ICS 35.030
CCS L 80

NATIONAL STANDARD OF THE PEOPLE'S REPUBLIC OF CHINA

GB/T 45654-2025

Cybersecurity technology - Basic security requirements for generative artificial intelligence service

网络安全技术 生成式人工智能服务安全基本要求

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

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: China Electronics Standardization Institute, National Computer Network Emergency Response Technical Processing Coordination Center, Zhejiang University, Beijing Zhongguancun Laboratory, Shanghai Artificial Intelligence Innovation Center, Fudan University, Beijing Baidu Netcom Technology Co., Ltd., Alibaba Cloud Computing Co., Ltd.

Beijing Kuaishou Technology Co., Ltd., Huawei Cloud Computing Technology Co., Ltd., Beijing University of Aeronautics and Astronautics, Lenovo (Beijing) Co., Ltd., Ant Technology Group Co., Ltd., iFLYTEK Co., Ltd., Peking University, China Cybersecurity Review and Certification and Market Supervision Big Data Center, Beijing DeepQuest AI Basic Technology Research Co., Ltd., Beijing Qihoo Technology Co., Ltd., and Automation Research Institute of the Chinese Academy of Sciences Institute of Software, Henan University of Science and Technology, China University of Political Science and Law,

Shanghai Jiao Tong University, Tsinghua University, Institute of Software, Chinese Academy of Sciences, OPPO Guangdong Mobile Communications Co., Ltd., China Mobile Communications Group Co., Ltd., Sangfor Technologies Co., Ltd., Beijing Mianbi Intelligent Technology Co., Ltd.

Beijing Ruilai Smart Technology Co., Ltd., National Industrial Information Security Development Research Center, the Third Research Institute of the Ministry of Public Security, the National Information Center, Shanghai Enflame Technology Co., Ltd., Shenzhen Luxi Technology Co., Ltd., Hangzhou NetEase Zhiqi Technology Co., Ltd., Beike Zhaofang (Beijing) Technology Co., Ltd.

Technology Co., Ltd., Beijing Topsec Network Security Technology Co., Ltd., Beijing Zero One Everything Technology Co., Ltd., Shanghai Xiyu Technology Co., Ltd., Guangzhou Dongyue Information Technology Co., Ltd. and Tianyi Security Technology Co., Ltd.

The main drafters of this document are: Yao Xiangzhen, Hao Chunliang, Zhang Yanting, Zhang Zhen, Ren Kui, Liu Yong, Yang Min, Qin Zhan, Hu Ying, Xia Wenhui, Chen Zhong, Wang Yingchun, He Min, Zhang Linghan, Xu Xiaogeng, Liu Jianwei, Luo Hongwei, Wang Fengjiao, Xu Ke, Chen Yang, Zhang Xiangzheng, Bao Chenfu, Xie Anming, Peng Juntao, Gu Chen, Zheng Zimu, Wu Shaoqing, Wang Jiao, Wang Bingzheng, Guo Jianling, Meng Lingyu, Xu Jia, Yang Ziqi, Wang Qinglong, Qiu Xipeng, Huang Qing, Shi Lin, Zhang Zongyang, Bian Song, Zhang Zhiyong, Zhang Mi, Hong Geng, Pan Xudong, Hu Yongqi, Lin Guanchen, Liu Junhua, Qiao Yuping, Mei Jingqing, Jia Kai, Zhao Jing, Zhang Yan, Quan Gaoyuan, Tan Zhixing, Yang Guang, Yao Long, Li Qi, Wang Hui, Zhu Guibo, Zhou Peng, An Qing, Shen Juncheng, Zhao Ruibin, Liu Dong, Ma Mengna, Wang Jun, Zhang Liyao, Jia Yumeng, Wang Haitang, Peng Tao, Li Gen, Qiu Qin, Jiang Weiqiang, Xu Yang, You Jianzhou, Zhou Chenghui, Liu Nan, Ding Zhiguo, Wang Rongshi, Li Dahai, Zhu Xiaofang, Wang Yuchen, Xue Zhihui, Xiao Bofeng, and Wei Jiaqi.

Introduction

At present, generative artificial intelligence technology is constantly developing, and related services have been widely used, providing convenience for social production and life.

It also brings a large number of new risks and challenges to network security, and there is an urgent need for standards and specifications to establish a security baseline.

This document focuses on generative AI services with public opinion attributes or social mobilization capabilities, supporting filing management, detection and evaluation When focusing on data annotation security, this document can be used together with GB/T 45674 "Network Security Technology Generative Artificial Intelligence This document can be used in conjunction with GB/T 45652 when focusing on the security of pre-training and optimized training data.

It is used in conjunction with the "Security Specifications for Generative Artificial

Intelligence Pre-training and Optimized Training Data for Cybersecurity Technologies".

Generative AI Services for Cybersecurity Technologies Basic safety requirements

1 Scope

This document specifies the requirements for generative AI services in terms of training data security, model security, and security measures.

This document is applicable to service providers carrying out activities related to generative artificial intelligence services, and is also intended for relevant competent authorities and third-party assessments.

Institutions provide references.

Note. The main security risks involved in training data and generated content are shown in Appendix A, and the reference method for security assessment of generative artificial intelligence services is shown in Appendix B.

2 Normative references

GB/T 25069

3 Terms and definitions

The terms and definitions defined in GB/T 25069 and the following apply to this document.

3.1

Use generative artificial intelligence technology to provide the public with services that generate text, images, audio, video and other content.

3.2 Service Provider serviceprovider

Organizations or individuals that provide generative artificial intelligence services in the form of interactive interfaces, programmable interfaces, etc.

3.3 Classification model

A machine learning model that outputs one or more categories to which the input data belongs.

[Source. GB/T 41867-2022, 3.2.6]

3.4 training data

All data directly used as input for model training.

Note. Includes pre-training data and optimized training data.

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".