ICS 35.240.50
CCS L 67

NATIONAL STANDARD OF THE PEOPLE'S REPUBLIC OF CHINA

GB/Z 203-2026

Artificial intelligence - Technical requirements of industrial foundation model

人工智能 工业大模型技术要求

Issued on: July 30, 2026 **Implemented on: July 30, 2026**

Issued by: State Administration for Market Regulation;
Standardization Administration of China

Table of Contents

Foreword

This document was drafted in accordance with the rules given in GB/T 1.1-2020, Directives for standardization - Part 1: Rules for the structure and drafting of standardizing documents.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. The issuing body of this document is not to be held responsible for identifying any or all such patent rights.

1 Scope

This document specifies the technical requirements for industrial large models, covering

pre-training, adaptation and fine-tuning, inference and interaction, and deployment and implementation.

This document applies to the development, adaptation and deployment of industrial large models.

2 Normative references

The following documents contain provisions which, through normative reference in this text, constitute indispensable provisions of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document, including any amendments, applies.

GB/Z 204-2026, Artificial intelligence - System architecture of industrial large models

GB/T 45288.1-2025, Artificial intelligence - Large models - Part 1: General requirements

3 Terms and definitions

For the purposes of this document, the terms and definitions given in GB/T 45288.1-2025 and GB/Z 204-2026 and the following apply.

3.1 industrial multimodal data - a complex data set generated in an industrial production and manufacturing environment and containing information of several types. Note: it includes industrial time-series data, CAD, CAE and CAPP files, machine instructions, industrial software code, industrial images, audio and video, and industrial documents.

3.2 industrial mechanism knowledge - in the industrial field, the understanding of the intrinsic operating laws and the causal relations of products, equipment, processes and production systems, formed on the basis of knowledge of fundamental disciplines such as physics, chemistry, materials science and mechanics, and validated by long production practice and experiment. Note 1: knowledge of fundamental disciplines includes physical laws, chemical equations, process specifications and equipment mechanisms. Note 2: domain knowledge includes equipment structure, failure modes and process flows.

4 Abbreviated terms

For the purposes of this document the following abbreviated terms apply. CAD: Computer Aided Design. CAE: Computer Aided Engineering. CAPP: Computer Aided Process Planning. CPU: Central Processing Unit. FPGA: Field-Programmable Gate Array. GPU: Graphics Processing Unit. LoRA: Low-Rank Adaptation. NPU: Neural Processing Unit. RAG: Retrieval-Augmented Generation. TPU: Tensor Processing Unit.

5 Technical requirements for the pre-training of industrial large models

5.1 Industrial self-supervised pre-training: a) self-supervised pre-training on single-modality industrial data shall be supported, including industrial time-series data, CAD, CAE and CAPP files, machine instructions, industrial software code and industrial images, audio and video; b) cross-sector pre-training capability shall be provided, with data from fields such as aviation, steel, petrochemicals and electric power, so as to strengthen the generalization of the model across fields; c) distributed training capability shall be provided, adapted to distributed training methods such as data parallelism, model parallelism and hybrid parallelism, so as to strengthen the perception of diverse data; d) multi-task learning, domain-adaptive learning and interference-resistant learning capabilities should be provided, so as to strengthen the adaptability and reliability of the model in many scenarios; e) the capability to augment small-sample industrial data should be provided, such as the generation and augmentation of industrial time-series data.

5.2 Industrial multimodal pre-training: a) an encoder-decoder for industrial multimodal data shall be provided; b) joint pre-training on industrial multimodal data shall be supported, covering at least two of industrial time-series data, CAD, CAE and CAPP files, machine instructions, industrial software code, industrial images, audio and video, and industrial documents; c) an efficient attention mechanism shall be provided, strengthening the semantic association among industrial multimodal data; d) contrastive learning and masked prediction should be supported, raising the capability to express industrial features.

5.3 Joint representation and alignment of industrial multimodal data: a) the ingestion of structured, semi-structured and unstructured industrial data in several formats shall be supported; b) the capability to align the different modalities within industrial multimodal data shall be provided; c) cross-modal feature fusion capability should be provided, supporting the fusion of more than two modalities and raising the capability of the model to represent cross-modal industrial scenarios.

6 Technical requirements for the adaptation and fine-tuning of industrial large models

6.1 Fine-tuning with industrial sector knowledge: a) the model parameters shall be adjusted through an adaptation layer, together with the sector knowledge base, so as to strengthen domain adaptability; b) incremental fine-tuning with sector-specific data and knowledge shall be supported, optimizing the performance of the model; c) cross-domain adaptation capability shall be provided, raising the generality of the model across sector scenarios; d) efficient parameter fine-tuning methods, such as LoRA and Adapter, shall be supported, raising the efficiency of adaptation.

6.2 Industrial task-oriented fine-tuning: a) at least one of the tasks of intelligent question answering, scenario cognition, process decision, terminal control, content generation and scientific discovery shall be supported; b) instruction sets shall be designed according to the needs of the different tasks, including but not limited to equipment health diagnosis and

maintenance, production scheduling optimization, product quality inspection, checking of personnel operating practice, detection of hidden hazards in safe production and the regulation of process industry parameters, and the model fine-tuned on that basis, with not fewer than five kinds of task instruction set; c) joint fine-tuning over more than two tasks should be provided, strengthening the task generalization of the model.

6.3 Embedding industrial mechanism knowledge: a) knowledge of the fundamental disciplines shall be converted into a form the model can understand; b) the introduction of industrial mechanism knowledge constraints into the loss function or the reward function shall be supported, ensuring that the output result conforms to scientific principles; c) the embedding of domain knowledge by means of a knowledge graph shall be supported; d) an explainability mechanism should be used, showing the part played by the industrial mechanism constraints during the inference of the model.

7 Technical requirements for the inference of industrial large models

7.1 Industrial retrieval-augmented generation. The industrial large model shall support integration with external knowledge bases, using techniques such as RAG to raise the accuracy, the currency and the explainability of the inference result, specifically: a) efficient connection to and information retrieval from structured knowledge bases, such as knowledge graphs and databases, and unstructured ones, such as documents and manuals, shall be supported; b) retrieval-augmented generation capability shall be provided, fusing real-time retrieval results from the knowledge base when generating an answer or a decision recommendation; c) a dynamic update and version management mechanism for the knowledge base shall be provided; d) an efficient vector retrieval mechanism based on semantic indexing should be supported, guaranteeing the relevance and the speed of knowledge retrieval; e) a knowledge provenance and citation mechanism should be provided, marking the source of the knowledge in the inference or generation result and so raising traceability and credibility.

7.2 Collaborative inference between large and small industrial models. The technology shall support a mechanism for the industrial foundation large model and dedicated small models to work together, selecting or combining models dynamically to carry out the inference task according to the complexity of the task and the computing resources on site, specifically: a) a model selection strategy based on the needs of the task, the complexity of the input content and the response latency requirement shall be supported, so that models are called as needed; b) a unified interface and scheduling framework shall be provided, supporting collaborative inference and output fusion across models of different scale; c) cascaded and parallel inference modes shall be supported, raising the response speed and the overall performance; d) a cloud-edge collaborative inference mode should be supported; e) collaboration strategies such as voting, weighting, arbitration and scheduling should be supported, fusing the inference processes and results of several models; f)