浪潮信息 CAICT 中国信通院

人工智能算力高质量发展评估体系报告

浪潮电子信息产业股份有限公司
中国信息通信研究院

2024年9月

版权声明

目 录

1 发展现状及挑战

1.1 发展现状

1.1.1 政策上：政策导向日益明确

全球各国通过政策支持、战略规划等手段，加速构建领先的算力竞争力。美国公布 2024 财年政府预算，包括国防部、能源部、国土安全部等多个机构，累计向 AI 领域计划投入超过 2511 亿美元，以推动 AI 研究和软硬件服务；欧洲陆续发布《塑造欧洲的数字未来》、《欧洲芯片法案》等文件，围绕数字化转型进行算力产业布局；日本近年来频繁强调振兴半导体产业，坚持以应用、绿色为导向发展算力，不断扩大国内尖端半导体生产。这些政策的实施加速了全球产业升级和科技创新，并提升了这些国家的算力竞争地位。

我国以算力基础设施建设为锚点，全面推动算力高质量发展。二十届三中全会提出，高质量发展是全面建设社会主义现代化国家的首要任务。我国通过加强算力基础设施建设，推动算力技术与产业的创新发展，为经济社会的高质量发展注入新动能。在国家层面，《数字中国建设整体布局规划》、《深入实施“东数西算”工程 加快构建全国一体化算力网的实施意见》、《算力基础设施高质量发展行动计划》、《数据中心绿色低碳发展专项行动计划》等提出我国算力高质量发展的具体要求；在地方层面，浙江、北京、上海、广东、贵州、山西等省市也纷纷发布相关政策明确未来几年算力高质量发展行动计划。

表 1 我国算力中心相关政策规划（部分）

发布时间	发布部委/省份	政策名称
2024 年 7 月	国家发改委、工信部、国家能源局、国家数据局	《数据中心绿色低碳发展专项行动计划》
2023 年 12 月	国家发改委、国家数据局、中央网信办、工信部、国家能源局	《深入实施“东数西算”工程 加快构建全国一体化算力网的实施意见》
2023 年 10 月	工信部、中央网信办、教育部、国家卫健委、中国人民银行、国务院国资委	《算力基础设施高质量发展行动计划》
2023 年 2 月	中共中央、国务院	《数字中国建设整体布局规划》
2024 年 5 月	浙江	《浙江省运力提升行动方案（2024—2027 年）》
2024 年 4 月	北京	《北京市算力基础设施建设实施方案（2024—2027 年）》
2024 年 4 月	江苏	《江苏省算力基础设施发展专项规划》
2024 年 3 月	上海	《上海市智能算力基础设施高质量发展“算力浦江”智算行动实施方案（2024-2025 年）》
2024 年 3 月	广东	《广东省算力基础设施高质量发展行动暨“粤算”行动计划（2024-2025 年）》
2024 年 2 月	贵州	《贵州省算力基础设施高质量发展行动计划（2024-2025 年)》
2024 年 1 月	山西	《山西省算力基础设施高质量发展实施方案》
2023 年 12 月	深圳	《深圳市算力基础设施高质量发展行动计划（2024-2025）》
2023 年 12 月	重庆	《重庆市算力网络发展“算力山城 强算赋能”行动计划(2023-2025 年)》
2023 年 12 月	安徽	《安徽省智能算力基础设施建设方案（2023-2025 年)》
2023 年 8 月	湖北	《湖北省加快发展算力与大数据产业三年行动方案》
2023 年 7 月	河南	《河南省重大新型基础设施建设提速行动方案（2023-2025 年)》
2023 年 4 月	天津	《关于做好算力网络建设发展工作的指导意见》
2023 年 3 月	宁夏	《全国一体化算力网络国家枢纽节点宁夏枢纽建设 2023 年工作要点》

（来源：公开资料）

1.1.2 技术上：生成式 AI 突破发展

人工智能以生成式 AI 技术为核心快速发展。以 ChatGPT 为代表的 AIGC 技术加速成为 AI 领域的最新发展方向，对经济社会发展产生了重大的影响。随着人工智能预训练大模型的不断进步、AIGC 算法的持续创新，以及多模态 AI 技术的日益普及，AI 已经能够生成包括文本、代码、图像、语音和视频在内的多样化内容。这些技术的发展提升了 AIGC 模型的通用性和工业化水平，AIGC 的商业潜力变得更加显著，如今大模型已成为企业在 AI 领域竞争的核心焦点。

算力成为推动生成式 AI 发展的关键。在大模型训练和生成式 AI 应用的推动下，GPU 和异构计算资源需求显著增长，算力的提升从简单的硬件扩展发展为涵盖算法优化、系统设计、资源调度和网络通信等多个层面的系统优化，算力性能和效率对模型推理、训练至关重要。在大模型训练中，通常采用多机多卡构建的算力集群进行分布式训练，而拥有大量的计算节点并不等同于拥有强大的计算能力。在分布式训练环境中，拥有数千亿至万亿参数的庞大模型通信时间可能占据整个训练过程的一半，网络通信和数据缓存等瓶颈问题会显著降低训练效率。另外，随着模型参数量增加，传统的训练方式可能会导致训练过程中算力利用率的降低。在大模型训练中，Checkpoint 机制常用于在训练中定期保存模型参数，然而对于参数量极大的模型，该训练方式可能会导致显著的写入延迟，如 GPT-3（1750 亿参数），以 15GB/s 的文件系统写入速度计算，完成一次 Checkpoint 需要 2.5 分钟，这不仅增加了训练时间，也降低了 GPU 的利用率。

1.1.3 市场上：算力投资持续加码

国家以直接投资或补贴方式推动算力产业投资建设。美国计划 5 年内投资 2800 亿美元以保持美国在芯片技术领域的领先地位；中国全面启动“东数西算”工程，截至 2024 年 6 月底，“东数西算”八大国家枢纽节点直接投资超过 435 亿元，拉动投资超过 2000 亿元；欧盟计划提供 12 亿欧元的公共资金用于“欧洲共同利益重要计划——下一代云基础设施和服务”；日本经济产业省拟为 5 家日本企业提供总额 725 亿日元的补贴，用于打造人工智能超级计算机。随着全球各国在算力领域的竞争愈发激烈，算力相关产业市场规模将呈现持续增长态势。以 AI 服务器为例，据 IDC 预测，未来几年全球人工智能服务器市场规模将持续增加。

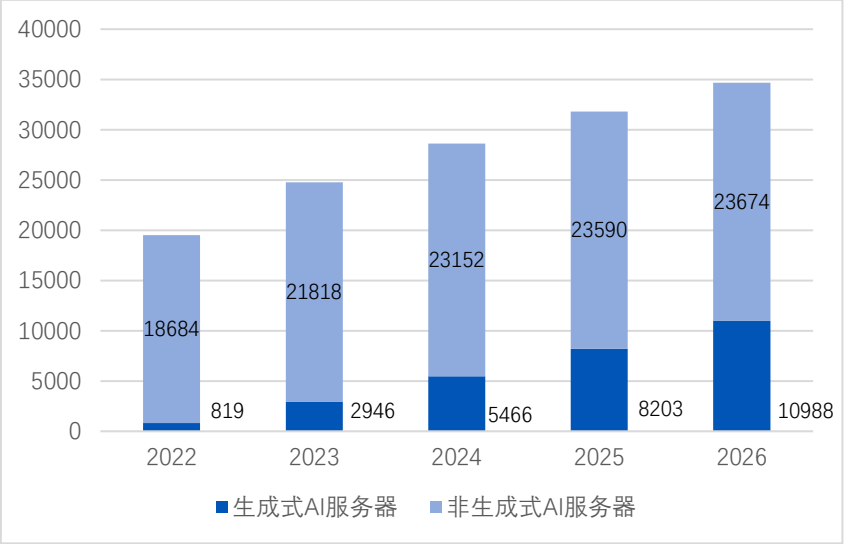

图 1 全球人工智能服务器市场规模预测（单位：百万美元）

（来源：IDC，2023）

科技巨头发力智能算力，万卡算力集群布局加快。2023 年以来人工智能市场持续保持高增长态势，成为推动各国经济增长和技术创新的关键因素。据 IDC 研究，预计 2022 年至 2032 年全球人工智能

产业规模的复合增长率高达 42%，2032 年将达到 1.3 万亿美元。基于人工智能的广阔前景，全球科技巨头纷纷加大对 AI 基础设施布局以维持行业竞争力。国际上 Meta、微软&OpenAI、xAI 等多家 AI 巨头陆续宣布或者完成 10 万卡集群建设，国内通信运营商、头部互联网、大型 AI 研发企业等均发力超万卡集群的布局。

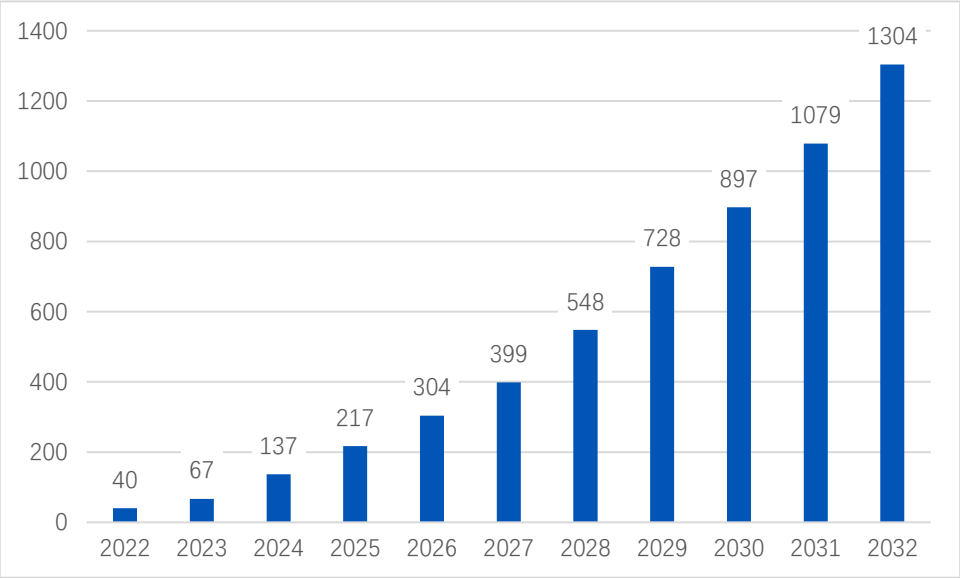

图 2 全球 AI 产业规模预测（单位：十亿美元）

（来源：IDC、Bloomberg、Mandeep Singh）

表 2 全球科技巨头智算布局（部分）

科技巨头	万卡智算集群布局进展
谷歌	2023 年 5 月，推出 AI 超级计算机 A3，搭载了约 26000 块 H100 GPU，为其在机器学习和深度学习研究中的应用提供强大的算力支持
Meta	2024 年初，Meta 建成了两个各含 24576 块 GPU 的集群，并设定目标：到 2024 年底，构建一个包含 35 万块 H100 GPU 的庞大基础设施，以支撑其元宇宙和 AI 研究
微软	早在 2020 年，微软便构建了一个覆盖 1 万块 GPU 的超级计算机，加速其在云计算和 AI 服务领域的发展
亚马逊	Amazon EC2 Ultra 集群采用了 2 万个 H100 TensorCore GPU，为用户在处理大规模数据分析和机器学习任务方面提供强大算力支持
特斯拉	2023 年 8 月，特斯拉上线集成 1 万块 H100 GPU 的集群，将极大提升特斯拉在自动驾驶和车辆智能化方面的研发速度
腾讯	推出的星脉高性能网络能够支持高达 10 万卡 GPU 的超大规模计算，网络带宽高达 3.2T，为未来的 AI 和大数据应用提供了广阔的发展空间
字节跳动	提出的 MegaScale 生产系统，支撑 12288 卡 Ampere 架构训练集群，为字节跳动在内容推荐、图像处理等 AI 应用方面提供了强大的算力保障
中国移动	计划今年商用哈尔滨、呼和浩特、贵阳三个万卡集群，总规模接近 6 万张 GPU 卡
中国电信	计划 2024 年在上海规划建设一个达到 15000 卡、总算力超过 4500P 的万卡算力池。2024 年 3 月，天翼云上海临港万卡算力池已正式启用
中国联通	计划今年内在上海临港国际云数据中心建成中国联通首个万卡集群，集群建成后将为中国联通在数据中心和云计算市场提供新的竞争优势

（来源：公开资料）

1.1.4 规模上：智算规模增速明显

全球算力规模稳步扩张，智算同比翻倍增长。以 AIGC 为代表的人工智能应用、大模型训练等新需求、新业务的崛起，推动全球智算规模呈现高速增长态势。据中国信通院测算，截至 2023 年底，全球

算力总规模约为 910EFLOPS[1]，同比增长 40%，智能算力规模达到 335EFLOPS，同比增长达 136%，增速远超算力整体规模增速。**我国智能算力占比显著增加，智算中心集聚分布。**据中国信通院测算，截至 2023 年底，智能算力规模占整体算力规模的比例近 30%，增效明显。国家及地方层面积极推进智算中心建设，北京、广东等多地提出 2025 年智算规模目标。从区域分布上来看，智算中心呈集群建设趋势，过半分布在我国东部地区。

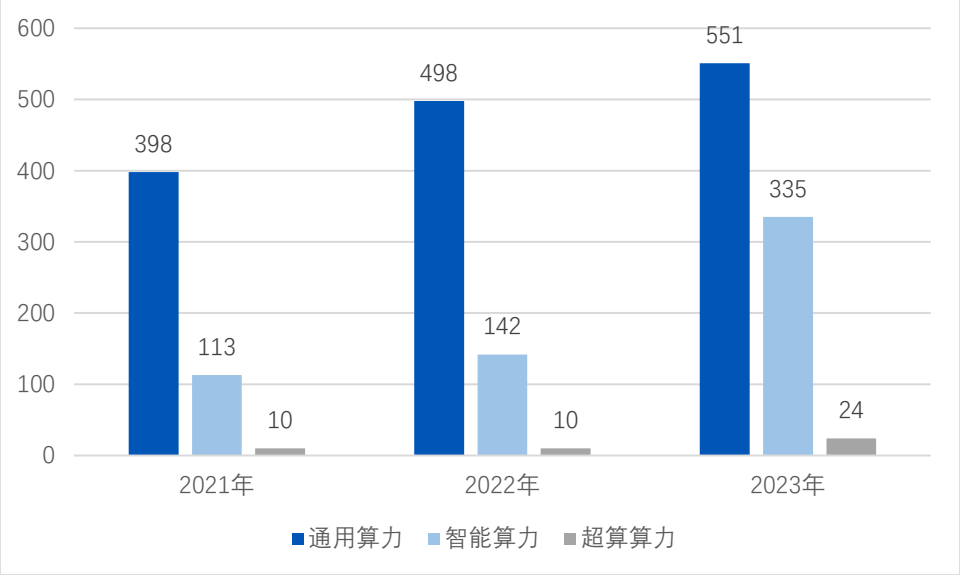

图 3 全球算力规模（单位：EFLOPS）

（来源：Gartner、IDC、中国信通院）

1.1.5 发展水平上：算力发展由“量”向“质”

我国算力发展正处在由“量的扩张”转向“质的提高”这一重要关口。我国算力产业规模扩张下开始以应用为导向，推动过去的重资产、重硬件模式向软硬协同、服务驱动转型等高质量发展方向转型升级，算力发展从规模速度型粗放增长转向质量效率型集约增长。在应用导向下，全国各地增加智能算力生产以提升算力在人工智能领域的适配

[1] 算力规模包含通用算力、智能算力、超算算力，边缘算力暂未纳入统计范围，表示方式皆为单精度（FP32）。

水平，建设超大规模算力中心集群，以匹配大模型训练需求。随着集群建设规模越来越大，算力效率问题引起关注。算力中心建设、运营开始重视超大规模组网互联、集群有效计算效率、训练高稳定性与可用性等算力处理效率相关的性能。另外，双碳目标日益紧迫，能耗要求日趋严格，算力行业的高耗能和碳排放问题引起诸多关注。我国出台了一系列政策对算力产业节能降碳提出更加严格的要求，相关企业通过技术创新、绿色管理等措施不断开展节能降碳行动，绿色低碳成为算力产业重要发展方向。

与此同时，算力中心作为算力的主要载体，承载功能逐渐多元化。算力中心可为服务购买方提供多元化业务支撑和多样化算力服务，不再只是提供计算、存储等服务的场所，数据、算力、算法、生态合作等服务内容不断拓宽算力中心能力边界。算力提供方越来越注重提升算力服务的品质，整合各类算力资源，为用户提供一站式服务，降低用户获取算力资源的成本，推动算力普适普惠。

1.2 面临挑战

1.2.1 挑战一：算力供给不足，供需匹配不平衡

一方面，AIGC 带动算力需求总量不断增长，大模型训练亟需大量 AI 算力支撑。从整体需求规模上看，模型训练引起 AI 算力需求暴增。据《新一代人工智能基础设施白皮书》表明，过去几年，大模型参数量以年均 400%复合增长，AI 算力需求增长也超过 15 万倍，远超摩尔定律。在单个大模型训练需求上，模型越大算力需求越大，以参数规模达到 4050 亿的 Llama3.1 大模型为例，其单次训练算力需

求相较于 700 亿的 Llama2 翻了 50 倍。另外伴随模型不断迭代，训练数据集规模将不断增大，未来的大模型的算力需求将呈现指数级爆发式增长。

另一方面，现有供给结构与用户实际的算力需求不匹配造成资源浪费。一是供需错位问题，国内算力产业链企业相对分散，众多芯片厂商和大模型企业技术路径不同容易造成芯片和模型之间不适配，且大多数智算服务仍是‘裸金属租赁’的粗放式经营方式，无法精准满足不同企业的多元化需求。二是资源利用率不足问题，据清华大学研究表明，大模型在处理大量数据时，由于算力调度、系统架构、算法优化等诸多问题，很多大模型企业的 GPU 算力利用率低于 50%，造成了巨大的资源浪费。

1.2.2 挑战二：算力智能水平较低，难以满足多元应用场景

人工智能、大数据、物联网等数字技术不断发展，多元应用场景对算力的智能水平和计算能力要求不断提升。从需求上看，算力应用场景的复杂化导致数据量和算法复杂度急剧增加，这要求算力具备更高的智能化水平。算力是算法自主学习的基础，能够灵活处理和分析大规模的数据集，有助于满足更大参数量模型的训练需求，不断提升模型的自主学习和泛化能力。从技术上看，传统芯片架构面临着“存储墙”和“功耗墙”的问题，难以满足现阶段人工智能应用的低时延、高能效、高可扩展性的需求，需要先进的计算架构将更多算力单元高密度、高效率、低功耗地连接在一起，提高异构多核之间的传输速率，从而为人工智能大模型提供强大计算能力保障。

1.2.3 挑战三：算力面临能源考验，节能降碳刻不容缓

双碳目标下，算力产业面临节能降碳挑战。算力中心是算力的主要载体，是公认的高耗能基础设施。据中国信通院数据表明，截至2023年底，我国算力中心耗电量达1500亿千瓦时，预计到2030年将超过4000亿千瓦时，若不加大可再生能源利用比例，2030年全国算力中心二氧化碳排放或将超2亿吨。人工智能模型训练的能耗远高于常规计算能耗，根据Digital Information World数据，训练AI模型产生的能耗是常规云工作的三倍。OpenAI曾发布报告称，自2012年以来，AI训练的电力需求每3-4个月就会翻一倍。据浪潮信息测算，一个10万亿参数大模型训练需要10万卡H100集群，训练1193天，所消耗的电量约40亿千瓦时，约1.4亿美国家庭1天用电量。面对人工智能对算力的旺盛需求，算力产业如何在高速发展的同时实现“碳中和”，是当下整个行业需要解决的重要问题。

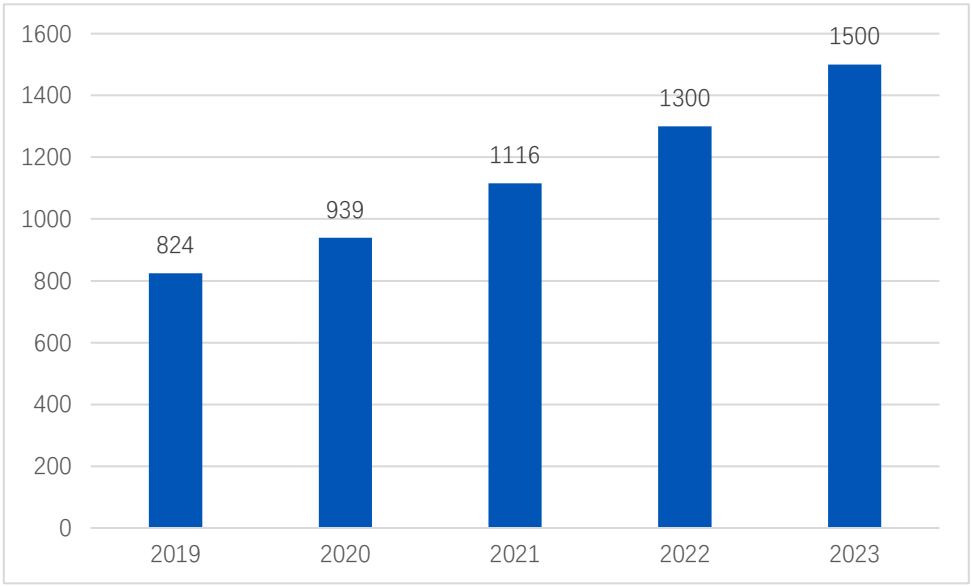

图4 近5年我国算力中心耗电量（单位：亿千瓦时）

（来源：中国信通院）

1.2.4 挑战四：多样化算力需求提升，普适普惠水平较低

算力资源获取成本有待降低，多元算力匹配能力有待加强。一是算力资源获取上，据斯坦福《2024 年人工智能指数报告》估算，OpenAI 的 GPT-4 预计使用了价值 7800 万美元的计算资源进行训练，而谷歌的 Gemini Ultra 耗费了高达 1.91 亿美元的计算资源成本。目前大模型研发已进入万卡时代，一家企业如果想拥有自己的大模型，至少需要几十亿投资，对于中小企业来说算力成本过高。**二是算力应用上，**大模型训练、推理等业务场景的出现促使企业业务对多样化算力需求提升，如今产业界不论是模型还是算力芯片，正处于百花齐放、创新并存的阶段，算力资源多元并用，多元算力与多种模型及框架的适配难度较大。另外，大模型应用能够帮助企业更高效率的实现商业目标，但对于绝大多数企业而言，大模型的应用开发流程繁琐，模型设计、训练、调优等环节需要专业开发人员，自研大模型成本高且研发门槛过高。

1.2.5 挑战五：供应链完备性不足，生态构建待完善

算力供应链完备性不足，亟需强化算力保障能力。一是随着多元异构算力的发展，不同 OS、固件、整机、芯片平台兼容性问题突出，不同硬件生态系统封闭且互不兼容，给算力使用方带来一系列技术挑战。二是算力服务商资源采购受各厂商芯片生态影响，存在应用与硬件紧耦合、难迁移问题。如一些芯片厂商为了维护自身利益，会构建相对封闭的生态系统，限制其他厂商或第三方开发者的接入。这种封闭性导致应用开发者只能针对特定厂商的芯片进行优化和定制，进一步加剧了应用与硬件之间的紧耦合关系。

1.2.6 挑战六：性能评价简单，算力实测性能欠缺

现有算力评估体系评估场景简单，无法全面、深入地反映算力应用的实际效果。如今算力基础设施面临着更高的建设与发展要求，应对算力质量进行系统评估，保障算力安全稳定运行和资源高效利用。在万亿参数模型的训练过程中，软硬件组件需精密配合，一旦出现问题，其定界与定位过程极为复杂。根据公开资料显示，业界在硬件故障定位上通常需要 1-2 天，复杂应用类故障的定位时间则更长。节点故障不仅会导致训练时间大幅延长，还会对算力资源造成巨大浪费。然而，当前算力评估体系由于评估场景相对单一，往往难以全面、深入地揭示算力应用在实际复杂环境中的真实效果，测试评估指标主要以单芯片性能测试为主，测试结果偏理论，参考价值有限，对于多场景下算力的性能评估也缺乏深入研究与重视，这严重限制了评估体系在指导高效能、多元化算力资源配置中的应用广度和效果。

2 定义、内涵及特征

全球及我国算力发展态势显示，在人工智能时代，算力产业政策导向日益明确、算力领域相关技术突破发展、算力投资持续加码、智算规模不断增加、整体发展水平日益提升，然而算力发展也面临着供给不足、供需匹配难、能耗激增等挑战。算力成为衡量国家综合实力和国际话语权的重要指标，亟需更高水平的发展变革以应对新阶段的发展机遇和挑战。

2.1 定义

人工智能时代，高质量算力是基于最新人工智能理论，采用先进

的人工智能计算架构，与算法、数据深度结合的高水平计算能力。高质量算力是驱动“算法”赋能劳动者、加速“数据”优化劳动对象、激发“算力（设备）”升级劳动资料，从而推动生产力要素发生根本性质变，实现全要素生产率提升的一种新质生产力，有效推动数字经济与实体经济的深度融合，实现经济社会高质量发展。

表 3 生产力变迁

	农业时代	工业时代	人工智能时代
劳动者	**人** 缓慢增长（马尔萨斯陷阱）	**人** 线性增长（医疗、粮食进步）	**人+算法** 算法产生智能，劳动者能力增强
劳动对象	**农作物** 比较原始	**+工业产品** 延伸至一切可利用的有形物质，总量越用越少	**+数据** 从有形到无形，生生不息，越用越多
劳动资料	**农业工具** 生物能驱动	**+工业设备** 化石能、电能驱动	**+算力（设备）** 计算力驱动、生产工具智能升级

（来源：浪潮信息、中国信通院）

人工智能时代，高质量算力具备技术创新的“主引擎”、生产要素配置的“优化器”、产业深度转型升级的“催化剂”三大内涵；具备高算效、高智效、高碳效、可获得、可持续、可评估六大特征；以系统设计、协同驱动、全生命周期管理、基建先行、繁荣生态、多元评估为六大主要发展路径，可充分响应数字经济快速增长的计算需求，应对人工智能时代算力发展机遇和挑战。

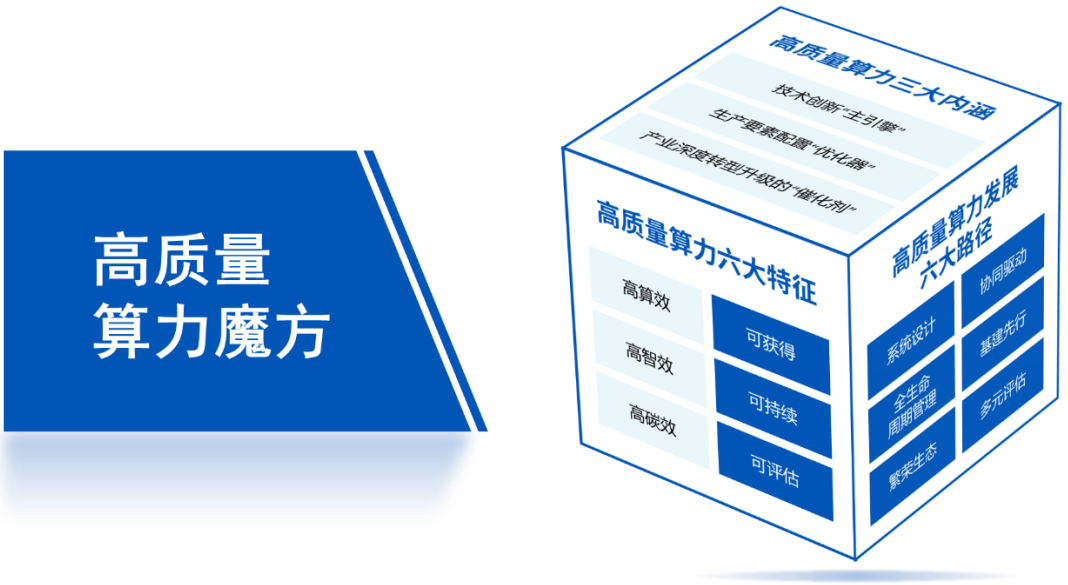

图 5 高质量算力魔方

（来源：中国信通院）

2.2 内涵

高质量算力是人工智能时代的新质生产力。新质生产力本质是生产力，由技术革命性突破、生产要素创新性配置、产业深度转型升级而催生。人工智能时代背景下，高质量算力已逐渐融合到生产生活的方方面面，为社会、产业智能化转型提供基础动力，具体体现在以下三点：

一是技术创新的“主引擎”。高质量算力推动人工智能模型训练和应用，在 AI 模型训练和部署上更具优势，推动预训练大模型在海量数据的学习训练后具有良好的通用性、泛化性和高效率，用户基于大模型通过零样本、小样本学习即可获得领先的效果，能够显著降低 AI 应用的门槛。在科学研究方面，高质量算力强大的计算能力能够提高大规模数据处理的速度，缩短模型训练的时间，进一步提升科研效率，降低科研成本，同时加速前沿科学问题的探究，如新药创制、基因研究、新材料研发等，推动科学技术创新发展。

二是生产要素配置的“优化器”。数据是人工智能的三要素之一，高质量算力提供强大的计算资源对数据进行分析、加工、处理，将海量数据转化为先进生产要素，畅通生产、分配、流通、消费各环节，构建数据流通交易体系，实现数据资源的有效配置和价值最大化。另外，高质量算力通过智能化的资源管理和调度系统，根据实际需求动态调整资源分配，对计算、存储和网络资源进行优化配置，通过虚拟化和云计算技术实现资源的弹性伸缩，提高资源利用效率。

三是产业深度转型升级的“催化剂”。在改造提升传统产业上，高质量算力推动前沿科技与传统行业相结合，为传统产业注入新活力，催生新的商业模式和服务，如智能制造、智能医疗、智慧金融等。在培育壮大新兴产业上，高质量算力推动人工智能、大数据、区块链等数字产业发展，加快数字产业化步伐。在推动未来产业建设上，高质量算力瞄准前沿领域，促进元宇宙、人形机器人、脑机接口、量子信息等新产业发展，加速重大科技成果产业化。

2.3 特征

2.3.1 高算效：设计与运行计算效率“双优”体现

高算效指的是在提高算力理论算效的同时考虑更高的实测性能和资源利用率，是综合考虑设计、运行等维度的计算效率。

理论算效是设计维度上的考量，是计算系统算力与功率的比值，即“每瓦功率所产生的算力”，是同时考虑计算性能与功率的一种效率指标。理论算效数值越大，代表单位功率的算力越强，效能越高。2023 年年底，全国在用算力中心平均算效为 11.8GFLOPS/W，达到

GB/T 43331-2023《互联网数据中心（IDC）技术和分级要求》中通用算力算效第三等级，我国算效水平仍有较大的提升空间。

实测性能是短期运行维度上的考量，在充分考虑理论算效的基础上，关注的是应用场景下的实测性能，如单位时间内处理的 Token 数量、时延、模型训练时间、数据处理质量等指标。在实测性能方面，高质量算力致力于提升整机系统输出能力，消除网络、存储等集群性能瓶颈，兼顾软件生态建设、应用适配。如今算力集群实测性能和理论性能差距过大这一问题逐渐凸显，部分算力实测性能不足 10%，大量算力资源被浪费，算力系统实测性能亟需优化提升。

资源利用率是长期运行维度上的考量，指的是算力系统实际运行过程中的平均资源利用率，避免算力堆砌及大量资源闲置。资源利用率的“高”体现在不断优化算力资源实际应用水平。如可通过优化 GPU 平均利用率来提升算力资源利用率。据公开数据统计，传统模式下的智算中心 GPU 利用率较低，平均数值低于 30%，英伟达 GTC 2022 公布数据显示 Google 云平均 GPU 利用率为 25%，算力资源利用率存在较大优化空间。

2.3.2 高智效：兼备高效和智能的 AI 业务支撑能力

高智效指的是算力具备高效处理 AI 业务的能力和较高的自动化、智能优化水平。

高智效体现算力面向人工智能业务的高效处理能力。在以大模型为代表的人工智能技术上，更高的模算效率是体现高智效的主要指标。模算效率正比于模型精度与模型的计算效率，模型本身精度越高、在

对应软件上对硬件性能利用率越高、推理及训练所需算力越小，模算效率越高，整体反映被测模型在人工智能业务应用中的综合处理效率。

高智效体现算力较高的自动化水平和智能优化能力。这种高效能不仅体现在硬件的计算速度和处理能力上，更在于整个系统对资源的智能管理和优化调度上。通过高度自动化的任务调度、资源分配和故障恢复机制，以及智能优化算法、能耗管理和自适应学习技术，高智效的算力系统能够实现更强的可靠性和可用性，为推动智能化应用的发展提供有力支持。

2.3.3 高碳效：最低碳排放前提下实现最大化算力输出

高碳效指的是算力在最低碳排放前提下实现最大化算力输出，是兼顾碳排放量和算力性能的综合指标。

高碳效不仅关注算力单纯的能源消耗、算力输出水平，更注重算力系统从采购到回收整个过程的全生命周期碳足迹管理。算力碳效是有效衡量高碳效水平的一个关键指标，指设备使用周期内产生的碳排放与所提供的算力性能的比值。据《数据中心算力碳效白皮书》测算表明，对于服务器来说，CPU 性能越好，可提供的计算能力更优越，所消耗的能量越多，带来的温室气体排放也越多，但服务器算力碳效即单位算力性能的碳排放量反而会降低。**全生命周期碳足迹管理**主要通过绿色采购、绿色设计、清洁生产、绿色包装和运输、绿色运营、回收处理等降低整个算力系统的碳排放。如在绿色设计环节采用高效的服务器和存储设备、先进的制冷和散热技术。根据中国电子节能技术协会的数据，液冷 PUE 低于传统风冷 PUE 至少 50%，意味着算力

中心的能耗可至少减少 50%，碳排放量也将减少 50%。

2.3.4 可获得：普适应用需求和普惠使用成本的极致追求

可获得指的是算力能够满足普适应用需求和普惠使用成本要求。

普适性表现在算力能够广泛满足各种应用场景的需求。普适性保障算力系统在智慧医疗、智能制造、自动驾驶、金融服务、科研计算、智慧城市等多个领域的广泛应用，是推动这些领域发展的关键因素。在使用门槛上，算力像水电类的公共资源一样，具备好用易得等特点，各行各业用户无需深入了解复杂的技术细节，只需简单的操作即可获得所需的计算资源。

普惠性主要表现在可保障用户以合理、可负担的成本使用算力资源。普惠算力通过优化资源配置、提供灵活计费方式、推动技术创新等手段不断降低用户成本、扩大服务覆盖范围、提升用户体验，可实现各行各业低成本使用，无论是大型企业还是中小企业，甚至是个人开发者，都可以方便地获取和使用算力资源。

2.3.5 可持续：技术兼容、供应链完备、产业生态开放的共同选择

可持续指的是算力具备技术兼容、供应链完备、产业生态开放的特性。

在技术上，算力的可持续特征体现在可向“前”兼容，向“后”持续迭代升级，软件配套支持能力及开放水平高，不同品类、不同技术路线的芯片、算法、模型、应用可实现兼容适配。

在供应链上，算力的可持续特征体现在从核心部件到专用芯片、从电子元器件到基础软件、应用软件的全方位、多层次的供应链条强

大且完备，并以完善的产业链带动算力产业上下游的发展。

在产业生态上，算力的可持续特征体现在算力服务生态开放，算力系统分层解耦，提供可替代的算力支撑能力，可有效打破封闭和垄断现象，降低应用迁移门槛。

2.3.6 可评估：反映算力实际应用水平的多元评估

可评估指的是面向人工智能应用场景，算力可通过完整、有效的评估体系得到反映实际应用情况的特性。

当前算力评估体系存在评估场景简单、同质化严重、评估维度单一、全面性不足等问题，亟需拓展系统性能评价维度，以更全面、深入地反映算力应用的实际效能与发展潜力。高质量算力的评估体系能够满足市场对高效、可衡量的计算力解决方案的评估需求，评估体系主要具有以下特点：

一是评估体系全面多元，适用范围广。评估体系指标多元，综合考虑人工智能不同业务场景、多环节的质量评估需求，侧重评估算力在人工智能方面的创新能力和对新技术的支持程度，评估结果为算力技术和成本优化提供依据，提升算力资源的人工智能支撑能力。

二是评估指标以应用为导向，科学合理。重点体现在从单纯的理论计算效率评估升级为对实际应用效率等的多维评估。通过标准化的评估方式得出准确数据，通过数据结果辅助用户决策，提出贴合实际的指导建议，促进技术、产品功能不断升级，引导产业向更高效、规范的方向发展。

3 发展路径及展望

3.1 发展路径

发展具有高算效、高智效、高碳效、可持续、可获得、可评估六大特征的高质量算力是迎接新阶段重大发展机遇、应对严峻挑战的关键途径。**聚焦高质量算力“三高三可”特征，可推动算力发展由“量”向“质”**。高质量算力以系统设计提升算效、协同驱动提升智效、全生命周期管理提升碳效、基建先行推动算力普适普惠、繁荣生态推动算力可持续发展、多元评估加速算力规范化发展为主要发展路径，可全面赋能数字经济、智能社会发展和新型科技创新。

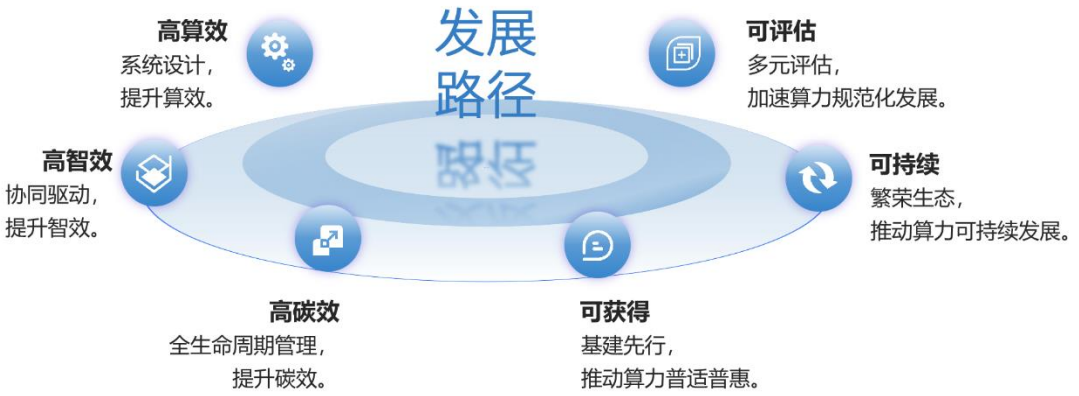

图 6 高质量算力发展路径

（来源：中国信通院）

3.1.1 系统设计，提升算效

以系统设计为核心，围绕生产、聚合、调度、释放算力四大关键环节全面提升算效。将算力的生产、聚合、调度和释放视为一个整体，通过精心设计的系统架构和流程来优化每个环节，从而实现算力资源的有效利用和性能的最优化。**在生产算力方面，提升算力输出水平。**聚拢核心部件、专用芯片、电子元器件、基础软件、应用软件等国内外产业链领先技术方案，整合高性能部件，打造多元异构的强大算力机组。**在聚合算力方面，提升集群算力使用效率。**运用系统工程方法，

构建高效的算力中心集群，通过卡间和节点间的互联网络、软件和硬件的适配调优等提升集群算力效率，对大规模集群网络进行精细化设计，减少算力资源空闲度。**在调度算力方面，实现算力高效调度。**通过硬件重构和软件定义对 GPU、AI 芯片等进行聚合池化，再利用先进的资源管理技术进行切分、调度、分配，提升算力资源利用率。**在释放算力方面，提升算力应用水平。**聚焦于模型算法、框架、工具等方面不断优化完善，提高模型框架与底层 AI 芯片适配度，开发更加高效、易用的模型训练工具，降低用户开发大模型的算力资源开销，充分激活算力资源性能。

3.1.2 协同驱动，提升智效

（1）算力、算法、数据协同创新，强化算力智效水平

算力层面，推动全新计算架构的创新应用，如通过数据流架构、存算一体、Chiplet 技术等改进芯片的架构、互联、封装，满足人工智能业务对算力高效率和高可靠性的需求；**算法层面，**加强模型算法的研究，改进算法结构，减少冗余计算，提高算法的运行效率，推动模型算法创新，研究更高效的并行算法、更智能的任务调度处理，使算力能更灵活的适应不断变化的应用需求；**数据层面，**通过加强数据清洗、创新数据合成等手段构建高质量数据集，充分激活数据要素价值，使得系统能够更好地理解和处理复杂的业务逻辑，从而提升智能化水平。

（2）软硬件联合设计与优化，提升算力 AI 业务应用水平

软硬件联合设计与优化的核心在于将软件算法与硬件架构紧密

结合，以实现更高效的计算性能和更智能的业务处理能力。一是通过软硬件协同设计消除软硬件之间的瓶颈，提高整体的算力应用水平。通过定制化的硬件设计和针对性的软件优化，提高整体系统的应用性能。例如通过专门为 AI 算法设计的硬件加速器，可以显著提高数据处理速度。二是根据特定业务需求提供定制化的解决方案，软硬件联合设计提高业务处理的灵活性和适应性。开展软硬件协同的高效微调方案研究，如结合企业专有数据进行模型微调，可使通用模型升级为企业模型，增强 AI 算力在特定应用场景和任务中的智能处理能力。

3.1.3 全生命周期管理，提升碳效

（1）全生命周期低碳管理，推动全产业链节能降碳

一是提升算力系统可再生能源利用率，减少对化石燃料的依赖，降低碳足迹，从源头打造绿色算力体系。二是梳理供电、制冷、服务器、网络、存储等各环节碳足迹，建立绿色算力供应链，优先选择环保材料和生产过程，推动全产业链节能降碳。三是建立全生命周期低碳管理制度，在算力系统的规划设计、施工建设、运行维护直至最终退役和废弃处置的全过程充分考虑绿色发展，减少算力全生命周期的环境影响。

图 7 全生命周期碳足迹管理探索

（来源：浪潮信息）

（2）采用节能低碳新技术，降低算力中心能耗

算力系统的建设、设备选型、平台运营等方面均采用绿色低碳技术，通过材料、产品、工艺创新等手段全方位降低能耗。如算力基础设施能耗主要来自于计算、制冷和配电系统，在供配电以及制冷散热方面，可通过高效的不间断电源（UPS）和电源转换设备减少电力转换过程中的能量损失、采用液冷技术提高散热效率、利用 AI 和大数据技术对算力制冷系统智能化管理，降低整体能耗。

图 8 算力绿色低碳技术

（来源：中国信通院）

3.1.4 基建先行，推动算力普适普惠

（1）以智算中心为建设重点，强化智算供给能力

生成式 AI 带来的智能时代需要大量增加智能算力的供给才能满足更大参数量的模型训练需求，应分层次、分类别建设布局智算中心，缓解如今智算资源紧张、算力成本高的问题。一是集中建设大规模智算集群，着重满足人工智能大模型对海量数据和复杂计算的需求，确保高效、稳定的运算服务；二是区域建设小规模算力中心，以适应不同行业和场景下的多样化计算需求，形成对大型算力节点的有力补充。如今智算中心的建设和服务市场高度竞争且不断变化，智算中心市场的参与者应重视智算中心技术研发、商业模式、服务模式和市场策略的创新，把智算中心“建好”更要“用好”。

（2）积极推动算力平台建设，提升算力供需匹配能力

我国算力产业存在算力市场分散、供需匹配能力不足、计算框架不同等问题，算力平台可实现对算力资源的统计、监测、匹配和分析，提升算力供给水平与资源匹配能力。算力资源需求方和提供方应积极促进算力平台的建设，以平台为依托进行供需对接，充分连接算力资源供给，实现算力的一键式订购和灵活调整，推动算力普适普惠。

图 9 中国算力平台

（来源：中国信通院）

3.1.5 繁荣生态，推动算力可持续发展

（1）算力技术，开放解耦

算力技术应采用多元开放的架构，兼容成熟主流的软件生态，支持主流的 AI 框架、算法模型、数据处理技术、广泛的行业应用等，CPU、服务器、云操作系统、平台层、应用层等分层解耦，消除单一技术路线依赖，构建开放技术生态。算力技术的开放解耦可通过标准化、模块化的设计实现，使算力技术的各个组件能够独立发展、灵活组合，从而实现技术间的互联互通和资源共享，这种模式有助于打破传统封闭系统的限制，促进技术创新和产业融合。

（2）算力产业，标准化建设

建立完善的标准体系，推进不同 OS、固件、整机、芯片平台兼容、统一算力度量标准，推动低代码无代码开发平台标准化。另外，从技术、接口、设备、平台等多个维度，全面采用标准化的设计，不同层次间可通过虚拟化、API 等标准化方式对接，实现产业链整体标

准规范，提高技术的兼容性和设备的通用性，降低集成和迁移的成本。

（3）算力生态，合作共创

加强交流合作，积极参与算力产业生态建设。依托行业协会、技术联盟等合作组织，加强和产业的交流合作，促进技术成熟推广，实现技术路线、目标架构、标准体系的统一，软件层、硬件层到模型层、应用层等全产业链合作共建，形成行业共识和最佳实践。另外，积极推动建构资源共享、平台共建、价值共创的产业生态。将企业个体向有机融合的产业链条聚集，促进算力上、下游产业及应用生态市场协同发展，充分发挥算力对其他相关行业的赋能价值。

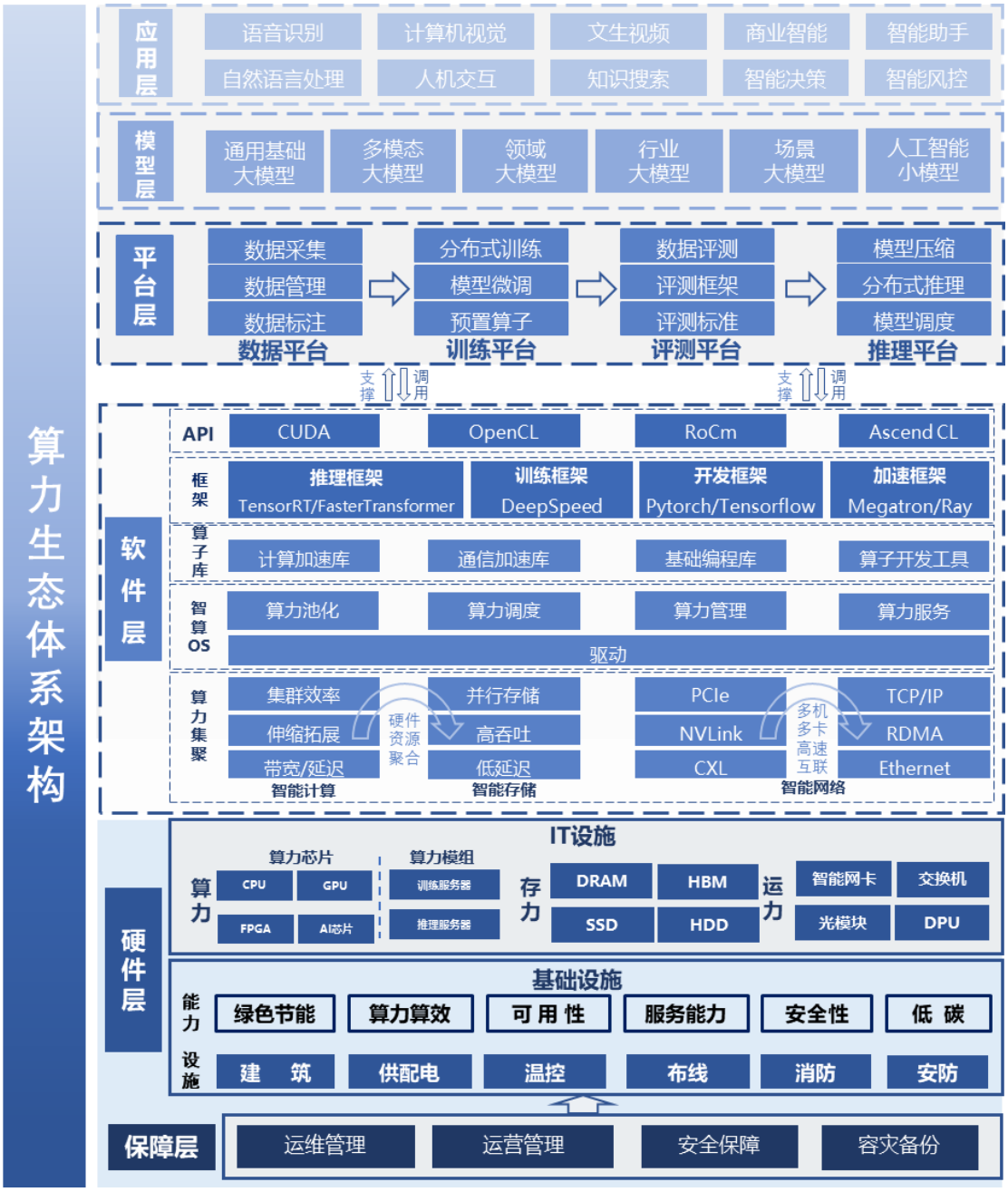

图 10 算力生态体系架构

（来源：中国信通院）

3.1.6 多元评估，加速算力规范化发展

（1）实测性能评估指导算力建设，提升算力利用率

算力的真实应用性能需要综合考虑计算、存储、网络以及平台软件各层协调所呈现的综合业务水平，传统算力度量常关注 IT 计算设备的单台设备理论性能，无法完全体现集群系统或者算力中心整体性

能。未来应以应用为导向，深入分析算力实际应用场景下的关键性能指标，通过评估真实业务性能表现，如实测性能、模算效率等，衡量算力对业务的支撑效果，引导算力提升实际利用率。

（2）建立全面算力评估体系，开展算力评估评测

产业各主体积极参与面向实际应用场景下的算力系统（小到单机大到算力中心集群）评估，完善相关技术要求和测试方法等；建立多维、全面算力评估体系，如针对算力中心的节能降耗，构建算力全生命周期节能低碳水平评价体系，推动整个产业链的绿色发展；算力相关生产者、使用者、集成者、测试者积极推动算力质量测试评估，依据测试结果不断优化技术、产品，提高新技术、新产品的广泛认可度，促进产业链成熟，规范加速算力产业高质量发展。

3.2 展望

根据新质生产力发展新方向和新要求，未来数年，高质量算力在推动人工智能全面发展、加速产业升级转型等方面的价值将被持续发掘，不断激发高质量发展新动能，并深度影响社会发展、产业变革、人类生产生活。聚焦高质量算力发展，将呈现如下趋势：

（1）市场环境将成为支撑我国算力产业发展的最大优势

我国拥有以人工智能、智能制造、自动驾驶等为代表的庞大应用市场，也是全球体量最大、用户最活跃的数据市场，丰富的数据量和应用场景为算力产业发展提供广阔土壤，对计算资源的需求巨大。如今国家层面高度重视算力产业的发展，将其纳入国家发展战略，各地方政府也纷纷出台相关政策，提供税收优惠、资金补贴等支持措施，

鼓励算力产业集聚发展。在这样的背景下，算力产业正迎来前所未有的发展机遇，将迸发出蓬勃的活力与生机。

（2）算力普适普惠将推动人工智能在各领域释放潜力

在人工智能、数字经济的拉动下，算力普适普惠化是大势所趋，产业生态也将逐步完善。依托完善的高质量算力基础设施，各行业将不必从零开始开发，只需结合领域数据进行调整和增量学习，即可形成具有良好精度和性能的下游应用。高质量算力的进一步普及将为AI在医疗、教育、交通、金融等多个领域的应用提供强大的支持，推动AI技术更深入地融入到各行各业的业务流程中，人工智能在各个领域将展现出巨大的潜力和价值。

（3）算力智能升级将推动经济社会深层次发展

随着AI大模型等新兴技术和应用的快速发展，算力资源将加速整合，形成规模化发展，高质量算力实现对经济发展效能的放大、叠加、倍增，推动算力经济蓬勃发展。算力投入将带动制造业、工业、交通等其他行业更高的经济增长，高质量算力建设将进一步激发数据要素创新活力，加快数字产业化和产业数字化进程，加速新旧动能转化，有效改善民生，为生产端、流通端、消费端对数字化、智能化的多样化需求提供坚实保障。

4 评估体系探索

在人工智能时代，多模态数据挖掘、智能化业务处理、海量数据分布式存储调度、人工智能模型开发、模型训练和推理服务等场景的不断涌现，对算力要求不断提高，我国算力发展开始走向了由“量”

向“质”的重要阶段，对高质量算力的需求正日益增长。然而，只有通过精准的性能评估与测试，发现算力系统的瓶颈，才能促进算力技术创新和产品优化，不断提升算力质量发展水平。因此，本报告结合人工智能时代算力发展面临的机遇与挑战，初步提出面向人工智能的算力高质量发展评估体系。

4.1 评估体系构建背景

4.1.1 评估体系构建现状

目前我国算力评估主要可分为规格算力评估和算力综合评估两大类。其中，规格算力评估主要关注硬件设备的计算性能，评估方法通常采用标准化的测试程序，对硬件设备进行基准测试，以获取其计算水平。算力综合评估通常采用多种测试方法和工具，对算力系统进行多元的性能测试和分析，由于综合考虑多个因素，评估过程相对复杂。规格算力评估和算力综合评估各有优缺点，适用于不同的应用场景和需求。

（1）规格算力评估

规格算力评估主要以芯片的标称算力为基准，一般可分为部件级算力评估、单机算力评估、算力中心/集群算力评估三个维度。部件级算力评估主要测试部件的规格算力，针对部件标称的算力指标进行测试，如内存的 Stream 测试软件，硬盘测试软件 IOzone 等，反映硬件设施的基本性能；单机算力评估多采用实际业务中的计算密集型业务场景作为测试负载，得到该设备的整机算力，如通用算力评测工具 SPEC、CPUBench 等，只关注 IT 计算设备的单台设备性能，无法完

整体现集群系统性能；算力中心/集群算力评估按照通用算力中心、智算中心、超算中心分类检测，这类测试侧重于对算力系统的单一性能评价，如超算算力评测工具比较成熟的有 Linpack（集群环境下多采用 HPL 基准）等，重点呈现系统在稠密矩阵求解方面的能力。

（2）算力综合评估

现阶段算力综合评估一般涉及算力的多维度、多指标分析，不同的评估模型体系分析角度不同，侧重点也有一定偏差。部分评估体系侧重于算力计算能力的综合展现，针对性解决基于单一指标难以全面评估算力水平的问题。如算力五力模型综合考虑通用算力、智能算力、算效能力、存储能力、网络能力，结合双向投影法和 TOPSIS 方法对算力综合情况进行评估。部分评估体系充分响应国家政策趋势和发展需求，针对某一方面进行系统评估。如《数据中心全生命周期绿色算力指数白皮书》针对算力绿色低碳发展方面构建了数据中心全生命周期绿色算力指数体系，从安全高效性、绿色低碳性、管理智能性和全生命周期绿色管理四个维度评价数据中心的绿色算力水平。整体而言，现阶段算力综合评估重点在于检测算力单一维度的能力，无法广泛覆盖算力从建设到应用的综合性评估。

4.1.2 评估体系构建建议

从规格算力评估和算力综合评估体系的指标中可以看出，算力评估指标从单点部件能力逐步过渡到更全面的系统能力，也逐渐更加响应产业发展趋势和国家政策要求。结合以往算力评估体系及当前算力需求特点，本报告认为算力评估体系应做出系列优化：

第一，应以应用为导向，增加实际业务性能指标检测。规格算力并不能准确反映实际计算能力，因为算力的发挥需要算力系统各个部件的协作，任何性能上的薄弱环节都会对整个计算系统产生影响。另外，虽然理论算效衡量方式较为简洁，但不能完整反映真实的网络、存储等系统能力。人工智能时代的高质量算力如何进行评估，对应的标准体系如何建立，需充分考虑应用实际情况，因此应增加实测性能指标直观地反映计算系统在特定作业上的、用户可获得的计算能力。

第二，应聚焦人工智能，强化 AI 业务支撑能力评估。人工智能驱动算力走向“重应用”阶段，算力加速向政务、工业、交通、医疗等各行各业渗透，成为传统产业智能化改造和数字化转型的重要支点，提升算力在垂直行业领域的智能支撑水平是未来算力高质量发展的重要着力点。对于算力质量的评估应聚焦算力在 AI 业务中的性能表现，为人工智能时代下算力高质量发展提供指导。

第三，应全面节能降碳，注重全生命周期碳足迹管理。PUE 及耗电量通常被认为是算力绿色评估的核心指标，如 2020 年底全国数据中心平均 PUE 为 1.62，总耗电量为 939 亿 kWh，2023 年底平均 PUE 和总耗电量分别为 1.48 和 1500 亿 kWh。算力产业耗电量急剧增加，但 PUE 优化空间不断缩小，单纯从 PUE 角度进行评估优化已无法适应算力产业的绿色发展需求。因此，应从全生命周期角度对算力碳足迹进行优化管理，评估算力从采购、设计到运营、回收等全产业链的节能降碳水平，响应国家双碳目标。

第四，应以系统为核心，算力设施与算力资源利用整合评估。从

评估体系现状可以看出，现有评估体系多是以上架率、PUE、WUE 等指标为主，无法综合反映算力资源应用时的系统性能。以上架率为例，据中国信通院统计，2023 年底全国在用数据中心上架率为 66.7%，该指标通过简单的计算就可以得出当前全国算力资源的利用情况，然而上架率主要关注物理层面的资源占用情况，如机架空间、电源插座等，忽略了服务器的实际性能和负载情况。如果数据中心的服务器配置不合理或存在大量闲置资源，即使上架率很高也可能无法满足实际业务需求。因此，算力评估指标方面应将算力基础设施和算力资源利用情况统一考量，注重算效水平、智效水平、碳效水平等效率的综合评估，弥补现有评估体系过于注重算力单一方面性能指标的缺点，满足日益复杂多元的算力应用需求。

4.2 评估体系构建原则

结合评估体系构建现状和建议，本报告在此基础上，致力于构建全面、实用的算力质量评估体系，客观评价算力质量发展水平，尝试提出高质量算力评估体系。评估体系评估对象主要为算力系统，在评价指标的筛选上强调以下六个原则：

一是导向性原则，确保评估指标与政策目标和区域发展需求保持一致；**二是系统性原则**，要求评估体系全面覆盖高质量算力的关键特征，确保评价结果能够全面反映算力质量；**三是针对性原则**，强调选择与高质量算力特征紧密相关的指标，使评估更具针对性和准确性；**四是全面性原则**，确保评估体系综合考虑数据的可获取性和量化的可行性，以实现全面、高效的评价；**五是可操作性原则**，要求评估体系

的设计既要理论合理，也要实际可行；**六是可拓展性原则**，要求评估体系具备适应未来技术迭代和政策变化的能力。这些原则共同确保高质量算力评估体系既符合当前需求，又能够灵活适应未来的发展。

4.3 评估体系构建实践

4.3.1 评估体系

根据建立评估体系系统性、全面性等原则，征求专家意见，梳理高质量算力内涵、特征及关键影响因素，从算效水平、智效水平、碳效水平、可获得水平、可持续水平 5 个维度形成“五位一体”高质量算力评估体系，指标包括理论算效、实测性能、模算效率等 12 个指标。

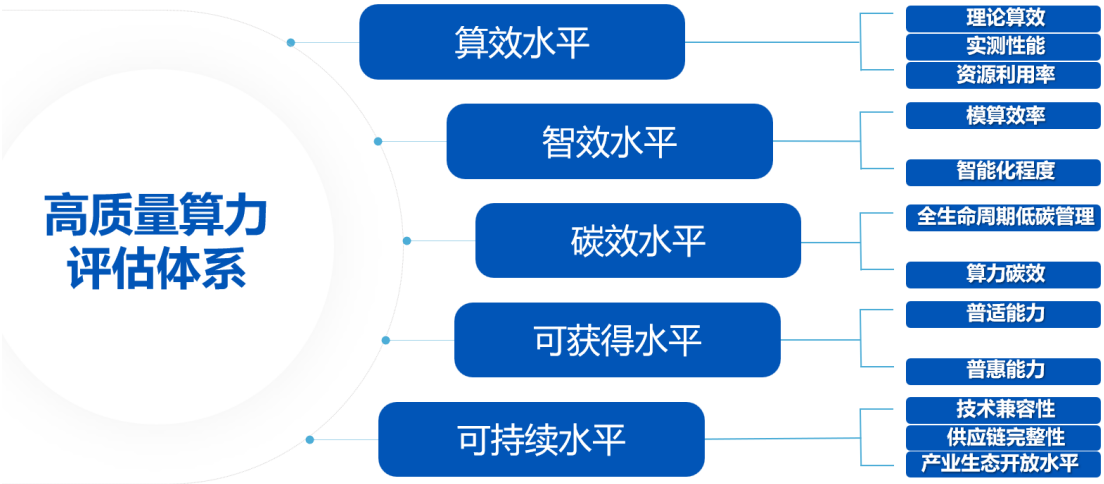

图 11 “五位一体”高质量算力评估体系

（来源：中国信通院）

4.3.2 算效水平

算效水平主要通过理论算效、实测性能、资源利用率三大指标评估，反映实测计算性能和算力资源利用率。

（1）理论算效：理论算效（CE）是算力（CP）与所有 IT 设备功耗的比值。数值越大，代表单位功率的算力越强，效能越高。计算

公式为：CE=CP/∑IT。理论算效的单位为 FLOPS/W，即单位功率的算力。提升算效可以同时降低 IT 设备本身的运行成本和制冷设备的运行成本，从而降低算力系统整体供电负担，降低整体的能耗。根据理论算效公式，可以通过提升算力输出或者降低能耗来提升算效。

（2）实测性能：实测性能反映算力系统对特定 AI 业务的处理能力。将代表性 AI 应用负载的实测性能综合起来，最终得出算力系统的综合实测计算能力，通过几何加权平均的方式获得具体数值，有助于进行定量、对比性分析。通过实测性能评估可准确了解算力系统在实际运行中的性能表现，避免仅依赖理论计算或模拟结果导致的偏差。

（3）资源利用率：通过算力系统实际运行过程中的平均资源利用率来综合评估。如可通过 IT 负载率或 GPU 平均利用率来衡量。IT 负载率可以反映算力系统 IT 设备的有效工作效率，是服务器总实际功率与总额定功率的比值，从设备层面衡量服务器的利用水平。GPU 平均利用率可以确保 GPU 资源得到充分利用，减少额外的硬件投资需求，降低运营成本。

4.3.3 智效水平

智效水平主要评估算力系统的人工智能业务支撑能力，体现在能根据 AI 业务的需求实现灵活自主、智能化的高效处理。通过模算效率和智能化程度两大指标评估。

（1）模算效率：模算效率正比于模型精度与模型的计算效率，模型本身精度越高、在对应软件上对硬件性能利用率越高、推理及训练所需算力越小，模算效率越高。模算效率评测对象是大模型训练和

推理的整机系统，包括模型本身、软件框架、算力底座，用于反映被测试模型在硬件和软件框架下的系统性能。

（2）智能化程度：通过算力系统自动化水平和智能优化能力来评估。自动化水平衡量算力系统在任务调度、资源分配、故障恢复等方面的自动化程度。高自动化程度意味着算力系统能够减少人工干预，提高运维效率。智能优化能力衡量算力系统是否具备根据实时负载和性能数据进行自我优化的能力。具备智能优化能力的算力系统能够更好地适应业务变化，提高整体性能和效率。

4.3.4 碳效水平

碳效水平追求最低碳排放情况下的最大化算力输出，主要通过全生命周期低碳管理能力和算力碳效两大指标来评估。

（1）全生命周期低碳管理：主要从算力的采购、设计、建设、运营、回收等全生命周期角度评估算力的低碳性。可通过算力系统碳足迹和全生命周期绿色发展战略来评估。算力系统碳足迹是指算力系统全生命周期过程中产生的温室气体排放总量，反映算力系统在整个生命周期中的环境影响程度。全生命周期绿色发展战略意味着算力系统的采购、设计、建设、运行维护直至最终退役和回收的全过程都要充分考虑绿色发展，保障全产业链节能减排。

（2）算力碳效：是兼顾算力 IT 设备的碳排放量和算力性能的综合指标，指设备使用周期内产生的碳排放与所提供的算力性能的比值。最新发布的《数据中心绿色低碳发展专项行动计划》提出到 2030 年底，全国数据中心单位算力碳效达到国际先进水平，未来算力碳效或

将成为算力系统设备设计、选型的重要指标。

4.3.5 可获得水平

可获得水平考察算力是否能够满足普适应用需求和普惠使用成本要求。主要通过普适能力和普惠能力两大指标综合评估。

（1）普适能力：主要考量算力对多样化应用场景的广泛支撑能力，可根据内部业务支持、区域业务支持、全国范围内业务支持能力来分级评估。应用门槛低、使用灵活的普适算力有助于满足不同行业和领域的多样化算力需求，推动人工智能、大数据、物联网等技术与实体经济的深度融合。

（2）普惠能力：主要考量算力是否满足各行各业低成本使用需求，通过算力的市场价格优势进行综合评估。主要进行成本效益分析和定价策略分析，对比算力提供者的服务价格与其竞争对手的价格，评估其性价比。

4.3.6 可持续水平

可持续水平考察算力系统技术兼容、供应链完备、产业生态开放的程度。通过技术兼容性、供应链完备程度、产业生态开放水平三大指标评估。

（1）技术兼容性：通过模拟实际应用场景，对算力系统进行兼容性测试，以验证其在协同工作时的稳定性和性能表现，进行分级评估。评估算力系统技术兼容性有利于确保不同系统、设备和应用之间能够高效地交互和协作，从而提升整体性能和用户体验。

（2）供应链完备程度：评估供应链中供应商的数量和质量，包

括供应商的稳定性、交货能力、质量控制等方面；评估合作伙伴数量与质量，如是否拥有广泛的合作伙伴网络，包括硬件供应商、软件开发商、服务提供商等。

（3）产业生态开放水平：分析算力系统的标准化建设程度，如采用的技术标准是否与行业主流标准兼容。另外评估算力系统是否构建开放的平台生态系统等。产业生态开放有助于确保不同厂商的设备和技术能够无缝集成和协同工作。

表 4 高质量算力评估体系指标说明

评估维度	评估指标	指标描述
算效水平	理论算效	算力系统中算力与所有 IT 设备功耗的比值，即“IT 设备每瓦功耗所产生的算力”。
	实测性能	将 AI 应用负载的实测性能综合起来，加权平均，最终得出算力系统的综合实测计算能力。
	资源利用率	算力系统实际运行过程中的平均资源利用率，如 IT 负载率、GPU 平均利用率。
智效水平	模算效率	模算效率正比于模型精度与模型的计算效率，模型本身精度越高、在对应软件上对硬件性能利用率越高、推理及训练所需算力越小，模算效率越高。
	智能化程度	通过算力系统自动化水平和智能优化能力来评估。
碳效水平	全生命周期低碳管理	通过算力系统碳足迹和全生命周期绿色发展战略来综合评估。
	算力碳效	IT 设备使用周期内产生的碳排放与所提供的算力性能的比值。
可获得水平	普适能力	根据内部业务支持、区域业务支持、全国范围内业务支持能力来分级评估。
	普惠能力	综合评估算力系统市场价格优势。
可持续水平	技术兼容性	通过模拟实际应用场景，对设备和系统进行兼容性测试。
	供应链完备程度	评估供应链中供应商的数量和质量、合作伙伴数量和质量。
	产业生态开放水平	评估标准化建设程度及开放平台系统建设情况。

（来源：中国信通院）

4.4 评估体系构建意义

“五位一体”高质量算力评估体系规范加速我国算力产业高质量发展。在算力产业由“量”向“质”的关键阶段，算力发展面临大规模、高要求、异构化等多重挑战，如何准确评估算力的质量是算力建设者和使用者同时面临的问题，从算效水平、智效水平、碳效水平、可持续水平、可获得水平五个方面构建完整、准确的高质量算力评估体系，可规范加速算力产业高质量发展。从算效水平上，提升算力资源利用率和实测计算性能；从智效水平上，提升算力人工智能业务支撑能力；从碳效水平上，促进算力全生命周期节能降碳；从可获得水平上，推动算力应用普适普惠；从可持续水平上，促进算力技术、产业、生态良性发展。

“五位一体”高质量算力评估体系为我国算力产业的技术创新与基础设施建设提供指引。该评估体系涵盖算效、智效、碳效、可获得、可持续等算力系统建设运营关键因素，多维度客观评估我国算力质量情况。通过评估标准的建立，可帮助企业用户识别和优化资源配置，提高算力资源的使用效率，并且可激励企业进行技术研发和创新，以满足更高的评估标准，从而推动整个行业的技术进步。在算力相关项目的规划期、建设期、运营期等不同阶段，该评估体系可为算力实现高质量、全生命周期可持续发展提供指导，推动算力产业的标准化进程，为行业的长远发展奠定基础。

4.5 评估体系应用建议

（1）加快配套标准及工具研制，推动评估体系落地实施

评估体系配套标准及工具的研制和使用是评估体系有效实施的关键，保障评估体系落地过程中有标准可依，有数据可查。完备的工具能够支持并辅助被测系统执行标准所规定的测试流程，确保测试实现公正性审核、过程监控、结果收集与管理发布。另外，在高质量算力评估体系使用过程中要健全算力指标数据采集及监测制度，明确数据采集测试的边界、内容、方法和时限，推动评估体系的落地应用。

（2）开展典型应用场景评估测试，并拓展理论技术研究

高质量算力评估体系仍处于建设初期，存在巨大的发展空间，可在人工智能典型应用场景下开展先行先试，以评估结果作为产业高质量发展改进依据。另外，应在服务器等关键部件设计、制造、运行等各环节开展技术研究，丰富不同架构（X86、ARM 等）、不同业务场景下算力算效、算力智效、算力碳效的模型构建和测试分析方法，在理论上为算力产业高质量发展奠基。

（3）评估算力质量相对水平，探索算力高质量发展新模式

高质量算力评估体系将参照业界算力相关评价规范，将算力高质量发展情况进行综合性分级，不同级别代表不同的算力高质量发展程度，直观反映高质量算力先进性。可将评估体系作为人工智能算力基础设施企业实现高水平、可持续发展的指南，并探索算力高质量发展挂钩贷款等新发展模式。如金融机构与政府合作建立算力高质量发展的监测和评估体系，引导算力行业朝着高算效、高智效、高碳效、可获得、可持续的高质量发展方向前进。

（4）引导算力相关方积极参与，扩大评估结果影响力

在相关机构的指导下，以权威第三方机构为主导，联合产业生产者、使用者、集成者、测试者等诸多参与方，组成测试工作组进行专题运作，加强人工智能高质量算力评估体系的应用推广。积极构建评价考核体系和应用结果奖励机制，引导社会资源、人力资源、债权资金、股权资金的持续投入。


DC TECH