ICS 35.240.01
CCS L 72

中华人民共和国国家标准

GB/T XXXXX—XXXX

信息技术　藏语语音识别数据标注

Information technology—Data labeling for Tibetan speech recognition

（本草案完成时间：2026 年 11 月 18 日，征求意见稿）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX-XX-XX 发布　　XXXX-XX-XX 实施

国家市场监督管理总局
国家标准化管理委员会　发布

目　　次

前言 III
1 范围 1
2 规范性引用文件 1
3 术语和定义 1
4 藏语语音数据标注流程 2
5 标注任务前期准备要求 2
5.1 待标注数据的采集及确定 2
5.2 语音数据标注任务 3
5.3 语音时间边界定位规则 3
5.4 文件标注规则 4
5.5 标注需求变更流程 8
5.6 标注环境要求 8
6 标注任务执行过程的要求 8
6.1 标注任务创建要求 9
6.2 标注任务分发要求 9
附录A （资料性）标注人员管理要求 10
A.1 标注人员范围 10
A.2 标注人员培训 10
A.3 标注人员能力档案 10
参考文献 11

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由全国信息技术标准化技术委员会（SAC/TC 28）提出并归口。

本文件起草单位：西藏大学、电子科技大学、中国电子技术标准化研究院、青海师范大学、中国科学院声学研究所。

本文件主要起草人：尼玛扎西、于永斌、仁增多杰、黄姗姗、陈晓研、洛桑嘎登、王向向、头旦才让、班玛宝、仁青东主、赵庆卫、群诺、范满平、朱宇雷、冯箫、盛天颖、张子玥、道吉扎西、拉毛吉、钱木吉、蔡郁青、王超、李臣波、郑智文、刘羽童、张瑾、才让卓玛。

信息技术　藏语语音识别数据标注

1　范围

本文件规定了信息技术领域藏语语音识别数据标注的要求。

本文件适用于藏语卫藏方言、康巴方言和安多方言的自然语言处理、语音翻译、同声传译、实时翻译等语音处理过程。

2　规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 13000—2025　信息技术 通用编码字符集（UCS）

GB/T 45674—2025　网络安全技术 生成式人工智能数据标注安全规范

3　术语和定义

下列术语和定义适用于本文件。

3.1

藏语语音数据标注　Tibetan speech data labeling

为藏语语音数据样本指定目标变量和赋值的过程。

3.2

藏语语音标注任务　Tibetan speech labeling task

按藏语语音数据标注说明对数据进行标注的活动。

3.3

藏语语音数据标注方　Tibetan speech data labeler

承担藏语语音数据标注任务的人员或机构。

3.4

藏语语音数据需求方　Tibetan speech data user

提出藏语语音数据标注需求的人员或机构。

3.5

藏语语音标注管理方　Tibetan speech data labeling administrator

管理藏语语音数据标注任务评估、分发、交付、验收以及质量控制的人员或机构。

3.6

藏语语音标注工具　Tibetan speech labeling tool

具备音频波形与时间边界编辑、三层结构化标注、标准藏文编辑、专用标签录入、任务与日志管理、多角色权限及安全管控功能，并能用于标注结果验收的软件工具或辅助系统。

3.7

标注任务说明 labeling task description

藏语语音数据需求方用于向藏语语音标注管理方以及藏语语音数据标注方明确标注任务的文字表达。

注：标注任务说明通常包含对要执行的标注任务的描述、标注方法、正反示例、验收方法与验收指标等内容。

4 藏语语音数据标注流程

藏语语音数据标注的流程主要涉及数据需求方、标注管理方和数据标注方三方人员，藏语语音数据标注流程见图1。

在标注前期，数据需求方提出语音标注需求；标注管理方根据需求确定标注任务，包括语音数据的采集方案和标注规范的制定（标注规范依据本文件规定的标注规则和数据需求方的具体要求制定），并根据语音标注数据需求确定标注人员的专业范围与培训需求，以及对可能存在缺失值、噪声或重复数据等质量问题进行数据清洗处理，检证数据质量达标；清洗后的数据进入标注阶段，这是流程中最重要的部分，其中标注管理方根据不同需求进行任务拆分与标注人员分配，并交由标注方执行具体的语音数据标注任务。

标注完成后，由标注管理方对标注结果进行检查，若不符合本文件规定的标注规则，则将任务退回重新标注，直至使数据准确、一致。标注任务执行阶段，三方人员应按控制流程，使标注任务质量符合要求。最终，标注后的数据经过内部质检，若符合需求则交付给数据需求方，否则进行后期优化直至达到数据需求方确认的验收要求。

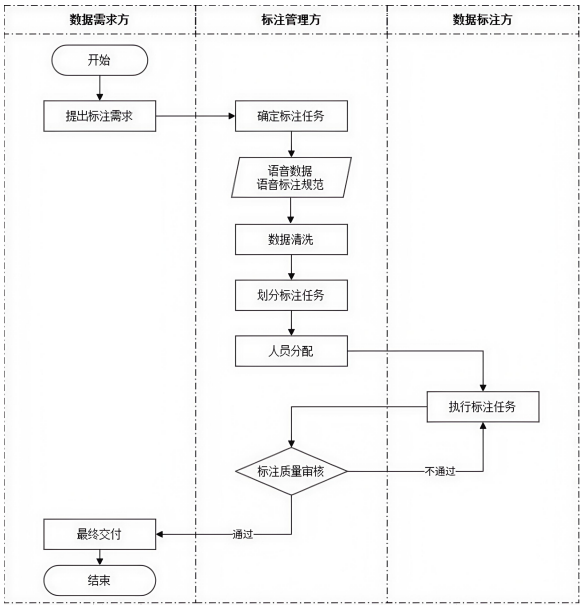


图 1 藏语语音数据标注流程

5 标注任务前期准备要求

5.1 待标注数据的采集及确定

待标注数据按以下要求采集及确定。

a) 数据采集：标注管理方按数据需求方的要求采集各类藏语语音数据，如近场语音、远场语音、自然场景对话、电视媒体等类别。

b) 待标注数据确定。包括：

1) 待标注数据整理：数据标注前，数据需求方应对已采集的待标注数据进行分析、数据核查及数据整理，并根据待标注数据的内容对数据进行聚类，为后续标注做准备。

2) 数据安全确定：数据需求方、标注管理方及数据标注方应按照 GB/T 45674—2025 给出的要求，检查数据在采集、传输、存储和标注过程中的安全性。对涉及敏感信息的数据（如个人隐私、商业机密等），需进行脱敏处理或加密保护。标注任务中应明确数据安全责任，各方需定期检查数据内容及存储环境，防止数据泄露或滥用。

5.2 语音数据标注任务

5.2.1 概述

语音数据标注任务包括语音时间边界定位和文件标注两方面任务。

5.2.2 语音时间边界定位

在藏语语音识别中，语音时间边界定位通常指的是对语音信号中每个音素、词汇或句子的开始和结束时间进行标注。

5.2.3 文件标注

文件标注按层次分为：

a) 全局（GLOBAL）层的标注；

b) 说话人（SPEAKER）层的标注；

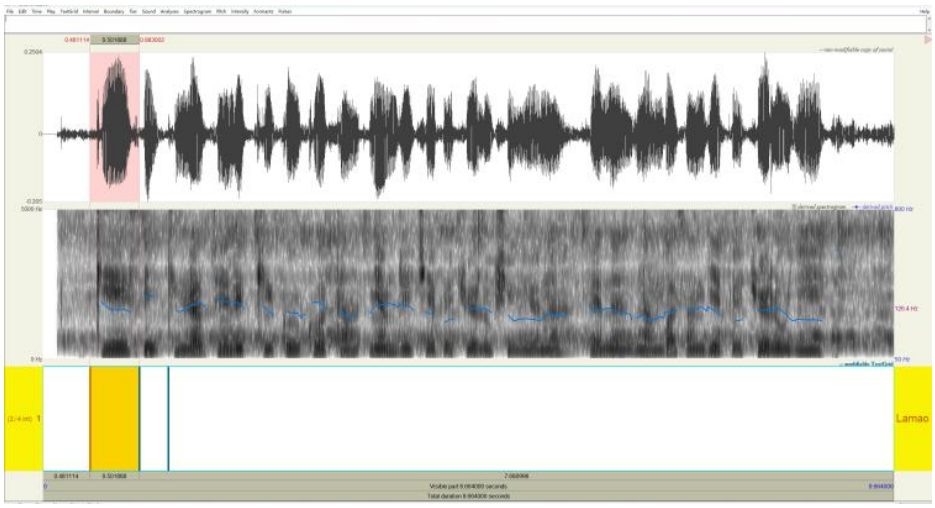
c) 内容（CONTENT）层的标注。

5.3 语音时间边界定位规则

5.3.1 缩放波形

宜放大3~4倍标注，可视时长约7.5s~15s。

示例：



5.3.2 听音

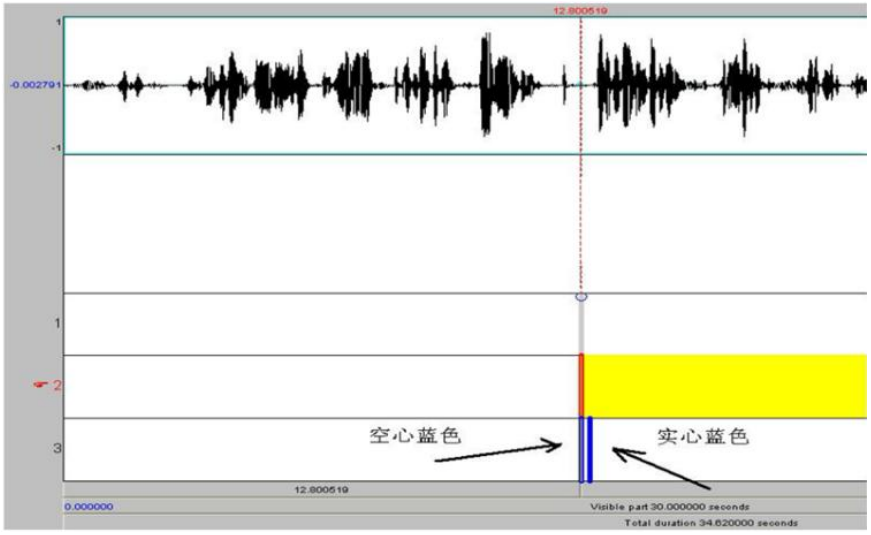
在整段语音数据样本的基础上，根据语义和停顿时间等因素，在音频信号中每一句话的句首和句尾分别添加时间边界。一般时长不小于2s，不超过15s。

5.3.3 添加时间边界

添加时间边界时不应切掉语音的首和尾，语音信号中每个音素、词汇或句子的开始和结束时间前后波形一般进行500ms留白。如果相邻的两段语音的留白小于500ms，时间边界应打在两段语音中间的位置。

在整个语音文件中，SPEAKER层和CONTENT层的时间边界的数目是等同的，每一对时间边界应保持一致。

示例：



5.4 文件标注规则

5.4.1 全局（GLOBAL）层的标注

全局（GLOBAL）层主要标注语音文件的一些全局信息，包括说话人性别信息、语种信息（方言区）和角色信息。

标注格式如下：

[speaker]:[空格]1[空格]性别,[空格]2[空格]性别;[空格][language]:[空格]1[空格]方言区,[空格]2[空格]方言区;[空格][role]:[空格]1[空格]角色1,[空格]2[空格]角色2,[空格]3[空格]角色3

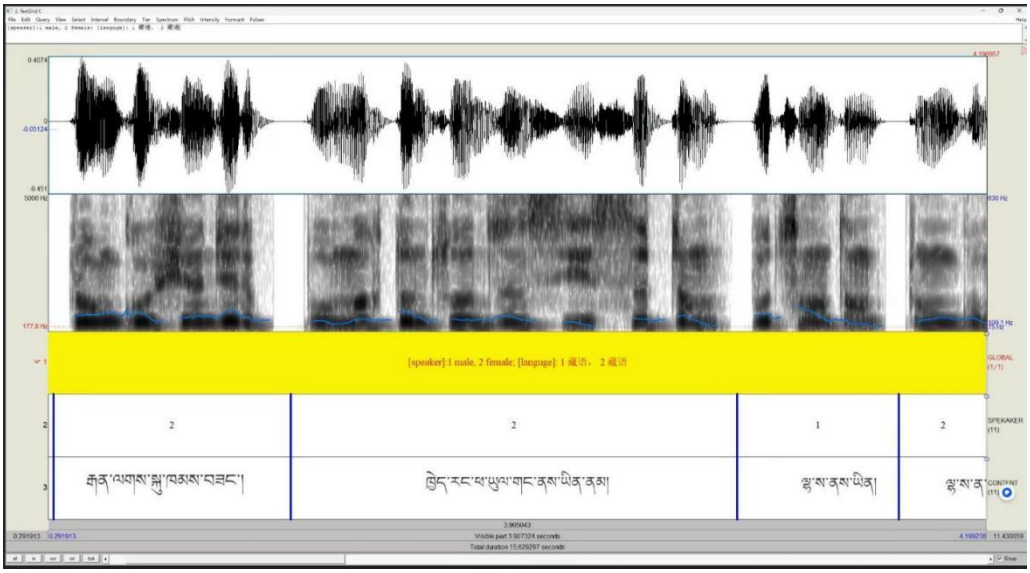
示例：

```
[speaker]: 1 Male, 2 Female; [language]: 1 Weizang, 2 Kangba, 3 Anduo; [role]: 1 Teacher, 2 Student, 3 Other
```

5.4.2 说话人（SPEAKER）层的标注

说话人（SPEAKER）层用于标注说话人信息。“说话人”的取值对应GLOBAL层[role]字段中定义的编号。通常情况下，编号为1和2，分别表示SPEAKER1和SPEAKER2。若GLOBAL层定义了其他说话人（例如：“3 other”），则SPEAKER层在对应语音段也可标注为相应的编号。

示例：



5.4.3 内容（CONTENT）层的标注

5.4.3.1 内容（CONTENT）层语音标注

5.4.3.1.1 常规标注

内容（CONTENT）层用于标注语音所对应的文本序列，标注要求如下。

a) 对于语音中简单的藏文词汇，在能听懂的情况下，直接标出即可。藏文文本标注应采用 GB/T 13000—2025 编码，不应使用其他藏文编码字符集。标注工具应具有显示与编辑 GB/T 13000—2025 藏文区字符的功能，且藏文字形正确。

b) 如果此语段为多人对话语音，应根据说话人的变化在 SPEAKER 层和 CONTENT 层标注时间边界，SPEAKER 层应标注对应说话人编号，CONTENT 层则输入相应的藏文文本。

c) 如果同一说话人说话时间较长，应根据语义合理地增加时间边界，每个时间段的长度一般不超过 15s，但断句不应过于零散或过短。

d) 针对 CONTENT 层的藏文文本，如果一个时间段内文本较长，应根据语意添加标点符号。若时间段的结束，恰好是句子的结束，可省略标点，只标注时间边界，以提高标注效率。句中的标点符号应符合藏语文的书写规范，为该语种状态下的标点符号。

e) 每个时间边界的最佳位置应在音频能量的最低点，如果某些字词无法完整包含在时间段内，则宜舍弃这些字词，并按本文件规定的标注规则标注舍弃的部分。

5.4.3.1.2 叠加

关于重叠（交叉）的语音，是指主要说话人同时说话的情况。即：某个人的一句话未完，另一个人已经开始说话。

对叠加的标注要求如下。

a) 当主要说话人音量相当时，在 SPEAKER 层和 CONTENT 层标注时间边界，SPEAKER 层不标注内容，CONTENT 层标注[+]。

b) 当主要说话人一人音量特别小（几乎和环境音音量相当）时，在 CONTENT 层标注音量较大且清晰的那个说话人的语音内容，并以[+]号结尾，SPEAKER 层标注该说话人编号。

c) 当一个说话人持续说话，而另一个人只表示在听的发出“ཡ།（译：好）”，“བ།（译：噢）”和“རྡ།（译：对）”等单字，并与持续说话人的语音交叠，则只标注持续说话人的语音内容。
d) 某个语音段中较小的叠加。如果这一段 A 说了“བཟཡ།（译：好的）”，B 说了“བ།（译：噢），那么这时因两人语音长度相当，则应标注叠加。同理，可延伸至：A 说了 3 个字而 B 说了 4 个字。
e) 对于叠加，应为真实的。不能将大段的听不清语音和叠加混在一起。叠加段内的非叠加部分，前后最多不得超过 1 个字。由于添加叠加的时间边界导致的半个语音（即切掉头或者尾）可用[*]或[UNK]来表示。

5.4.3.1.3 拖音

一般只针对于固有名词或者关键词的拖音进行标注。

对拖音的标注要求如下。

a) 如果一个字的大部分音都没有发出来，或者吞掉了，则可忽略此发音。比如特别短的磕巴，则可忽略前面短暂的磕巴音，直接正常标注即可。
b) 如果磕巴的音拖了很长，那么该字用括号括起来，并标注[-]，表示发音不全。
c) 如果不能确认声母是什么字，则标注成：[UNK]，代表前面有个不知道的音。

5.4.3.1.4 吞音

一般只针对于固有名词或者关键词的吞音进行标注。

对吞音的标注要求如下。

a) 对于因语速过快而导致的全吞音现象，如果影响了关键词的理解，应按吞音处理。
b) 如果是发音错误，则按实际发音标注。
c) 若发音不明确（模棱两可），则优先标注关键词。

5.4.3.1.5 情绪

情绪类别包括中性、高兴、愤怒、悲伤和其他情绪，分别使用 [NEUTRAL]、[HAPPY]、[ANGRY]、[SAD] 和 [OTHER] 标签进行标注。

对情绪的标注要求如下。

a) 当语句中出现明显的高兴、愤怒或悲伤情绪时，应在相应语句末尾分别添加 [HAPPY]、[ANGRY] 或 [SAD] 标签。
b) 当能够判断情绪类别，但情绪表现较弱或不明显时，应在相应情绪标签中使用“-”表示弱情绪，例如：[HAPPY-]、[ANGRY-] 或 [SAD-]。
c) 当语句中无明显情绪倾向时，应标注为 [NEUTRAL]。
d) 当语句中存在明显情绪，但难以归入高兴、愤怒或悲伤类别时，可标注为 [OTHER]。
e) 当情绪出现在叠音部分时，应在相应情绪标签前添加“+”，例如：+[HAPPY]、+[ANGRY]、+[SAD]、+[HAPPY-]、+[ANGRY-] 或 +[SAD-]。

5.4.3.1.6 非主要说话人语音

非主要说话人语音是指不在对话中的语音。例如：说话人A与说话人B进行对话，出现说话人C的非对话相关语音。

对非主要说话人语音的标注要求如下。

a) 若 C 的语音清晰且较长，应标注出 C 的语音内容。其处理流程如下。

1) GLOBAL 层定义：按 5.4.1 的规定，在 GLOBAL 层的[role]字段中为此类非主要说话人定义一个角色及编号。例如：“3 other”。
2) SPEAKER 层标注：按 5.4.2 的规定，在 SPEAKER 层对应的时间段标注该非主要说话人的编号（在此例中为 3）。
3) 统一处理：若音频中出现多个不同的远讲声音，只要它们都属于非主要说话人范畴，均可使用在 GLOBAL 层定义的同一个 other 角色及其编号进行标注。例如，若 GLOBAL 层已定义“3 other”，则后续出现的某个自然人的远讲语音和清晰的电视音，在 SPEAKER 层均可标注为 3。

b) 若 C 的语音清晰但短暂，可视为环境噪音处理。
c) 若 C 的语音不清晰，也按环境噪音处理。

5.4.3.1.7 合成语音

对于来源为合成的语音，应标注出语音内容，并在句首以[TTS]开头。
SPEAKER层编号1代表男，编号2代表女。

5.4.3.1.8 录制语音

对于来源为录制的语音，应标注出语音内容，并在句首以[RECORD]开头。
SPEAKER层编号1代表男，编号2代表女。

5.4.3.2 内容（CONTENT）层短暂噪音标注

短暂噪音是指非常短暂的突发声音,所有此类标注都是中括号与语音内容的组合,不标注时间边界。对短暂噪音的标注要求如下。

a) 听不清的藏文字词：直接在句子中标注[UNK]。
示例 1：“བོད་ལྗོངས་སློབ་ཚན་[UNK] ཚན་རིག་སློབ་གྲྭ།（译：西藏大学[UNK]科学技术学院）”。
b) 短暂的笑声：直接在句子中标注[LAUGH]。
c) 短暂的由说话人发出的干扰浊音：直接在句子中标注[SONANT]。
示例 2：咳嗽声、打喷嚏、清嗓子、呼吸声等。
d) 短暂的环境噪声：直接在句子中标注[GAR]。
示例 3：拍桌子、关门、敲击、鼓掌等。
e) 系统提示音，即系统播出的语音提示：[PROMPT]+系统自动播放的语音内容，而非说话人的语音内容。
示例 4：“[PROMPT]འདི་ཀུང་ཧེ་ལས་བར་གཏོང་བར་དགའ་བསུ་ཞུ། དེ་ཉུ་ཁལས་ཀྱ་བ་ཨང་རྟགས་དང་པོས་ཚོང་ལ་ཁལས་ཞུ་ཐུབ།（译：[PROMPT]欢迎致电我公司，现在由一号客服为您服务）”。

5.4.3.3 内容（CONTENT）层持续噪音标注

持续噪音是指一段时长大于500ms的非语音声音，所有此类标注都是单独的噪音类型，应标注中括号和时间边界，SPEAKER层不标注内容。

对持续噪音的标注要求如下。

a) 明显的静音段（大于 500ms）:[SIL]。
b) 表示多说话人：[MORE]。
c) 非目标标注语音段：[NO]。
示例 1：当需标注的是藏语，而此语音段是普通话、四川话或粤语等其他语种或方言时，标注为[NO]。
d) 各种垃圾语音：[*]。

示例 2：听不清的长句、小地方方言等。

e) 各种垃圾环境噪音：[ENS]。

示例 3：连续的拍桌子、连续的敲击声、持续的各种环境噪音（大于 500ms）等。

f) 持续的音乐声：[MUSIC]。

示例 4：唱歌声（有歌词和旋律）、哼唱（没有歌词，但有旋律）、口哨声、可能是由背景的电视或收音机发出的音乐和歌声等。

g) 录音及电信系统引起的噪声，应区分标注如下：

1) 彩铃：[COLORING]；
2) 回铃音：[BELL]；
3) 忙音：[BUSY]；
4) 传真音：[FAX]；
5) 按键音：[DTMF]；
6) 不属于以上情况的系统音：[SYS]。

h) 如果不属于以上情况，则用[#]进行标注。

5.5 标注需求变更流程

数据需求方提出需求变更时，应按以下流程。

a) 需求变更通知：数据需求方及时将变更需求以书面形式通知标注管理方，包括变更的具体内容和原因。

b) 评审变更：

1) 变更评审：标注管理方对变更需求进行详细评审，评估变更对标注任务的影响，包括对标注规则、任务可行性和资源需求的影响；
2) 可行性分析：确认变更后的需求是否可行，并评估所需的调整和额外工作量。

c) 更新标注任务：

1) 任务调整：根据评审结果更新标注任务，包括修改标注规则、调整时间安排和资源配置；
2) 文档更新：更新相关文档和标注规范（含本文件规定的标注规则及数据需求方的具体要求），检查确认所有参与者掌握最新的标注要求和流程。

d) 重新进入评估阶段：

1) 预标注测试：在更新后的标注任务上进行小规模的预标注测试，验证变更是否有效且操作性强；
2) 结果验收：将测试结果提交给数据需求方进行验收，根据反馈进行必要的调整。

e) 正式执行：在数据需求方确认更新后的标注任务无误后，正式启动变更后的标注任务。

5.6 标注环境要求

5.6.1 选择标注工具

选择适宜的标注工具，完成数据标注任务。

5.6.2 选择标注场地

数据标注方应根据标注任务中必要的数据安全要求和项目需求，选择或搭建适宜的数据标注场地。

6 标注任务执行过程的要求

6.1 标注任务创建要求

6.1.1 任务包创建要求

任务包创建过程的要求如下：

a) 数据分组：数据需求方应将数据按类型、内容或标注难度合理分组，检查每个任务包中的数据是否具有一致性，以提高标注的准确性和效率；

b) 质量控制：任务包创建时应考虑标注质量的监控措施，设置质量检查点和反馈机制，以检验数据标注的准确性和一致性；

c) 任务说明：为每个任务包提供明确的标注说明和标注规范（含本文件规定的标注规则及数据需求方的具体要求），以指导标注人员进行正确的操作，减少标注错误。

6.1.2 任务账户创建要求

任务账户创建过程的要求如下：

a) 账户配置：根据标注人员的角色和职责创建账户，包括标注人员、审核人员和管理员，检查每个账户是否具备适当的权限和功能；

b) 权限管理：为不同角色分配相应的权限，包括数据访问、标注编辑和质量审核权限，以保护数据安全；

c) 账户规则：制定账户使用规则，包括登录安全、操作规范和数据保护要求。

6.2 标注任务分发要求

6.2.1 标注管理方应根据标注任务的具体要求、环境和工具选择合适的任务分发方式。

常见的方式包括：

a) 主动领取：标注人员自选任务；

b) 系统自动分发：系统根据任务需求和人员状态自动分配。

6.2.2 标注管理方应制定优化策略，平衡标注质量和速度。

示例：针对任务复杂度设置合理的优先级。

6.2.3 标注任务分发应采取以下保障措施：

a) 安全管理制度：制定并实施标注任务分发的安全管理制度，检查分发过程的安全性和合规性，包括数据传输安全、访问控制和权限管理等方面；

b) 日志记录与监控：

1) 明确任务分发日志的内容，包括分发时间、任务分配情况和操作记录；

2) 实施实时监控，以便及时发现并解决可能的安全问题或异常情况。

附 录 A
（资料性）
标注人员管理要求

A.1 标注人员范围

数据标注方应根据标注任务内容中规定的标注人员资质和专业需求，选拔符合要求的人员进入标注人员培训环节。标注人员应具备必要的背景知识和技能，能够胜任指定的标注任务。

A.2 标注人员培训

数据标注方应制定并实施针对性的岗前培训计划，根据标注任务说明对标注人员进行系统培训。培训内容应包括标注标准、工具使用方法、数据处理流程等。培训后应进行标注能力考试，只有通过考试的人员方可参与实际标注任务。

A.3 标注人员能力档案

数据标注方应建立详细的标注人员能力档案，记录每位标注人员的培训经历、考试成绩、任务参与情况以及标注质量评价。承担任务结束后，应及时更新标注人员的能力档案，用于进行标注人员能力评估与标注质量追踪，并作为后续培训和任务分配的依据。

参 考 文 献

[1] GB/T 42755—2023 人工智能 面向机器学习的数据标注规程