Translation

The following Chinese government announcement lists 22 compute-related technologies that it wishes to make progress on by 2026, and invites qualified companies and research institutes to apply for subsidies and other government support in developing these technologies. The overall goal of this effort is to make China's data centers faster, more powerful, more reliable, and more energy-efficient, chiefly to support AI model training and inference.

Title
Notice of the Office of the Ministry of Industry and Information Technology on Organizing and Launching the Computing Power Infrastructure Enhancement Challenge Campaign
工业和信息化部办公厅关于组织开展算力强基揭榜行动的通知

Author
The Office of the Ministry of Industry and Information Technology (工业和信息化部办公厅)

Source
Website of the Central People's Government of the People's Republic of China (中国政府网). The notice was formulated on February 18, 2025 and publicized on February 20, 2025.

This translation is a combination of two related documents: the notice on the launching of the challenge campaign, and the list of tasks for the challenge campaign.
The Chinese source texts for the notice and the list are available online, respectively, at:
https://www.gov.cn/zhengce/zhengceku/202502/content_7005004.htm
https://www.gov.cn/zhengce/zhengceku/202502/P020250222303340603758.pdf
Archived versions of these two Chinese source texts are available online, respectively, at:
https://perma.cc/2SZX-WWCP and https://perma.cc/7UTU-A6BH

Translation Date	Translator	Editor
May 29, 2025	Etcetera Language Group, Inc.	Ben Murphy, CSET Translation Manager

Notice of the Office of the Ministry of Industry and Information Technology on Organizing and Launching the Computing Power Infrastructure Enhancement Challenge Campaign

Communication of the Office of the Ministry of Industry and Information Technology [2025], No. 55 (工信厅通信函〔2025〕55号)

To the main oversight departments (主管部门) for industry and information technology of, and to the communications administrations (通信管理局) of, all provinces, autonomous regions, and province-level municipalities, as well as the relevant central enterprises (中央企业):

In order to consolidate the foundation for the development of the computing power (compute) network, accelerate the application of innovative technologies and products, promoting the systemic development of the "points, chains, networks, and the totality"[1] of the compute network, we hereby launch the Computing Power Infrastructure Enhancement Challenge Campaign (算力强基揭榜行动). The relevant matters are detailed as follows:

I. Challenge (揭榜) Tasks

Focusing on six key areas of the compute network—computing, storage, networking, applications, low environmental impact (绿色), and security—this campaign seeks to identify a group of enterprises and public institutions[2] that possess key and core technologies (关键核心技术) and demonstrate strong innovation capabilities in order to achieve breakthroughs in a number of landmark technology products and solutions.

In the area of compute, efforts will focus on tackling technologies such as intelligent compute[3] management and compute acceleration to improve compute performance and efficiency. In the area of memory, technologies such as multimedia storage device management and cross-domain memory resource pool coordination will be developed to enable reliable and flexible storage of massive data. In the area of networking, breakthroughs will be made in technologies such as intra-compute

[1] Translator's note: China's Ministry of Industry and Information Technology (MIIT; 工业和信息化部; 工信部) uses the slogan "points, chains, networks, and the totality" (“点、链、网、面”) to refer to its vision for improving China's computing power ("compute") infrastructure. "Points" refers to improving the quality of each individual data center, including reducing overcapacity, meeting the needs of the market, and improving energy efficiency. "Chains" refers to improving the technology used across data centers, enhancing software and hardware ecosystems, and ensuring conformity with common standards. "Networks" refers to engaging Chinese telecom providers to improve the speed and stability of networks. And "the totality" refers to promoting new use cases and application scenarios for compute across all key industries.

[2] Translator's note: "Public institutions" (事业单位) are organizations created and led by Chinese government departments that perform a variety of functions. Notably, universities, research institutes, and hospitals in China tend to be public institutions. Unlike state-owned enterprises (SOEs), public institutions do not create material products and are non-profit. Public institutions are not considered government agencies, and their employees are not civil servants. Most public institutions are fully or partially government-funded, but some fully privately funded (but still government-led) public institutions exist.

[3] Translator's note: "Intelligent computing power" ("intelligent compute"; 智能算力; 智算) typically refers to computing power specifically designed and optimized for artificial intelligence (AI) model training, inference, or use.

networking (算内网络) and inter-compute networking (算间网络) to promote high-speed interconnection of compute resources. In the area of applications, deeper integration between compute and industries will be pursued to enable convenient compute usage across multiple scenarios. In the area of low environmental impact, technologies such as new cooling methods and carbon emissions-aware optimization (碳排放感知优化) will be developed to drive energy conservation and carbon emission reduction in compute infrastructure. In the area of security, technologies such as intelligent monitoring and operations and maintenance (O&M) robots will be promoted to ensure the reliable operation of compute centers.

II. Applications and Recommendations

(1) Applicant entities must be enterprises or public institutions registered within the People's Republic of China (PRC),[4] possessing independent legal person status, and demonstrating strong capabilities in technological innovation and industrialized applications. Applicants shall select challenge tasks from the *Computing Power Infrastructure Enhancement Challenge Campaign Task List* (see appendix) and must commit to completing the selected tasks within the designated time frame. Each entity may apply for no more than three projects. In the case of joint applications by relevant enterprises, universities, or research institutions, there shall be one lead unit and no more than four jointly participating units.

(2) The industry and information technology main oversight departments and communications administrations of all provinces, autonomous regions, and province-level municipalities, as well as relevant central enterprises, shall organize relevant units to actively apply for the challenge in accordance with the principles of voluntary enterprise participation under government guidance. As recommending units, they shall follow the principles of openness, fairness, and impartiality in reviewing, selecting, and recommending projects that possess outstanding innovation capacity, strong industrialization prospects, and significant industry-driving effects, and report them to the Ministry of Industry and Information Technology (Department of Information and Communications Development).

III. Work Procedures and Requirements

(1) Applicant entities shall apply through the designated system (https://gs.hcp.ac.cn). After completing registration, they shall fill out the required application. The application deadline is March 15, 2025.

---

[4] Translator's note: The Chinese word 境内 *jìngnèi*, translated throughout as "within the People's Republic of China (PRC)," literally means "inside the borders [of mainland China]." China considers Hong Kong, Macao, and Taiwan to be part of China but not to be "within the PRC."

(2) Industry and information technology departments and communications administrations of all provinces, autonomous regions, and province-level municipalities, as well as relevant central enterprises, acting as recommending units, shall log in to the system and confirm the list of recommended applicants by March 31, 2025 (the account password can be obtained from the contact person). In principle, each recommending unit may recommend no more than three projects per direction, with the total number of recommended projects across all directions not exceeding 20. Recommending units are encouraged to provide increased support to the applicants they recommend, in terms of policies, funding, and resource allocation, in accordance with actual conditions.

(3) The Ministry of Industry and Information Technology (MIIT) will organize the selection process and announce the list of selected challenge participants. Upon completing their challenge tasks (within no more than two years from the date of announcement), MIIT will commission third-party professional institutions to conduct evaluations and will select the winning participants based on merit (in principle, no more than three winners will be selected per challenge direction). MIIT will coordinate and leverage various resources to support the selected and winning participants and promote the demonstration and application of outstanding achievements.

IV. Contact Information

Contact person and telephone:

MIIT Department of Information and Communications Development (信息通信发展司): 010-68206162

Technical support unit:

China Academy of Information and Communications Technology (CAICT): 13701380040

Office of MIIT

February 18, 2025

Appendix

Computing Power Infrastructure Enhancement Challenge Task List

I. Compute

(1) Cloud–Edge–Terminal Compute–Network Coordinated Management System

Challenge Task: Develop a compute–network coordinated application management system for multi-tiered compute environments spanning clouds, edges, and terminals. Design management mechanisms tailored to different application software architectures to enable unified management across various types of software. Develop automated build and deployment capabilities for application software within compute–network coordination, supporting automatic builds and distribution. Study integrated observation capabilities for compute–network coordinated application systems to reduce O&M complexity and improve the operational stability and reliability of complex application software.

Expected Outcomes: By 2026, develop an application software management system that supports full lifecycle management of no fewer than five types of application software, including traditional software, cloud-native applications, artificial intelligence (AI) applications, and big data applications. Research distributed build and deployment technologies based on compute–network coordination to enable automatic distribution and deployment of the above software across compute nodes with zero manual intervention. Develop integrated observation functions for compute–network applications with white-box dynamic analysis and intelligent error origin pinpointing capabilities. Complete pilot verifications in no fewer than three industries.

(2) Integrated Intelligent Compute Platform for Training and Inference of Ultra-Large-Scale Parameter Models

Challenge Task: In response to the compute resource demands of training and inference for large-scale AI models, develop an integrated intelligent compute platform that supports ultra-large-scale parameter models. This platform should include resource scheduling strategies and training–inference acceleration toolkits, support multiple hardware architectures, abstract away underlying hardware differences, and improve the stability, resource utilization, and operational efficiency of ultra-large-scale models during training and inference processes.

Expected Outcomes: By 2026, develop an integrated intelligent compute platform that supports models with trillions of parameters. In ten-thousand-card[5] environments, ensure stable training durations of no fewer than 30 days and effective training times of no less than 95 percent. Training efficiency should be improved by no less than 30% over current mainstream levels, and inference efficiency should be improved by no less than 50%. The platform should support mainstream deep learning frameworks, be compatible with multiple hardware architectures, and provide a unified programming interface and development environment. The solution should be validated through deployment by no fewer than 10 industry users.

(3) Cross-Domain Task Orchestration System for Heterogeneous Compute

Challenge Task: Develop a cross-domain heterogeneous compute management system to enable coordination and application of heterogeneous compute resources across domains. Develop standardized and open interconnection functions for diverse compute types, supporting unified abstraction and encapsulation of different types of heterogeneous compute models. Develop cross-domain heterogeneous compute management functions to support unified management and coordination across domains. Research security authentication and control methods for multi-entity compute across domains to ensure secure cross-domain coordination.

Expected Outcomes: By 2026, develop no fewer than six cross-domain coordinated scheduling algorithms. Support deployment of compute tasks in no fewer than three scenarios, such as data processing, function compute (函数计算), and machine learning. Achieve unified management of no fewer than five cross-domain compute centers. Develop security authentication methods for multi-entity cross-domain compute to meet security requirements for coordination across compute tiers such as cloud, edge, and terminal. Complete pilot verification in no fewer than two industries.

(4) Integrated Compute All-in-One Solution for Training and Inference

Challenge Task: For AI training and inference scenarios, develop a high-performance integrated solution for training and inference based on Infrastructure as a Service (IaaS) and Platform as a Service (PaaS). The solution shall cover the entire process of large model development, training, evaluation, and deployment for inference. At the same time, it shall support capabilities such as large model encryption and cyber defense to address security issues and risks such as data leakage and instruction attacks (指令攻击) against large models.

---

[5] Translator's note: "Ten-thousand-card" (万卡) refers to clusters of more than 10,000 accelerator cards—such as graphics processing units (GPUs), tensor processing units (TPUs), or other specialized AI accelerator chips—that are used to accelerate AI model training and inference.

Expected Outcomes: By 2026, develop an all-in-one training–inference compute solution that supports chips based on at least three types of instruction set architectures. Carry out applications of the all-in-one solution in no fewer than five industries, provide users with diversified all-in-one training–inference services, and achieve deployment of the all-in-one AI training-inference solution in no fewer than ten different scenarios.

(5) Inference Acceleration Technologies for Large-Scale Heterogeneous Compute Clusters

Challenge Task: Develop coordinated optimization technologies for memory, networking, and computing and employ methods such as model acceleration and scheduling acceleration to accelerate inference of large models in large-scale heterogeneous compute clusters and thereby support larger models, longer context lengths, higher performance, and lower energy consumption and promote the improved application of compute chips in large model inference.

Expected Outcomes: By 2026, achieve at least a 500% increase in cluster goodput (有效吞吐量), at least a 100% increase in the number of requests processed in actual application scenarios, at least a 100% improvement in time to first token, and at least a 50% improvement in chip utilization. Break through limitations in inference acceleration for heterogeneous clusters exceeding 1,000 accelerator cards by optimizing the allocation of compute, storage, and network resources in compute centers, as well as topology structures and system scheduling strategies.

II. Storage

(6) Magneto-Optical-Electric Integrated Storage System

Challenge Task: In light of the current difficulties that prevent a single storage medium from meeting diverse data storage needs, leverage the differentiated characteristics of magnetic, optical, and electrical storage technologies in terms of performance, lifespan, and power consumption to develop a magneto-optical-electric integrated storage system (磁光电融合存储系统). Construct a multi-tiered storage architecture based on solid-state drives (SSD), hard disk drives (HDD), and optical storage. Based on business requirements, store data on different tiers of storage devices to realize centralized and unified management of massive data and support efficient, low-carbon, secure, and sustainable development of compute centers.

Expected Outcomes: By 2026, develop a magneto-optical-electric integrated storage system that supports at least three types of storage, such as distributed file, distributed block, and distributed object. The system shall support custom tiering strategies based on data access time, access frequency, and file attributes, and dynamically adjust data migration according to business loads. It shall reinforce underlying security capabilities through media security, system security, and software security, and enhance data security through ransomware protection, encryption algorithms, remote monitoring, and optical storage early warning and detection. Build application demonstrations of the integrated storage system, complete application deployment in no fewer than 20 business systems, and achieve data migration from at least four eastern regions to western-region integrated storage systems, with a cumulative data volume of no less than 10 PB.

(7) Storage Scheduling Management and Application Technologies

Challenge Task: In response to challenges posed by massive data storage and compute silos, develop systems for cross-domain multi-compute (多算) storage capacity (存立) scheduling, storage–network orchestration, and integrated storage–compute–network systems. Such systems should enable intelligent tiering of data into hot and cold categories, cross-domain seamless access to applications, and reductions in costs, improvements in performance, and enhanced business support. The systems should feature capabilities for resource planning and strategy adjustment, enabling optimization and dynamic reconfiguration of data storage layouts across the entire network to meet continuously changing demands.

Expected Outcomes: By 2026, develop a highly efficient and scalable storage system that uses intelligent algorithms to analyze and schedule data, enabling seamless application access and intelligent data mobility. Study storage capacity scheduling strategies to keep data recall rates below 30 percent. Research scheduling algorithms based on tidal network patterns to improve network bandwidth utilization by more than 50% and achieve the integration of storage and networking. Integrate capabilities across storage, compute, and networking to support unified scheduling of storage–compute–network resources, with implementation in compute center resource pools.

III. Network

(8) High-Performance Data Processing Unit (DPU)

Challenge Task: Launch research on integrated hardware–software DPU chip technologies based on chiplet architecture and fifth-generation reduced instruction set computer (RISC-V) technologies. The DPU shall support scenarios such as compute centers, intelligent compute centers, and supercomputing centers that require ultra-high bandwidth and ultra-low latency. Make breakthroughs in key technologies including heterogeneous chiplet packaging, high-speed Serializer/Deserializer (SerDes) communication, large-scale lossless network congestion algorithms, hardware cryptographic algorithms, high-performance virtualization, and hardware programmability. Enable DPU applications based on heterogeneous cores, such as ARM, x86, and RISC-V, to enhance compute center infrastructure processing capability and data transmission energy efficiency.

Expected Outcomes: By 2026, complete the development of an ultra-high-performance DPU chip with a throughput capacity of 400 Gbps and one-way traffic latency not exceeding 30 μs. The DPU shall support adaptation to mainstream domestic and international central processing unit (CPU) and graphics processing unit (GPU) platforms, compatibility with mainstream operating systems, and programmability of datagram processing logic in hardware.

(9) Intelligent Compute Network Based on RoCE

Challenge Task: Carry out research and development (R&D) on devices and control systems for Remote Direct Memory Access (RDMA) over Converged Ethernet (RoCE) networks. Improve RoCE network throughput and latency performance by increasing device bandwidth, optimizing load balancing algorithms, and enhancing network traffic planning and O&M capabilities. Develop a new generation of intelligent control tools incorporating large AI model capabilities to simplify the deployment and configuration of RoCE networks, and achieve global, multidimensional visualized O&M. Under conditions such as network fluctuations, service changes, or errors, enable automatic adjustment of network parameters and rapid traffic switching, thereby improving network efficiency and reducing O&M costs.

Expected Outcomes: By 2026, realize commercial deployment of a new RoCE network solution that improves network performance by more than 10%. Through intelligent management and O&M tools, significantly reduce the difficulty of network deployment and improve O&M efficiency by more than 50%, enabling support for larger-scale deployments and applications.

(10) Research and Validation of Optical Switching Technologies for Intelligent Compute Networks

Challenge Task: In response to the low power consumption, high bandwidth, and low latency requirements of intelligent compute clusters, carry out research and validation on key technologies for optical switching–based networking within such clusters. Focus on breakthroughs in optical switching networking and routing protocol adaptation for intelligent compute clusters. Conduct research on optical topology mapping, optoelectronic hybrid routing, and multipath load balancing to meet the requirements for functionality, performance, reliability, and scalability.

Expected Outcomes: By 2026, realize an optical network that supports intelligent compute clusters with ease of operation, high reliability, and smooth transitions and upgrades. The network shall support critical services such as AI. Optical switching equipment shall support per-port transmission rates of 100GE, 400GE, and 800GE, with elastically scalable switching capacity. The system shall support the interconnection of no fewer than three types of heterogeneous compute resources, complete validation in no fewer than two intelligent compute clusters, and complete validation for no fewer than three types of intelligent compute workloads.

(11) Research and Validation of Key Networking Technologies for Distributed Intelligent Compute Centers

Challenge Task: In response to the trend of intelligent compute clusters evolving from centralized to distributed deployment, carry out R&D on inter-center networking technologies. Develop highly reliable transmission equipment for communication between intelligent compute centers, and construct an ultra-high-capacity, ultra-low-latency, ultra-high-reliability optical–electrical coordinated network to achieve high-speed and reliable interconnection between intelligent compute centers.

Expected Outcomes: By 2026, achieve breakthroughs in key technologies for ultra-high-capacity, ultra-high-reliability data transmission between intelligent compute centers. Develop transmission equipment for inter-center networks with a single-wavelength transmission rate of no less than 1.6 Tbps and device latency not exceeding 30 μs, thereby supporting highly reliable data transmission between distributed intelligent compute centers.

IV. Applications

(12) Cross-Domain Interconnection Applications for Intelligent Compute Centers

Challenge Task: Optimize the layout of AI compute infrastructure to build ultra-large-scale AI compute service capabilities characterized by cross-regional complementarity and coordinated compute scheduling. Strengthen compatibility and adaptation with AI chip vendors, construct a large-scale, high-performance heterogeneous compute pool, and provide AI compute services optimized for large model training and inference scenarios, featuring elastic scheduling, elastic fault tolerance, and high resource utilization.

Expected Outcomes: By 2026, establish AI compute centers covering more than five nationwide key compute hubs and nodes (全国重点算力枢纽节点), capable of sensing and dynamically scheduling compute demand across regions and across clouds. Complete adaptation for more than three types of compute chips. Focus on large model training and inference scenarios to build a large-scale, high-performance compute resource pool with integrated training and inference capabilities, elastic scheduling, and high fault tolerance. The system should possess minute-level post-interruption training resumption (分钟级断点续训) capabilities and support graded parallel training on ten-thousand-card clusters.

(13) Compute–Power Coordination Applications

Challenge Task: Develop a multi-cloud heterogeneous compute–power coordination management platform based on compute scheduling technologies and energy large models. Construct a data-driven model of electrical load characteristics for compute clusters, as well as an energy large model based on the spatiotemporal transfer characteristics of compute tasks. Promote the application of compute forecasting and scheduling technologies in intelligent compute centers to improve overall resource utilization. Conduct coordinated optimization of compute load and power systems using new energy sources and new-style energy storage systems, to enable precise, dynamic, and real-time energy scheduling and trading and realize deep coordination between compute and various forms of energy such as electricity.

Expected Outcomes: By 2026, achieve end-to-end data penetration and process integration between energy and compute in intelligent compute scenarios. Establish direct control and indirect guidance mechanisms where "compute follows the pull of power" (“算”随“电”动). Achieve compute demand prediction accuracy of 70% and increase the cluster effective load rate by more than 25%. Improve overall cluster resource utilization in intelligent compute centers by 10%. Based on compute cluster electricity usage data, time cycles, weather data, and workloads, implement energy efficiency optimization and compute efficiency enhancement where "power follows compute use" (“电”随“算”用). Achieve decision-making accuracy for infrastructure energy use of over 85% and improve response timeliness by reducing response time on the order of 15 minutes. Improve compute energy efficiency in intelligent compute centers by 30%, and reduce electricity costs of compute centers by more than 5%.

(14) Compute Applications in Large-Scale Telecommunications Service Scenarios

Challenge Task: Focusing on network function virtualization (NFV) system architecture, address challenges in NFV related to network performance, resource utilization, and deployment flexibility. Develop technologies and systems for high-performance virtualization, intelligent network management, and resource orchestration algorithms tailored to the NFV architecture. Achieve breakthroughs in coordination between the virtualization layer and hardware accelerators such as field-programmable gate arrays (FPGAs), DPUs, and GPUs.

Expected Outcomes: By 2026, achieve intelligent scheduling of virtualized network functions within the NFV compute platform system. Support heterogeneous cluster deployment, dynamic expansion, and dynamic resource allocation, increasing virtualization resource utilization by more than 20%. Support virtualization scheduling of hardware accelerators such as GPUs and FPGAs, boosting network processing performance to Tbps levels. Support intelligent management of network virtualization functions, enhancing the automated O&M capabilities and management efficiency of the NFV system, and reducing error recovery time by no less than 30%.

V. Green and Low-Carbon Technologies

(15) Research and Application of Green Compute Technologies

Challenge Task: Focus on breakthroughs in green and energy-saving technologies in the compute field. Target key factors such as task scheduling characteristics, energy usage patterns, and load balancing requirements in compute systems to develop dynamic resource scheduling algorithms for green computing, energy consumption optimization management systems, and multi-scenario collaborative energy-saving mechanisms. Make breakthroughs in the intelligence level of energy-saving algorithms and improve energy utilization efficiency across multiple nodes in compute networks.

Expected Outcomes: By 2026, the energy consumption management system shall enable real-time monitoring and energy-efficient scheduling of compute centers and network nodes. Support dynamic frequency scaling and dynamic voltage adjustment at the compute node level, reducing average energy consumption per node by more than 30% while meeting the requirements of applications such as AI inference.

(16) Enterprise Green Computing Carbon-Awareness Platform

Challenge Task: Establish a carbon emissions measurement system for enterprises' compute centers that is capable of real-time, accurate calculation of carbon emissions across all of an enterprise's compute centers. The system shall also allocate carbon emissions to specific business departments, application scenarios, and workloads to enable fine-grained (精细化) carbon emissions management. Based on carbon emissions data, develop carbon-aware scheduling capabilities that shift workloads to lower-carbon compute centers without compromising service experience or continuity, thereby further reducing overall carbon emissions.

Expected Outcomes: By 2026, build an enterprise-grade green computing carbon-awareness platform for cross-domain compute centers at the scale of tens of millions of cores. Establish a set of industry-standard technical methods and metrics to precisely measure carbon emissions for different types of workloads. Through ecosystem co-construction (共建), form a standardized system for green metrics and benchmarks. Develop carbon-aware scheduling capabilities to achieve the goal of sourcing 30% of the energy used in compute centers from renewable sources.

(17) Cold-Plate Liquid-Cooled Native Rack-Scale Server

Challenge Task: For new-generation liquid-cooled compute centers, develop cold-plate liquid-cooled rack-scale systems, including liquid-cooled server nodes and passive liquid-cooled doors. Achieve breakthroughs in cooling technologies and architectural requirements for high-density and diverse compute loads. The equipment shall support blind-mate O&M for the power supply bus, network interconnection bus, and liquid cooling pipelines. It shall feature multi-level leak detection capabilities across the liquid-cooled rack and liquid-cooled servers to effectively reduce the scope and impact of interruptions in operations.

Expected Outcomes: By 2026, the liquid-cooled rack-scale server shall achieve 100% heat dissipation through liquid cooling, with a cooling power usage effectiveness (PUE) below 1.15.[6] The internal design of the rack-scale server shall have full blind-mate capability. The management module shall enable functions such as rack-scale power consumption management, leak detection, and asset management. Per-server (单柜) power output for general-purpose compute shall be no less than 20 kW, and per-server (单机柜) power output for intelligent compute shall be no less than 30 kW. Deployment of no fewer than 500 liquid-cooled nodes shall be achieved at scale.

(18) Energy Optimization and Tuning Platform for Compute Centers

Challenge Task: Develop a compute center PUE simulation platform with high accuracy, high simulation efficiency, and multi-scenario coverage. Achieve breakthroughs in key technologies such as physical mechanism model construction, simulation engine clustering, and automatic model generation, enabling fast and accurate assessment of compute center PUE at granular time segments under different operational states. Develop an AI-based energy conservation optimization system for compute center cooling systems using big data analytics technologies. Through automated data governance and automated inference, accurately match cooling demands and achieve overall dynamic real-time optimization of cooling systems under reliability constraints, thereby optimizing compute center PUE.

---

[6] Translator's note: Power usage effectiveness (PUE) is a metric used to determine the energy efficiency of a data center. PUE is the ratio between the total energy a facility consumes and the energy specifically used by the facility's IT equipment. The closer PUE is to 1, the more efficient the data center.

Expected Outcomes: By 2026, support energy efficiency optimization for at least two typical cooling scenarios, such as liquid cooling and water cooling. Support joint simulation of cooling systems and power distribution systems. The system shall output PUE operating curves and simulated equipment operating conditions under different load and operational scenarios, with a PUE simulation accuracy of 97% or higher. Based on the energy efficiency optimization platform, support AI-based automatic inference and hourly strategy dispatch, enabling visibility, management, and control of compute center energy consumption. Through AI-driven energy efficiency optimization, improve compute center PUE by more than 5%. By linking infrastructure and information technology (IT) for energy saving, achieve a total energy consumption reduction of more than 5%. Deploy the solution in more than five compute centers.

(19) New-Generation Cooling System

Challenge Task: Develop an AI-based energy-saving system for the operational regulation and environmental monitoring of compute center infrastructure. Come up with a brand-new adaptive algorithm to overcome the limitations of existing conventional algorithms, enhance data analysis and processing performance, and build an AI algorithm database based on expert knowledge to improve the energy efficiency, reliability, and cost-effectiveness of functions such as energy consumption management, energy scheduling, security monitoring, failure diagnosis, and O&M assistance.

Expected Outcomes: By 2026, on the basis of meeting cooling requirements, improve the reliability and adaptability of cooling systems, and enhance the efficiency of energy use, water use, and O&M, with a power conservation rate of more than 10%. Support capabilities such as data collection, labeling, governance, and storage for cooling systems. Provide functions including abnormal system operation alerts, alert aggregation, automatic diagnosis, remote communication, and automatic control. Support core capabilities for intelligent tuning and intelligent control of cooling systems. Implement AI-based energy-saving tuning solutions in no fewer than five real-world business scenarios.

VI. Security and Reliability

(20) Intelligent O&M Robots for Compute Centers

Challenge Task: Develop intelligent O&M robots for compute centers, along with an intelligent robot management platform. Based on a cloud–edge–terminal three-layer architecture, optimize technologies and algorithms for intelligent robots in areas such as equipment and facility identification across multiple floors and rooms within building data centers, multimodal environmental sensing, precise spatial positioning, intelligent human–computer coordination, and multi-task joint scheduling. Support the deployment of robots in typical scenarios such as facility O&M and IT operations within compute centers to improve inspection quality and promote O&M cost reduction and efficiency gains.

Expected Outcomes: By 2026, achieve coordinated deployment of intelligent robots across multiple rooms and floors within large-scale compute centers. The robot inspection task success rate shall not be lower than 95%; equipment recognition accuracy shall reach 97%; and the recall rate for environmental inspections shall be no less than 90%, ensuring continuous operation of inspection tasks in compute centers. Achieve cloud–edge–terminal coordinated scheduling, support autonomous operations in diverse scenarios, improve concurrent task execution efficiency, and promote the construction of a stable, secure, reliable, and controllable intelligent O&M system for compute centers.

(21) Integrated Cloud–Edge–Terminal Intelligent Monitoring Platform

Challenge Task: Develop a high-performance integrated system spanning cloud, edge, and terminal. Design a centralized monitoring, management, and O&M inspection solution that combines hardware—such as intelligentized (智能化) terminals or robots—with software that integrates multiple algorithmic models and platform tools. Achieve breakthroughs in key technologies including multi-level automated O&M, multi-dimensional diagnostics, multi-platform coverage, and multi-model quantization. Establish a digital evaluation system and model for comprehensive O&M and health status information, enabling total life cycle digital management of compute infrastructure across planning, design, construction, deployment, operation, and maintenance.

Expected Outcomes: By 2026, establish intelligentized O&M capabilities for large-scale clusters. Devices shall be capable of identifying cross-platform and system stability risks and security risks. Using technologies such as comprehensive video recognition, enable structured alert aggregation and dispatch, with an accuracy rate exceeding 98%. Realize digitalized linkage across the total life cycle of compute infrastructure. The platform shall support automated workflow advancement and enable cloud-terminal direct control (云端直控) over more than 10 compute centers. Implement digital health evaluations of compute centers, autonomous inspections by intelligentized terminals or robots, linkage between video stream recognition and alerting, and intelligentized Q&A for O&M. Ensure an operational service-level agreement (SLA) compliance rate of 99% or higher.

VII. Other

(22) Other specialized technologies, products, services, or platforms in the compute domain that demonstrate advanced technology, a high level of technological maturity, and strong industrialization prospects.

Appendices: 1. Recommendation Form for Compute Infrastructure Enhancement Challenge Applicants

2. Application Materials for Compute Infrastructure Enhancement Challenge Applicants

Appendix 1

Recommendation Form for Compute Infrastructure Enhancement Challenge Applicants

Recommending Unit (official seal):

No.	Organization Name	Task Category	Challenge Product	Reason for Recommendation	Contact Person	Contact Method
1						
2						
3						
...						
...						
...						
...						

Notes: 1. This form shall be completed by recommending units such as local governments or central enterprises.

2. Recommending units shall rank submissions in order of priority.

3. “Task Category” refers to one of the 22 key tasks listed in the Challenge Task List.

Appendix 2

Application Materials for Compute Infrastructure Enhancement Challenge Applicants

Challenge Direction:________________________

Applicant Entity: ________(official seal)_______

Recommending Unit: ____(official seal)_______

Date of Application: ____(MM/DD/YYYY)_____

Instructions

1. The applicant entity shall carefully read the relevant instructions for the Application Materials for Compute Infrastructure Enhancement Challenge Applicants and complete each section truthfully and in detail.

2. Unless otherwise specified, no fields in the application form may be left blank. Where supporting materials are required, please attach the corresponding documents.

3. The applicant must possess intellectual property rights for the product that is the subject of the application. The applicant is responsible for the authenticity of all submitted materials and must make a valid commitment to completing the key challenge task as planned by signing an Enterprise Commitment Statement (see the Challenge Task Commitment Letter template).

I. Basic Entity Information				
Organization name				
Challenge lead	Name		Title/position	
	Email		Contact method	
Application contact person	Name		Contact method	
	Email		Fax	
Legal representative			Registered capital	
Organization address				
Organization code / three-in-one license code				
Entity type	☐ State-owned enterprise ☐ Private enterprise ☐ Foreign-funded enterprise ☐ Public institution ☐ Other (please specify):________________________			
Publicly listed?	☐ No ☐ Yes			
Total employees			Number of R&D personnel	
Brief introduction of the applicant entity	Include basic information such as founding date, main business areas, main products, technological capabilities and development history, as well as relevant patents, standards, intellectual property rights, and awards from competitions (Attach relevant supporting documentation) (This section shall not exceed 500 words)			
Participating entity 1				
Participating entity 2				
......				
II. Challenge Task Information				

Challenge task direction	**Compute** ☐ Direction 1: cloud–edge–terminal compute–network coordinated management system ☐ Direction 2: integrated intelligent compute platform for training and inference of ultra-large-scale parameter models ☐ Direction 3: cross-domain task orchestration system for heterogeneous compute ☐ Direction 4: integrated compute all-in-one solution for training and inference ☐ Direction 5: inference acceleration technologies for large-scale heterogeneous compute clusters **Storage** ☐ Direction 6: magneto-optical-electrical hybrid storage system ☐ Direction 7: storage scheduling management and application technologies **Network** ☐ Direction 8: high-performance data processing unit (DPU) ☐ Direction 9: intelligent compute network based on RoCE ☐ Direction 10: research and validation of optical switching technologies for intelligent compute networks ☐ Direction 11: research and validation of key networking technologies for distributed intelligent compute centers **Applications** ☐ Direction 12: cross-domain interconnection applications for intelligent compute centers ☐ Direction 13: compute–power coordination applications ☐ Direction 14: compute applications in large-scale telecommunications service scenarios **Green and low-carbon technologies** ☐ Direction 15: research and application of green compute technologies ☐ Direction 16: enterprise green computing carbon-awareness platform ☐ Direction 17: cold-plate liquid-cooled native rack-scale server ☐ Direction 18: energy optimization and tuning platform for compute centers ☐ Direction 19: new-generation cooling system **Security and reliability** ☐ Direction 20: intelligent O&M robots for compute centers ☐ Direction 21: integrated cloud–edge–terminal intelligent monitoring platform **Other** ☐ Direction 22: other technologies, products, services, or platforms in the compute domain
Overview of the challenge product	Include a description of the product/service submitted for the challenge, overview of investment and financing, the maturity of the relevant R&D and applications, and the expected technical and industrial application level to be achieved by 2026 (If multiple products are involved across different domains, please describe them separately). (Limit: 1,000 words)

Challenge Task Application

I. Challenge Task Introduction

(1) Name and Brief Description of the Challenge Task

(2) Scope of Application / Intended Use

Describe the expected scope of application, usage scenarios, and target users.

(3) Task Value and Benefits, Etc.

Include anticipated economic benefits, social utility, and other relevant outcomes.

II. Existing Capabilities and Relevant Progress of the Applicant Entity

(1) Existing Capabilities

Describe the applicant's position in the industry, scientific research qualifications (e.g., high-tech enterprise, enterprise technology center, key laboratories), technological foundation, talent and team capabilities, and main advantages.

Summarize the applicant's innovation capabilities, such as publications, patents, software copyrights, standards, monographs, and awards won at competitions.

Include the qualifications and professional experience of the challenge lead.

Describe any national-level projects previously undertaken by the project team.

(2) Relevant Progress

Outline the applicant's current technological level (in comparison with international standards), innovation and application status, relevant R&D personnel, and capital investments.

III. Key R&D Objectives and Plan

(1) Expected Outcomes by 2026

Include key technical indicators, functional indicators, and performance indicators, and their definitions, as well as testing scenarios and evaluation methods.

(2) Implementation Plan for Key Tasks

Include timelines, milestone-related tasks, and detailed objectives.

(3) Organizational Support Mechanisms

Describe the project team, organizational structure, coordination mechanisms, any industry-academia-research institute-user (产学研用) collaboration (e.g., existing capabilities and support capabilities of participating entities), and collaborative innovation capacity (e.g., team member cooperation on projects, joint laboratories).

(4) Potential Problems and Mitigation Measures

IV. Other Relevant Information

Note: In general, this application form should not exceed 5,000 words, with a focus on the objectives and implementation plan. If applying for projects in multiple fields, please submit separate applications for each field, based on this template.

Supporting Documents for the Applicant Entity

1. Proof of R&D capabilities (supporting documents related to patents, standards, intellectual property rights, and participation in national key R&D programs or "winner-takes-all open competition"[7] projects)

2. Proof of honors and recognitions (supporting documents proving status as a high-tech enterprise, enterprise technology center, or key laboratory, and awards won at competitions)

3. Proof of current performance indicators and application outcomes for the proposed product/service (such as third-party testing materials)

Challenge Task Commitment Letter

In accordance with the requirements set forth in the *Notice of the Office of the Ministry of Industry and Information Technology on Organizing and Launching the Computing Power Infrastructure Enhancement Challenge Campaign*, our entity has submitted the XXX task for evaluation.

We hereby make the following commitments:

1. Our entity is responsible for the authenticity of all submitted materials. We guarantee that we hold intellectual property rights to the submitted products and application solutions and that the submitted products and services comply with relevant national laws, regulations, and industrial policy requirements.

2. The products and services submitted by our entity comply with national secrecy protection (保密) regulations and do not involve state secrets, personal privacy,

[7] Translator's note: The idea behind "winner-takes-all open competition" (揭榜挂帅), in the context of Chinese science and technology projects, is that the government openly lists the technological breakthrough(s) it desires. Any individual or group in society, not just a select few, are then eligible to win a cash award if they succeed in making the breakthrough.

or other sensitive information.

3. The text and images in the relevant materials have been reviewed by our entity and have been confirmed to be accurate.

Our entity shall bear full legal responsibility for any consequences arising from violations of the above commitments.

Our entity will, in accordance with the requirements of the challenge task implementation plan, enhance overall awareness, and sincerely assume primary responsibility. We will organize diligently, focus on making advancements, and strengthen assurance during the implementation of the challenge task. We will make every effort to accomplish the key R&D objectives and strive to achieve substantive progress by the end of 2026, reaching or exceeding the expected outcomes.

Contact person:

Telephone number:

Legal representative: (Signature)

Company (Official seal)

______________________ 2025 (Date)