个指标赋值，并对所有指标数值 x 进行归一化处理，得到每个指标的标准化数值 x'。

3、评估指标权重设计

关于评估指标权重的确定采用主客观结合的方式进行，保证评估结果的专业性和客观性。对于一、二级指标，涉及指标全面性的确定，需专家参与判定，采用 AHP 的评判矩阵来确定指标的权重。对于三级指标，在已经确定指标全面性的前提下，采用熵权法确定指标权重，确保结果的客观性。

（一）一、二级指标权重确定

基于 AHP 方法对一、二级指标进行权重设计，借助评判矩阵得出一、二级指标的权重，权重确定流程如下：

1）根据指标分类制定评断矩阵模板。

附表 2. 智算发展潜力评估指标评判矩阵模板

	指标 1	指标 2	…	指标 n
指标 1	a_{11}	a_{12}	…	a_{1n}
指标 2	a_{21}	a_{22}	…	a_{2n}
…	…	…	…	…
指标 n	a_{n1}	a_{n2}	…	a_{nn}

备注：n 是一或二级指标的个数。矩阵中的值为对应纵向指标比横向指标重

要程度，例如，$a_{ij}=\frac{l}{m}$ 是第 i 个指标与第 j 个指标比较对智算发展重要程度比值，其中$l,m \in (0,9)$。0 到 9 表示两个指标比较对智算发展的重要程度，数值越大重要程度越大。

2）业内智算专家按步骤 1 规则对需要评估的 n 个指标进行打分，分别给出相应的$n \times n$阶评判矩阵，我们将这些评判矩阵记为$A_1, A_2, A_3, \ldots, A_m$。

3）通过公式$CR=\frac{\lambda-n}{(n-1)*RI}$，对评判矩阵进行一致性验证。

4）若评判矩阵通过一致验证，计算最大特征值 λ，对应的特征向量，即为指标对应的权重。

（二）三级指标权重确定

基于三级指标对应的 31 省数据，采用熵权法确定三级指标的权重，主要思路是**根据指标变异性的大小来确定客观权重**。流程如下：

1）根据三级指标的 31 省数据，构造矩阵B,模板如下。

	指标 1	指标 2	…	指标 24
省份 1	b_{11}	b_{12}	…	$b_{1,24}$
省份 2	b_{21}	b_{22}	…	$b_{2,24}$
…	…	…	…	…
省份 31	$b_{24,1}$	$b_{24,2}$	…	$b_{24,24}$

2)对矩阵B数据进行标准化处理，对于正向指标采用

$$b'_{ij}=\frac{b_{ij}-\min(b_{ij})_{j=1,2\ldots,24}}{\max(b_{ij})_{j=1,2\ldots,24}-\min(b_{ij})_{j=1,2\ldots,24}}$$

对于负向指标采用

$$b_{ij}{}'=\frac{\max(b_{ij})_{j=1,2\ldots,24}-b_{ij}}{\max(b_{ij})_{j=1,2\ldots,24}-\min(b_{ij})_{j=1,2\ldots,24}}$$

3）算每个指标 j 的熵值

根据矩阵B标准化后的数值计算信息熵：

$$H_j = -\frac{1}{\ln 31}\sum_{i=1}^{31}\frac{b_{ij}}{\sum_{i=1}^{31} b_{ij}}.\ln\sum_{i=1}^{31}\frac{b_{ij}}{\sum_{i=1}^{31} b_{ij}}$$

备注：**信息熵**是对一个信源所含信息的度量，即信息量的期望。

4）计算指标 j 对应的权重值

$$w_j = \frac{1 - H_j}{24 - \sum_{i=1}^{31} H_j}$$

4、各省评估得分

根据以上方法确定的一、二、三级指标权重和 31 省的 24 个指标的标准化分值，为各省进行综合评分，并分别根据对应的二、三级指标为各省的一级指标外部环境、基础设施、服务应用三个类别进行评分。

九、参考文献

[1] 中国信通院. 中国算力白皮书（2022 年）[R]. 2022

[2] 国家信息中心信息化和产业发展部，浪潮.智能计算中心规划建设指南[R].2020

[3] 毕马威, 联想集团. “普慧”算力开启新计算时代[R]. 2023.

[4] https://baijiahao.baidu.com/s?id=1759777944172036313&wfr=spider&for=pc

[5] https://baijiahao.baidu.com/s?id=1772916034401019850&wfr=spider&for=pc

[6] https://baijiahao.baidu.com/s?id=1762579606694238632&wfr=spider&for=pc

[7] https://baijiahao.baidu.com/s?id=1771713668789917788&wfr=spider&for=pc

[8] https://it.sohu.com/a/683393111_121124373

[9] https://baijiahao.baidu.com/s?id=1766948778589069131&wfr=spider&for=pc

[10] https://baijiahao.baidu.com/s?id=1769579795147577688&wfr=spider&for=pc

[11] https://baijiahao.baidu.com/s?id=1765501244888740606&wfr=spider&for=pc

[12]
https://baijiahao.baidu.com/s?id=1765851289022520868&wfr=spider&for=pc
[13]彭卉，申红梅. 全球主流行业大模型发展跟踪[EB/OL]. 天翼智库，2023 年 7 月.
（https://mp.weixin.qq.com/s/Re6X0L_ugn9fGOkoLd2YZQ）
[14]
https://baijiahao.baidu.com/s?id=1759036970210481596&wfr=spider&for=pc
[15]
https://baijiahao.baidu.com/s?id=1740115176154968385&wfr=spider&for=pc
[16] https://www.sohu.com/a/629046397_121207965
[17] 中邮证券. 国产 AI 芯片的创业裂变[R]. 2023.
[18] 中国科学技术信息研究所, 科技部新一代人工智能发展研究中心. 中国人工智能大模型地图研究报告[R]. 2023.
[19] 麦肯锡. 生成式人工智能的经济潜力[R]. 2023.
[20] 国家工业信息安全发展研究中心信息政策所. 智能计算中心 2.0 时代展望报告[R]. 2023.
[21]
https://baijiahao.baidu.com/s?id=1766332110002351924&wfr=spider&for=pc
[22] A. Zhang, Y. Liu, L. Feng, et al. “Record 46.2Pbit·km/s real-time optical transmission over 1050-km G.652.D SSMF utilizing 400Gbit/s transponder with a symbol rate of 91.6-Gbaud”, *Optoelectronics and Communications Conference,* 2023.
[23]
https://baijiahao.baidu.com/s?id=1768593075913091747&wfr=spider&for=pc

r=pc
[24] omdia. Network Simplification in the Digital Era Through Distributed Disaggregation[R]. 2023.
[25]
https://baijiahao.baidu.com/s?id=1763394886903697355&wfr=spider&for=pc
[26]
https://baijiahao.baidu.com/s?id=1763492289705886258&wfr=spider&for=pc
[27]
http://www.chinatelecom.com.cn/news/03/202304/t20230428_74053.html.
[28] https://h5.drcnet.com.cn/docview.aspx?docid=7008871