

ICS 35.240
CCS L 70



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

GB/Z 237-2026

**Artificial intelligence - Trustworthy datasets
(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.Dataset trustworthiness requirements	3

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".

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 is applicable to developers, users, and third-party evaluation organizations in the field of artificial intelligence, and related organizations that conduct research on artificial intelligence data.

The reliability of the data is used to carry out planning, construction, and evaluation work.

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 25069-2022 Terminology for Information Security Technologies

GB/T 35273-2020 Information Security Technology - Personal Information Security Specification

GB/T 41867-2022 Terminology for Information Technology and Artificial Intelligence

3 Terms and Definitions

The terms and definitions defined in GB/T 25069-2022, GB/T 35273-2020, and GB/T 41867-2022, as well as the following terms and definitions, apply.

This document.

3.1

dataset

A collection of data that has a specific theme, can be identified, and can be processed by computers.

Note: The datasets mentioned in this document mainly refer to those in specific stages of their lifecycle, such as collection, storage, processing, analysis, model training, or application, and serve...

The data set for the activities regulated by this document.

[Source. GB/T 36344-2018, 2.6, with modifications]

3.2

The ability to meet stakeholder expectations in a verifiable manner.

Note: Trustworthy characteristics include compliance, privacy protection, fairness, traceability, timeliness, bias mitigation, security, relevance, consistency, and auditability.

[Source. GB/T 41867-2022, 3.4.2, with modifications]

3.3

security

The characteristics of data sets in terms of confidentiality, integrity, and availability during storage, transmission, and use.

3.4

Privacy protection

Throughout the entire lifecycle of datasets-from acquisition and storage to processing, sharing, and destruction-appropriate technical and management measures should be implemented to prevent [data loss/damage].

The process by which personal information is illegally accessed, leaked, misused, or linked to a specific natural person through means such as re-identification.

Note: The purpose of privacy protection is to safeguard the legitimate rights and interests of

individuals during data processing, including their right to know and their right to decide (control) their personal information.

END: Draft Version (GBZ237-2026)

chinastandards.org - PREVIEW