

深度智控
DeepCtrls®

CAICT
中国信通院

《面向绿色算力的人工智能应用研究报告》

2026年9月

深度智控 中国信通院云大所

版权声明

本报告的版权归编写单位共同所有。未经许可，任何机构或个人不得以任何形式对本报告的全部或部分内容进行复制、转载、摘编、发行或用于商业用途。

若需引用、转载或使用本报告内容，必须注明来源为《面向绿色算力的人工智能应用研究报告》，且不得对内容进行歪曲或篡改。

违反上述声明者，编写单位有权追究其相关法律责任。

本报告所载内容仅供参考，编写单位对因使用本报告内容而导致的任何直接或间接后果不承担法律责任。

编制说明

本报告由南京深度智控科技有限公司和中国信息通信研究院云计算与数字化研究所牵头撰写，并在撰写过程中得到了多家单位的大力支持，在此特别感谢以下参编单位：

参编单位（排名不分先后）：

富士康工业互联网股份有限公司、北京秦淮数据有限公司、中国联合网络通信有限公司上海市分公司、宁德时代新能源科技股份有限公司

前 言

算力作为数智时代的重要新质生产力，是驱动数字经济提质增效的关键基础。立足“双碳”战略全局，我国围绕“东数西算”、全国一体化算力网建设以及人工智能与能源双向赋能持续完善政策部署，算力基础设施加快向绿色、高效、可靠、智能方向演进。如何统筹算力供给能力、能源利用效率和低碳发展要求，已成为推动人工智能规模化应用和算力产业高质量发展的重要课题。

基于此，本报告立足人工智能赋能绿色算力的核心需求，研究构建以物理 AI 为核心的技术体系，围绕 AI 能效优化、算电动态协同、预测性运维、液冷智控等关键领域，挖掘人工智能在算力全生命周期的应用价值。报告从技术发展和工程实际出发，贯穿规划设计、运行调控和运维保障等环节，系统梳理了人工智能应用的关键技术，并结合典型项目实践，探索相关技术由研究验证走向工程应用的可行路径。

本报告旨在为行业主管部门、算力中心建设运营单位、技术服务机构及相关研究人员提供参考，促进政策、技术、工程和应用之间的有效衔接。期待产业各方进一步加强跨专业、跨系统和跨主体协作，推动人工智能与算力基础设施深度融合。

绿色算力与人工智能相关技术与产业仍处于快速演进与迭代之中，报告中的认识与实践难免存在局限性，敬请各界专家和读者批评指正。

目 录

一、 绪论	1
(一) 算力中心行业发展概况	1
(二) 绿色算力发展背景与现状	3
(三) 行业痛点与需求分析	9
二、 面向绿色算力的 AI 应用技术体系	13
(一) 方法层：面向绿色算力的 AI 技术底座	14
(二) 能力层：AI 实现赋能的关键技术模块	15
(三) 业务层：基于全生命周期的场景应用	16
三、 面向绿色算力的 AI 数字孪生设计	18
(一) 基于物理 AI 的高精度建模	18
(二) 构建数字孪生设计仿真平台	21
(三) 数字孪生的验证与反馈闭环	26
四、 面向绿色算力的 AI 能效优化控制	28
(一) 一体化智慧监控	28
(二) AI 节能优化控制	30
(三) 边缘计算与端侧 AI	33
(四) AI 控制的可靠性、鲁棒性与韧性	37
五、 面向绿色算力的 AI 算电协同	40

(一) 算电协同综合调度	40
(二) 算冷协同优化控制	43
(三) 储能系统 AI-EMS	45
六、 面向绿色算力的 AI 智慧运维	48
(一) 健康管理 with 预测性维护	48
(二) 故障诊断与根因定位	50
(三) 自动化闭环运维	52
(四) 专属知识库与运维助手	54
七、 案例实践	57
(一) 超算中心系统级协同寻优实践	57
(二) 水冷冷站系统节能及 PUE 诊断实践	58
(三) 风冷精密空调末端按需供冷实践	60
(四) 间接蒸发制冷数据中心智慧运维实践	62
(五) 风液协同数据中心分层寻优实践	63
(六) 高效机房仿真设计优化实践	65
八、 前沿展望：未来技术趋势与演进路径	67

图目录

图 1 我国算力中心机架规模图	1
图 2 全国算力中心用电量占比	2
图 3 我国在用算力中心 PUE 情况	3
图 4 数据中心按地区划分的电力消耗（2020-2030 年）	8
图 5 绿色算力 AI“方法-能力-业务”总体架构	14
图 6 机理框架与数据驱动融合建模	20
图 7 AI 数字孪生设计流程	25
图 8 AI 数字孪生设计平台功能示例	26
图 9 数字孪生设计的验证与反馈	27
图 10 监控体系架构演进图	29
图 11 数据中心制冷控制能力演进	30
图 12 AI 节能闭环控制原理	33
图 13 云边协同三层架构	34
图 14 端侧 AI 应用：CDU 快速响应与节能控制	35
图 15 算电协同 AI 调度体系图	42
图 16 冷负荷预测与冷却系统控制逻辑	44
图 17 面向算电冷协同的水蓄冷系统协同调度	45
图 18 园区级“源网荷储”协同优化平台架构图	46
图 19 设备健康度评价示例	49
图 20 自动化闭环运维体系	54

图 21 算力中心专属知识库	55
图 22 冷机能效模型训练	57
图 23PUE 诊断界面展示	59
图 24 风冷机房温度场调试前后对比	61
图 25 FMA 失效模式库	63
图 26 风液协同数据中心分层寻优架构	65
图 27 系统模拟过程运行策略	66

表 目 录

表 1 绿色算力相关政策情况 5

表 2 不同建模技术路径对比 20

表 3 传统仿真平台与数字孪生仿真平台技术特征对比 24

表 4 节能寻优的主要对象、变量与目标 31

表 5 端侧 AI 部署不同硬件形态对比 36

表 6 AI 控制系统安全保障措施 37

表 7 项目典型问题与工程响应 60

一、绪论

（一）算力中心行业发展概况

我国算力基础设施建设持续推进，算力供给能力和资源规模不断提升。截至 2026 年 3 月底，我国在用算力中心机架总规模达 1445 万标准机架，截至 2026 年 6 月底智能算力规模达到 2185EFlops(FP16)，我国已建成万卡以上智算设施 52 个。算力基础设施规模和水平持续提升。随着人工智能、大模型训练推理和行业智能化应用加快发展，算力供给结构正在从以通用算力为主，逐步向智能算力快速增长、多元算力协同供给转变。算力中心也不再仅以单体数据中心形式存在，而是逐步嵌入全国一体化算力网、算力调度平台和区域算力集群体系之中，成为支撑算力资源统筹供给和跨区域协同的重要基础设施。同时，智算中心建设加快推进，算力中心建设形态由通用化、低密度部署向智算化、高密度部署演进，在提升算力供给能力的同时，也对供电容量、制冷能力、网络互联、资源调度和运行管理提出更高要求。



图 1 我国算力中心机架规模图

算力中心用电需求随行业规模扩大同步增长，能源保障和运行成本正在成为影响行业发展的重要因素。2019 年至 2024 年，我国算力中心用电量由 824 亿千瓦时增长至 1660 亿千瓦时，过去五年年均复合增长率约为 15%。据中国信通院数据，2025 年我国算力中心用电量为 1960 亿千瓦时，同比增长 300 亿千瓦时，增速为 18.1%，同期全社会用电量增速仅 5%，算力中心用电增速仍高于全社会平均水平。从发展趋势看，智算中心高功率密度部署和人工智能负载增长，将进一步提高算力中心对连续稳定供电、容量弹性配置和用能成本控制的要求。算力中心作为高可靠、连续运行的数字基础设施，其建设布局和运行效率不仅取决于算力资源规模，也与电力供应能力、绿色电力获取能力、能源利用效率和算电协同水平密切相关。



图 2 全国算力中心用电量占比

我国算力中心能效水平持续提升，但传统设施节能的边际空间正在收窄。根据中国信通院数据测算，我国算力中心 PUE 由 2019 年的 1.65 降至 2024 年的 1.46，但 2024 年较上年仅下降 0.02，PUE 降幅总体趋缓，反映出传统设施能效优化空间已趋于收窄。在高密度算力部

署、人工智能动态负载、液冷与风液混合制冷、绿电接入和碳排放管理等因素共同作用下，算力中心能效表现不再仅取决于供配电、制冷等单一系统效率，而是受到 IT 负载变化、冷却策略、能源供给、运行控制和运维管理等多种因素影响。面向绿色算力发展要求，算力中心需要在规划设计、运行调优、能源协同和运维管理等环节引入数字孪生、人工智能算法、智能控制和自动化运维等能力，推动绿色低碳发展从单点节能向系统优化、从静态管理向动态调优、从经验运维向智能决策转变。

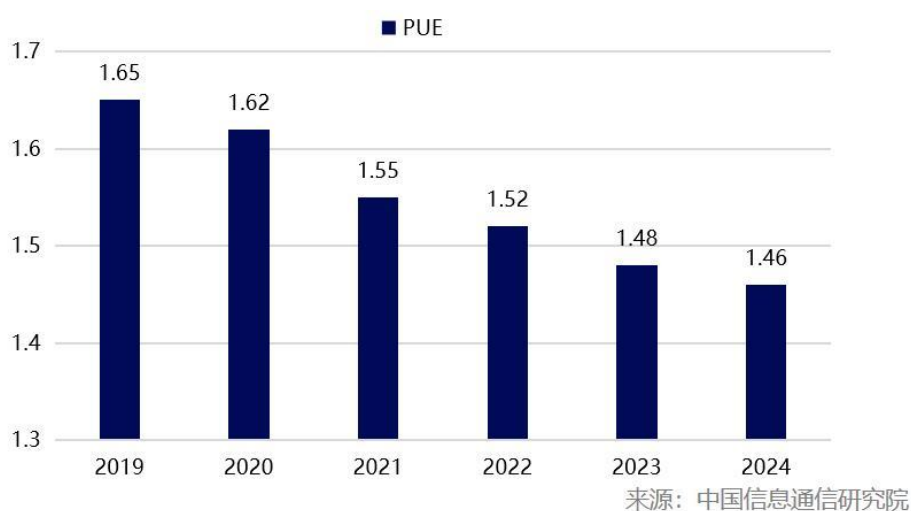


图 3 我国在用算力中心 PUE 情况

（二）绿色算力发展背景与现状

1. 绿色算力发展背景与政策环境

近几年，我国绿色算力政策重点已由数据中心节能降耗，逐步转向绿色低碳算力基础设施建设和算力电力协同发展。2023 年以来，《算力基础设施高质量发展行动计划》《关于深入实施“东数西算”工程 加快构建全国一体化算力网的实施意见》等政策相继提出促进绿色低碳算力发展、优化算力设施电能利用效率和碳利用效率、推动算力与绿色电力一体化融合，并明确国家枢纽节点新建数据中心绿电占

比要求。2024 年，《数据中心绿色低碳发展专项行动计划》进一步围绕数据中心布局优化、上架率提升、PUE 约束、可再生能源利用、单位算力能效和碳效提升等提出目标任务；同年，《加快构建新型电力系统行动方案（2024—2027 年）》提出实施一批算力与电力协同项目，统筹数据中心发展需求和新能源资源禀赋，科学整合源荷储资源，探索新能源就近供电、聚合交易、就地消纳的“绿电聚合供应”模式，提升算力与电力协同运行水平。2025 年，《关于促进可再生能源绿色电力证书市场高质量发展的意见》提出逐步提高绿色电力消费比例并使用绿证核算，明确国家枢纽节点新建数据中心绿色电力消费比例在 80%基础上进一步提升；《关于深入实施“人工智能+”行动的意见》提出优化国家智算资源布局，完善全国一体化算力网，加大数、算、电、网等资源协同，推动智能算力供给普惠易用、经济高效、绿色安全。上述政策进一步强化了绿色电力消费、绿证核算和智能算力协同供给要求。

进入 2026 年，国家层面进一步将人工智能、智算集群和算电协同纳入新型基础设施建设重点。《政府工作报告》提出，深化拓展“人工智能+”，推动重点行业领域人工智能商业化规模化应用，实施超大规模智算集群、算电协同等新基建工程，加强全国一体化算力监测调度。在此基础上，国家发展改革委、国家能源局、工业和信息化部、国家数据局印发《关于促进人工智能与能源双向赋能的行动方案》，提出强化能源对人工智能发展的基础支撑作用，发挥人工智能对能源转型的叠加倍增作用，促进能源、算力、场景、数据、模型等要素高效协同。围绕算力设施建设运行，该行动方案进一步提出提升算力设施多元电力供给能力，推进算力设施绿电消费占比统计和碳排放核算，

持续提升算力设施能效碳效，促进算力电力高效经济协同。由此看，绿色算力政策正在从数据中心绿色化和算力基础设施低碳化，进一步向人工智能、算力、电力、能源和碳排放协同管理方向深化。

表 1 绿色算力相关政策情况

发布时间	政策文件名称	具体内容
2023 年 10 月	《算力基础设施高质量发展行动计划》	提出完善算力综合供给体系、提升算力高效运载能力、强化存力高效灵活保障、深化算力赋能行业应用、促进绿色低碳算力发展、加强安全保障能力建设等重点任务；到 2025 年，算力规模超过 300EFLOPS，智能算力占比达到 35%。
2023 年 12 月	《关于深入实施“东数西算”工程 加快构建全国一体化算力网的实施意见》	提出统筹通用算力、智能算力、超级算力一体化布局，推动东中西部算力一体化协同、算力与绿色电力一体化融合；到 2025 年底，算力电力双向协同机制初步形成，国家枢纽节点新建数据中心绿电占比超过 80%。
2024 年 7 月	《数据中心绿色低碳发展专项行动计划》	提出到 2025 年底，全国数据中心布局更加合理，整体上架率不低于 60%，平均电能利用效率降至 1.5 以下，可再生能源利用率年均增长 10%，平均单位算力能效和碳效显著提高；到 2030 年底，全国数据中心平均 PUE、单位算力能效和碳效达到国际先进水平。

2024 年 8 月	《加快构建新型电力系统行动方案（2024—2027 年）》	提出实施一批算力与电力协同项目，统筹数据中心发展需求和新能源资源禀赋，科学整合源荷储资源，开展算力、电力基础设施协同规划布局，探索新能源就近供电、聚合交易、就地消纳的“绿电聚合供应”模式。
2025 年 3 月	《关于促进可再生能源绿色电力证书市场高质量发展的意见》	提出依法稳步推进绿证强制消费，逐步提高绿色电力消费比例并使用绿证核算；加快提升数据中心等重点用能单位和行业的绿色电力消费比例，国家枢纽节点新建数据中心绿色电力消费比例在 80%基础上进一步提升。
2025 年 8 月	《深入实施“人工智能+”行动的意见》	提出优化国家智算资源布局，完善全国一体化算力网，加大数、算、电、网等资源协同，加强智能算力互联互通和供需匹配，推动智能算力供给普惠易用、经济高效、绿色安全。
2026 年 3 月	《政府工作报告》	提出深化拓展“人工智能+”，实施超大规模智算集群、算电协同等新基建工程，加强全国一体化算力监测调度，支持公共云发展。
2026 年 4 月	《关于促进人工智能与能源双向赋能的行动方案》	提出促进能源、算力、场景、数据、模型等要素高效协同，强化能源对人工智能发展的基础支撑作用，发挥人工智能对能源转型的叠加倍增作用，加快构建协同高效、安全可靠、绿色低碳、开放融合

		的“人工智能+”能源发展新格局。
--	--	------------------

2.国内外数据中心绿色算力产业现状

全球数据中心产业正在呈现算力需求快速增长、用电规模持续上升、能效缓解压力加大的发展态势。国际能源署研究指出，2024 年全球数据中心用电量约为 415TWh，占全球用电量约 1.5%，自 2017 年以来，全球数据中心用电量年均增长约 12%；在人工智能和数字服务需求增长带动下，全球数据中心电力消费预计到 2030 年将翻番至约 945TWh，占全球用电量比例接近 3%。从区域分布看，美国、中国和欧洲仍将是全球数据中心电力需求最主要区域，其中中国和美国预计贡献 2030 年前全球数据中心用电增长的近 80%，反映出数据中心用电增长具有较强的区域集聚特征。大型数据中心和近年来新建数据中心的 PUE 水平相对更优，但大量存量数据中心、区域气候条件和传统制冷系统约束仍限制全球平均 PUE 进一步快速下降。国际数据中心绿色算力产业的关注重点正在从单一 PUE 优化，逐步转向清洁能源供给、高效冷却、液冷适配、水资源管理、余热利用、碳排放核算和客户侧可持续发展披露等多维度能力建设。

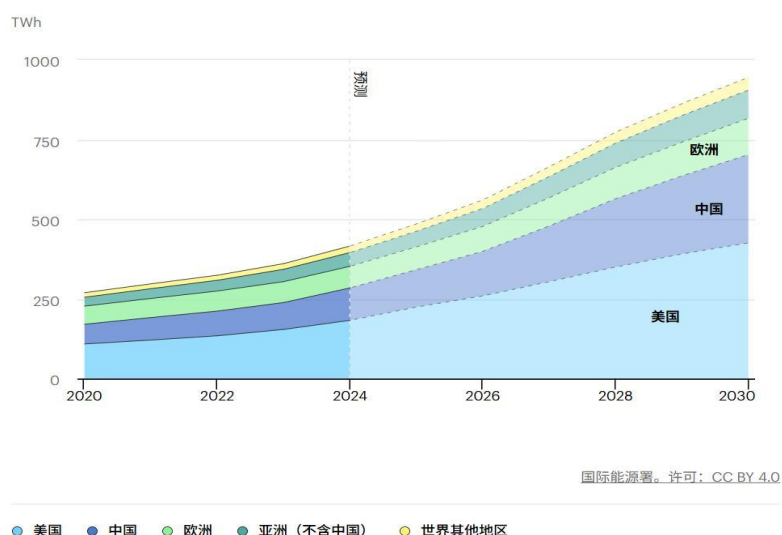


图 4 数据中心按地区划分的电力消耗（2020—2030 年）

国际头部云服务商、互联网企业和数据中心运营商正从能源供给、冷却方式、资源利用和客户服务等方面完善绿色算力能力。头部云服务商 Google 正从能源供给与水资源管理两方面降低算力基础设施的运行影响，持续推进 24/7 无碳能源目标，2024 年其全球数据中心及办公园区 24/7 无碳能源比例提升至 66%，并在 20 个拥有并运营数据中心的电网区域中有 9 个区域实现至少 80% 的无碳能源供给，同时通过清洁能源采购和水资源补偿等方式降低数据中心运行影响。互联网企业 Meta 正通过扩大风电、光伏等清洁能源项目，为算力业务增长提供绿色能源支撑，截至其 2025 年可持续发展披露，Meta 支持的风电和太阳能项目正在为全球电网新增近 29GW 清洁和可再生能源。数据中心运营商 Equinix 正通过提高可再生能源覆盖率、改善设施能效和拓展绿色服务，推进数据中心低碳运营，2025 年其全球零售 IBX 数据中心可再生能源覆盖率达到 96%，平均年度 PUE 为 1.37，并通过绿色债券、绿色电力报告和余热输出项目支持客户低碳运营。国际数据中心绿色算力实践正在从早期碳中和承诺和可再生能源采购，进

一步转向面向 AI 算力增长的能源供给、冷却系统、水资源利用和客户碳管理等综合能力建设。

国内运营商、互联网企业和第三方数据中心服务商正围绕基础设施节能、绿色能源利用、高效冷却和智能调优等方向拓展绿色算力实践。运营商方面正通过绿色数据中心建设和专项节能行动，持续提升大规模算力基础设施的能源利用效率。中国移动持续推进绿色基站和绿色数据中心建设，2025 年开展“绿智无线”“绿智算力”等专项行动，全年节电 150 亿度，大型、超大型数据中心总体 PUE 值降至 1.285。互联网企业正将 AI 能效调优、间接蒸发冷却、液冷部署和绿色电力利用协同应用于数据中心运行。腾讯在 2024 年将 AI 节能算法模型应用于租赁数据中心，并扩大间接蒸发冷空调部署，全年节省用电约 10825.0 兆瓦时，同时针对 GPU 高功率密度需求完成高效液冷方案试点部署；在绿色电力方面，腾讯 2024 年采购绿色电力 1364625.9 兆瓦时，数据中心可再生能源设施总装机容量累计达到 63.8 兆瓦，可再生能源占比提升至 22.0%。第三方数据中心服务商正通过提高可再生能源使用率、持续改善 PUE 和扩大绿色认证覆盖，推进数据中心规模化低碳运营。万国数据 2024 年可再生能源使用率达到 40%，42 个数据中心获得绿色数据中心认证，平均 PUE 由 2023 年的 1.28 改善至 2024 年的 1.24。整体来看，国内绿色算力企业实践正在从传统节能改造和 PUE 优化，向绿电采购、液冷部署、AI 调优、可再生能源利用和全生命周期低碳运营等综合路径拓展。

（三）行业痛点与需求分析

在规划建设环节，智算中心项目落地与实际需求之间仍存在一定脱节，突出表现为供电制冷适配不足、算力架构支撑乏力和异构资源

协同不畅。一方面，智算中心业务变化快、客户需求迭代快，部分已建成智算中心在建设初期主要依据当时的负载规模、设备形态和功率密度进行规划，后续面对企业对高密度、高性能算力部署的需求时，原有供电制冷架构难以灵活适配，容易出现供电容量、散热能力、管路空间、机柜承载和系统冗余等方面的约束。另一方面，随着大模型训练推理、全精度算力和多场景智能应用发展，部分低精度、单一化的算力架构难以支撑未来全精度、全场景、超大规模算力需求。同时，国产与进口异构算力资源分散部署于不同中心，生态体系相对封闭，技术栈、调度方式和软件适配存在壁垒，也进一步加大了算力资源统一管理 and 高效利用难度。规划建设环节需强化弹性设计、供电制冷预留、算力架构兼容和异构资源协同能力，提升智算中心对高密度、高性能和多元化算力需求的适配能力。

在运行优化环节，智算中心面临动态工况下能效调节难度提升的问题。与传统数据中心相比，智算中心业务负载波动更明显，训练、推理、存储和网络通信等任务对供电、制冷和环境控制的影响更加复杂，运行工况不再是相对稳定的单一状态。部分项目仍主要依赖固定阈值、人工经验和分系统控制方式开展运行调节，供配电、制冷、IT负载和动环监控之间缺少统一联动，容易出现局部系统已优化、整体能效未最优的问题。例如，在高密机柜、液冷系统、CDU、冷源和末端设备协同运行场景下，若缺少对负载变化、进出水温、流量、压差、环境参数和设备状态的连续分析，运行策略就难以及时匹配实际需求，可能造成过度制冷、局部热点、设备低效运行或能耗波动。运行优化的核心需求不是简单增加监测点位，而是建立面向动态工况的能效调节与闭环控制能力，通过数据采集、状态识别、策略优化和执行反馈，

实现 IT 负载、制冷系统和供配电系统之间的协同调节，推动算力中心从“被动监控”向“主动预测”转变。

在能源协同环节，算力中心用电需求增长使电力供给、绿电消纳和负荷调节等问题更加集中，算电协同压力进一步凸显。算力中心对电力具有充足可靠、安全稳定、经济节约、绿色低碳等多重诉求，尤其是智算中心负载集聚度高、连续运行要求强，对局部电网容量、供电可靠性和用能成本控制提出更高要求。一方面，算力需求和用电负荷呈现集聚化发展趋势，大型、超大型算力中心向重点区域集中，局部地区电网扩容和供电保障压力有所加大；另一方面，算力中心对绿色电力需求持续提升，绿电直供、源网荷储一体化、绿电绿证交易等模式仍需结合区域资源条件、电力市场机制和工程建设条件进一步推进。此外，算力中心作为兼具时间和空间调节潜力的特殊柔性负荷，可通过算力负载调度调整任务执行时间和地点，但算力调度平台、负载用电特性预测、异构算力调度机制以及电力与算力联动价格激励机制仍需完善。UPS、蓄冷、备用电源、余热利用等资源也具备一定调节潜力，后续需进一步纳入算力中心与电力系统协同运行体系。

在运维管理环节，智算中心系统复杂度持续提升，对运维管理的数字化、智能化和协同化能力提出更高要求。与传统数据中心相比，智算中心运行对象更加多元，既包括服务器、网络、存储等 IT 资源，也包括供配电、制冷、液冷、动环监控、安防和运维流程等基础设施系统，任何单一系统异常都可能影响整体算力服务连续性。当前部分项目仍以人工巡检、经验判断、分系统告警和事后处置为主，存在告警信息分散、故障根因定位慢、跨系统联动不足、运维知识难以沉淀等问题，难以适应高密度、高可靠、快速响应的运维要求。尤其在液

冷、异构算力等新技术加快应用的背景下，运维对象和故障模式更加复杂，传统依赖人工经验的方式容易出现响应不及时、处置不一致和风险预判不足等问题。运维转型的核心需求，是推动运维管理从“人工巡检、被动响应、单点处置”向“状态感知、智能分析、预测预警”转变，通过 AI 算法、知识库、根因分析和自动化流程等能力，提升故障识别、风险预判、策略推荐和闭环处置水平，支撑智算中心长期稳定、高效、安全运行。

二、面向绿色算力的 AI 应用技术体系

人工智能技术经历了由知识规则驱动向数据学习驱动、再向通用模型与融合智能演进的发展过程。随着机器学习和深度学习的发展，人工智能逐步具备从大数据中提取规律、识别状态和预测趋势的能力，呈现出数据、知识、机理与模型融合的趋势。同时，人工智能为绿色算力转型提供新路径，推动运维从依赖人工经验的被动响应向智能感知、分析决策与自主优化转变，由单点优化向跨系统协同优化转变。然而，人工智能对绿色算力的赋能并非若干算法或应用功能的简单叠加，而需要形成从底层方法、共性能力到业务场景的系统性技术体系，使不同人工智能方法能够转化为可复用的建模、预测、诊断、决策和控制能力，并进一步贯通设计优化、运行调控、能源协同和智能运维全过程。围绕这一要求，绿色算力 AI 技术体系由以下三个层面展开。

方法层：以物理 AI 为核心，整合知识表示与推理、机器学习、强化学习、生成式人工智能等主要方法。在面向绿色算力的场景下，需结合算力中心设备机理、系统拓扑和运行约束，实现 AI 技术的适配。

能力层：通用人工智能方法需要经过数据治理、对象建模、模型验证和系统集成，才能转化为绿色算力所需的专业能力。在这一过程中，逐步形成系统建模、预测分析、异常诊断、决策控制和知识服务等能力，为具体业务提供可验证、可执行的技术支撑。

业务层：是 AI 能力向算力中心全生命周期价值转化的落地界面。将前两层所聚合的能力，嵌入到从规划设计、运行调控到运维管理的完整决策链条，形成“场景定义问题、目标驱动优化、闭环验证效果”的持续改善机制，推动算力中心建设运营从依赖经验判断的离散行为，演进为数据驱动、模型支撑的系统化工程实践。

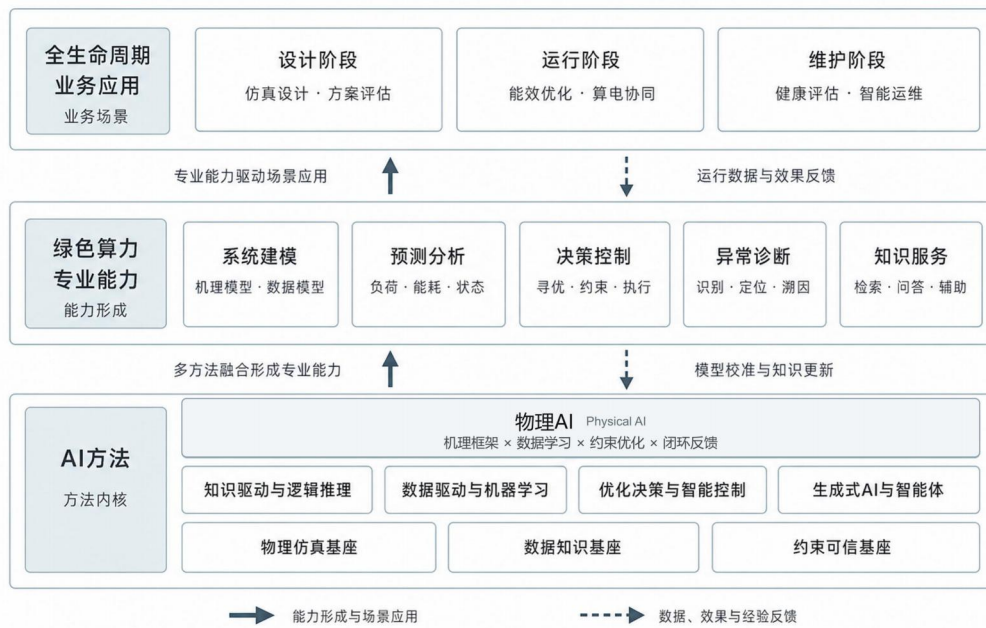


图 5 绿色算力 AI “方法—能力—业务” 总体架构

（一）方法层：面向绿色算力的 AI 技术底座

方法层以物理 AI 为核心，融合知识表示与推理、机器学习、强化学习与优化控制、生成式人工智能等方法，共同支撑智能体在物理世界完成“感知—推理—规划—行动”闭环。物理 AI 是面向真实物理系统，能够理解物理规律、识别系统状态，并在动态真实环境下进行推理、决策与闭环控制的人工智能形态。与传统的人工智能方法相比，物理 AI 具有机理数据融合、物理一致性、数据高效性、工况适应性、可解释可信和闭环安全可控等特点，有助于在数据有限、状态

不完全可观测、工况动态变化和多重运行约束条件下,提高系统建模、状态预测与运行决策的可靠性。在绿色算力场景中,计算、供配电、制冷等多个物理系统耦合构成,能量流、热流与信息流相互作用,运行状态受业务负载、环境条件和安全边界等多重因素影响。面向这类多系统耦合、强约束和动态不确定的运行环境,物理 AI 将物理机理、系统状态与运行约束融入数据学习、状态推演和决策控制过程,为智能体理解系统运行规律、判断状态演化趋势并形成可执行的控制策略提供基础。在物理 AI 核心框架下,知识表示与推理将设备机理、运行规则、约束条件和专家经验转化为可计算的规则和关系;机器学习从多源运行数据中识别状态、发现规律并预测系统演化;强化学习与优化控制面向能效、可靠性等目标,在设备能力和安全边界等约束下形成运行策略并实施闭环调节;生成式人工智能可增强智能体的知识获取、任务理解和人机交互能力。上述方法围绕物理 AI 形成能力互补,共同构成面向绿色算力人工智能应用的技术底座。

(二) 能力层: AI 实现赋能的关键技术模块

能力层主要通过系统建模、预测分析、决策控制、异常诊断和知识服务等关键技术模块实现。系统建模是数字化能力构建的基础,可以系统描述冷却、供配电和储能等设备在不同工况下的性能,以及算力负荷、IT 功耗、热负荷、冷量需求与能源供给之间的耦合关系,并为后续仿真、预测和控制提供数字底座。预测分析根据历史数据、实时状态和外部条件,基于模型仿真判断系统未来状态或变化趋势,主

要预测 IT 负荷、热负荷、冷量和能源需求、新能源出力、储能状态及设备性能变化，为设备运行、能源调度和维护计划提供依据。**决策控制**根据系统状态、预测结果、运行目标和约束条件，基于模型预测控制和强化学习等技术实现设备设定值与调度策略的主动寻优。**异常诊断**从测点数据和告警中识别偏离正常模式的状态，再结合设备拓扑、运行规则和历史案例判断异常位置、影响范围和可能原因。**知识服务**将设备资料、运行规程、技术标准、告警记录、工单和故障案例整理成可检索、可关联的知识资源，并结合知识图谱、通用大模型和检索增强生成开展查询、归纳和问答。

（三） 业务层：基于全生命周期的场景应用

业务层贯穿仿真设计、能效控制、算电调度和智能运维。**数字孪生仿真设计**通过构建设备级数字模型，结合全年逐时仿真、典型工况分析和多方案比选，深度分析冷源形式、设备选型、系统容量、运行模式及控制参数等因素对系统效能的影响，从而实现精细化设计。**能效闭环优化控制**基于一体化监控与高精度性能模型，融合负荷预测、系统仿真和全局寻优算法，对冷源、输配及末端系统的设备组合、运行参数和控制设定值进行动态优化。在满足热安全、设备能力和运行稳定性等约束的前提下，将优化策略下发至现场执行，并根据实际能耗、环境参数和设备响应持续评估控制效果、修正模型与策略，形成“监测—分析—寻优—执行—反馈”的闭环调控过程。**算电协同综合调度**基于 IT 负荷、电价和新能源出力等预测分析，统筹算力任务执行

时序、制冷系统运行方式以及能源系统调控策略，协调“算、冷、电、储”之间的供需关系；在满足计算任务时延、热安全和业务连续性等约束的前提下，引导可转移算力负荷向绿电充裕或电价较低时段迁移，并利用储能、蓄冷等资源削峰填谷，提高新能源就地消纳和能源系统调节能力。**智能诊断与运维闭环**通过设备分级管理、健康评估和性能退化预测从运行数据的变化中识别设备风险，并据此安排巡检、检修和备件准备，实现预测性维护。异常发生后，基于 FMEA 失效模式库还原故障链路、定位根因，诊断结果进入运维工单和处置流程，构建运维自愈体系，实现运维自动化闭环。

三、面向绿色算力的 AI 数字孪生设计

（一）基于物理 AI 的高精度建模

算力中心能源系统的 PUE 设计预测，往往与投运后的实际数据存在偏差。仿真结果能否贴近真实运行状态，很大程度上取决于底层模型对设备性能和系统耦合关系的刻画精度。全年逐时动态仿真、控制策略的优化迭代，都以可靠的系统模型为前提。回顾建模技术的发展，可以看到行业一直在计算效率、工程安全和模型可信度之间寻找最优的解决方案。

纯机理仿真阶段。基于热力学、流体力学和电工学的机理模型在专业领域有着较早的应用。这类模型通过非线性代数和微分方程组描述系统运行状态，物理含义清楚，可解释性强。以冷水机组为例，可通过描述制冷剂在蒸发器和冷凝器中的相变、换热和流动过程来构建相应性能模型。纯机理模型也存在明显的短板：首先，求解过程复杂，非线性方程组需要离散化和迭代计算，复杂场景的仿真可能需要数十分钟乃至数小时。其次，参数标定成本高。模型涉及物性参数、传热系数、阻力系数等大量变量，通常要依赖实验数据或厂家性能曲线。不同品牌、不同型号设备差异较大，工程落地时很难快速复用。

纯数据驱动阶段。近年来，AI 模型开始用于算力中心能源系统的能效评估和负荷预测。常见方法包括神经网络、支持向量机、随机森林等，主要依靠历史运行数据学习输入变量与输出结果之间的关系。相比复杂机理仿真，数据驱动模型的在线推理速度通常更快，适合实

时监测和快速预估。但纯数据驱动模型也存在明显局限，它缺少物理约束，本质上仍是从历史数据中拟合统计关系。当运行工况超出训练样本范围时，预测结果可能出现明显偏差，甚至给出违背热力学规律的结果。此外，它还高度依赖数据覆盖度，需要大量高质量、全工况数据训练；一旦遇到设备老化、工况偏移或系统改造，模型稳定性会下降。因此，这类模型不适用于数据中心关键设备系统的闭环控制。

物理 AI 融合建模。基于物理信息神经网络（PINN）的核心理念，其基本思路不是用 AI 替代物理模型，而是把物理规律作为约束引入深度学习过程，让模型既能学习非线性特征，又不脱离工程边界。融合建模保留了数据模型处理复杂输入输出关系的能力，同时将热力学定律、传热传质模型等物理约束嵌入训练过程，使输出结果符合工程常识。与单纯机理或单纯数据模型相比，这一技术路线主要改善以下三个方面。

一是提升可解释性，物理定律的引入赋予了模型明确的物理意义。

在冷水机组建模时，模型不仅要拟合历史数据，还要满足相关物理规律，如运行能效与工作温度的关系必须符合热力学原理。这种约束消除了传统 AI 数据驱动的“黑箱”隐患，使设计人员能够理解模型的决策依据，增强了对模型输出结果的信任度。**二是降低数据依赖，机理框架的支撑减少了对单一项目海量历史运行数据的依赖。**在数据稀缺的工况下，物理约束能够起到“兜底”作用，保证模型输出的合理性。这意味着新建设的算力中心即使缺乏长期运行数据，也可以基于

设备铭牌参数和物理原理建立可靠的模型，大幅缩短了模型开发周期。

三是增强可迁移性，标准化的物理边界支持模型在不同场景、多品牌设备间的快速配置与参数校准。通过建立统一的设备类型编码、参数模板和模型接口，可以实现模型的模块化复用。当更换设备品牌或调整系统配置时，只需更新模型的参数权重系数，无需更新模型本体。

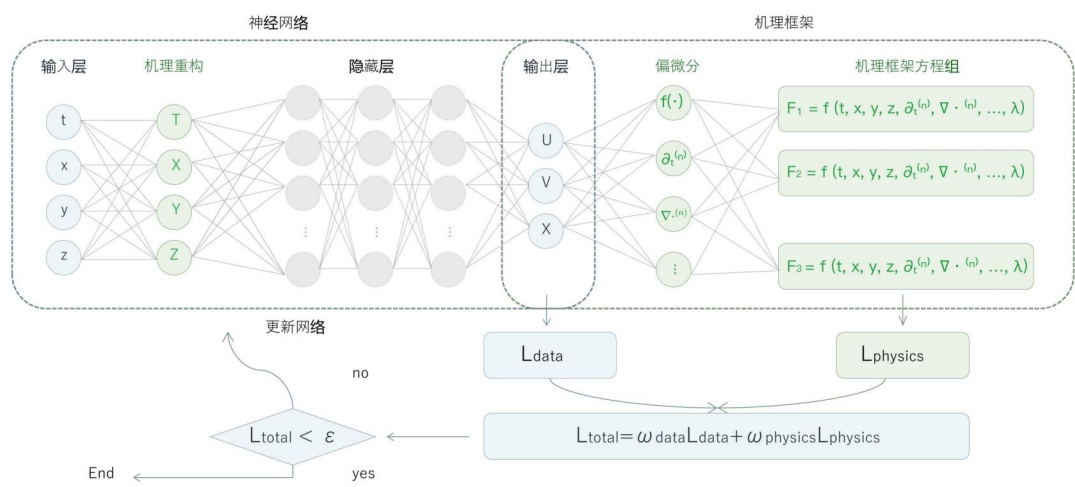


图 6 机理框架与数据驱动融合建模

表 2 不同建模技术路径对比

分析维度	纯机理模型	纯数据驱动模型	物理 AI 模型
核心原理	基于热力学、流体力学、传热学等物理原理构建偏微分方程组进行数值求解	基于机器学习算法,通过历史运行数据拟合输入输出映射关系	融合物理机理框架与数据驱动方法进行建模
物理可解释性	高,机理透明,完全符合物理规律	低,缺乏物理约束,存在输出违背物理规律的安全隐患	较高,受物理边界约束,输出结果符合工程常识
数据依赖程度	低,主要依赖设备基础参数与	高,需要海量、全工况覆盖的	中等,物理约束降低了对纯数据

	物理常数	历史运行数据进行训练	量的依赖
计算效率	低,数值求解复杂,单次仿真 耗时通常要数小时	高,模型训练完成后可在毫秒 级完成预测	较高,通过降阶模型可将计算压 缩到秒级
跨场景迁移能力	弱,设备更迭或管路改造时需 重新推导方程与标定参数	弱,模型与特定数据分布强耦 合,工况偏移、设备老化或更 换品牌时,数据分布改变导致 模型泛化能力差、易失效	强,跨场景或更换设备时,只需 少量新数据微调 AI 参数,物理 框架依然适用,实现快速迁移。
工程应用	广泛应用于学术研究的热力/ 流体稳态与瞬态仿真,但受限 于计算耗时,难以用于实时控 制与高频寻优	多用于离线分析与辅助预测, 因缺乏物理机理约束,在核心 工程设计与强安全约束的闭 环控制中应用受限	正从头部企业试点向行业推广 演进;可应用于高密度智算中 心、复杂能源系统的精细化设计 与 AI 实时控制

(二) 构建数字孪生设计仿真平台

高精度建模为算力中心设计提供了数字孪生底座,但模型本身并不能直接回答工程决策问题。算力中心的能效表现取决于全年 8760 小时内负荷、气象与设备状态的持续耦合变化,常规的静态设计校核难以充分覆盖这一动态过程。通过构建数字孪生动态仿真平台,可以将高精度模型置于逐时变化的运行条件中进行连续求解,模拟不同设备组合与控制策略下制冷系统的动态响应与能耗分布,使设计决策建立在真实工况的量化评估之上,从而实现面向动态运行的 AI 智慧设计。

1.传统仿真设计工具的应用局限

算力中心能源系统仿真逐步由经验计算和局部模拟向系统级联合仿真演进。早期设计主要依靠手工计算和电子表格，此后 EnergyPlus、DeST 等建筑能耗模拟工具逐步应用于仿真计算，特别是暖通系统的动态负荷计算，为设备初选提供了更精细的计算依据。近几年，6SigmaDC 等 CFD 工具进一步将仿真延伸至机房热环境层面，对于局部空间的热分布优化提供了有效的仿真手段。随着仿真范围向系统层面拓展，Modelica 等面向对象物理建模语言开始被用于多子系统的联合仿真，尝试在统一框架下描述不同设备与回路之间的交互关系。

传统仿真工具在工况覆盖、控制策略表达和跨系统计算方面仍存在局限。一是相对依赖于典型工况，传统仿真多基于设备铭牌参数，以及标准典型工况进行计算分析。全年大多数部分负荷时段，设备并不在设计工况下运行，仿真结果也难以反映真实的性能变化。二是控制策略表达不足。很多传统工具只能设置固定控制逻辑，难以模拟变频寻优、冷冻水温动态设定等策略对系统能效的影响。设计阶段的仿真结论和后期自控群控逻辑因此容易脱节。自控工程师缺少仿真数据支撑，只能更多依赖经验编写控制程序，建成后再通过现场调试修正。三是跨系统计算能力有限。传统工具往往面向单一水系统或单一风系统。随着机柜功率密度提升，风冷、水冷、液冷混合使用越来越常见，跨介质换热和多能流耦合成为常态。若仍按单系统拼接设计，难以科

学反映系统的协同关系。

2.数字孪生仿真设计技术路径

精细化负荷预测能够为数字孪生仿真设计提供全工况的运行输入。传统负荷预测多聚焦典型工况的峰值负荷，即使采用动态逐时负荷计算，其变量也侧重于反映处理气象和建筑热惰性的影响。然而，算力中心的冷负荷 80%以上来源于 IT 设备散热，围护结构负荷占比极小。此外，算力中心的业务负载具有明显的潮汐效应，特别是智算中心在进行大模型训练时，功耗会呈现剧烈的阶跃式波动。**因此，AI 数字孪生设计的前提，是充分结合建筑特性参数、气象时序和业务负载数据，通过机器学习与算法推演，形成高精度的全年逐时动态负荷预测。**相比单点峰值计算，负荷预测能够更准确地刻画不同季节、不同业务负荷下的能源需求，为设备方案提供更贴近实际的依据，从源头提供系统全工况能效，减少设备投资冗余。

融合控制策略的运行仿真能够在设计阶段验证不同控制方式的能效表现。常规仿真软件通常只能模拟固定控制逻辑，难以反映实际运行中变频、动态重置等策略的影响，而 AI 数字孪生仿真平台需要把设备性能模拟扩展到系统运行策略模拟。平台需要结合项目条件生成节能控制建议，并将策略嵌入设备运行模拟。在仿真环境中，设计人员可以看到不同负载率和气象条件下的关键运行参数，例如冷机出水温度、水泵频率、冷却塔风机转速和阀门开度。**通过设计阶段的策略比选，平台可以量化不同控制方式的能耗差异，例如将冷冻水出水**

温度固定为 7℃的传统策略，与随室外湿球温度和业务负载动态调整的策略进行对比。行业实践表明，这种动态水温策略通常可使全年制冷系统能耗降低约 5%至 10%。这类结果不仅能优化设备选型边界，也能形成自控编程的初始逻辑，减少后期现场调试成本。

表 3 传统仿真平台与数字孪生仿真平台技术特征对比

	传统仿真平台	数字孪生仿真平台
负荷计算	基于典型气象年和静态峰值计算，动态负荷主要考虑天气和围护结构影响	全年 8760 小时逐时动态负荷预测，融合 IT 业务特征与气象时序波动
控制策略模拟	仅支持固定设定点或简单的控制逻辑	支持动态控制策略仿真，可仿真变频寻优，智能群控等控制策略
物理约束	一般基于机理模型	物理 AI 建模，将机理框架作为约束条件，兼顾效率和工程安全

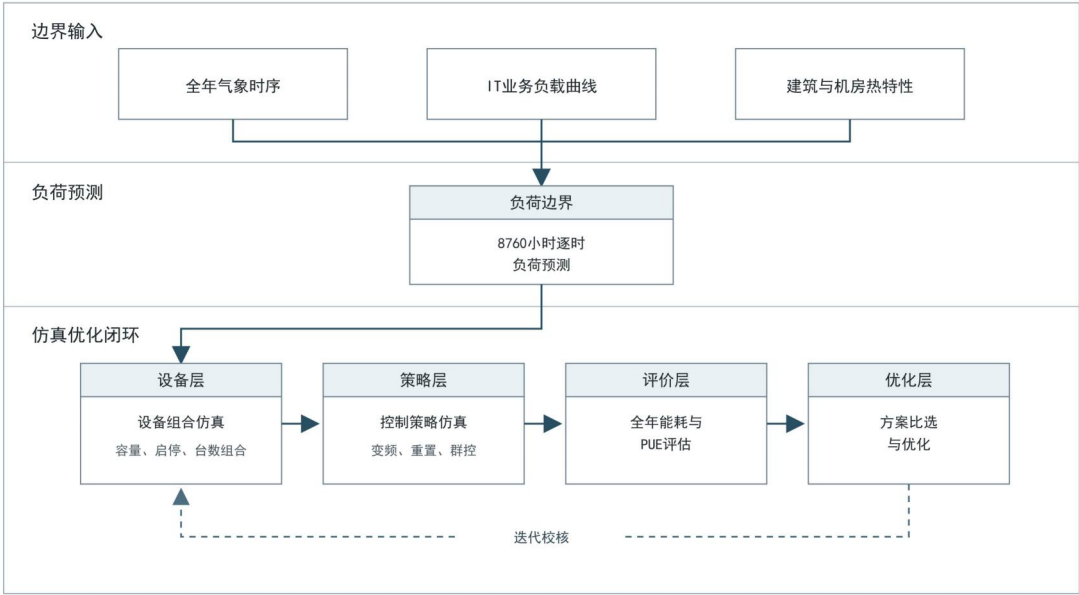


图 7 AI 数字孪生设计流程

数字孪生仿真平台可真正实现面向全工况的系统方案比选、设备选型与运行策略设计。基于仿真平台，设计人员可依据设计模型和项目参数建立仿真对象，选择典型日、全年 8760 小时或自定义时段进行计算，分析设备组合、运行状态和系统能效随负荷变化的情况，并在设备运行范围和系统约束内调整控制参数。各方案采用相同的边界条件和评价指标，提供系统能耗、负载率分布、设备运行小时和启停状态等核心仿真结论，从而全面支撑以运维为导向的精细化设计。例如，在高效制冷机房的设计中，平台可模拟不同冷机容量组合方案的运行效果，通过量化评估各方案在全年部分负荷区间的年运行小时数、负载率分区、设备启停频次及主辅匹配度，评估不同容量配置的全年

能效，为制冷站系统设计的方案组合选择提供依据。



图 8 AI 数字孪生设计平台功能示例

(三) 数字孪生的验证与反馈闭环

运行数据反馈是数字孪生从设计阶段持续延伸至运营期的关键。

设计阶段模型所依据的负荷预测结果、设备样本参数、设计工况与工程经验，本质上都是对未来运行状态的模拟假设。项目进入实际运营后，IT 业务负载的波动、室外气象的变化、设备性能的衰减以及运维策略的差异，都会使实际运行边界偏离这些假设。若缺乏运行数据的校正，模型精度将随时间衰减，数字孪生也会逐渐失去决策参考价值。只有将实际产生的运行数据回灌模型，完成在线校正与假设校核，数字孪生才能与物理系统保持同步，从而为运营期的控制优化与能效诊断提供可信的量化基准。

运行经验回传设计端，可为后续新建与扩容项目提供更可靠的设计依据。运行数据的反馈与验证，是数字孪生仿真平台实现迭代优化的重要依据。

一是实现模型校验。基于运行数据与模型预测值的偏差，对负荷预测模型和设备性能模型逐项修正：负荷侧依据实测冷负荷及其时间分布特征校正预测参数，设备侧依据实测运行点重新标定冷机效率、水泵效率、冷却塔散热特性等性能曲线，提高模型精度。

二是支撑扩容设计。通过既有系统运行后的实测数据，可用于验证修正原先的设计模型，然后将新增容量、接入方式和运行策略放入修正后的设计模型中进行仿真决策。

三是检验控制策略有效性。现场运行数据还可检验冷冻水温度优化、水泵变频、冷机群控、冷却塔调节以及风液协同等策略的实际效果，用于完善新项目的测量与计量点位、设备启停条件、设定值调整范围和工况切换逻辑，使自控设计更贴近实际运行需求。

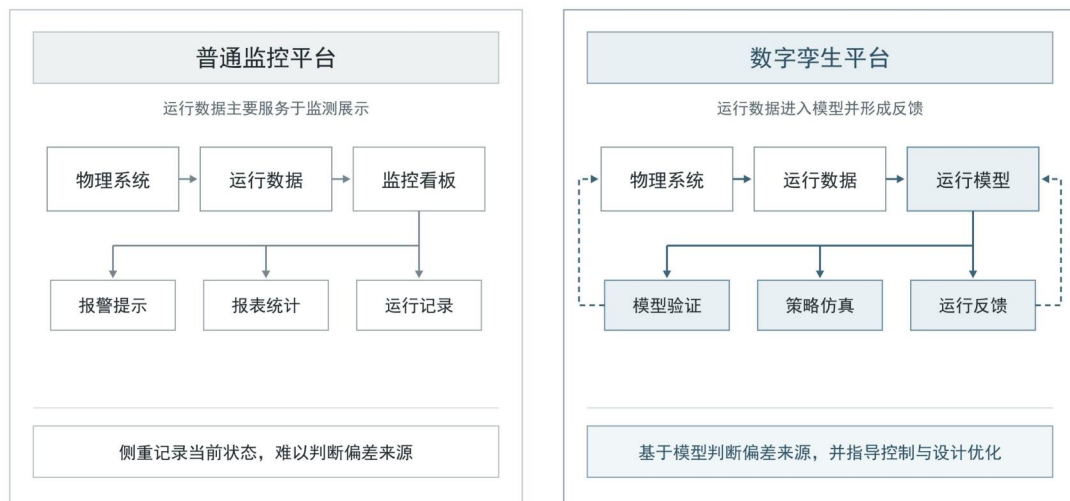


图 9 数字孪生设计的验证与反馈

四、面向绿色算力的 AI 能效优化控制

（一）一体化智慧监控

平台具备一体化智慧监控功能是 AI 能效优化闭环体系的基础支撑。其核心价值在于向下打通异构子系统、向上为智能引擎提供高质量数据供给，围绕全域采集、统一底座、智能分析与标准化集成四大能力模块，逐步构建可观测、可分析、可闭环的智能化运营体系。

传统数据中心监控体系长期分立建设，跨系统数据融合与集中监控能力不足，建设一体化平台已成为行业共识。动力环境、楼宇自控、制冷群控、电力监控、安防消防及 IT 带外管理等系统，通常由不同厂商在不同建设阶段分别承建，彼此在数据格式、通信协议和时间基准上缺乏统一规范。全局能效的评估需跨多套系统人工提取与拼接数据，故障发生时各系统独立推送告警，告警之间的因果关联难以自动辨识。一体化平台关键在于采集、存储、告警、呈现各层的全面统一，实现跨系统数据融合与集中监控。部分头部数据中心运营方已在新一代运营平台中率先实践了这一方向，将供电、暖通、安防等子系统统一接入，故障响应效率获得显著提升。在此基础上，一体化平台还需为上层 AI 优化引擎提供高质量的数据供给。行业先行者已引入“机理框架+数据驱动”的双底座架构，在传统数据总线上叠加物理模型的校验与增强能力，为后续的高精度仿真、全局寻优和闭环控制提供可靠的数据基础。

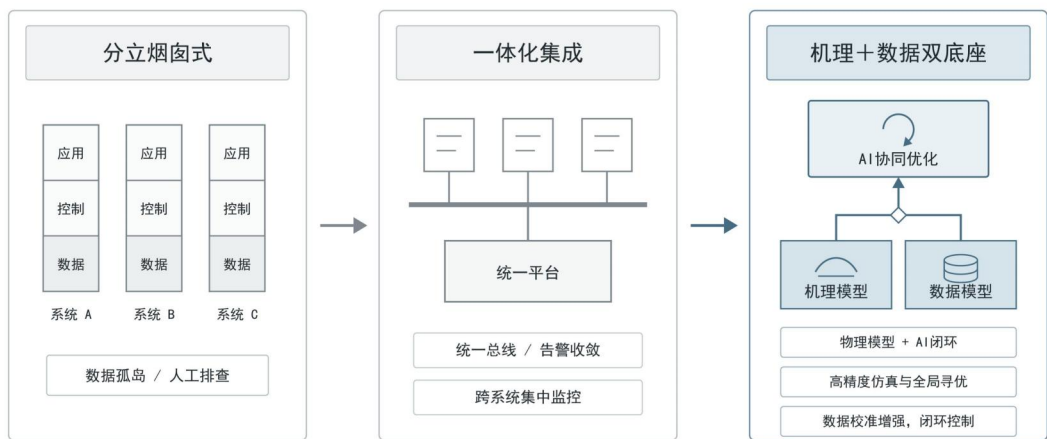


图 10 监控体系架构演进图

智能分析与集成的核心，在于将运行数据转化为状态认知、分析决策和协同控制能力，支撑上层应用实现跨安全、优化与运维业务的联动闭环。例如在智算中心中，高密高功耗的运行环境使安全、能效与设备可靠性之间的耦合程度显著加深，任何单一维度的管理手段都难以独立奏效。一体化平台将运行数据实时输出至寻优引擎，控制指令经安全校验后下发执行，效果再通过数据回流验证，使能效优化成为持续运行的闭环过程。与此同时，告警收敛与设备健康评估同步运行，在优化过程中持续监控安全边界与设备状态，异常趋势在影响运行之前即被识别和干预。优化动作的边界由安全模型限定，设备状态的变化反过来修正优化策略，安全事件的发生也触发运维流程的自动衔接。常规调节由系统依据策略自动执行，涉及安全的操作经人工确认后下发，由此形成安全、优化、运维一体化的自动闭环运行体系。

（二）AI 节能优化控制

数据中心制冷控制由传统的单设备局部控制向系统协同、预测优化与自适应寻优演进，逐步形成基础控制、预测优化与安全保障相互协同的技术体系。PID 和可编程逻辑控制器能够实现温度、压力、流量及设备启停的自动调节，是保障设备稳定运行的基础，但主要用于单台设备或局部控制回路。专家规则和设备群控扩大了控制范围，可根据负荷及环境条件协调多台设备运行，但控制效果较为依赖规则设计和现场调试经验。随着监测、计量及数据分析能力逐步完善，机器学习开始用于能耗预测和运行参数分析。模型预测控制（MPC）进一步将状态预测、约束处理和滚动优化引入控制闭环，使冷机、水泵、冷却塔和末端设备能够协同运行。强化学习和物理 AI 为复杂策略搜索及可信决策提供了新的研究方向。从能力演进看，制冷控制可分为定值控制、规则群控、数据辅助、预测优化和自适应优化五个阶段，整体呈现由单设备局部控制向系统协同优化、由事后响应向负荷预测、由固定规则向动态适应、由单一数据模型向物理机理与运行数据融合发展的趋势。各阶段并非相互替代，较高阶段仍需以底层控制和人工管理为基础。



图 11 数据中心制冷控制能力演进

智算场景下的节能需求，集中体现在负荷快速波动、冷却系统耦合增强以及算力负荷与冷却资源协同等方面。智算中心单位机柜功率和热流密度较高，训练、推理及资源调度会引起 IT 负荷快速变化和空间分布不均的情况。GPU 训练和推理任务的启停、并行度变化及资源迁移可在较短时间内改变机柜功率和空间热负荷分布。按照设计峰值持续供冷，容易造成冷量冗余；仅根据当前温度进行反馈调节，则可能无法及时应对负荷突变和局部热点。若冷量供给长期与热负荷不匹配，可能增加热点和设备热降频风险，进而影响有效算力输出与算力交付效率。传统控制主要依据局部温度、压力和流量等现场反馈进行调节，对上层算力任务、资源利用率和负荷变化趋势的感知相对有限。AI 节能寻优可与算力调度系统建立信息接口，获取任务启停、资源利用率、功率变化趋势及运行计划等信号，提前预测冷量需求和热点迁移，并依据设备响应特性实施多时间尺度调节，实现算力负荷与冷却资源的动态协同，同时综合研判 IT 负荷、室外气象条件、制冷设备性能及末端冷却需求，在保障机柜入口温区、液冷供液参数与设备运行安全的约束下，实现冷量供给与实时热负荷动态匹配。

优化对象	主要对象	典型控制变量	主要目标
冷源系统	冷机、冷却塔、板式换热器	启停组合、负载分配、供水温度	提高冷源效率，扩大自然冷源利用
输配系统	冷冻水泵、冷却水泵及管网	流量、压差、运行频率	降低输送能耗，满足末端需求

风冷末端	精密空调、行级空调及风机	送风温度、风量、风机转速	控制机柜入口温度，减少过度供冷
液冷系统	CDU、换热器及液冷末端	供液温度、流量、压差	保障芯片散热与换热效率
系统整体	冷源、输配、末端与IT 负荷	多设备联合运行策略	降低综合能耗并保障算力性能

表 4 节能寻优的主要对象、变量与目标

基于物理 AI 的 AI 节能系统采用“感知—预测—寻优—控制—反馈”的闭环架构，实现面向全局的协同优化。以制冷系统为例，冷机、水泵、冷却塔、末端空调及液冷设备高度耦合，AI 寻优应以减少系统综合能耗和满足末端冷却需求为依据，避免仅以提高单台设备效率作为控制目标。一是冷源与输配系统协同寻优，应对冷机启停组合、负载分配、水温、流量、压差及冷却塔风机转速进行多变量协同优化。二是末端环境与冷量供给动态匹配，传统末端控制多采用固定送风温度、固定压差或最不利点定值的策略，易导致多数常规工况下的过度供冷。AI 寻优可结合机柜入口温湿度、冷热通道状态与 IT 负荷分布，精准识别热点与冷量冗余，动态调节末端送风温度、风机转速与运行台数，实现差异化按需供冷。三是风冷、液冷及混合制冷场景适配，如在风液协同场景中，AI 系统还应根据 IT 设备冷却需求动态分配风冷与液冷承担的热负荷，并统筹公共冷源和换热设备运行，避免不同冷却回路独立控制造成效率损失

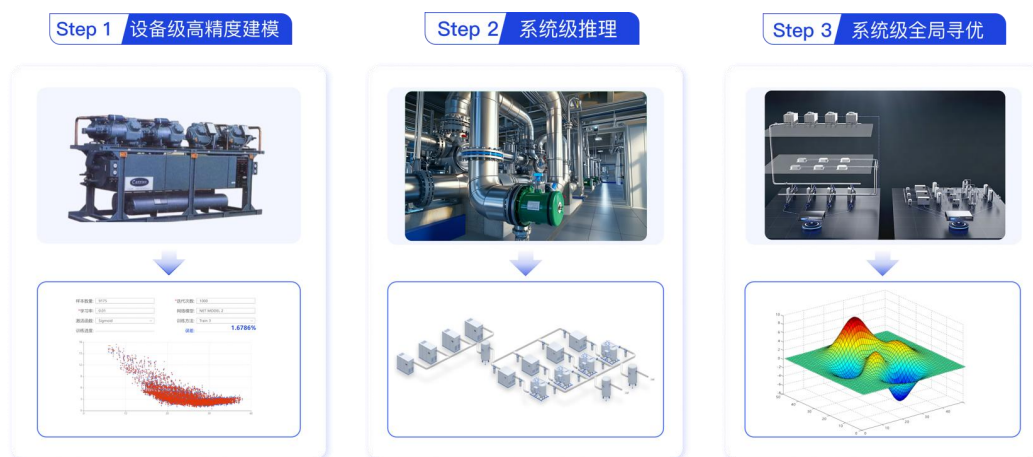


图 12 AI 节能闭环控制原理

(三) 边缘计算与端侧 AI

云边协同架构由云端平台、边缘节点和现场设备三层组成，各层通过模型下发、数据回传和策略协同形成闭环。云端平台负责多站点统一监控与数据分析、AI 模型训练与版本管理、全局策略优化与调度、运行效果评估与报表生成、安全策略统一下发；边缘节点部署在数据中心现场，负责实时感知与本地控制，主要功能包括：多源数据采集与预处理、本地 AI 推理与控制决策、指令下发与执行确认、设备状态监测与异常告警、本地数据缓存与断点续传；现场设备层包括各类传感器、执行器、冷机、水泵、冷却塔和末端空调等，通过工业协议与边缘节点通信，完成实际的数据采集和控制执行。云端将优化策略下发到边缘节点；边缘节点基于本地数据实时推理并执行控制；运行效果和关键数据周期性回传云端，用于模型评估、再训练和策略迭代。

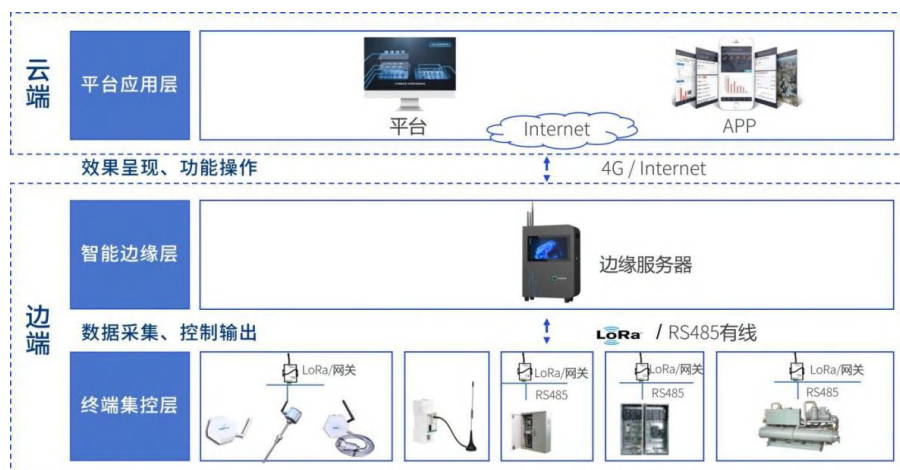


图 13 云边协同三层架构

边缘节点承担数据处理、实时控制、安全校验和本地自治等现场侧功能。数据采集与预处理方面，边缘节点通过 Modbus、BACnet、OPC UA、MQTT 等工业协议接入冷机、冷却塔、精密空调等设备，完成数据采集、时间同步、异常值识别、缺失值填补和降采样处理，高质量的数据预处理可减轻云端计算负担并缩短处理链路。**本地推理与实时控制方面**，部署在边缘节点的轻量化 AI 模型基于实时数据进行推理，生成设备设定值和运行策略，并通过控制器下发执行。与云端远程控制相比，本地推理消除了网络传输延迟，能够更好地适应负荷快速变化的场景。**安全校验与边界保护方面**，边缘节点内置物理安全约束模块，对 AI 输出的所有控制指令进行合法性校验，确保参数调节不超出设备安全运行范围。当检测到传感器异常、通信故障或运行参数越界时，立即触发保护机制，冻结优化策略并回退至安全运行状态。**本地自治与故障容错方面**，边缘节点具备独立运行能力，在与云端通信中断时可切换至本地自治模式，依靠内置模型和规则继续执行控制，保证节能优化不中断。同时，边缘节点之间可实现冗余备份，

单点故障时其他节点可接管控制。

端侧 AI 应通过模型轻量化满足边缘侧实时推理需求，并依靠设备、数据和模型安全措施保障控制执行。云端训练的深度学习神经网络模型通常参数量大、计算复杂度高，难以直接在资源受限的边缘设备上运行。模型轻量化技术通过在可控精度损失范围内压缩模型体积和计算量，使 AI 推理能够高效地在边缘侧执行。例如，在数据中心液冷 CDU 系统中，可将经过等轻量化处理的温度预测与控制模型部署至 CDU 边缘控制器，基于供回液温度、流量、压力及 IT 负载等实时数据，动态调节泵速、阀门开度和换热能力，在降低冷却系统能耗的同时，减少云端通信与决策时延，实现对负载波动和温度异常的快速响应。

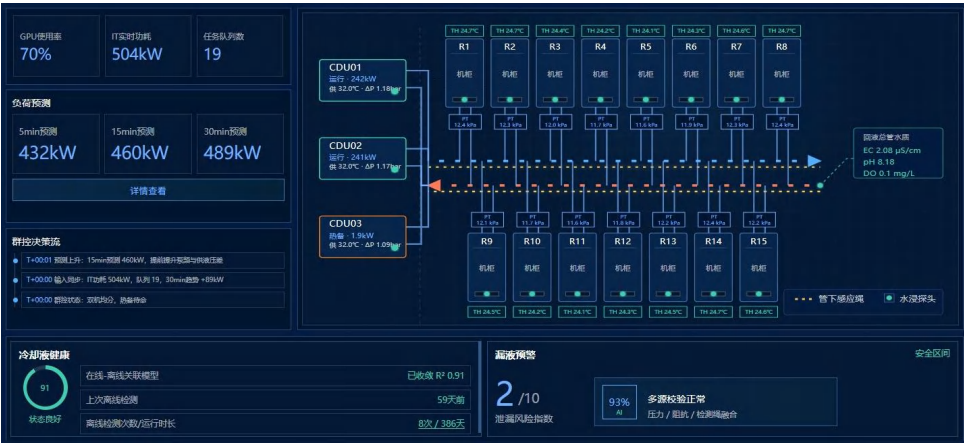


图 14 端侧 AI 应用：CDU 快速响应与节能控制

端侧 AI 的部署可以结合场景选择边缘网关、边缘控制器、边缘服务器等不同硬件形态。边缘网关属于轻量级边缘设备，通常采用 ARM 架构处理器，配备多路工业通信接口，主要承担协议转换、数据采集和轻量级计算任务。其特点是体积小、功耗低、环境适应性强，适合部署在控制柜或设备旁，作为末端设备接入网络的入口。边缘控制器具备更强的计算能力和控制功能，通常采用 x86 或高性能 ARM 处理器，集成 PLC 控制功能和 AI 推理能力，可同时完成数据采集、逻辑控制和智能优化。边缘控制器适合作为冷站或机房级的控制核心，直接对接现场设备并执行闭环控制。边缘服务器算力资源较为充沛，通常采用标准服务器架构，支持运行更复杂的模型和多任务并行处理。边缘服务器适合部署在数据中心机房或监控中心，作为区域级的边缘计算节点，统筹管理多个子系统的智能优化。

表 5 端侧 AI 部署不同硬件形态对比

硬件形态	典型算力	主要功能	部署位置	典型功耗
边缘网关	低（单核至四核 ARM）	协议转换、数据采集、轻量计算	设备层/控制柜	5~30W
边缘控制器	中（高性能 ARM/低功耗 x86）	本地控制、AI 推理、数据处理	冷站/机房控制层	30~100W
边缘服务器	高（标准 x86+可选加速卡）	区域优化、模型运行、数据管理	机房/监控中心	100~500W

(四) AI 控制的可靠性、鲁棒性与韧性

AI 控制系统的安全保障重点在于应对模型不确定性、可解释性和多设备强耦合风险。数据中心制冷系统的稳定运行直接关系 IT 设备安全和业务连续性，一旦 AI 控制出现异常决策或运行失控，可能导致热环境恶化、设备降频甚至业务中断。一是模型的输出具有一定的不确定性，当运行工况超出训练样本覆盖范围时，模型可能产生不可预测的输出。数据中心运行环境复杂，负荷变化、设备故障、气象异常等均可能导致模型进入未知区域。二是纯数据驱动模型可解释性不足，深度神经网络等复杂模型的决策过程难以直观解释，运行人员难以快速判断 AI 输出的合理性和安全性，也给故障排查和责任认定带来困难；为此，行业逐步转向物理 AI 融合建模路径，将物理规律和设备性能特征嵌入模型，使决策过程可追溯、可校验。三是系统耦合性强，AI 节能控制通常涉及冷机、水泵、冷却塔、末端空调等多类设备的协同调节，各子系统之间存在强耦合关系，一个控制变量的调整可能引发连锁反应，局部异常可能扩散为系统级问题。

表 6 AI 控制系统安全保障措施

安全保障措施	核心问题	主要保障能力
模型可信	AI 是否理解当前系统	不确定性量化、分布外检测、模型有效域管理、漂移监测
决策鲁棒	不确定条件下是否安全	物理边界保护、运行约束限幅、鲁棒优化、安全过滤、分级熔断

安全保障措施	核心问题	主要保障能力
执行安全	正确决策能否安全执行	本地指令校核、通信监护、无扰切换、分层控制、分阶段上线
系统韧性	故障后如何维持核心功能	故障诊断定位、虚拟传感冗余、控制重构、分级降级、快速恢复

AI 节能控制系统必须构建多层级安全约束体系。一是**物理边界保护**，物理边界通常基于设备额定参数、设计规范和运行经验确定，具有明确的物理意义和不可逾越性。AI 最终下发的指令都必须经过物理边界校验，超出范围的值将被强制裁剪到安全区间内。二是**运行约束与安全限幅**，在物理边界之内，还需要根据当前运行状态设置更严格的运行约束，以确保系统平稳运行并留有足够的安全裕度。对设定值的调整幅度和变化速率加以限制，避免参数大幅突变对系统造成冲击。三是**分级熔断机制**，根据异常严重程度采取不同强度的保护措施，包括策略冻结、策略回退、切换控制等，既避免过度反应影响节能效果，又确保严重异常得到有效处置。

AI 节能控制系统必须具备稳定运行和故障恢复能力。一是**控制稳定性保障**，AI 输出的设定值应采用渐进式调整，限制单次变化幅度和调整频率，并持续监测温度、压力、流量等参数的波动趋势；当出现周期性振荡或偏差持续增大时，应降低调整力度或暂停优化。二是**通信与数据容错**，系统还需对接入数据进行实时校验，对异常数据采取标识、替代计算、保持旧值或触发安全保护等措施。通信中断时，可根据中断范围和持续时间依次保持当前输出、切换至安全运行策略

或退出 AI 控制，并在通信恢复后完成数据确认和渐进式恢复。关键通信链路、边缘设备和重要测点可通过双网口、双电源、主备控制器、多传感器或软测量等方式形成冗余。**三是故障自愈与韧性设计**，当传感器、设备、通信或软件发生故障时，系统应及时诊断并定位异常，通过排除故障设备、调整控制策略或采用软测量替代等方式维持必要功能；故障排除后，再由降级状态逐步恢复至正常运行模式，避免一次性恢复对系统造成冲击。

工程化安全验证应贯穿 AI 控制系统上线前、分级投运和全生命周期运行。在正式投入闭环运行之前，利用历史运行数据和仿真模型，在离线环境中对 AI 控制策略进行大批量测试。仿真测试可以覆盖各种典型工况和极端场景，包括满负荷、低负荷、设备故障、气象突变、传感器异常等，验证 AI 输出的合理性、安全性和节能效果。随后开展影子运行，接入真实现场数据并实时计算控制策略，但不实际下发执行，只将计算结果与当前运行策略进行对比。现场测试需对所有接入的传感器和计量点位进行准确性核验，包括数据范围、更新频率、响应特性等，确保输入数据的可靠性。上线过程应遵循“由点到面、由浅入深、逐步放开”的原则，由建议模式逐步过渡至部分闭环和全功能闭环，并为各阶段设置相应的评估标准和审批流程。系统投运后，还需构建面向全生命周期的安全保障和应急管理机制。

五、面向绿色算力的 AI 算电协同

算力电力协同优化调度是数据中心由单一用电负荷转向电网互动参与者的核心技术路径，在保障 IT 业务质量前提下，通过算力时空调度、储能蓄冷调控与多能源配合，实现降本增效、提升绿电消纳、获取电网服务收益。在分时电价、“东数西算”、电力辅助服务市场等政策市场驱动，叠加算力柔性调度、储能蓄冷配置等技术支撑下，算电协同从规则策略逐步演进为 AI 驱动动态寻优，由行业试点探索走向广泛应用。

（一）算电协同综合调度

协同感知与联合预测是算力电力协同优化调度开展上层决策的重要前置基础。在负荷侧，行业在算力调度平台集成功耗采集能力，通过服务器管理接口获取节点级功耗，结合任务类型、资源申请量、历史运行数据构建任务至功耗的映射模型，实现细粒度功耗标定。集群层面融合任务队列、历史负载、气象信息，借助时序预测输出未来数小时至全天的总功耗与制冷需求，大型云服务商 24 小时功耗预测平均误差可控制在 5% 以内，为日前优化调度提供可靠边界条件。在电源侧，依托气象数据完成风光出力预测。在电网和市场侧，接入省级电力市场平台获取历史电价、系统负荷、新能源出力等数据，提取价格时序、峰谷及供需特征，滚动融合最新市场信息完成 24 小时电价预测，并同步获取调峰、调频等辅助服务需求。

面向多目标、强不确定性的复杂算电协同场景，AI 调度技术逐步

取代传统规则式调度，形成机理约束与数据驱动相结合的分层闭环调控体系。早期协同调度主要依靠规则与启发式方法，虽实现简便，但难以应对新能源波动、电价跳变、算力负载突变叠加的复杂工况。模型预测控制凭借前馈补偿、滚动优化、可处理多变量约束的特性，成为行业主流基础框架，通过有限时域求解优化问题、滚动迭代更新控制指令，适配各类预测偏差与系统动态变化。工程实践形成共识，AI调度必须以物理机理作为安全边界，数据驱动承担优化引擎功能，AI输出策略需经过机理模型校验方可下发。完整的算电协同 AI 调度体系分为以下四层：

感知融合层：汇聚“源—网—荷—储—冷—算”多系统实时数据，统一状态监测、数据治理与特征融合，形成算力中心全局运行状态视图。

物理 AI 模型层：构建设备机理模型、算电热耦合模型及负荷预测模型，刻画算力、电力、冷却与储能之间的动态关联，为决策提供高精度数字底座。

算电协同决策层：综合电价、碳排、热安全、设备容量及算力任务约束，开展多目标全局寻优，生成储能充放电、冷却调节与算力负荷迁移策略。

控制下发通道：通过 BA、PLC、PCS 等控制协议及 K8s、Slurm 等算力调度接口，将优化指令下发至能源、冷却和算力系统，并依据执行反馈形成闭环调控。

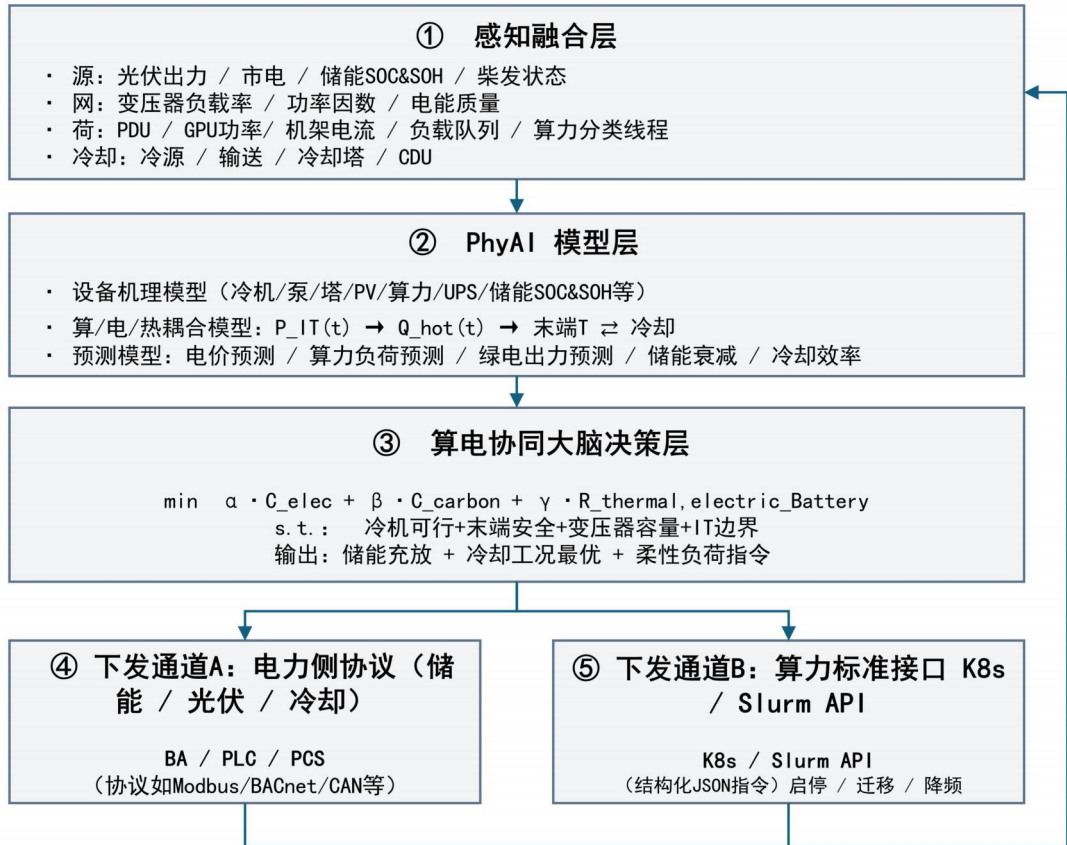


图 15 算电协同 AI 调度体系图

算力电力协同调度工程普遍采用分层实施架构，通过经济性决策与安全性执行解耦，兼顾全局优化能力与底层控制的可靠性。上层为十五分钟至一小时周期的滚动优化层，以能量管理系统为核心搭建多源能量平衡模型，围绕用电成本、碳排放、辅助服务收益开展多目标寻优，同时约束业务 SLA、储能寿命、配电容量等边界条件，部分项目嵌入机理仿真模型校准能耗预测，缩小计划与实际运行偏差。下层为秒-分钟级实时闭环控制层，接收上层下发指令，依托 PID、自适应算法完成现场快速调节，并具备安全阈值保护能力。算力柔性调度根据上层生成的任务调度计划，通过 K8S、Slurm API 将调度指令下发至算力集群，并由原生调度机制完成资源校验与任务执行。储能与

蓄冷充当功率缓冲器，电化学储能平抑短时功率波动，蓄冷系统在高峰释放冷量降低制冷耗电，二者协同优化需量电费，提升整体电能质量。

（二）算冷协同优化控制

随着智算中心向万卡集群快速演进，单机柜功耗从传统数据中心的 4-6kW 跃升至 40-120kW 以上，高密度运行特征凸显，算冷协同优化调度正成为适配高密算力场景、解决算冷供需错配的核心技术之一。其依托 AI 预测与智能调控技术，通过精准冷负荷预测、智能制冷控制与蓄冷柔性调度，弥补传统制冷滞后缺陷，保障算力热安全，同时释放冷源调节能力，兼顾算力中心节能降耗与电网需求响应的协同价值。

算力—冷负荷双层映射模型结合预测前馈控制，可有效破解智算系统热惯性带来的温控难题，实现算力中心精细化、自适应智能热管理。冷负荷精准预测依托算力任务、IT 设备、液冷系统、设施环境等多维数据支撑，通过双层因果映射机制实现精准推演：先依据算力任务参数、GPU 运行数据拟合算力负载与 IT 功率的对应关系，再基于设备功率预判冷负荷与温度波动趋势，为制冷前馈调节提供依据。模型采用多源融合多时间尺度机器学习架构，短时可依托 LSTM 网络捕捉实时的负载瞬时波动与周期特征，中长期通过 XGBoost、LightGBM 模型保障小时级预测的稳定性与可解释性，同时依托历史数据与滚动窗口迭代更新，适配工况动态变化。在此基础上搭建“预

测前馈+实时反馈+安全保护”的 CDU 复合控制模式，提前预判冷负荷变化并预调泵阀参数、联动上级冷源，同时设置多重设备与温度安全约束，异常工况下自动切换保守运行模式，实现安全前提下的精细化节能调控。

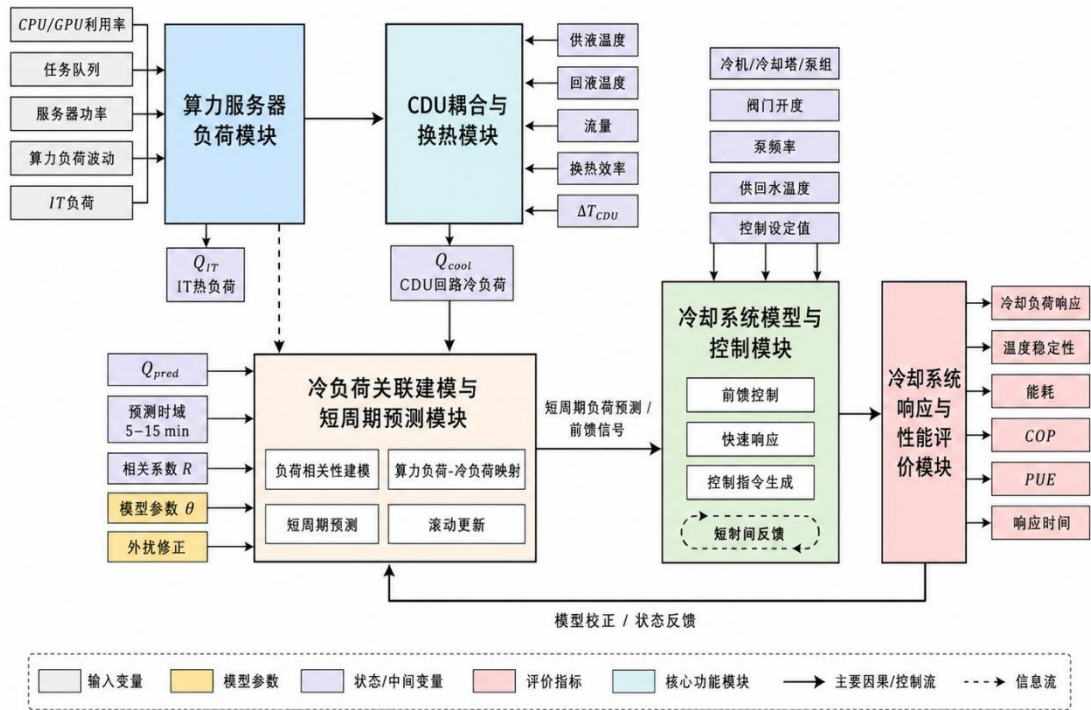


图 16 冷负荷预测与冷却系统控制逻辑

蓄冷系统作为关键的柔性冷量调节资源，需融入全域算电冷一体化调度体系开展协同优化，而非局限于独立制冷系统优化，通过冷量时序解耦，实现节能降耗、安全保障与电网互动的综合效益。蓄冷系统可脱离制冷实时供需限制，实现冷量错峰存储与释放，既可作为应急冷源保障算力设备稳定运行，也能平抑冷负荷波动，还可转化为可调负荷支撑虚拟电厂需求响应，已有数据中心依托冰蓄冷设施实现高峰电网负荷调节。在算电协同体系下，蓄冷调度彻底跳出单一制冷维

度优化，深度融入源网荷储全局架构，参与算力、电力、冷却多维度一体化协同调控。

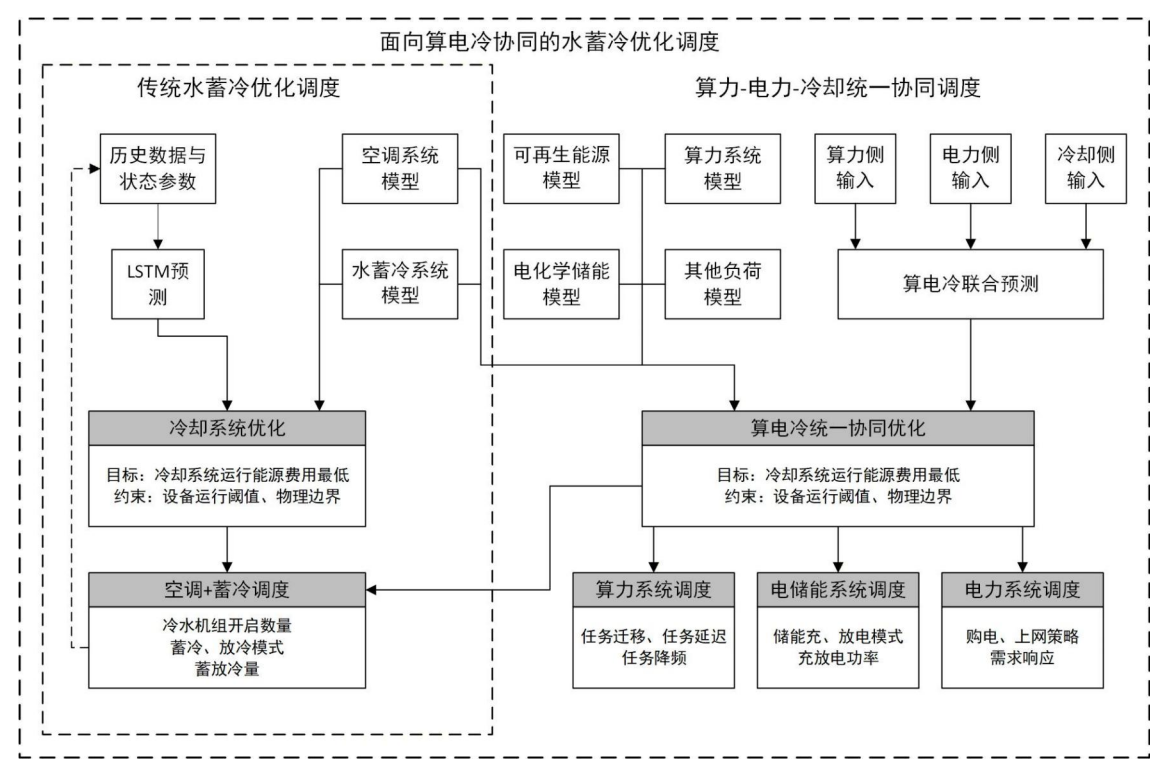
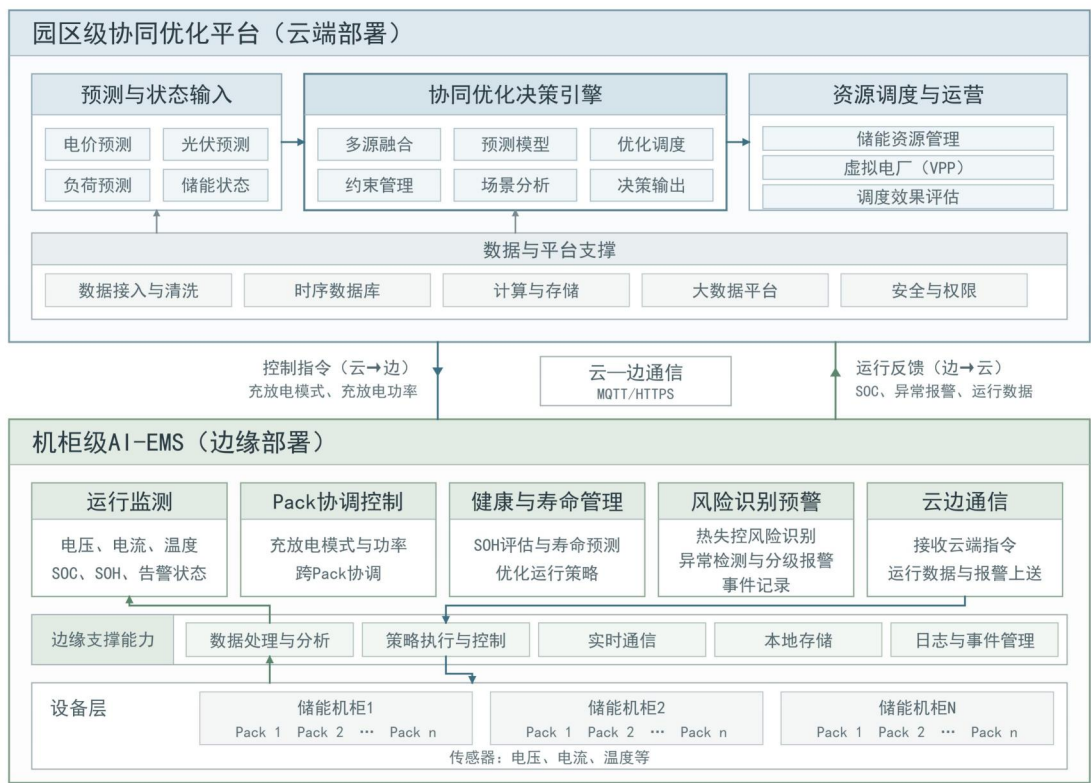


图 17 面向算电冷协同的水蓄冷系统协同调度

（三）储能系统 AI-EMS

传统储能能源管理系统固化的规则化调控模式，已无法适配算电协同场景下多约束、多目标的动态调度需求，AI-EMS 通过云边端一体化智能架构，实现储能系统从被动监测向主动预测、自主优化、精细调控升级。依托轻量化 AI 部署能力，可打通园区全局调度与设备

级实时控制壁垒，兼顾储能运行安全、电池寿命、绿电消纳与用电经



济性，支撑算电协同体系高效落地。

图 18 园区级“源网荷储”协同优化平台架构图

储能 AI-EMS 依托标准化、模块化的全链路技术架构，可适配不同储能规模灵活部署，实现数据接入、边缘处理、智能决策、控制执行与上层协同的一体化闭环管控。系统通过以太网、RS485 等通信接口接入 BMS、PCS、计量仪表、消防及环境安防设备，构建完整多源数据底座；边缘节点负责数据清洗、本地持久化存储、实时安全校验，在上层通信中断时维持本地自治运行，保障储能基础功能不受影响；AI 决策模块完成储能调度能力评估、多目标优化、Pack 级功率分配、运行风险识别；优化调度指令经过多重安全校验后下发底层设备，硬件保护装置独立承担刚性安全防护。小规模储能场景采用单柜集中式

部署，AI-EMS 本地统一管控柜内全部设备；大型储能集群采用分布式架构，各储能柜部署本地 AI-EMS 边缘节点，站级设置集群协调节点，统一汇总运行数据、协调柜间功率分配，实现分层分级管控。

AI 技术贯穿储能状态感知、精细调度、风险处置全过程，可实现基于设备差异化特征的智能调度，解决传统均质化功率分配带来的安全与寿命损耗问题。在运行状态识别层面，AI-EMS 融合储能功率、可用容量、电池健康度、告警信息及历史运行数据，动态评估系统可调能力与安全边界，精准判定可调度范围与受限设备，为上层调度提供可靠的设备侧能力依据。在功率分配层面，系统摒弃传统平均分配模式，依据各储能 Pack 的 SOC、SOH、温差、一致性与寿命风险差异计算调度权重，差异化分配功率指令，对优质设备提升调度占比，对异常设备限流降权，并通过实时反馈持续修正策略，兼顾调度经济性与电池寿命。在风险管控层面，AI 可自动识别温度异常、单体压差偏移、SOC 失衡等隐患，分级判定风险等级并给出优化处置建议，动态调整设备调度权限。同时 AI-EMS 仅负责智能分析与辅助优化，过充过放、高温短路等刚性安全保护仍由底层硬件系统兜底，保障储能运行绝对安全。

六、面向绿色算力的 AI 智慧运维

算力规模指数级扩张与设备形态快速迭代，使传统人工运维模式难以适配万卡级智算中心的运维需求。AI 智慧运维以数据驱动、智能研判、闭环处置为核心，通过设备健康管理、故障智能诊断、自动化闭环运维与大模型知识赋能，实现从被动抢修向预测性维护、从经验依赖向平台赋能转型，为算力中心安全、高效、低碳运行提供全生命周期保障。

（一）健康管理 with 预测性维护

建立设备动态分级机制，以健康度量化评价为抓手，差异化分配运维资源。算力中心设备体量庞大，不同设备承担业务功能、故障影响范围、冗余配置条件、维修难度存在明显差异。若所有设备采用统一巡检频次、维护标准，将造成关键算力设备保障力度不足，普通配套设备巡检冗余、部件提前更换等资源浪费。分级管理综合考量设备功能定位、故障影响等级、冗余能力、维修难度、备件保障水平，差异化设定监测粒度、告警响应优先级、巡检周期、维护策略与备件储备方案，优先将运维资源倾斜至高风险、高价值设备。健康度评价持续采集整机与核心部件运行偏差、历史故障记录、运维处置结果、业务负载重要性开展综合打分。当设备状态持续恶化、故障频次上升或承载算力任务加重时，及时提升管控等级；对长期运行稳定、风险水平较低的设备，可适度降低运维频次。部件分项评分能够精准定位异常发生环节，为后续故障根因排查缩小范围，健康度评分成为维护计

划编制、异常溯源的统一基础。



图 19 设备健康度评价示例

结合工程机理与正常工况数据搭建设备性能模型，依靠多维度偏差特征识别长期性能退化。设备性能模型的核心作用是标定不同工况条件下设备应当具备的运行水平。建模过程融合性能曲线、能量传递关系、设备控制边界等工程机理，叠加长期稳定运行的历史数据，建立不同负载、环境温度、控制策略下的基准运行指标。实测运行参数短时间偏离基准值，大多由算力负载波动、设备切换、测量噪声引发；当偏差持续存在、偏离幅度逐步扩大，并且得到多项关联参数相互印证时，预示设备性能持续下降。开展退化预测不能仅依靠单一测点或者单次异常信号，需要综合评估偏差幅度、持续时长、参数变化速率。比如，液冷系统的冷却液健康管理同样遵循这一逻辑，通过在线监测指标与离线检测指标之间的关联模型，将电导率、pH 值等关键参数纳入统一的性能基准框架，当在线监测值相对模型基准持续偏移且多项参数交叉印证时，即可判断冷却液性能正在退化，进而动态评估离线检测的必要性，在降低检测频次的同时保持健康状态透明可控。

打通风险预测与运维工单链路，构建“风险识别—任务派发—现场处置—模型迭代”预测性维护业务闭环。设备潜在风险识别完成后，

需要依托标准化流程调动人员、备件资源完成隐患治理。系统持续汇聚设备状态、环境参数、能耗数据，融合规则判定、趋势分析、机器学习模型共同识别潜在故障隐患。当风险等级达到设定阈值，自动生成预警信息、现场检查建议或者正式维护工单，工单附带设备位置、异常部件清单、风险判定依据、推荐处置方案以及所需备件信息。运维班组完成现场作业、验收确认后，上传现场现象、真实故障原因、处置结果。平台利用反馈数据持续修正设备健康基线与风险判别逻辑，提升后续预测精准度。针对核心关键设备，系统结合故障发生概率、性能退化速度、业务影响程度排序维护优先级，让巡检、检修、备件备货工作贴合设备真实运行状态，有效解决过度维护与维护不足两类问题。

（二）故障诊断与根因定位

规则、趋势、模型、传感器数据校验四类诊断策略协同应用，提升故障诊断覆盖率与准确率。针对算力中心故障模式复杂、单一诊断手段适用范围有限的问题，智能诊断体系综合发挥规则诊断快速判定、趋势诊断识别渐进劣化、模型诊断发现工况偏离以及传感器校验甄别异常数据的优势，形成多策略互补、多证据交叉验证的诊断机制。上述策略在液冷系统漏液预警中的协同应用，展示了智能诊断的完整推理路径：液冷系统通过总管与支管的双层感知网络，时序模型学习压力与流量差基线以识别系统级趋势偏移，机器学习模型实时比对各机柜支管压力差异并输出置信度评分；在此过程中，规则诊断对异常机

柜的工况偏离进行快速判定，趋势诊断追踪压力差的持续变化方向，模型诊断在正常工况基线之上识别超出动态置信区间的异常模式，传感器校验则通过多源数据一致性交叉验证剔除虚假信号，多条证据链相互印证后，平台输出包含异常位置、置信度评分和排查建议的事件信息，使运维人员从单纯接收告警转向对故障成因的深度理解。

FMEA 实现故障知识标准化沉淀，动态故障树实现智能化路径推理，支撑高效根因分析。设备手册、运行规程、历史工单、运维经验蕴藏大量故障知识，但资料来源分散、记录格式不统一，无法直接支撑自动化故障诊断。平台以系统功能、设备结构为主线，从部件失效场景切入，梳理失效诱因、故障影响，关联监测测点、告警特征、运行征兆与处置措施，按照“系统功能—部件结构—失效模式—失效原因—失效影响—可观测信号—处置方案”逻辑搭建 FMEA 失效模式库。风险评价参照 IEC 60812 标准，结合算力中心设备特征，对各类失效模式的发生度、严重度、探测度进行打分，计算风险优先数（RPN），明确监测与处置优先级。重点围绕冷机、水泵、冷却塔、CDU、精密空调、供配电设备完善知识库，实现实时测点、告警信息与已知故障特征自动匹配。故障树以现场真实故障现象作为顶层事件，向下拓展多重诱因分支，涵盖设备本体退化、传感器采集异常、控制策略偏差、环境扰动、人为配置失误等类型，逐层拆解为可验证中间节点与底层事件。系统内置智能排查引擎，结合 RPN 数值、故障发生频次、探测难度、实时监测数据动态推荐最优排查路径。流量计读数等可自动

核验的节点由系统直接判定；硅脂干涸等需要现场拆检的问题，自动生成作业工单。故障树与工单闭环联动，处置结果持续优化推理路径，后续同类故障无需重复全量排查。

绿色算力运维需要同时覆盖安全风险与能效损失两类异常，实现可靠运行与节能降耗协同推进。传统监控告警体系重点关注参数超限、设备停机等安全风险，面向绿色算力的运维体系，还需要识别“参数未超限但运行不经济”的工况。PUE 持续波动、冷机运行 COP 偏离合理区间、冷却塔换热效率下滑、冷热通道气流失衡等现象，诱因既包含设备性能退化，也涉及运行策略不合理、算力负载动态变化。诊断过程按照园区、机房、系统、设备、部件层级逐层下钻分析，区分设备故障、测量误差、低效运行三类状态。诊断结论清晰界定问题属性，针对性输出设备维修、传感器校准、运行策略优化等差异化建议。通过安全、能效双维度诊断，在守住基础设施运行底线的基础上持续挖掘节能潜力，契合绿色算力中心低碳运营需求。

（三）自动化闭环运维

安全闭环保障基础设施稳定运行，能效闭环支撑低碳运营，管理闭环持续优化运维体系。安全闭环以故障告警受理为起点，整合物联网采集数据、IBMS、自控系统、日常巡检记录，全面感知设备与环境状态。异常发生后，依据问题等级、影响范围启动对应工单处置流程；边界清晰、风险较低的标准化操作可在授权范围内自动联动执行；涉及核心设备、业务中断、重大安全风险的操作，保留人工确认与权限

审批流程。处置结束后持续监测关联参数，确认设备与环境恢复稳定状态。能效闭环围绕异常识别、根因研判、优化效果核验构建链路，综合算力负载、室外环境、供回水工况分析 PUE、CLF、冷机、泵组、冷却塔运行效率。当运行工况偏离合理区间时，区分设备性能衰减、控制策略缺陷、正常工况波动，采取参数调整或者节能优化策略，并且在相近工况条件下复核优化成效，避免将短期能耗波动误判为节能成果。管理闭环依托工单串联故障处置全流程，将告警信息、诊断依据、任务分派、现场操作、备件使用、验收结论、复盘意见归集至对应设备与事件。利用响应时长、处置时长、重复故障数量、人工介入频次等指标，识别排班机制、备件储备、外包管理短板。针对反复出现的共性问题追溯深层根因，修订控制策略、维护计划与运维操作规程。

自愈能力循序渐进分阶段落地，在安全可控前提下稳步提升运维自动化水平。自愈体系建设不能急于追求全自动控制，需要匹配数据基础、系统成熟度有序推进。工程落地分为多个阶段：第一阶段实现异常自动发现与告警推送；第二阶段输出诊断依据与处置方案，由运维人员决策是否执行；第三阶段在授权边界内自动开展故障隔离、设备冗余切换、运行参数恢复等标准化操作；更高等级自愈需要跨系统协同联动，并且持续验证调控效果。当前行业成熟应用集中在辅助决策与受限自愈区间，故障隔离、冗余切换、限流降级等风险低、置信度高的场景可以启用自动执行。自愈能力建设必须坚守两条底线：安

全底线与可解释底线。涉及停送电、大型设备启停、可能扩大业务中断、损伤硬件设备的操作，强制设置人工确认环节。所有自动化操作完整记录触发条件、执行过程、运行结果，支持全程追溯审计。AIOps建设遵循由易到难路径，首先统一资产、测点、告警、工单数据，再完善诊断规则、性能模型、FMEA 库与故障树，逐步接入预测性维护、运维知识助手、能效诊断；数据基础完善、控制边界清晰、策略充分验证后，再上线自动联动与局部自愈功能。各算力中心设备条件、管理基础存在差异，建设顺序结合可用数据与核心运维痛点灵活调整。长期来看，闭环运维将与数字孪生、边缘智能、算电协同深度融合，成为绿色算力中心降耗增效、提升系统韧性、支撑低碳运营的核心能力。

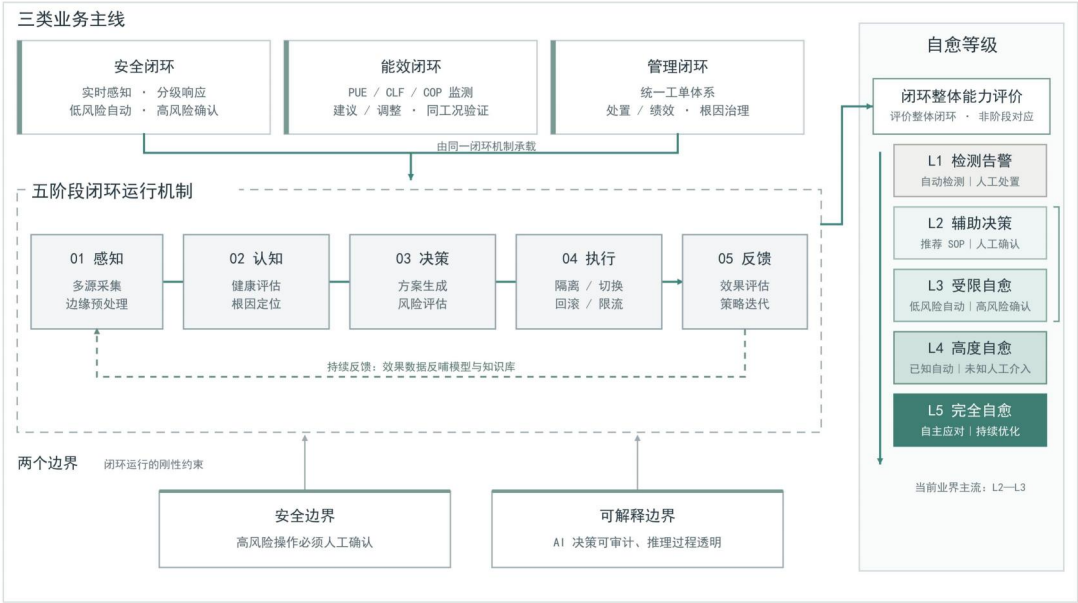


图 20 自动化闭环运维体系

（四）专属知识库与运维助手

采用 RAG 技术构建知识库, 实现运维知识精准召回、来源可追溯,

保障大模型输出内容可信。专属知识库主流建设路线为检索增强生成（RAG）。大模型对上传的各类技术文档进行解析，采用级联切分、自适应拆分策略，将长文档拆解成语义完整的知识片段，经过向量化处理存入向量数据库。运维人员发起查询时，系统融合关键词检索、语义检索、表格多路检索召回相关知识片段作为上下文，生成附带资料来源、可追溯的回答。行业实践证明，该技术路线在运维知识问答场景下，文档出处召回准确率可达 90% 以上，相比传统关键词检索效果提升显著，为运维助手提供可靠知识底座。

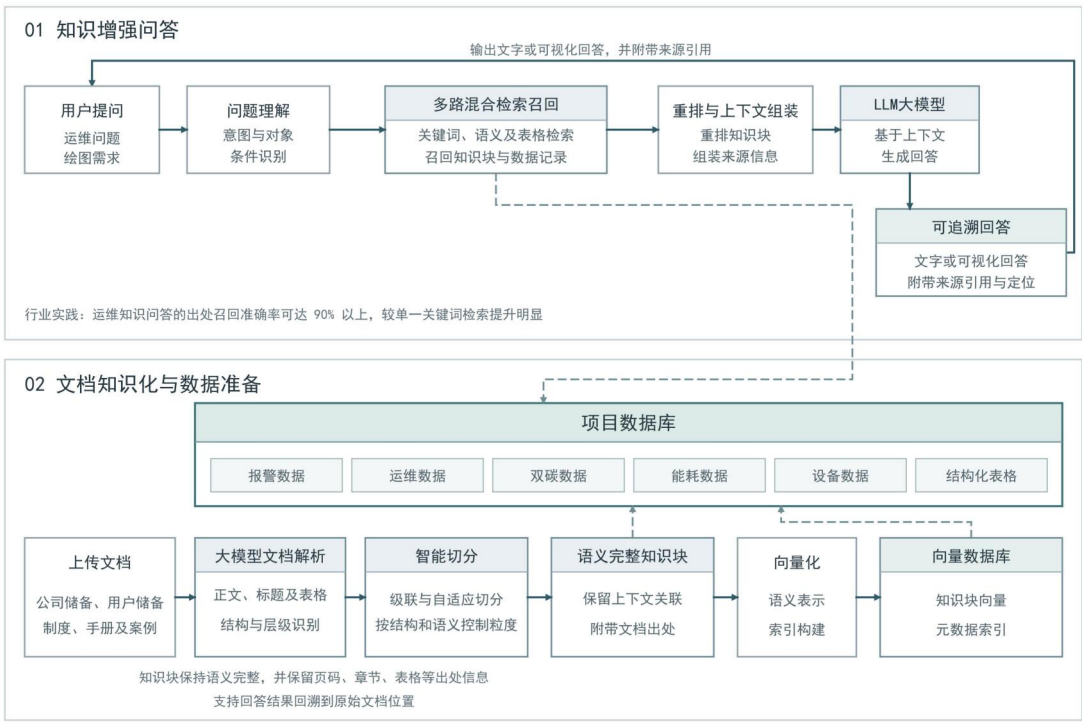


图 21 算力中心专属知识库

运维助手融合知识库与实时业务数据，覆盖操作指导、态势查询、告警分析三大典型场景。大模型运维助手核心价值是将静态知识转化为现场可用的交互能力，典型应用分为三类。一是 SOP 指导问答：

运维人员现场面对陌生设备、不熟悉操作时，通过自然语言发起提问，系统依托知识库推送操作步骤、风险注意事项，关联对应技术文档章节，相当于为每位运维人员配备在线专家。二是运维态势自然语言查询：打通工单系统、监控平台后，管理人员可使用自然语言查询工单数量、超时情况、班组工作负荷，分析响应时长、处置时长分布，系统自动汇总信息并输出优化建议，快速掌握整体运维态势。三是设备告警与测点数据分析：对接实时监控数据流，自动归纳周期性告警模式，推送排查方向，帮助运维人员从海量告警中识别关键风险，缓解告警疲劳。大模型应用严格遵循可解释、可追溯、可管控原则。对于停送电、设备启停、控制参数修改等高风险操作场景，大模型仅提供资料依据、风险提示、备选方案，最终操作由授权人员或者经过验证的自动化策略执行。知识库配套版本管理、审核发布、权限隔离、过期提醒机制，避免错误经验、过期规程流入生产流程。长期来看，每一次告警处置、复盘总结都能够沉淀为案例资产，持续更新 FMEA 条目、故障树、诊断规则与知识库，推动运维知识体系由静态资料库，持续进化为具备自主学习能力的智能底座。

七、案例实践

（一）超算中心系统级协同寻优实践

1.项目概况

某国家级超算中心制冷站共配置 9 台变频冷水机组、18 台冷冻水泵、8 台冷却水泵及 3 座自然冷却冷却塔。改造前冷机出水温度与水泵频率均采用固定值运行，缺乏群控策略，全年以人工经验控制为主，PUE 为 1.28。

2.AI 技术应用

1) 设备级性能建模

基于物理 AI 建立冷机、水泵等核心设备性能模型，充分考虑主机性能特点及实际水力特性，模型预测误差 $RRMSE < 3\%$ ，为全局寻优提供精确的虚拟仿真环境。

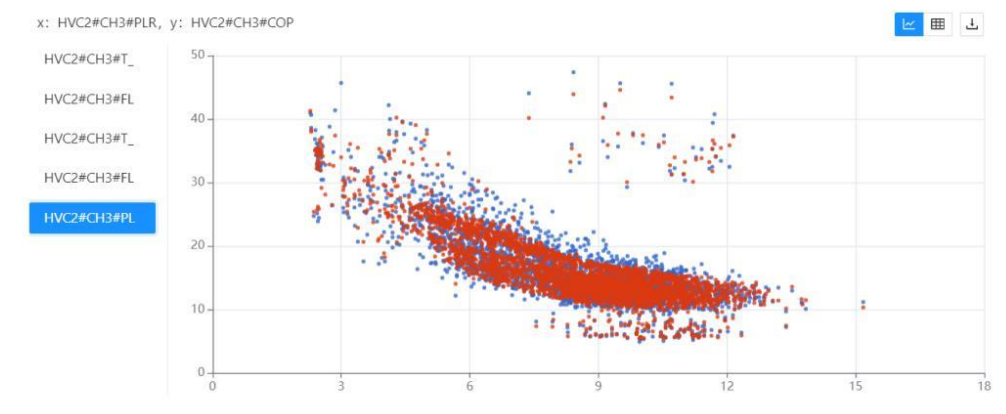


图 22 冷机能效模型训练

2) 系统级协同寻优

构建模型预测控制策略，基于系统能效最优原则，协调优化冷水

机组开启台数与冷冻水出水温度设定、冷冻/冷却水泵台数与频率、冷却水温差等关键参数，同时兼顾冷机安全运行边界。

3) 自然冷却系统优化

将自然冷却系统接入节能优化控制系统，通过算法预测自然冷却开启时间及运行参数设定，以数据为基础提升自然冷却系统运行效率，减少冷机冬季运行时间。

3.项目收益

以 2018 年用电数据为基准，改造后中心全年 PUE 从 1.28 降至 1.21，空调系统能耗降低约 23%。2019 年 9 月进入分享期以来，累计总节电量达 1128 万 kWh，平均节能率 24.6%，中心能源利用效率提升 5.8%。

（二）水冷冷站系统节能及 PUE 诊断实践

1.项目概况

某华东数据中心 A1 楼制冷站配置 4 台变频离心机组、4 台冷冻水泵、4 台冷却水泵及 4 组冷却塔，系统管路双供双回。改造前冷机与水泵以人工经验控制为主，缺少全局寻优策略，制冷站能耗存在较大优化空间。

2.AI 技术应用

1) 系统级节能优化

基于性能模型与历史工况构建节能优化算法，协调优化冷机开启台数、冷冻/冷却水泵频率、冷却塔风量等关键参数。经现场测试稳定

后系统自动闭环下发，实现实时计算、自动调参，使制冷站持续运行在最优工况。

2) PUE 性能诊断

构建 PUE 根因分析引擎,基于树状节点结构与 280+类诊断规则，覆盖压缩机、水泵、风机、阀门、传感器、板换、CDU、冷却塔等设备类型；针对 PUE 异常事件，从冷机侧（COP 下降、冷媒泄漏、冷凝器结垢、流量不足等）与系统侧（管网失衡、水泵偏离等）逐级下钻定位根因，输出故障类型、根因概率排序与处置建议，实现从事后修到计划修的转变。



图 23PUE 诊断界面展示

3.效益分析

以 23 年能耗数据为基准，按照 GB/T 26759-2011 节能测试方法，制冷站全年节能率达 16.1%，系统上线至今保持稳定安全运行。

(三) 风冷精密空调末端按需供冷实践

1.项目概况

某华东通信数据中心局站共 33 个机房、318 台风冷精密空调，机房面积合计 18119m²。传统控制依据单台空调回风温度统一设定运行，存在局部热点与冷热抵消等问题，年 PUE 在 1.60~2.20 之间。

现场送风形式多样，部分区域采用弥散式送风，机柜负荷密度及排列方式差异较大。传统控制主要依据单台空调回风温度和统一设定值运行，难以反映不同区域的实际热负荷。

表 7 项目典型问题与工程响应

典型问题	工程响应
回风温度与服务区域实际温度不一致	补充环境测点，建立空调—区域影响权重和温度场响应模型
负荷分布不均，但多台空调采用相同设定	按区域冷量需求协同优化运行台数、温度设定和风机转速
过度供冷触发再热或除湿，形成能量抵消	联合判断制冷、再热、加湿和除湿状态，减少冷热抵消

2.AI 技术应用

1) 机房温场建模与空调影响力分析

补充机房环境温湿度传感器，建立空调—区域影响权重模型和温度场响应模型，精准刻画每台空调对不同区域温度变化的影响力，为

全局协同优化提供基础。

2) 按需供冷协同优化

基于机房负荷预测与温场模型，按区域冷量需求协同优化空调运行台数、温度设定和风机转速，联合判断制冷、再热、加湿和除湿状态，减少过度供冷或过度除湿产生的冷热抵消能耗。

3) 安全保障与自动闭环

闭环运行保留不少于 20% 的制冷量冗余，单次停机不超过 1 台，总停机数量不超过机房空调数量的 30%；发生高温、设备故障或数据中断时，系统自动回退至预定义安全策略。



调试前：局部热点较多

调试后：温场趋于均衡

图 24 风冷机房温度场调试前后对比

3.效益分析

通过温场建模与按需供冷策略，实现空调运行状态的可视化监控与智能调控，降低运维人员手动调节工作量，提升机房环境管理的精细化水平。项目调试结果显示，原有高温区域得到抑制，机房温度场由局部冷热不均向相对均衡状态收敛；项目阶段性测算节能率为 12%。

（四）间接蒸发制冷数据中心智慧运维实践

1.项目概况

某西北大型数据中心，规划 13 期、占地 2500 亩，已建成投产 1 至 9 期共 8 栋数据中心，可容纳 20 万台服务器，全部建成后 IT 总容量超 500MW。本项目对基地配置的 144 台间接蒸发冷却空调实施 AI 节能寻优与智慧运维改造。

2.AI 技术应用

1) 性能建模与运行诊断

基于数据驱动与物理机理相结合的方法，针对干工况、湿工况及压缩机辅助工况分别建立 AHU 性能映射模型，输出制冷量预测与能效基线。以性能模型为基准，建立模型预测与实测值的动态对比机制，识别换热效率下降、控制策略偏差等性能退化类型，为设备健康评估提供量化依据。

2) FMEA 失效模式分析与知识库针对 AHU 六大功能系统建立结构化 FMEA 数据库与故障树，支持从性能偏差检测到故障模式解释再到维修建议生成的闭环分析。系统自动匹配失效记录，根据风险优先级输出分级诊断方案，实现运维经验的数字化沉淀与复用。

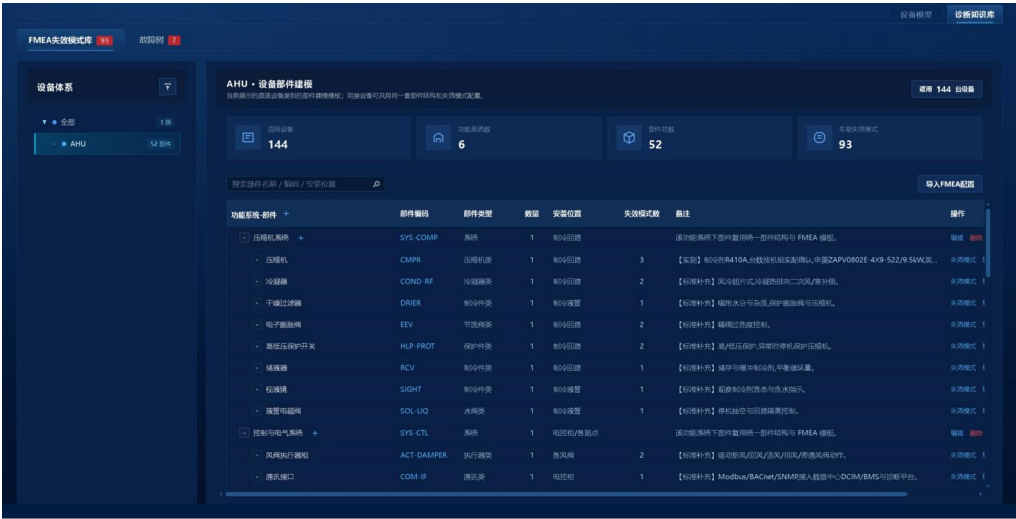


图 25 FMA 失效模式库

3) 多工况运行模式调优

基于预测负荷与性能模型，构建模型预测控制策略，实现干工况、湿工况及压缩机辅助工况间的智能切换与参数自适应优化。系统向运维平台下发经约束校验的设定值，现场控制器继续承担基础闭环和安全保护，形成 AI 全局寻优与本地安全保护相结合的分层控制架构。

3.效益分析

144 台 AHU 全部纳入在线诊断体系，通过规则矩阵实现运行控制、设备性能、传感器故障三类问题的自动识别；构建了标准化知识底座，实现从被动维修向预测性维护转变；系统可实现 AI 自动化控制闭环，AI 推荐设定值经校验后自动下发执行，同时保留人工最终决策权。预计 AHU 节能率可达 13.6%，年节电量约 253.3 万 kWh。

（五）风液协同数据中心分层寻优实践

1.项目背景

某风液协同数据中心配置 4 台液冷机柜、4 台气冷机柜、两台 500

kW 级 CDU 及 5 台水冷精密空调。液冷主要承担芯片热量，气冷系统负责非液冷器件和纯气冷机柜散热。项目设置冷、热通道温湿度测点各 24 个，并配置压差、流量、漏液及水质监测，为两条冷却链路的协同优化提供运行边界。

2.AI 技术应用

本项目采用分层物理 AI 控制，通过冷源、CDU、机柜和气冷各层级的联动实现风液协同，即以系统负荷和末端温度状态为共同依据，协调液冷回路、精密空调及冷源设备，使系统供冷能力与实际散热需求相匹配。具体而言，冷源层依据室外湿球温度和系统负荷优化冷却塔、冷却泵及冷水机组运行；CDU 层根据换热需求、一次侧压力和二次侧供液温度协调阀门与水泵；机柜层依据支路流量、压差和机柜温度调节动态平衡阀；气冷层根据冷热通道温度场优化精密空调台数、设定值和风量。漏液、水质、结露风险、最小流量和设备冗余作为刚性安全约束，AI 优化不覆盖底层保护逻辑。

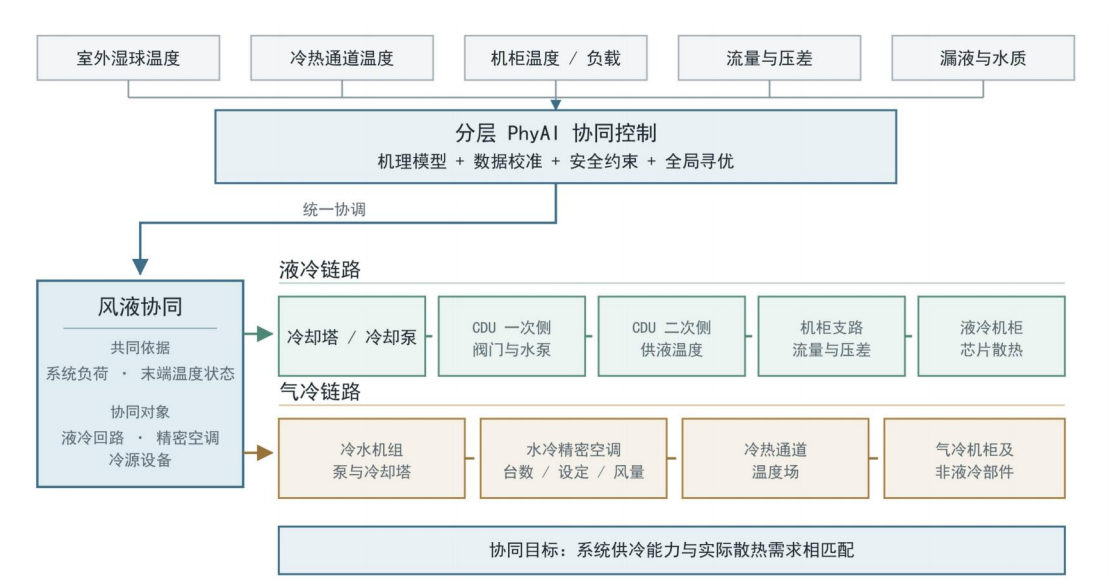


图 26 风液协同数据中心分层寻优架构

3.效益分析

根据短周期对照测试，优化工况与常规工况的室外平均温度差为 0.1~0.6℃，各组平均节电率达到 10.3%。

（六）高效机房仿真设计优化实践

1.项目概况

某新建制冷站房项目，拟建设采用 2 台 1000RT 定频离心冷机、1 台 1000RT 变频离心冷机和 1 台 500RT 变频螺杆冷机；冷冻侧采用并联系统，冷却侧采用串联系统并设置联通管路备用，冷却塔集中布置。项目在设计期引入高效机房仿真优化设计技术，对全年负荷、设备性能、系统运行策略和能效水平进行数字化仿真。

2.AI 技术应用

1) 全年 8760h 逐时负荷模拟与数字化建模

通过将设计工况由单一峰值负荷扩展为全年动态负荷曲线，使后续设备运行模拟能够覆盖不同季节和不同部分负荷工况，为设备选型及全年能效判断提供更接近实际运行边界的数据基础。结合既有物理 AI 通用模型框架，分别建立冷水机组、冷冻水泵、冷却水泵、冷却塔以及系统管路模型，构建可仿真的数字孪生冷站系统。

2) 常规控制与 AI 优化控制策略对比

在同一负荷、设备和系统边界下，项目分别配置常规控制模式和

优化控制模式。常规模式按照设计控制及常规工控逻辑运行，优化模式采用基于物理 AI 的节能优化控制算法，以综合比较不同设备组合及运行参数，寻找系统总能耗更低的运行方案。

3) 系统级仿真与过程自动寻优

完成负荷、设备、管路和策略配置后，设计仿真工具对常规运行与优化运行进行对比计算，并支持单日、全年及全年典型日等不同模拟范围。过程分析功能能够展示特定时刻冷站的全局自动寻优过程，用于校核模拟数据合理性，并输出综合最优运行参数配置。

3.效益分析

基于数字孪生仿真设计优化分析，在承担相同总冷量的条件下，常规模式系统综合 EER 为 4.87；采用优化控制方案后，系统综合 EER 提升至 5.31，设计能耗降幅约 8.47%。



图 27 系统模拟过程运行策略

八、前沿展望：未来技术趋势与演进路径

在大模型与多模态 AI 技术持续迭代驱动下，算力需求持续增长，算力基础设施加速向高密度、低碳化、智能化、规范化演进。绿色算力发展不再局限于单点技术节能改造，而是进入技术创新与制度供给双向驱动、协同赋能的全新阶段。行业正从设备级节能优化，迈向技术革新、应用升级与产业制度体系完善的系统性低碳转型。

绿色算力产业未来将依托全链条制度供给，实现从技术节能向制度赋能、从经验管控向标准量化、从单点优化向系统治理的进阶升级。

当前行业已具备能效、碳排、绿电消纳等多元评价维度，但整体体系较为零散、尚未形成闭环治理机制。未来产业制度建设将重点构建覆盖算力基础设施、硬件设备、软件算法平台的全生命周期评价标准体系，突破传统单一 PUE 评价局限，将设备能效、算法调度效率、全周期碳足迹、余热利用、资源复用率、平台优化水平纳入综合评价框架，建立适配技术迭代的动态能效阈值机制。在核算确权层面，行业将打通电力、算力设施、服务产出三级核算链条，构建从算力中心级、任务级至单次调用级的精细化碳强度核算方法，依托绿证小时级溯源、绿色 Token 分层测算，形成可落地、可对接国际的确权体系。同时依托生产者责任延伸制度、绿色金融工具完善设备更新与退役回收机制，配套跨区域电力交易、输配电价、算力差异化收费等制度，支撑绿电权益与算力负载跨域协同流转，并稳步推进国内标准与国际软件碳强度规范对接互认，构建标准化、市场化、国际化的绿色算力产业治理

体系。

超高密度万卡、十万卡级智算集群规模化落地，持续推动算力基础设施冷却与协同调控技术体系系统性革新。现阶段主流 CDU 间接液冷技术逐步触及超高密算力散热瓶颈，存在散热均匀性不足、输配能耗偏高、运维成本攀升等问题，倒逼冷却技术向极致散热与能源耦合利用方向演进。未来冷却技术将逐步向浸没式冷却、相变散热、微通道散热、芯片级直冷等技术方案升级，通过缩短热传导路径、降低风泵能耗，持续逼近硬件散热极限。冷却系统功能也将实现从单一散热配套向“散热+储能+余热资源化”一体化单元转型，深度融合余热采集、输配与复用链路，实现算力废热向园区供暖、工业热源高效转化。与此同时，算电冷协同、储能 AI 智慧调控技术持续成熟，依托模型预测控制、AI 动态寻优、云边端分层管控，实现算力负载、制冷工况、储能充放电、新能源出力的多变量联动优化，形成高密度、高可靠、高效能、可资源化的新一代绿色算力底层技术架构。

绿色算力应用逐步将摆脱传统人工建模、经验调参的静态管控模式，依托 AI for AI 范式打造具备自主感知、智能推演、动态寻优、持续迭代的自进化算力底座。大模型训练与推理任务具备强潮汐、高波动、强耦合特征，传统固定策略难以适配复杂动态工况，而 AI 赋能体系可依托数字孪生仿真、实时运行数据迭代与强化学习算法，自主探索算力调度、制冷控制、储能时序匹配的最优运行工况，挖掘人工经验难以覆盖的精细化节能空间。在业务应用层面，行业将建立单

位任务资源占用、算力减排效率等量化评价机制，推动算法优化、软件迭代与硬件更新协同降碳，实现同等业务产出下资源能耗持续下降。同时结合故障自主诊断、预测性运维、动态负载自适应调控，构建算力—电力—冷却—储能全域协同闭环，在严格保障业务 SLA 约束的前提下，持续降低 PUE 与碳排放，最终实现绿色算力从被动节能、人工管控向主动寻优、自主进化的应用形态跃迁，支撑 AI 产业长期低碳高质量发展。