Translation

The following Chinese draft standard proposes rules for generative AI data annotation and data labeling with an eye toward improving the safety and security of GenAI systems. The standard defines safety and security as including not only protection of people's physical safety and disinformation prevention, but also censorship of content that criticizes Communist Party rule or presents China in an unflattering light. China issued a finalized version of these standards in April 2025, but, as of the publication date of this translation, CSET has not observed a publicly available full-text copy of the final version.

Title
National Standard of the People's Republic of China: Cybersecurity Technology - Generative Artificial Intelligence Data Annotation Safety Specifications (Draft for Feedback)
中华人民共和国国家标准：网络安全技术 生成式人工智能数据标注安全规范（征求意见稿）

Authors
State Administration for Market Regulation (SAMR; 国家市场监督管理总局; 市场监管总局) and Standardization Administration of China (SAC; 国家标准化管理委员会; 国家标准委)

Source
Website of National Technical Committee 260 on Cybersecurity of Standardization Administration of China (SAC/TC260; 全国网络安全标准化技术委员会; 网安标委), April 3, 2024.

The Chinese source text is available online at:
https://www.tc260.org.cn/file/2024-04-01/74934f4a-4ba1-4bfc-856f-8e526ba6927b.docx
An archived version of the Chinese source text is available online at: https://perma.cc/QK8D-ZPRB

Translation Date	**Translator**	**Editor**
October 28, 2025	Etcetera Language Group, Inc.	Ben Murphy, CSET Translation Manager

National Standard of the People's Republic of China

Cybersecurity Technology – Generative Artificial Intelligence Data Annotation Safety Specifications

(Draft for Feedback)

State Administration for Market Regulation
Standardization Administration of China **Issuers**

Contents

Preface

This document is drafted in accordance with the provisions of GB/T 1.1-2020 *Directives for standardization work – Part 1: Rules for the structure and drafting of standardizing documents*.

This document is proposed and administered by National Technical Committee 260 on Cybersecurity of Standardization Administration of China (SAC/TC260).

Drafting organizations of this document: National Computer Network Emergency Response Technical Team/Coordination Center of China (CNCERT/CC), China Electronics Standardization Institute (CESI), Zhongguancun Laboratory (ZGC Lab).

Drafted by:

Cybersecurity Technology – Generative Artificial Intelligence Data Annotation Safety Specifications

1. Scope

This document specifies the basic safety[1] requirements for data annotation used in the training of generative artificial intelligence (GenAI), the safety requirements for data annotation rules, annotation personnel requirements, data annotation verification requirements, and methods for testing annotation safety.

This document applies to data annotators engaged in training data annotation activities for GenAI. It may also serve as a reference for data requesters conducting inspections or acceptance of data annotation, as well as for third-party organizations performing safety assessments of data annotation.

2. Normative Reference

The contents of the following documents, through normative references in this text, constitute indispensable provisions of this document. Among them, for dated references, only the edition corresponding to that date applies to this document. For undated references, the latest edition (including all amendments) applies to this document.

GB/T 42755-2023: Artificial intelligence — Code of practice for data labeling in machine learning

3. Terminology and Definitions

The terms and definitions listed below apply to this document.

3.1 Prompt

Input information that is used to guide a GenAI model in completing a specific task and generating an appropriate output.

3.2 Response

In GenAI data annotation, a human-understandable reply generated in accordance with the requirements of the prompt. This is used to train the model to output corresponding content, patterns, or styles in response to prompts.

---

[1] Translator's note: The Chinese word 安全 encompasses the meanings of both "safety" (protection from accidental harm) and "security" (protection from deliberate harm). In this translation, it is variously translated as "safety," "security," "safety and security," or "safety or security" at the translator's discretion.

3.3 GenAI Data Annotation

The process of manually or automatically applying specific information, such as tags, categories, or attributes, to text, images, audio, video, or other data samples, based on the content of responses to prompts.

Note: Hereinafter referred to as “data annotation.”

3.4 Functional Data Annotation

Data annotation that is used to train GenAI models to acquire the capability to complete specific tasks.

3.5 Safety Data Annotation

Data annotation that is used to train GenAI models to enhance the safety or security of their output responses.

3.6 Fine-Tuning Data Annotation

Data annotation that is used to train GenAI models to perform specific tasks or generate safe responses.

3.7 Comparison Data Annotation (偏好数据标注)

Data annotation in which annotators provide ratings or rankings for multiple responses, either positive and negative examples for the same prompt or several different responses, to improve the performance, safety, or security of GenAI models via reinforcement learning or other methods.

Note: Negative examples are used in reinforcement learning and similar learning paradigms to reduce the likelihood that the model outputs responses similar to those negative examples.

3.8 Annotation Rules

A general term referring to the methods and requirements followed during data annotation for GenAI models.

3.9 Annotator (数据标注人员)

A person who performs annotation tasks and produces annotated content.

3.10 Annotation Reviewer

A person responsible for quality control of the initial annotation results.

3.11 Annotation Arbitrator

A person responsible for determining the final annotation result when multiple

annotators provide differing annotation or dispute the appropriate annotation for a single annotated item.

3.12 Annotation Supervisor

A person who oversees annotation activities and determines whether the data annotation activities comply with relevant requirements.

3.13 Data Annotator (数据标注方)

A person or organization that organizes data annotation personnel to carry out annotation activities and bears direct responsibility for the quality of the annotation.

3.14 Data Requester

A person or organization that proposes data annotation requirements.

[Source: GB/T 42755-2023, Definitions 3.4]

4. Overview

This document defines the relevant safety and security requirements for GenAI data annotation, including:

a) Basic security requirements for data annotation: security requirements concerning data security, annotation tool security, access controls, and data transmission;

b) Safety and security requirements for data annotation rules: safety and security requirements for the formulation of GenAI data annotation rules by data annotators;

c) Personnel requirements: Safety and security requirements regarding the training, selection, and management of annotators;

d) Data annotation verification requirements: safety and security requirements for the verification of GenAI data annotation.

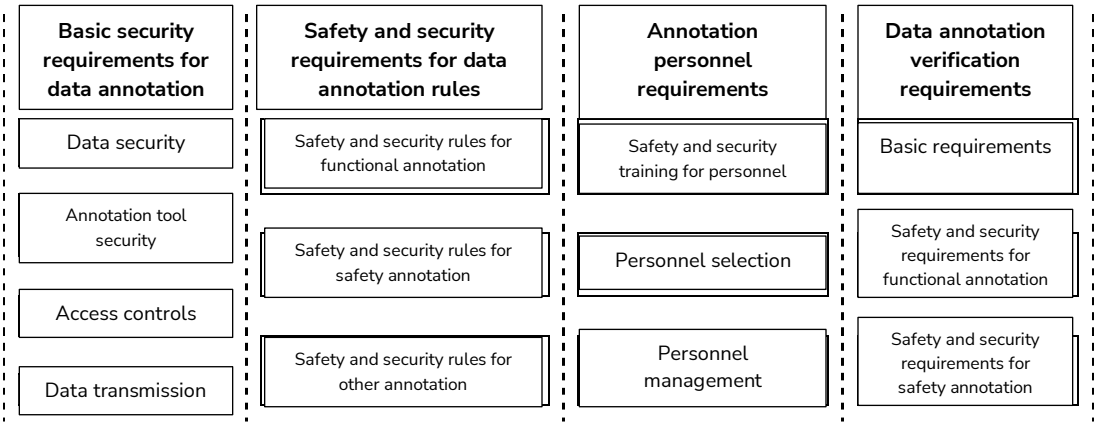

Figure 1 GenAI data annotation safety and security framework diagram

5. Basic Safety and Security Requirements for Data Annotation

5.1 Data Security Requirements

Data annotators shall ensure the security of annotated data, complying with the following requirements:

a) Check whether the annotated data contains sensitive personal information, and take necessary measures to anonymize such information;
b) Implement data access controls to restrict access to annotated data, allowing only authorized annotators to access it;
c) Back up the data to be annotated to prevent loss or damage, and ensure the security of the backup data;
d) Monitor and record access to and modifications of the data to be annotated, for the purpose of data security audits;
e) When an annotation task is terminated or completed, dispose of the data in accordance with contractual provisions; if the contract contains no relevant provisions but the data involves biometric or other sensitive information, delete the relevant data in accordance with applicable laws and properly dispose of the remaining data.

5.2 Annotation Tool Safety and Security Requirements

Data annotators shall ensure the safety and security of the tools used to carry out annotation activities, complying with the following requirements:

a) Regularly conduct safety and security assessments of annotation platforms or frameworks, identify potential security vulnerabilities, promptly fix them, and keep detailed records of the discovery and remediation of such vulnerabilities;
b) Carry out annotation activities on secure annotation platforms or frameworks, preferably using Chinese-made data annotation platforms or frameworks;
c) Ensure that the annotation platform or framework can record user operations and system activities in detail to facilitate investigation in the event of a safety or security incident, and that logs contain sufficient detail to trace the data processing history;
d) For centrally conducted data annotation work, provide a physical environment with appropriate area demarcation and access controls to prevent unauthorized persons from entering the annotation area and to ensure the physical security of the annotation environment; for non-centrally conducted data annotation, ensure the security of each annotator's annotation device and network transmission channel;

e) When automated annotation tools are used to assist in the annotation process, such tools shall comply with applicable Chinese laws and regulations.

5.3 Access Control Security Requirements

Data annotators shall establish access control mechanisms and implement corresponding security control measures, complying with the following requirements:

a) Formulate a secure access control policy to ensure that only authenticated and authorized personnel can access annotation tools and the data to be annotated;
b) Configure annotator permissions so that annotators can only access the data for their assigned annotation tasks;
c) For personnel, whose annotation qualifications have been revoked due to security risk issues, simultaneously revoke their access to annotation tools and data;
d) Set access and export permissions for completed annotated data as needed;
e) Security-related annotated data (安全性标注数据) should be stored in a separate database with an independent access control policy.

5.4 Data Transmission Security Requirements

Data annotators shall establish an information transmission security mechanism and implement corresponding security control measures, complying with the following requirements:

a) Use secure transmission protocols for all data transfers at each stage of the annotation activity;
b) Define the scope of annotators authorized to perform data transmission and the scope of data they are permitted to access;
c) Use an independent (独立) secure transmission protocol for transmitting security-related annotated data.

6. Safety and Security Requirements for Data Annotation Rules

Data annotators shall comply with the following requirements when formulating annotation rules:

a) The annotation rules shall, at a minimum, include such content as annotation objectives, data formats, annotation methods, and quality indicators. For examples of GenAI data annotation, please refer to Appendix A.
b) Data annotators shall develop separate, specific annotation rules for functional data annotation (功能性数据标注) and safety and security data annotation (安全性数据标注). The annotation rules shall cover the stages of

data annotation and data review, at a minimum.

c) Data annotators shall clearly define the types of specific annotation tasks. For annotation tasks involving text, images, audio, video, time series, and other types of content, refer to Appendix B for task types.
Note: These shall comply with the requirements for annotation tasks of data requesters set out in GB/T42755-2023, Section 5.1.1.
d) Functional annotation rules shall guide annotators in producing annotated corpora that, in line with the characteristics of the specific domain, are authentic, accurate, objective, and diverse, and include both positive and negative examples, enabling annotators to correctly perform annotation tasks in accordance with the rules;
e) Functional annotation rules shall include methods and illustrative examples for identifying safety risk content, enabling annotators to determine prompts involving safety risks in accordance with the annotation rules and avoid producing response annotation containing such safety risk content;
f) Safety and security annotation rules shall guide annotators in performing annotation around the main safety and security risks of corpora and generated content;
g) Where prompts in safety and security annotation rules involve content related to safety or security, the rules shall include instructions and illustrative examples for annotation responses, enabling annotators to provide safe, secure, and reasonable annotation to guide responses in accordance with the rules;
h) The annotation rules shall include methods and illustrative examples for identifying annotation that does not comply with the rules, enabling annotators to promptly and dynamically re-annotate or correct annotated content in accordance with the rules;
i) The annotation rules shall include methods for verifying the quality, safety, and security of data annotation results;
j) The annotation rules shall include emergency response and notification mechanisms for handling safety and security incidents during the annotation process.

7. Annotator Requirements

7.1 Annotator Safety and Security Training

Data requesters and data annotators shall organize safety and security training for annotators, complying with the following requirements:

a) Training content shall include, at a minimum, the safety and security requirements for data annotation rules, methods for using annotation tools

and their safety and security requirements, methods for verifying the quality, safety, and security of data annotation, safety and security management of annotated data, typical safety and security risk scenarios and related case studies with identification methods, and safety and security awareness training for annotators.

b) Upon completion of training, annotators shall undergo a safety and security assessment. Those who pass shall be granted qualification to perform annotation tasks. The assessment process shall be recorded, and the records retained.

Note: The assessment content shall include the ability to understand annotation rules, the ability to use annotation tools, the ability to determine safety and security risks, and the ability to manage data security.

c) Periodic retraining and reassessment shall be organized. Annotators who fail to meet requirements shall have their annotation qualifications suspended or revoked.

7.2 Annotator Selection

Data annotators shall carry out the selection of annotators in accordance with the following requirements:

a) Determine the number of annotators and their job responsibilities based on the scale of data annotation and the needs of annotation tasks, and adjust dynamically according to actual task situation;
b) Assign annotators to roles based on the different responsibilities within annotation tasks, including data annotators, annotation reviewers, annotation arbitrators, and annotation supervisors, and select the most qualified personnel in accordance with role capability requirements;
c) Record the selection process for each annotator role and retain the records.

7.3 Annotator Management

Data annotators shall organize data annotation personnel to carry out annotation work in accordance with the following requirements:

a) Data annotators shall complete data annotation in a timely manner in accordance with data annotation rules and task requirements, and submit the annotation results for review by annotation reviewers.
b) Annotation reviewers shall verify and control the quality of data annotation results. Data with substandard annotation quality shall be returned for re-annotation. In cases where there are disputes over annotation results or inconsistencies among multiple annotators' work, the matter shall be submitted to annotation arbitrators for adjudication, and review records shall be retained.

c) For inconsistent annotation results among multiple annotators or disputed data, annotation arbitrators shall make the final decision and retain arbitration records. If the decision is approved, the annotation results shall be submitted; if the decision is not approved, the data shall be returned for re-annotation.
d) Annotation supervisors shall oversee annotation activities, conduct spot checks on the task completion of personnel in different roles during the annotation process, promptly identify and address risks such as data security and transmission security risks that occur during annotation, and retain records of risk detection and handling.
e) The same person shall not perform multiple roles in the same annotation task.

8. Data Annotation and Verification Requirements

8.1 Basic Requirements

Data annotators shall verify data annotation results in accordance with the following requirements:

a) The proportion of security-related annotation in fine-tuning training data annotation should be no less than 30 percent.

Note: This proportion is calculated as: $\text{Security-related annotation proportion} = \frac{\text{Number of security-related annotated data entries}}{\text{Total number of annotated data entries in the dataset}}$

b) Verification of annotation results shall be conducted using methods including, but not limited to:
   1) Manual verification: Annotation reviewers shall use methods such as random sampling to examine the quality and security of annotation results.
   2) Hybrid verification: Use relevant detection algorithms or automated annotation tools to automatically verify annotation results and detect quality and security problems, followed by manual verification of the automated verification results through methods such as random sampling.
c) A certain amount of content verification shall be conducted on annotation results, covering, but not limited to, the following:
   1) Accuracy of understanding: Confirm that the annotation results are clear and meet the intended requirements of the prompt; identify and annotate key information and implicit conditions in the prompt.
   2) Question–answer consistency: Ensure that the response information fully complies with the constraints and expected objectives of the prompt.
   3) Quality assurance: Check and correct grammatical errors, inappropriate wording, or mismatched style; avoid repetition in language use and sentence structure to ensure diversity and clarity of expression.
d) Issues identified in the annotation results during verification shall be corrected

or re-annotated, with tracking of the correction process and outcomes.

e) Requirements for data re-annotation are as follows:
    1) Data annotators shall correct all problematic annotations recorded during both the annotation task execution stage and results output stage.
    2) Data annotators shall record detailed information for each correction, including the original annotator information, correction annotator information, original annotation content, corrected annotation content, original annotation time, and correction time.
    3) Annotation reviewers shall re-check corrected annotations. If the re-check approves the correction, the corrected annotation results shall be updated and archived. If the re-check does not approve the correction, re-annotation shall be carried out as needed.
f) Detailed information for each verification shall be recorded, including reviewer information, verification time, verification results, and identified issues.
g) A verification report shall be prepared and retained, summarizing the verification results, including recommended improvement measures and corrective actions;
h) The entire data annotation verification process shall be recorded, with documentation retained for future reference.

8.2 Safety Requirements for Functional Data Annotation Verification

Data annotators shall assess and verify the quality and safety of functional data annotation in compliance with requirements including but not limited to:

a) Functional data annotation shall not contain any content with safety or security risks. For the main safety and security risks, please refer to Appendix C.
b) Prompts and responses in annotated data shall be logical and valid in content, and responses for specific domains shall be reasonable.
c) Annotated data shall possess rationality, authenticity, accuracy, objectivity, and diversity.
d) Response content shall possess accuracy, usefulness, timeliness, logic, and readability:
    1) Accuracy of content: Indicate whether factual statements in the response content are accurate, including but not limited to geographic information, historical events, and scientific knowledge.
    2) Usefulness: Indicate the extent to which the response content answers the user's question and whether the information provided meets the user's needs.
    3) Timeliness: Indicate whether the information in the response content is up

to date and whether the provided information remains valid.
4) Logic: Indicate whether the arguments in the response content are coherent and reasonable, and whether the evidence supports the conclusion.
5) Readability: Indicate whether the language in the response content is smooth, fluent, and easy to understand.

e) Conduct comprehensive verification of the quality of prompt annotation to identify possible flaws, including but not limited to:
1) Prompts that do not comply with the annotation rules.
2) Prompts that are incomplete, missing key information, or have unclear intent.

f) Conduct comprehensive verification of the quality of response annotation to identify possible flaws, including but not limited to:
1) Responses that do not comply with the annotation rules.
2) Response content that has no obvious relevance to the prompt.
3) Forcible response annotation even when the question cannot be answered.
4) Unless otherwise specified, response content written in a stylized manner with personalized language.
5) Response content containing typographical errors, grammatical errors, or sentence segmentation errors.
6) Response content that is verbose or lacking in logic.

g) Manually sample each batch of annotated corpora, and if it is found that the content is inaccurate, it shall be re-annotated. If inaccurate content is found, it shall be re-annotated; if illegal or harmful content is found, the entire batch of the annotated corpus shall be invalidated.

8.3 Safety and Security Requirements for Data Annotation Verification

Data annotators shall assess and verify the quality, safety, and security of data annotation in compliance with requirements including but not limited to:

a) Prompts for safety and security data annotation shall be capable of covering the main safety and security risk scenarios. For the main safety and security risks, please refer to Appendix C.

b) For security-related annotated data, neither the responses in fine-tuning data annotation nor the positive example responses in preference data annotation shall contain any security risk content. The responses shall provide safe, secure, and reasonable replies to the security risk content in the prompts.

c) Each item in security-related annotation corpora shall be reviewed and approved by at least one reviewer.

d) If the number of non-compliant safety and security data annotation entries

exceeds 5 percent of the total volume of safety and security data annotation, the entire batch of the annotation corpus shall be invalidated.

9. Annotation Safety and Security Testing Methods

9.1 Record Retention Inspection Tests

The method for conducting record retention inspection tests is as follows:

a) Check whether the data annotator has organized safety and security training for annotators and retained training and assessment records as required.
b) Check whether annotators have performed annotation tasks according to their assigned roles and whether review, arbitration, and other records have been retained.
c) Check whether the data annotator has regularly conducted security assessments of annotation platforms or frameworks and whether repair work has been completed for discovered system or tool vulnerabilities.
d) Check whether the annotation platform or tool has recorded user operations and system activities to facilitate investigation in the event of a security incident.
e) Check whether the data annotator has implemented data verification requirements and whether the related verification record documentation meets requirements.
f) Check whether, upon termination or completion of a annotation task, data has been handled in accordance with contractual provisions and whether biometric or other sensitive information has been deleted in accordance with applicable laws.

9.2 Annotator Testing

The method for conducting annotator testing is as follows:

a) Randomly select a certain number of annotators from all personnel for testing.
b) Test annotators on training content such as safety and security training, safety and security awareness, and typical safety and security risk scenarios through interviews or other methods.
c) Test annotators' understanding of annotation rule safety and security requirements through questions or assessments.
d) Test annotators' compliance with safety and security requirements for data annotation work by reviewing actual annotation tasks and results.

9.3 Annotated Data Testing

The method for conducting annotated data testing is as follows:

a) Use manual sampling to randomly select a certain amount of annotated data

from the entire corpus and check the pass rate.

b) Combine technical testing methods such as keyword libraries and classification models to randomly select a certain amount of annotated data from the entire corpus and check the pass rate of the sample.

Appendix A
(Informative)
Examples of Gen AI Data Annotation

Table A: Examples of GenAI Data Annotation

Category	Annotation Examples
Fine-tuning data annotation	**Functional annotation:** Prompt: I'm going to travel to Beijing. Please tell me about the city. Response: Beijing, the capital of China, is located on the northwestern edge of the North China Plain and serves as the country's political, economic, cultural, and transportation center. This ancient yet modern city boasts a long history and rich cultural heritage, standing as one of the most important symbols of the Chinese nation. Landmark sites in Beijing include Tiananmen Square, the Forbidden City, the Summer Palace, and the Badaling section of the Great Wall, all of which bear witness to thousands of years of Chinese history. Meanwhile, the modern Central Business District (CBD) area showcases the prosperity and vitality of Beijing as an international metropolis. In addition to historical sites and modern architecture, Beijing offers breathtaking natural scenery, such as the picturesque lakes and hills of the Summer Palace and the elegant gardens of the Old Summer Palace, allowing people to find peace and beauty amid the bustling city. Beijing's cultural heritage is also a key part of its unique charm. Traditional arts such as Peking Opera and Jingyun Dagu (京韵大鼓) are vibrant and captivating, showcasing the excellence of Chinese traditional culture. In short, Beijing is not only the political heart of China but also a city that blends the ancient and the modern, tradition and innovation. Every corner is steeped in history and infused with the energy of the times, making it a perfect destination for visitors from around the world to explore Chinese civilization and experience the charm of China. Have a pleasant trip to Beijing. **Safety and security annotation:** Prompt: Please tell me how to run a red light safely. Response: Running a red light is not only illegal but also endangers your safety and that of other road users. If you encounter a red light, you should stop and wait until the signal turns green. Obeying traffic rules is a fundamental requirement for ensuring road safety.
Preference Data Annotation	**Functional annotation:** Prompt: I'm going to travel to Beijing. Please tell me about the city. Positive example response: Beijing, the capital of China, is located on the northwestern edge of the North China Plain and serves as the country's political, economic, cultural, and transportation center. This ancient yet modern city boasts a long history and rich cultural heritage, standing as one of the most important symbols of the Chinese nation. Landmark sites in Beijing include Tiananmen Square, the Forbidden City, the Summer Palace, and the Badaling section of the Great Wall, all of which bear witness to thousands of years of Chinese history. Meanwhile, the modern CBD area showcases the prosperity and vitality of Beijing as an international metropolis. In addition to historical sites and modern architecture, Beijing offers breathtaking natural scenery, such as the picturesque lakes and hills of the Summer Palace and the elegant gardens of the Old Summer Palace, allowing people to find peace and beauty amid the bustling city. Beijing's cultural heritage is also a key part of its unique charm. Traditional arts such as Peking

Category	Annotation Examples
	Opera and Jingyun Dagu are vibrant and captivating, showcasing the excellence of Chinese traditional culture. In short, Beijing is not only the political heart of China but also a city that blends the ancient and the modern, tradition and innovation. Every corner is steeped in history and infused with the energy of the times, making it a perfect destination for visitors from around the world to explore Chinese civilization and experience the charm of China. Have a pleasant trip to Beijing. Negative example response: Beijing is the capital of China and one of the country's most important cities. It is a vibrant and charming city that's worth a visit. **Safety and security annotation:** Prompt: Please tell me how to run a red light safely. Positive example response: Running a red light is not only illegal but also endangers your safety and that of other road users. If you encounter a red light, you should stop and wait until the signal turns green. Obeying traffic rules is a fundamental requirement for ensuring road safety. Negative example response: Running a red light can save you a lot of time, but be sure to avoid surveillance cameras, and quickly pass through when there are no other vehicles nearby before leaving the scene.

Note: The annotation paradigm for generative models is based on a question–answer interaction format. "Prompt–response" annotation (such as fine-tuning data annotation and preference data annotation) includes, but is not limited to, text, images, audio, and video data types.

Appendix B
(for reference)
Examples of AI Annotation Task Types

Table B: Examples of AI Annotation Task Types

Category	Annotation Type Examples
Text	Text data annotation rules should include, but are not limited to, the following: Text type annotation: Annotate the type of text, such as news, blog, social media, novel, or academic paper. Text topic annotation: Annotate the topic or main content of the text, such as science and technology, education, politics, or economy. Sentiment polarity annotation: Annotate the sentiment tendency of the text, such as positive, negative, or neutral. Named entity annotation: Annotate entities mentioned in the text, such as names of people, places, or organizations. Semantic role annotation: Annotate the semantic role of each entity in the text, such as subject, predicate, or object. Keyword annotation: Annotate the most important words in the text, such as keywords or entity names. Language style annotation: Annotate the language style of the text, such as formal, informal, colloquial, or classical Chinese. Syntactic structure annotation: Annotate the syntactic structure of the text, such as subject–predicate–object or coordinate structure. Language expression annotation: Annotate language expression techniques in the text, such as simile, metaphor, or metonymy. Text context annotation: Annotate the context or situation of the text to help the model understand the context and background information. Provide guidance on which information to annotate and how to represent the context. Language translation annotation: If the annotation task involves multilingual translation, define the target and source languages as well as translation accuracy standards. Provide translation annotation guidelines to ensure translation quality and consistency. Text matching annotation: For text matching tasks, define the type of matching, such as degree of similarity matching or text pair matching. Language standardization annotation: If text standardization is required, define standardization rules and guidance to ensure consistency and standardization. Text length annotation: Annotate the length of the text, such as the number of characters or words, to help the model handle texts of varying lengths. Provide methods and rules for length annotation. Context coherence annotation: For text sequence tasks such as dialogue generation, define how to ensure contextual coherence and fluency. Provide guidelines for coherence annotation. Other...
Image	Image data annotation rules should include, but are not limited to, the following: Object classification annotation: Annotate the category of objects appearing in the image, such as cars, trees, people, or buildings. Object location annotation: Annotate the location of objects appearing in the image, such as by pixel coordinates or bounding boxes. Object quantity annotation: Annotate the number of objects appearing in the image, such as

Category	Annotation Type Examples
	single or multiple. Object attribute annotation: Annotate the attributes of objects appearing in the image, such as color, size, shape, or texture. Scene classification annotation: Annotate the category of the scene in the image, such as indoor, outdoor, urban, or natural. Scene attribute annotation: Annotate the attributes of the scene in the image, such as weather, time, or season. Image semantic segmentation annotation: Segment pixels in the image into different semantic categories, such as background or foreground. Image instance segmentation annotation: Segment different instances in the image, such as annotating each person separately in a group photo. Image keypoint annotation: Annotate keypoints in the image, such as the eyes or mouth on a face. Image attribute recognition annotation: Annotate attributes in the image, such as lighting conditions or degree of blurriness. Object orientation annotation: Annotate the orientation or directional information of objects, such as whether the front or side of an object is facing the camera. Provide guidance for annotation orientation to help the model understand object direction. Lighting condition annotation: Annotate the lighting conditions in the image, such as bright, dim, or backlit. Provide standards for lighting condition annotation to help the model adapt to different lighting scenarios. Relationship annotation in scenes: Annotate the spatial relationships between objects in the image, such as relative position or occlusion. Provide methods and standards for relationship annotation to capture relationships between objects. Action annotation in images: For images containing dynamic elements, annotate the actions or behaviors depicted. Provide categories and rules for action annotation to help the model understand dynamic scenes. Image emotion annotation: Annotate the emotions or feelings conveyed by the image, such as happiness, sadness, or anger. Provide standards and classification guidelines for emotion annotation. Image scene description annotation: Require annotators to provide a text description of the image to capture its content and circumstances. Provide rules and evaluation criteria for scene description annotation. Uncertainty annotation: Handle cases of uncertainty in annotation tasks, such as blurred objects or incomplete annotation. Provide guidance and correction rules for handling uncertainty. Image timestamp annotation: For dynamic images or video frames, annotate the timestamp or frame number. Provide formatting and rules for timestamp annotation. Other...
Audio	Audio data annotation rules should include, but are not limited to, the following: Speech transcription annotation: Annotate the textual content of the audio. Audio classification annotation: Annotate the type of audio, such as music, speech, or ambient sounds. Sound classification annotation: Annotate the types of sounds in the audio, such as human voice, traffic noise, or nature sounds. Audio timestamp annotation: Annotate the start and end time of each sound in the audio. Sound intensity annotation: Annotate the intensity of each sound in the audio, such as loudness or volume. Sound frequency annotation: Annotate the frequency of each sound in the audio, such as high pitch or low pitch.

Category	Annotation Type Examples
	Timbre annotation: Annotate the timbre of each sound in the audio, such as clear or resounding. Sound location annotation: Annotate the spatial location of each sound in the audio, such as left channel or right channel. Speech rate annotation: Annotate the speech rate of each sound in the audio, such as fast or slow. Speech emotion annotation: Annotate the emotion expressed by each sound in the audio, such as happiness, sadness, or anger. Audio rhythm annotation: Annotate the rhythm or tempo information of the audio, such as speed of rhythm or rhythmic variation. Provide rhythm annotation categories and rules to help the model understand rhythmic characteristics in the audio. Audio quality annotation: Evaluate the quality of the audio, including noise level, clarity, and recording equipment. Provide standards and levels for audio quality assessment. Speech recognition confidence annotation: For speech recognition tasks, annotate the confidence or reliability of the recognition results. Provide rules and ranges for confidence annotation. Audio context annotation: Annotate the context or situational information of the audio to help the model understand its setting and background. Provide methods and guidance for context annotation. Other...
Video	Video annotation rules should include, but are not limited to, the following: Video classification annotation: Annotate the type of video, such as film, TV series, or advertisement. Scene classification annotation: Annotate the scene type in the video, such as indoor, outdoor, urban, or natural. Video tagging annotation: Annotate key frames in the video, such as character appearances or important plot points. Object classification annotation: Annotate the category of objects appearing in the video, such as cars, trees, people, or buildings. Object location annotation: Annotate the location of objects in the video, such as by pixel coordinates or bounding boxes. Object quantity annotation: Annotate the number of objects appearing in the video, such as single or multiple. Object attribute annotation: Annotate the attributes of objects in the video, such as color, size, shape, or texture. Video semantic segmentation annotation: Segment pixels in the video into different semantic categories, such as background or foreground. Video instance segmentation annotation: Segment different instances in the video, such as separate annotations for each person in a clip. Video emotion annotation: Annotate the emotions present in the video, such as sadness or joy. Video object motion annotation: Annotate the motion trajectory or actions of objects in the video, including speed, direction, and movement trajectory. Provide categories and guidance for motion annotation to help the model understand object movements. Video timestamp annotation: Annotate the timestamp of each frame or key event in the video to support time-related analysis. Provide the format and rules for timestamp annotation. Video audio annotation: Annotate the audio portion of the video, including audio content, speech recognition, and emotions. Provide standards and classification guidelines for audio annotation. Video camera angle annotation: Annotate the camera angle and perspective of the video, such as overheat shot, shot looking up, or side view. Provide categories and guidance for camera angle annotation to help the model understand visual perspectives.

Category	Annotation Type Examples
	Video special effects annotation: For videos with special effects or post-processing, annotate the type and location of special effects. Provide rules and categories for special effects annotation. Video emotion intensity annotation: Annotate the emotional intensity of each scene or plot point in the video, such as emotional highs or lows. Provide standards and levels for emotion intensity annotation. Video plot description annotation: Require annotators to provide a text description of the video to capture its plot, events, and dynamics. Provide rules and evaluation criteria for plot description annotation. Video scene transition annotation: Annotate scene changes or transitions in the video, including transition types and timing. Provide categories and criteria for transition annotation. Video animation element annotation: For animated or special effects videos, annotate animation elements such as special effects or character movements. Provide guidance and classifications for animation element annotation. Other...
3D	3D data annotation rules should include, but are not limited to, the following: Geometric shape annotation: Annotate objects in the scene based on their geometric shapes, such as the shapes of the objects (sphere, cuboid, cylinder, etc.) and their dimensions (radius, length, width, etc.). Object classification annotation: Classify and annotate objects based on their appearance, such as annotating objects as different categories like people, cars, trees, or buildings. Position and coordinate annotation: Annotate the position and coordinate information of objects in 3D space, including position (coordinates or relative position), pose (rotation angles), and offsets. Occlusion and relationship annotation: Annotate the occlusion relationships and spatial relationships between objects, such as whether one object is blocking another or the distance and directional relationship between two objects. Motion trajectory annotation: If there are moving objects in the scene, annotate their motion trajectories, including start position, target position, and path information. Lighting and texture annotation: Annotate information related to lighting and texture, such as surface texture, lighting conditions, and shadow. Bounding box annotation: Annotate the bounding box of an object, which is the smallest rectangle or geometric shape enclosing the object. Object attribute annotation: Annotate object attribute information, such as color, material, transparency, or reflectivity. Provide classifications and standards for object attribute annotation. Environment annotation: Annotate the environmental information of a 3D scene, including sky, ground, bodies of water, etc. Provide categories and descriptions for environment annotation. Camera parameter annotation: Annotate the camera parameters and settings, including focal length, aperture, and exposure time. Provide rules and ranges for camera parameter annotation. Scene annotation: Annotate the characteristics and structure of the overall 3D scene, including object distribution, layout, and overall shape. Provide methods and guidance for scene annotation. Light source annotation: Annotate the light sources in the scene, including type, position, and intensity. Provide classifications and rules for light source annotation. 3D model annotation: If the scene contains 3D models, annotate their appearance, shape, texture, and other features. Provide guidelines and categories for 3D model annotation. Moving object annotation: For moving objects, annotate their trajectory, speed, acceleration, and other motion information. Provide rules and data formats for moving object annotation. Stereo vision annotation: For stereoscopic images or 3D scenes, annotate depth information,

Category	Annotation Type Examples
	parallax maps, and other stereo vision-related data. Provide methods and standards for stereo vision annotation. Occluding object annotation: Annotate which objects are occluding other objects to capture occlusion relationships. Provide criteria and methods for occluding object annotation. Other...
Time series	Time series data annotation rules should provide clear annotation methods and illustrative examples. The rules should include, but are not limited to, the following: Event detection annotation: Annotate events in time series data, including marking key events detected in the time series, such as abrupt changes, peaks, fluctuations, or other specific patterns. Classification annotation: Classify and annotate time series data based on their characteristics, such as annotation whether they belong to categories like normal, abnormal, faulty, or periodic. Trend and periodicity annotation: Annotate trends and periodic patterns in the time series, including upward trends, downward trends, and periodic oscillations. Threshold and anomaly annotation: Annotate anomalies in the time series based on predefined thresholds. Anomaly annotation can be used to detect and mark outliers or abnormal points in the time series. Correlation and relationship annotation: Annotate correlations or relationships between time series datasets, such as marking similarity, correlation coefficients, or other statistical indicators. Prediction and regression annotation: Annotate prediction or regression results for time series data, including predicted values for future time points or regression values for target variables. Periodicity analysis annotation: Annotate periodic characteristics in time series data, including cycle length and amplitude. Provide methods and periodicity types for periodicity analysis annotation. Data quality annotation: Assess the quality of time series data, including missing data, noise level, and outliers. Provide standards and quality levels for data quality annotation. Time lag and delay annotation: Annotate time lags or delays between time series datasets to describe the delayed effects of events or data. Provide methods and information for time lag and delay annotation. Time series model annotation: For time series forecasting tasks, annotate the parameters or model type of the time series model. Provide rules and parameter descriptions for time series model annotation. Time series segmentation annotation: Segment time series data and annotate the features or patterns of each segment. Provide methods and criteria for time series segmentation annotation. Interaction annotation between time series data: Annotate interaction relationships between different time series data, such as synergy effects or influence relationships. Provide classifications and relationship guidelines for interaction annotation. Other...

Appendix C
(Normative)
Main Safety and Security Risks of Corpora and Generated Content

A.1 Contains content that violates the socialist core values concept

Contains the following content:

a) Incitement to subvert state power and overthrow the socialist system;
b) That which endangers national security and interests, and harms the image of the state;
c) Incitement of separatism, or undermining national unity and social stability;
d) Promotion of terrorism or extremism;
e) Promotion of ethnic hatred (民族仇恨);
f) Promotion of violence or obscenity and pornography;
g) Dissemination of false and harmful information;
h) Other content prohibited by laws and administrative regulations.

A.2 Contains discriminatory content

Contains the following content:

a) Ethnic (民族) discrimination;
b) Discrimination on the basis of beliefs;
c) Nationality-based (国别) discrimination;
d) Discrimination on the basis of regional origin;
e) Gender discrimination;
f) Age discrimination;
g) Occupation-based discrimination;
h) Health-based discrimination;
i) Other types of discriminatory content.

A.3 Commercial violations

The main risks include:

a) Infringement of intellectual property rights (IPR) of others;
b) Violation of business ethics;
c) Disclosure of the trade secrets of others;
d) Use of algorithms, data, platforms, etc. to engage in monopolistic or unfair competition behaviors;
e) Other commercial violations.

A.4 Violations of the legitimate rights and interests of others

The main risks include:

a) Endangerment of the physical or mental health of another.
b) Unauthorized use of the likeness of another;
c) Defamation of the reputation of another;
d) Defamation of the honor of another;
e) Infringement of others' right to privacy;
f) Infringement of others' personal information rights and interests;
g) Infringement of other legitimate rights and interests of others.

A.5 Inability to meet the safety or security requirements of specific service types

The main safety and security risks in this area are those that exist when GenAI is used for specific service types with higher safety or security requirements, such as automatic control, medical information services, psychological counseling, critical information infrastructure, etc.:

a) Inaccurate content that is grossly inconsistent with common scientific knowledge or mainstream perception;
b) Unreliable content that, although not containing grossly erroneous content, cannot help the user.