

T/ACCEM

团 体 标 准

T/ACCEM XXXX—2025

数据标注流程与操作技术规范

Data annotation process and operational technical specifications

（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。

2025 – XX – XX 发布

2025 – XX – XX 实施

中国商业企业管理协会 发 布

目 次

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 总体要求 1

 4.1 准确性 1

 4.2 一致性 1

 4.3 完整性 1

 4.4 时效性 2

 4.5 安全性 2

5 数据标注流程 2

 5.1 任务准备 2

 5.2 数据分配 2

 5.3 数据标注 2

 5.4 标注审核 2

 5.5 结果验收 2

 5.6 过程记录 2

6 操作规范 3

 6.1 图像数据标注 3

 6.2 文本数据标注 3

 6.3 音频数据标注 3

 6.4 视频数据标注 3

7 质量管控 3

 7.1 质量标准 3

 7.2 质量检查方式 4

 7.3 质量问题处理 4

8 安全管理 4

 8.1 数据安全 4

 8.2 系统安全 4

 8.3 人员安全 4

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。

本文件由宁波博登智能科技有限公司提出。

本文件由中国商业企业管理协会归口。

本文件起草单位：

本文件主要起草人：

数据标注流程与操作技术规范

1 范围

本文件规定了数据标注的总体要求、术语定义、标注流程、操作规范、质量管控、安全管理及附录等内容。

本文件适用于各类数据标注任务，包括但不限于图像、文本、音频、视频等数据类型的标注工作，涉及数据标注项目的管理人员、标注人员、审核人员及相关技术支持人员。

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 25069-2022 信息安全技术 术语
- GB/T 35273-2020 信息安全技术 个人信息安全规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据标注

对原始数据进行处理、标记和注释，使其成为可供机器学习模型训练、验证和测试的结构化数据的过程。

3.2

标注任务

根据特定的业务需求和标注规则，对一批数据进行标注的具体工作安排。

3.3

标注规则

用于指导标注人员进行数据标注的具体说明和要求，包括标注对象、标注类型、标注格式等。

3.4

标注工具

用于辅助标注人员完成数据标注工作的软件或系统，具备数据加载、标注操作、结果保存等功能。

3.5

审核

对标注人员完成的标注结果进行检查和验证，确保标注结果符合标注规则和质量要求的过程。

4 总体要求

4.1 准确性

标注结果应准确反映数据的真实特征和含义，符合标注规则的要求，错误率应控制在规定范围内。

4.2 一致性

对于相同或相似的数据，不同标注人员的标注结果应保持一致；同一标注人员在不同时间对相同数据的标注结果也应保持一致。

4.3 完整性

标注应覆盖数据中需要标注的全部内容，不遗漏任何关键信息。

4.4 时效性

应在规定的时间内完成数据标注任务，确保标注工作的进度符合项目要求。

4.5 安全性

标注过程中应采取必要的安全措施，保护数据的机密性、完整性和可用性，防止数据泄露、篡改或丢失。

5 数据标注流程

5.1 任务准备

- 5.1.1 项目管理人员应明确标注任务的目标、范围、要求和时间节点，制定标注计划。
- 5.1.2 应收集和整理需要标注的原始数据，确保数据的完整性、可用性和合法性。
- 5.1.3 应制定详细的标注规则，包括标注对象的定义、标注类型的划分、标注格式的规范等，并对标注规则进行评审和确认。
- 5.1.4 应选择合适的标注工具，确保标注工具具备完成标注任务所需的功能，并对标注工具进行调试和配置。
- 5.1.5 应对标注人员进行培训，使其熟悉标注规则、标注工具的使用方法以及项目的相关要求，培训后进行考核，考核合格方可参与标注工作。

5.2 数据分配

- 5.2.1 项目管理人员应根据标注人员的数量、能力和工作量，将标注任务合理分配给各个标注人员。
- 5.2.2 在数据分配过程中，应考虑数据的多样性和代表性，确保每个标注人员分配到的数据具有一定的随机性和均衡性。
- 5.2.3 向标注人员下达标注任务通知书，应明确标注数据的范围、数量、完成时间和质量要求等。

5.3 数据标注

- 5.3.1 标注人员接收标注任务和数据后，应使用标注工具按照标注规则进行标注操作。
- 5.3.2 在标注过程中，标注人员应认真核对数据，确保标注结果的准确性和一致性。
- 5.3.3 对于标注过程中遇到的疑问或不确定的情况，标注人员应及时向项目管理人员或技术支持人员咨询，不得擅自进行标注。
- 5.3.4 标注人员应定期保存标注结果，避免因意外情况导致标注成果丢失。

5.4 标注审核

- 5.4.1 标注人员完成标注任务后，应将标注结果提交给审核人员进行审核。
- 5.4.2 审核人员按照标注规则和质量要求，应对标注结果进行全面检查和验证，包括准确性、一致性、完整性等方面。
- 5.4.3 审核过程中，对于发现的错误或不符合要求的标注结果，审核人员应及时反馈给标注人员，并说明错误原因和修改意见。
- 5.4.4 标注人员应根据审核意见对标注结果进行修改和完善后，再次提交审核，直至审核通过。

5.5 结果验收

- 5.5.1 所有标注结果审核通过后，项目管理人员应对标注任务进行整体验收，检查标注任务的完成情况、标注质量是否符合要求以及是否在规定时间内完成等。
- 5.5.2 验收合格后，应将标注结果进行整理和归档，形成最终的标注数据集。
- 5.5.3 对于验收不合格的标注任务，项目管理人员应分析原因，采取相应的整改措施，并重新组织标注和审核工作。

5.6 过程记录

在数据标注的整个流程中，项目管理人员应做好过程记录，包括任务分配情况、标注进度、审核结果、问题处理情况等，以便对标注过程进行追溯和总结。

6 操作规范

6.1 图像数据标注

6.1.1 目标检测标注

应使用矩形框、多边形等工具准确框选图像中的目标对象，确保目标对象完全包含在框内，且框的边缘与目标对象的轮廓尽可能贴合。标注目标的类别应准确无误，避免错标或漏标。

6.1.2 图像分类标注

应根据图像的内容和特征，将图像划分到相应的类别中。标注人员应熟悉各类别的定义和特征，确保分类的准确性。

6.1.3 语义分割标注

应对图像中的每个像素进行分类标注，明确每个像素所属的类别。标注应精细，边界清晰，避免出现模糊或错误的标注区域。

6.1.4 图像标注过程中，应保持标注的一致性，对于相同或相似的图像，标注方法和结果应统一。

6.2 文本数据标注

6.2.1 实体识别标注

识别文本中的实体，如人名、地名、组织机构名等，并标注出实体的类型和边界。实体的边界应准确，类型应正确。

6.2.2 情感分析标注

对文本的情感倾向进行标注，如积极、消极、中性等。标注人员应理解文本的含义和语境，准确判断文本的情感。

6.2.3 文本分类标注

根据文本的主题、内容等将文本划分到相应的类别中。标注应基于文本的实际内容，避免主观臆断。

6.2.4 文本标注应注意标点符号、大小写等细节的准确性，保持文本的完整性和可读性。

6.3 音频数据标注

6.3.1 语音转写标注

应将音频中的语音内容准确转写成文本，包括说话人的语音、语气、停顿等信息。转写应符合语言规范，错别字率应控制在较低水平。

6.3.2 声音分类标注

应根据音频的特征和内容，将音频划分到相应的类别中，如音乐、语音、环境音等。标注人员应具备一定的音频识别能力，确保分类的准确性。

6.3.3 音频标注过程中，应注意音频的时长、采样率等参数，确保标注结果与音频数据的一致性。

6.4 视频数据标注

6.4.1 视频目标跟踪标注

应对视频序列中的目标对象进行连续跟踪和标注，记录目标对象在不同帧中的位置和状态。标注应准确跟踪目标对象的运动轨迹，避免目标丢失或误跟踪。

6.4.2 视频行为识别标注

应识别视频中人物或物体的行为动作，并标注出行为的类别和发生时间。标注人员应理解行为的定义和特征，准确判断行为的类型和持续时间。

6.4.3 视频标注应注意视频的帧率、分辨率等参数，确保标注结果的准确性和连贯性。

7 质量管控

7.1 质量标准

7.1.1 准确性

不同数据类型的标注错误率应控制在一定范围内，例如图像目标检测标注错误率 $\leq 2\%$ ，文本实体识别标注错误率 $\leq 3\%$ 。

7.1.2 一致性

同一批数据由不同标注人员标注的一致性率应 $\geq 95\%$ ；同一标注人员在不同时间对相同数据标注的一致性率应 $\geq 98\%$ 。

7.1.3 完整性

标注覆盖率应达到 100%，不遗漏任何需要标注的内容。

7.2 质量检查方式

7.2.1 自检

标注人员应在完成标注任务后，对自己的标注结果进行自我检查，及时发现和纠正错误。

7.2.2 互检

应组织标注人员之间进行相互检查，互相发现和指出对方标注结果中的问题。

7.2.3 抽检

审核人员从标注结果中随机抽取一定比例的数据进行检查，根据抽检结果评估整体标注质量。抽检比例应根据标注任务的规模和重要性确定，一般不低于10%。

7.2.4 全检

对于重要性高、质量要求严格的标注任务，审核人员应对所有标注结果进行全面检查。

7.3 质量问题处理

7.3.1 对于检查中发现的质量问题，应及时反馈给标注人员，标注人员应在规定时间内进行修改和重新提交。

7.3.2 对多次出现质量问题的标注人员，应进行再次培训或调整其标注任务。

7.3.3 定期对标注质量进行统计和分析，总结质量问题的类型和原因，采取相应的改进措施，不断提高标注质量。

8 安全管理

8.1 数据安全

8.1.1 标注数据应采取加密存储和传输措施，防止数据泄露。

8.1.2 应严格控制数据的访问权限，只有授权人员才能接触和处理标注数据。

8.1.3 标注过程中不得将数据复制、传播或用于其他目的，标注任务完成后，应按照规定对数据进行处理，如删除或归还。

8.1.4 对于涉及个人信息的标注数据，应符合 GB/T 35273 的要求，采取匿名化或去标识化处理等措施，保护个人隐私。

8.2 系统安全

8.2.1 标注工具和相关系统应安装杀毒软件和防火墙，定期进行安全漏洞扫描和更新，确保系统的安全性。

8.2.2 标注人员应使用强密码登录系统，并定期更换密码，防止账号被盗用。

8.2.3 禁止在标注设备上安装未经授权的软件或连接未经安全认证的外部设备。

8.3 人员安全

8.3.1 应对标注人员进行安全意识培训，使其了解数据安全和系统安全的重要性，掌握基本的安全防护知识和技能。

8.3.2 应与标注人员签订保密协议，明确其在数据标注过程中的保密责任和义务。