

ICS 35.030  
CCS L 80



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

**GB/T 45674-2025**

---

**Cybersecurity technology - Generative artificial intelligence  
data annotation security specification**

网络安全技术 生成式人工智能数据标注安全规范

**Issued on: April 25, 2025**

**Implemented on: November 1, 2025**

---

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

## Table of Contents

|                                                                              |    |
|------------------------------------------------------------------------------|----|
| Preface                                                                      |    |
| Introduction                                                                 |    |
| 1 Scope                                                                      |    |
| 2 Normative references                                                       |    |
| 3 Terms and Definitions                                                      |    |
| 4 Overview                                                                   |    |
| 5 Data annotation platform or tool security requirements                     |    |
| 6 Data Labeling Rules Security Requirements                                  |    |
| 7 Data Labeling Personnel Requirements                                       |    |
| 7.1 Safety Training                                                          |    |
| 7.2 Task Assignment                                                          |    |
| 7.3 Personnel Management                                                     |    |
| 8 Data Labeling Verification Requirements                                    |    |
| 8.1 Basic Requirements                                                       |    |
| 8.2 Functional labeling verification safety requirements                     |    |
| 8.3 Safety labeling and verification of safety requirements                  |    |
| 9 Data Annotation Security Evaluation Method                                 |    |
| 9.1 Data Annotation Platform or Tool Security Requirements Evaluation Method |    |
| 9.2 Data Annotation Rules Security Requirements Evaluation Method            |    |
| 9.3 Data Labeling Personnel Requirements Evaluation Method                   |    |
| 9.4 Evaluation Methods for Data Annotation Verification Requirements         |    |
| Appendix A (Informative) Generative AI Data Annotation Examples              |    |
| Appendix B (Informative) Examples of AI Labeling Task Types                  | 14 |

## Foreword

This document is in accordance with the provisions of GB/T 1.1-2020 "Guidelines for standardization work Part 1: Structure and drafting rules for standardization documents" Drafting.

Please note that some of the contents of this document may involve patents. The issuing organization of this document does not assume the responsibility for identifying patents.

This document was proposed and coordinated by the National Cybersecurity

Standardization Technical Committee (SAC/TC260).

This document was drafted by: National Computer Network Emergency Response Technical Coordination Center, China Electronics Technology Standardization Institute, Beijing Zhongguancun Village Laboratory, Beijing Kuaishou Technology Co., Ltd., Beijing Baidu Netcom Technology Co., Ltd., Beijing Topsec Network Security Technology Co., Ltd., Alibaba Cloud Computing Co., Ltd., Peking University, Jiangsu Branch of the National Computer Network Emergency Response Technical Processing Coordination Center, and the Third Research Institute of the Ministry of Public Security Institute, Tsinghua University, Shanghai Artificial Intelligence Innovation Center, Beijing Municipal Public Security Bureau Artificial Intelligence Security Research Center, Xi'an University of Posts and Telecommunications, Zhejiang University, Institute of Information Engineering, Chinese Academy of Sciences, China Mobile Communications Group Co., Ltd., Xiaomi Technology Co., Ltd., Ant Group Co., Ltd.

Co., Ltd., Huawei Cloud Computing Technology Co., Ltd., Beijing Shuanxing Technology Co., Ltd., Beijing Qingshu Wisdom Technology Co., Ltd., Beijing Zero One Wanwu Technology Co., Ltd., Beijing Qihoo Technology Co., Ltd., iFlytek Co., Ltd., Lenovo (Beijing) Co., Ltd., Venusstar Information Technology Group Co., Ltd., AsiaInfo Technologies (Chengdu) Co., Ltd., Hangzhou EZVIZ Software Co., Ltd., Beijing Eastcom Network Technology Co., Ltd.

Co., Ltd., Guangdong Information Security Evaluation Center, Xiamen Mayu Co., Ltd., Beijing Ruilai Smart Technology Co., Ltd., Tianyi Security Technology Co., Ltd., Beijing Yuanjian Information Technology Co., Ltd., Shanghai SenseTime Intelligent Technology Co., Ltd., Suzhou Heshuju Information Technology Co., Ltd., Nanjing Lingxing Technology Co., Ltd., Jiangsu Manyun Software Technology Co., Ltd., Changan Communication Technology Co., Ltd., OPPO Guangdong Mobile DYNAMIC COMMUNICATIONS LIMITED.

The main drafters of this document are: Zhang Zhen, Tan Zhixing, Zhang Yanting, He Min, Liu Yong, Sun Xudong, Xu Ke, Chen Zhong, Du Jinhao, Hao Chunliang, Ren Kui, Liu Nan, Luo Hongwei, Ye Xiaojun, An Qing, Hu Ying, Wang Yan, Yao Long, Xie Anming, Ji Cheng, Jiang Weiqiang, Ding Zhiguo, Lei Xiaofeng, Dai Jiao, Gu Chen, Zhang Qingqing, Guo Jianling, Zhang Yong, Luo Lei, Liu Yuhong, Liao Shuangxiao, Jiang Hui, Zhao Yun, Zhang Feng, Xu Xiaogeng, Wang Wenyu, Chen Yang, Zhang Xia, Peng Juntao, Bao Chenfu, Wang Haitang, Meng Fanqin, Zhao Lili, Liu Junhua, Li Jiakun, Cui Tingting, Yu Hanyang, Li Fengfeng, Zang Jiaojiao, Lin Guanchen, Ding Xin, Wang Shijin, Han Han, Zhang Xiangzheng, Hu Songzhi, Xu Yiyue, Guan Ming, Zhang Tianyi, Huang Zhe, Liu Jun, Zhou Xue, Zheng Rong, Liu Dong, Luo Xupeng, Zheng Hongdong, Jiang Faqun, Ma Mengna, Tian Weili, Hu Yue, Huang Penghua, Zhang Xiaomin, Zhang Zhongwei, Zhou Cheng, Li Gen, Li Xiaoru, Zhang Bingsheng, Wang Hejun, and Liu Dongbin.

## Introduction

Data annotation is a key activity in generative artificial intelligence, which directly determines the quality and security level of training data and generated content.

Due to imperfect labeling rules, irregular personnel management, unclear verification standards and other reasons, generative human factors may also be used in the data labeling process.

Artificial intelligence introduces new risks and hidden dangers, and standards and specifications are urgently needed to improve the security level of data annotation.

The purpose of this document is to help service providers, data annotators and Organizers and data demanders should clarify the security baseline of data labeling and improve the level of service security. Cybersecurity Technology Generative AI Data Annotation Security Specification

## 1 Scope

This document specifies the security requirements for data annotation platforms or tools for generative AI training, data annotation rule security requirements, data The requirements for labeling personnel and data labeling verification describe the data labeling security evaluation method.

This document is applicable to generative artificial intelligence data annotation organizers to carry out training data annotation activities and to The demander can inspect and accept the data labeling, or a third-party organization can provide reference for security assessment of the data labeling.

## 2 Normative references

GB/T 42755-2023

GB/T 45654-2025

## 3 Terms and definitions

The following terms and definitions apply to this document.

### 3.1 prompt

Input information that guides generative AI models to complete specific tasks and provide reasonable output content.

### 3.2 Response information

In generative AI data annotation, the response information that conforms to human cognition is formed according to the prompt information requirements and is used to train

the model.

The ability to generate responses to prompts with output of appropriate content, mode, or style.

### **3.3 [Source. GB/T 45654-2025, 3.5]**

Through manual operation or the use of automated technical mechanisms, specific information such as labels, categories, etc. are sent to The process of adding attributes or properties to text, images, audio, video, or other data samples.

Note. Hereinafter referred to as "data annotation".

### **3.4 [Source. GB/T 45654-2025, 3.6]**

Data annotation used to train generative artificial intelligence models to be able to complete specific tasks.

### **3.5 [Source. GB/T 45654-2025, 3.7]**

Data annotation used to train generative artificial intelligence models to improve the security of output response information.