

ICS 35.240
CCS L 70



**NATIONAL STANDARD OF THE
PEOPLE'S REPUBLIC OF CHINA**

GB/Z 225-2026

**Artificial intelligence - Evaluation metrics and methods of
large-scale models for petroleum and petrochemical industry
(English Translation)**

Issued on: August 27, 2026

Implemented on: January 1, 1800

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

Table of Contents

Preface	
1.Scope	1
2 Normative References	1
3.Terms and Definitions	1
4.Abbreviations	2
5 Evaluation Framework	2
6.Evaluation Task	3
6.1 Industry Capabilities	3
6.2 Security Capability	14
7.1 Industry Capability Indicators	15
7.2 Safety Capability Indicators	21
8 Evaluation Data	22
8.1 Dataset Construction	22
8.2 Dataset Update	23
8.3 Data Quality Control	24
8.4 Evaluation Dataset Sampling	24
9.Testing Environment	25
9.1 Basic Requirements	25
9.2 Hardware Environment	25
9.3 Software Environment	25
9.4 Network Environment	25
9.5 Security and Backup	25
10 Evaluation Tools	25
10.1 Evaluation execution	25
10.2 Result score	25
10.3 Report generated	26
11 Evaluation implementation	26
11.1 Grading Assessment	26
11.3 Evaluation Methods	26
11.4 Comprehensive Assessment	27
11.5 Evaluation Management	27
Appendix A (Normative) Evaluation Index Calculation Formula	28

Foreword

This document is a standard or guiding technical document.

This document complies with the provisions of GB/T 1.1-2020 "Standardization Work Guidelines Part 1. Structure and Drafting Rules of Standardization Documents".

Drafting.

Please note that some content in this document may involve patents. The issuing organization of this document assumes no responsibility for identifying patents.

This document was proposed and is under the jurisdiction of the National Information Technology Standardization Technical Committee (SAC/TC28).

1 Scope

This document establishes the evaluation framework for a large-scale model of the oil and petrochemical industry, specifies the evaluation tasks, evaluation indicators, and evaluation data, and describes the evaluation process.

method.

This document is intended to guide assessment organizations in evaluating and testing the capabilities of large-scale oil and petrochemical models across multiple dimensions, covering industry capabilities.

Strength and safety capability assessment.

2 Normative references

The contents of the following documents, through normative references within the text, constitute essential provisions of this document. Dated citations are not included.

For references to documents, only the version corresponding to that date applies to this document; for undated references, the latest version (including all amendments) applies.

This document.

GB/T 8423 (all parts) Terminology for the Oil and Gas Industry

GB/T 41867 Terminology for Information Technology and Artificial Intelligence

GB/T 42755-2023 Data Labeling Procedure for Artificial Intelligence Oriented to Machine Learning

GB/T 45288.2 Large-scale models of artificial intelligence - Part 2.Evaluation metrics and methods

3 Terms and Definitions

The terms and definitions defined in GB/T 41867, GB/T 8423 (all parts), and GB/T 45288.2, as well as the following terms and definitions, apply to this document.

large-scale model

It has a large number of parameters and a complex computational structure, and is trained based on a large amount of data, requiring a large amount of data and computing resources for training.

Practice makes it possible to develop deep learning models that can handle complex tasks and have a certain degree of generalization.

3.2

Trained on large-scale text data from the oil and petrochemical industry, it possesses capabilities such as understanding oil and petrochemical terminology, semantic analysis, and text generation.

Capabilities, a large language model specifically developed for the oil and petrochemical industry.

3.3

Trained using large-scale visual data from the oil and petrochemical industry, it possesses visual capabilities including image recognition, object detection, scene understanding, and defect recognition.

Perception capabilities, a large visual model specifically developed for the oil and petrochemical industry.

3.4

Trained using multi-modal data from the oil and petrochemical industry, and through cross-modal feature fusion and learning, it possesses the ability to handle complex information in the oil and petrochemical field.

END: Draft Version (GBZ225-2026)