ICS 03.080.01
CCS A 00

团 体 标 准

T/CECC XX—2025

面向人工智能的数据标注合规指南

Compliance guidelines on data labeling for artificial intelligence

标准修订稿

【本修订稿完成时间：2025 年 7 月】

2025-XX-XX 发布 2025-XX-XX 实施

中国电子商会 发 布

目　　次

前　　言 II
引　　言 III
1 范围 1
2 规范性引用文件 1
3 术语和定义 1
4 基本合规原则 2
5 数据采集合规 3
5.1 数据来源合规性 3
5.2 用户授权与告知 3
5.3 数据内容合规性 4
6 数据管理合规 4
6.1 数据分类分级 4
6.2 数据监控记录 4
6.3 数据访问控制 4
6.4 数据清洗管理 4
6.5 数据投毒防范 4
7 数据标注流程合规 5
7.1 标注人员管理 5
7.2 标注质量控制 5
7.3 技术工具及场地要求 5
8 数据传输合规 6
8.1 数据存储与传输 6
8.2 数据共享 6
9 数据删除合规 6
9.1 数据删除触发条件 6
9.2 数据销毁技术要求 6
9.3 备份数据管理规范 6
10 合规管理保障 7
10.1 组织架构 7
10.2 制度建设 7
10.3 合规审计 7
10.4 偏见防控 7
10.5 认证监督 7
附录 A（资料性）数据标注配套文件及指南 8
附录 B（规范性）标注质量与数据分类规范 13
参 考 文 献 15

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。

请注意本文件的某些内容可能涉及专利，本文件的发布机构不承担识别专利的责任。

本文件由XXXX发起。

本文件由中国电子商会归口。

本文件起草单位：XXXX。

本文件主要起草人：XXXX。

引　言

随着人工智能、大数据等技术的快速发展，数据标注作为数据要素加工与价值释放的核心环节，已成为推动人工智能产业创新和企业数字化转型的重要基础。数据标注的规范性和准确性，直接影响数据资源应用的可靠性及算法的公平性。

为促进人工智能产业数据标注活动高质量健康发展，规范数据标注活动全流程，保障数据主体合法权益，依据《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等相关法律，《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》《人工智能生成合成内容标识方法》等相关部门规章，结合我国人工智能产业数据标注的实践发展，制定本文件。

面向人工智能的数据标注合规指南

1 范围

本文件规定了人工智能产业数据标注活动在数据采集、数据管理、数据标注流程、数据传输、数据删除等各数据标注环节中应遵循的合规原则和行为指南，以及可供参考的合规管理保障措施。

本文件适用于指导中华人民共和国境内人工智能数据标注方开展数据标注活动，也适用于人工智能数据需求方对于数据标注进行检查、评价及验收或第三方技术服务机构对数据标注进行合规性评估。

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 35273-2020 信息安全技术 个人信息安全规范
GB/T 41867-2022 信息技术 人工智能 术语
GB/T 42755-2023 人工智能 面向机器学习的数据标注规程
GB/T 43697-2024 数据安全技术 数据分类分级规则
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

GB/T 35273-2020、GB/T 41867-2022和GB/T 43697-2024界定的以及下列术语和定义适用于本文件。

3.1

人工智能 artificial intelligence；AI

〈学科〉人工智能系统（3.2）相关机制和应用的研究和开发。

［来源：GB/T 41867-2022，定义 3.1.2］

3.2

数据标注 data labeling

通过人工或自动化工具，对原始数据（如图像、文本、语音、视频、表格等）添加标签、注释，使其转化为机器可理解的结构化数据，从而支持机器学习模型训练、分析或决策的数据处理过程。

3.3

数据标注方 data labeler

承担数据标注任务的组织或个人。

3.4

数据需求方 data user

提出数据标注需求的组织或个人。

3.5

个人信息 personal information

以电子或者其他方式记录的与已识别或者可识别的自然人有关的各种信息，不包括匿名化处理后的信息。

［来源：GB/T 42574-2023，定义 3.1］

3.6

敏感数据 sensitive data

敏感数据是指因包含个人隐私、生物特征、医疗记录、财务信息等关键内容，一旦泄露或滥用可能对个人权益、社会公共利益或国家安全造成严重损害的数据。

3.7

数据匿名化 data anonymization

通过技术手段使数据无法识别特定个人且不可复原。

3.8

数据清洗 data cleaning

通过技术手段对原始数据进行规范性、准确性处理的过程，包括冗余数据去重、错误数据修正、无效数据过滤等，以提升数据可用性。

3.9

数据投毒 data poisoning

在训练数据中故意添加虚假、有害或恶意数据，以干扰模型训练、影响模型的输出结果。

3.10

联邦学习 federated learning

一种分布式机器学习技术，允许在不共享原始数据的前提下，通过加密的模型参数交换实现多方联合建模。

3.11

幻觉 hallucination

指生成式人工智能模型输出的内容基本符合逻辑与语法，但却与客观事实不符甚至完全虚构，或者不具备可理解性的现象。

3.12

偏见 bias

对待特定对象、人员或群体时，相较于其他实体出现系统性差别的特性。

［来源：GB/T 41867-2022，定义 3.4.10］

4 基本合规原则

4.1 合法合规原则

数据标注活动应严格遵守关于网络安全、数据安全以及个人信息保护相关法律法规，确保标注目的、数据来源、标注内容以及处理流程的合法合规性。涉及个人信息或敏感数据的标注行为，需依法取得数据主体授权或履行法定告知义务，不得标注非法获取或用途不明的数据。

4.2 安全可靠原则

数据标注方应建立与风险等级相匹配的管理制度和技术措施，保障数据标注全生命周期的安全性、规范性和可靠性。数据标注前应对数据标注方及数据需求方的资质进行核验，数据标注后应对标注结果形成电子存证，确保可追溯、可验证。应采取有效措施避免标注违背社会主义核心价值观的内容，避免标注具有歧视性的内容，避免标注虚假信息等法律法规禁止的内容。

4.3 最小必要原则

数据标注的数据采集范围、数据类型、数据数量及存储期限应严格限于实现标注目的所需的最低限度。数据标注方或数据需求方不应超范围收集与标注任务无关的数据，标注完成后应及时删除或匿名化冗余数据。涉及个人信息时，应确保标注内容与业务场景直接关联，避免过度处理。

4.4 透明可控原则

数据标注方应以清晰、易懂的方式向数据主体明示标注目的、数据类型、使用范围及权利行使路径，并取得其有效授权。标注规则、流程及结果应向利益相关方公开，接受监管机构和社会监督。数据主体有权查询标注记录、更正错误信息或撤回授权，以保障其知情权与控制权。

4.5 数字公平原则

数据标注应避免因标注规则、样本选择或人工主观因素导致算法偏见或歧视性结果。不同社会主体在数字技术上享有公平权利。数据标注方应遵循相关技术要求，确保标注结果在性别、年龄、地域等维度上体现公平性，不应损害特定群体在数字经济社会中的合法权利。

5 数据采集合规

5.1 数据来源合规性

5.1.1 数据来源类型及要求

数据标注方获取数据的途径包括但不限于以下类别，应确保数据来源合法：

a) 自行生产：在生产经营活动中自行产生的原生数据，应确保数据采集场景及技术手段合规，应建立原始数据溯源机制。
b) 间接获取：通过数据交易、数据共享、公共数据授权运营、可信数据空间、数据要素流通平台等方式间接获取的数据，应签订相关法律协议，并对提供方的数据来源合规性进行审核。
c) 公开采集：通过人工采集或自动爬取等方式公开采集的数据，应严格遵守目标网站的爬虫协议或相关声明规则，不应采用恶意技术手段违规采集；公开数据附有使用条件或使用限制的，在数据标注中应遵守相关约定。
d) 技术生成：通过人工智能模型或基于机器学习等技术生成的数据，应明确标识生成源并建立内容安全审核机制（如内容真实性校验、AI 生成内容鉴别）。

5.1.2 第三方合作要求

通过数据交易或合作获取数据时，应提供与数据来源类型对应的合法性证明文件，如用户授权协议、数据交易协议、公共数据开放许可、API 接口授权书、数据交易凭证等。

对于提供相关法律协议的，协议条款应包含数据来源合规声明、使用范围限制、数据主体权利保障及违约责任条款，且定期审核交易合作方的履约情况。

5.2 用户授权与告知

5.2.1 个人信息授权要求

数据标注方应根据个人信息的不同类别，建立分类授权机制：

a) 一般个人信息：应通过用户协议、隐私政策等明确处理目的及方式，获取个人明确的同意；
b) 敏感个人信息：除上一项内容外，还应取得个人通过明示、可验证方式作出的单独同意，并按照 GB/T 35273-2020 第 6.3 条进行个人信息安全影响评估，评估报告存档备查；处理不满 14 周岁未成年人信息时，还应取得监护人同意并制定专门处理规则。

5.2.2 告知内容规范

告知信息应包括个人信息处理者的名称或者姓名、联系方式、处理目的、方式、数据类型、保存期限、共享范围及权利行使途径，使用通俗语言确保用户充分理解。

5.2.3 用户权利保障

提供清晰、简便且高效的用户权利行权通道，涵盖数据查询、错误更正、撤回同意、数据删除等全流程操作功能，确保用户能够便捷、自主地行使个人数据权利。

5.3 数据内容合规性

5.3.1 知识产权合规

数据标注方在采集数据时，应对以下方面进行风险评估和合规审核：

a) 权属明确：通过法律文件核验、权属声明溯源等方式审核所采集的数据是否包含他人享有著作权的作品。
b) 授权审查：对尚在保护期内的作品，应取得著作权人的明确授权；对公有领域作品的使用，确保不侵犯署名权、修改权等权利主体的知识产权。
c) 技术识别：采用技术手段（如数字水印、元数据分析）检测数据是否包含未授权的知识产权内容。
d) 风险防控：鼓励建立知识产权侵权行为负面清单并定期更新，建立侵权投诉响应机制。

5.3.2 不良信息过滤

数据标注方可以通过以下措施进一步降低采集数据包含的不良信息：

a) 自动筛查：部署内容过滤工具对采集数据进行初步筛查，排除涉及暴力、色情、歧视等不良信息。
b) 人工复审：建立人工审核团队，对疑似不良信息进行二次核验和排除。
c) 记录追溯：留存过滤日志及处置记录，确保操作可追溯。
d) 动态更新：根据法律法规及舆情变化，定期更新不良信息过滤规则和样本库。

6 数据管理合规

6.1 数据分类分级

数据标注方应依据GB/T 43697-2024建立三级分类体系。具体分类体系见附录B。

涉及国家安全、军事监控、生物医药等特殊敏感场景的数据不得用于标注（获得法律明确许可或经有关主管部门批准的的除外）；涉及国家安全、核安全、军工数据等特殊敏感技术的数据，需在独立封闭且满足特定要求的环境下由具备保密资质的机构处理，应增加特殊监管措施（如权限隔离、政审机制）。

鼓励建立动态分级调整机制，根据具体的数据应用场景进行调整评估。

6.2 数据监控记录

数据标注方应建立流程监控和日志记录体系，定期追踪数据访问和修改行为，确保数据处理操作可溯源，以便进行数据安全审计。

6.3 数据访问控制

实施基于角色的最小权限原则，数据标注方应仅授权特定人员访问对应的敏感数据，标注岗位与审核岗位需职责分离，高密级数据需通过双重授权或生物识别验证后方可接触并处理。

6.4 数据清洗管理

数据标注方应通过标准化清洗流程提升数据准确性、规范性和可溯性，重点修正数据内容错误、冗余信息及异常值，以满足数据可用性要求。

6.5 数据投毒防范

数据标注方应防范标签污染、特征污染、后门攻击及数据分布偏移等数据投毒攻击手段，应建立多层审核机制和采用技术检测工具，确保训练数据映射关系的真实性和分布稳定性。数据投毒攻击手段参考例举如下：

a) 标签污染：修改数据标签（如将“良性肿瘤”标注为“恶性”，或反之），误导模型学习错误的映射关系。
b) 特征污染：篡改数据特征（如在房价数据中伪造“面积为0”的样本，或在图像中添加人眼不可见的噪声），破坏特征与标签的真实关联。
c) 后门攻击：在数据中植入特定触发模式（如图片角落的特定色块），使模型对包含该模式的输入输出预设错误标签（如所有带红色方块的图片都被分类为“飞机”）。
d) 数据分布偏移：通过注入大量异常数据，改变训练数据的整体分布（如在正常用户行为数据中混入大量虚假点击记录），导致模型泛化能力下降。

7 数据标注流程合规

7.1 标注人员管理

数据标注方在管理标注人员时，应符合以下要求：

a) 资质认证与任务匹配：建立标注人员技能分级认证体系和标准化考试制度，垂类领域（如医疗、金融、智能驾驶）应同时持有相关行业资质证书方可参与标注任务，可以通过关联认证、行业资格互认等形式核验其在数据标注领域及垂类行业中的资质。根据数据类别、敏感级别及应用场景匹配具备相应资质的人员。
b) 背景审查与培训机制：高保密等级标注人员需通过政治审查、犯罪记录核查及心理健康评估，并签署保密协议；对标注人员的合规培训内容应涵盖数据安全法律法规、个人信息保护法律法规以及数据标注规范等相关内容。
c) 权限管理与职责分离：标注、审核、复检、终检岗位独立，操作日志需实现跨岗位追溯。

7.2 标注质量控制

鼓励数据标注方制定标注规则手册，明确标签定义、标注流程、风险提示以及负面行为；鼓励建立质量量化指标体系，实施分层抽样质检，审核记录备份保存；鼓励采用动态质检机制，建立涵盖准确率、漏标率（≤1%）等关键评价指标在内的量化指标体系，定期开展虚假标注抽查与标注质量追溯审计。具体评估质量评估指标见附录B。

质检层级	抽样比例	验收标准
初审	100%	基础完整性检查
复审	一级数据≥30% 二级数据≥15%	标签准确率≥98%
终检	≥5%	通过对抗样本测试

表1 三级质检机制

数据标注方应当对质检中的所有问题进行纠正或或重新标注，并跟踪纠正情况和处理结果。应逐条记录标注纠正的详细信息，包括原始标注人员信息、标注纠正人员信息、原始标注内容、纠正后的标注内容、原始标注时间、纠正标注时间等。应对纠正后的标注进行复核，复核通过的将纠正标注结果进行更新和归档，复核未通过的按需进行重新标注。

7.3 技术工具及场地要求

数据标注方采用的技术工具及场地应符合以下要求：

a) 标注平台：应具备安全基座、账号风控、操作留痕功能，记录标注人员、标注时间及修改历史。具体操作举例如下：
1）确认标注人员资质，按分级体系和垂域资质将标注人员录入标注平台；
2）通过如 TLS/Https、短信验证等通讯方式确认标注人员身份的前提下，标注平台向标注人员派

发待标注内容；

3）应具备操作留痕功能，记录标注人员、标注时间及修改历史；

4）针对标注人员账号的标注情况，对标注频率、标注时段等账号行为进行监控，向标注平台报送异常预警信息。

b) 标注场地：应设置在安全可控的区域，具备门禁系统，限制非授权人员进入。标注场地需根据订单保密性进行分级，高保密等级需配备摄像头、独立网关、服务器和安保人员等。对低保密等级的任务，允许远程标注场景采用“虚拟桌面+行为审计软件”替代传统门禁，适应人工智能行业未来分布式办公的发展趋势。

8 数据传输合规

8.1 数据存储与传输

对于不同级别的数据，数据标注方应采取差异化的存储与传输管理措施：

a) 存储数据加密：根据 GB/T 43697-2024 对数据存储介质进行分类管理。一级敏感数据存储应采用国密局认证的密码模块，二级数据应采用符合 CSA STAR 标准的云存储，三级数据可分布式存储。在存储期限阈值方面，一级数据原则上不超过模型训练所需周期（一般不超过 2 年）。

b) 定期存储审计与灾备机制：应每季度对存储系统的权限日志加密状态进行审计，鼓励建立异地灾备的数据备份和恢复机制。

c) 传输协议要求：数据传输应使用 SSL/TLS 加密协议，以保障敏感数据不会受到未经授权的访问或篡改；不得通过公共云盘、即时通讯工具传输敏感数据。

8.2 数据共享

8.2.1 第三方共享规范

数据标注方与第三方共享数据时，应签署相关法律协议，明确约定数据共享的目的、用途、期限、授权范围、数据安全以及违约责任；建议通过联邦学习、隐私计算或可信数据空间等可控技术环境实现数据共享。

8.2.2 跨境传输管理

涉及跨境数据传输时，若涉及重要数据或一定规模个人信息等特定情形，应当依法依规开展数据出境安全评估，或采用经国家网信部门制定的个人信息出境标准合同与境外接收方订立合同，亦或按照国家网信部门规定经专业机构进行个人信息保护认证。

9 数据删除合规

9.1 数据删除触发条件

数据标注方应基于以下情形启动数据删除流程：

a) 任务周期结束：标注任务提前终止、合同履行完毕或数据处理目的已实现。

b) 用户权利行使：数据主体明确提出行使删除权。

c) 服务合作终止：标注方、云服务商等解除合作关系。

d) 法律法规要求：涉及非法采集、违反最小必要原则或超期留存。

9.2 数据销毁技术要求

数据标注方应按以下要求进行数据销毁：

a) 当标注任务提前终止或完成时，应按照合同约定对原始数据进行相关处置；合同中如无相关约定，但涉及生物特征等敏感个人信息的，应按照法律法规要求删除原始数据。

b) 销毁数据应采用不可逆方式（如物理粉碎、多次覆写）。

9.3 备份数据管理规范

数据标注方应在合同中约定第三方备份数据的同步删除责任；建议约定关于云服务商提供数据彻底删除证明的相关条款，以实现数据删除覆盖云端存储场景。

10 合规管理保障

10.1 组织架构

数据标注方的组织架构应符合以下要求：

a) 设立数据合规官（DPO）：独立行使合规监督职能，直接向公司管理层汇报，并负责统筹制定合规策略体系，监督合规制度执行；掌握数据分类分级结果，管理数据处理全流程合规；处理数据泄露事件与监管调查。
b) 成立数据合规管理委员会：由法务、技术、业务、审计部门负责人组成，定期召开数据安全风险研判会议，协调解决数据合规冲突问题。
c) 标注人员资质管理：针对涉密或高风险的标注任务，实施标注人员准入考试或认证制度，并建立黑白名单动态调整机制。

10.2 制度建设

数据标注方应遵循以下制度建设措施：

a) 采用数据分类分级机制：对不同敏感等级的数据匹配差异化的存储加密与访问控制措施，如一级敏感数据强制采用同态加密技术。
b) 建立标注质量监控体系：设置标注错误率（如≤0.1%）、漏标率（如≤0.05%）等量化指标，通过分层随机抽样实现定期质量核查（如日检、周检、月检）。
c) 建立数据标注追溯制度：保留完整的标注记录与修改日志，保障标注流程可追溯。
d) 制定数据泄露应急预案：要求在信息系统发生异常访问时自动触发三级响应，包括局部断网、镜像取证及监管通报等程序。

10.3 合规审计

数据标注方应每年度进行内部合规审计，重点核查标注数据权属、用户授权文件、跨境传输审批记录，审计报告应至少留存3年。

10.4 偏见防控

数据标注方应遵循以下偏见防控措施：

a) 数据均衡：对标注数据集进行性别、地域年龄分布等多维度多样性审查。
b) 算法公平：建立涵盖数据选取、测试场景设计、多维度指标设定及动态验证的标准化算法公平性测试流程，系统性检测模型输出在性别、种族、年龄等关键维度是否存在歧视性结果。

10.5 认证监督

鼓励数据标注方自愿申请第三方合规认证，如通过DSMM、ISO 27701等认证；鼓励各地数据相关行业协会建立黑白名单管理制度，对多次违规的数据需求方或数据标注方采取公开通报等公示举措，以促进行业健康规范发展。

附 录 A

（资料性）

数据标注配套文件模版及指南

数据标注服务协议和用户授权同意书模板见表A.1。

表A.1 数据标注配套文件模版及指南

序号	名称	内容示例
1	数据标注服务协议模板	**合同名称：人工智能数据标注服务协议** 数据需求方（甲方）： 数据标注方（乙方）： **第一条 服务内容** 1.1 标注数据类型：甲方委托乙方对以下数据进行标注（示例）： 1.1.1 数据类型：图像/文本/音频/视频（具体描述）。 1.1.2 标注要求：明确标注类别、格式、精度标准等。标注过程应严格遵循乙方制定并经甲方认可的《数据标注规则手册》（附件一）执行。 1.2 交付标准：乙方应按照甲方提供的《数据标注质量标准》（附件二）完成标注，并通过甲方验收。 1.3 交付时间：乙方应于____年____月____日前完成全部标注任务并交付成果。 1.4 验收流程：乙方完成标注任务并向甲方交付成果后，甲方应在7 日内完成验收。验收合格的，双方签署验收确认书；甲方逾期未提出书面异议的，视为验收合格。 1.5 未经甲方书面同意，乙方不得将本合同项下的标注任务分包或转包给任何第三方。 **第二条 费用及支付** 2.1 服务费用：总计人民币___元（大写：_____）。 2.2 支付方式： 预付款：本合同签订之日起___日内，甲方向乙方支付合同金额的___%； 尾款：本合同项下数据标注任务验收合格之日起___日内，甲方向

		乙方支付合同金额的___%。 **第三条 数据来源合规性** 3.1 乙方承诺其标注数据来源合规： (a) 自行生产的数据，应确保采集场景及技术手段符合法律法规，并建立原始数据溯源机制； (b) 间接获取的数据，应提供数据交易协议、公共数据授权等合法性证明文件，并对数据提供方合规性进行审核； (c) 公开采集的数据，应严格遵守目标网站爬虫协议及公开数据使用限制； (d) 技术生成的数据，应明确标识生成源并建立内容安全审核机制。 3.2 乙方与第三方合作获取数据时，应提供原始数据权利证明，并定期审核合作方履约情况。 3.3 若标注数据包含个人信息，乙方应根据个人信息的不同类别，建立分类授权机制： (a) 一般个人信息：应获取个人明确同意； (b) 敏感个人信息：应获取个人单独同意，并依据完成个人信息安全影响评估。处理不满 14 周岁未成年人信息时，应取得监护人同意并制定专门处理规则。 3.4 乙方应提供清晰、简便且高效的用户行权通道，确保用户能够便捷、自主地行使个人数据权利。 **第四条 数据内容合规性** 4.1 乙方应确保数据不侵犯第三方知识产权，采集数据时审核数据权属及授权，采用技术手段识别侵权内容，并建立侵权投诉响应机制。 4.2 乙方应部署工具筛查暴力、色情等违规内容，人工二次复核，并留存过滤日志至少【2】年。 **第五条 数据管理合规** 乙方应履行数据管理合规义务，具体包括以下要求： (a) 依据国家标准建立三级分类体系（见附录 B），禁止使用涉及国家安全、军事监控等特殊场景数据； (b) 建立全流程操作日志及监控体系，记录数据访问、修改行为，留存日志不少于【2】年； (c) 敏感数据仅限授权人员访问，标注与审核岗位职责分离，高密

		级数据接触需通过双重授权或生物识别验证； (d) 通过标准化清洗流程修正数据错误、冗余及异常值； (e) 防范数据投毒攻击，建立多层审核机制并采用技术工具检测异常数据。 **第六条 标注人员管理与工具、场地要求** 6.1 垂类领域（如医疗、金融）标注人员须持行业资质证书；高保密任务人员须通过背景审查并签署保密协议。 6.2 标注、审核、终检岗位独立，高密级数据需双重授权。全流程操作日志留存【2】年，甲方可随时调阅。 6.3 乙方采用的技术工具及场地应符合以下要求： (a) 标注平台：应具备操作留痕功能，记录标注人员、标注时间及修改历史。 (b) 标注场地：应设置在安全可控的区域，具备门禁系统，限制非授权人员进入。标注场地需根据订单保密性进行分级。 **第七条 数据标注质量控制** 7.1 乙方标注准确率应≥98%，漏标率≤1%（按附录 B 标准验收）。 7.2 乙方应实施三级质检机制，最高密级数据终检比例≥5%（具体要求见附件二）。 **第八条 数据安全与共享** 8.1 存储要求：一级数据乙方应采用 AES-256 加密及物理隔离；跨境传输须完成网信办安全评估。 8.2 禁止通过公有云盘传输敏感数据，共享数据应通过联邦学习或隐私计算环境进行。 **第九条 数据删除与销毁** 9.1 乙方应在以下情形发生时立即启动数据删除流程： (a) 标注任务提前终止或数据处理目的已实现； (b) 数据主体依法行使删除权并提出书面请求； (c) 与云服务商等合作方解除合作关系； (d) 数据涉及非法采集、违反最小必要原则等法定情形。 9.2 标注任务终止或完成后，乙方应按合同约定采取不可逆方式销毁原始数据；合同未约定的，涉及生物特征等敏感个人信息的，应按照法律法规要求删除。 9.3 乙方须与第三方备份服务商约定同步删除责任，确保云端、本

		地备份数据同步清理。 9.4 乙方完成数据删除或销毁后，应在【5】日内向甲方提供书面证明材料. **第十条 标注成果知识产权** 10.1 标注成果知识产权归甲方所有，乙方不得擅自使用或转让。 10.2 标注成果涉及生成式数据的，应明确生成内容的标识及版权合规责任。 **第十一条 违约责任** 11.1 乙方未按时交付或质量不达标的，甲方有权要求免费整改；逾期超过____日的，甲方有权解除合同并要求乙方支付违约金。 11.2 发生数据泄露事件，乙方应【24 小时】内通知甲方，并承担实际损失赔偿。 11.3 虚假标注比例超【1%】时，甲方有权解除合同并要求乙方承担实际损失赔偿。 11.4 因乙方违反本合同义务引发纠纷的，乙方承担相应法律责任。 11.5 甲方未按约定支付款项的，每逾期 1 日按应付金额【0.05%】支付违约金。 11.6 乙方违反本合同第六条 1.5 约定，未经甲方同意擅自分包或转包的，甲方有权立即解除合同，乙方应支付合同总金额【20%】的违约金，并赔偿甲方因此造成的全部损失。 **第十二条 争议解决** 因本合同产生的争议，双方应协商解决；协商不成的，提交原告方所在地有管辖权的人民法院诉讼解决。 **第十三条 其他约定** 13.1 本协议一式两份，甲乙双方各执一份，具有同等法律效力。 13.2 本协议自双方签字盖章之日起生效。 附件： 附件一《数据标注规则手册》 附件二《数据标注质量标准》 甲方：（授权代表签名及公司盖章） 乙方：（授权代表签名及公司盖章）

		签署日期：_____
2	用户授权同意书模板	**文件名称：数据采集与标注授权同意书** 授权方：_____ 被授权方：_____ 授权方已充分理解并自愿作出如下授权，同意被授权方按照本同意书约定处理其个人数据。 **一、授权内容** 1. 授权类型与范围：授权【企业名称】采集、标注本人的【数据类型】（如面部图像、语音记录等）。 2. 数据使用范围：仅限于【企业名称】指定的机器学习项目。 3. 存储期限：自标注完成之日起【2】年；存储期限届满后，【企业名称】将对数据进行妥善删除或匿名化处理。 **二、用户权利** 1. 授权撤回：本人可随时通过【企业指定渠道，如官网单独入口/客服邮箱】撤回授权，企业须在【15】日内删除或匿名化相关数据； 2. 知情权：有权查询数据标注内容、修改日志及使用记录； 3. 更正与删除：发现数据错误或存在侵权内容，可要求企业更正或删除。 **三、使用限制** 1. 不得将数据用于用户画像、自动化决策或其他商业用途； 2. 不得向第三方共享原始数据（合作科研机构需签署保密协议）； 3. 数据出境须单独取得数据主体书面同意。 **数据主体签署确认** 本人已完整阅读、理解并自愿接受本授权书全部条款。 签名：_____ 日期：_____ **被授权方确认** 承诺严格按照本授权书约定处理数据，保障数据主体合法权益。

		企业盖章：_____ 法定代表人/授权代表签字：_____ 日期：_____

附 录 B

（规范性）

标注质量与数据分类规范

标注质量评估指标见表 B.1，数据三级分类体系见表 B.2。

表B.1 标注质量评估指标

指标	标准	检测方法
准确率	≥98%	随机抽样人工核验
漏标率	≤1%	对抗样本测试及全量扫描
时效性	任务交付延迟≤3天	项目管理系统日志追踪

表B.2 数据三级分类体系

级别	数据类型	保护要求
一级	生物特征、医疗记录、金融信息等敏感个人信息、重要数据	强制匿名化处理，采用AES-256及以上加密算法存储，物理隔离环境标注，严格限定访问权限，保证收集、访问、加工和传输可溯源
二级	政务数据、法律文书等公共数据、一般个人信息	实施动态访问控制，日志记录保存不少于3年。采用校验技术控制访问权限并制定操作规程，严格按照授权范围和数据处理协议进行数据处理

三级	公开文本、通用图片等不包含个人信息、重要数据、公共数据的数据	基础访问鉴权机制，定期备份验证

参 考 文 献

[1] 中华人民共和国网络安全法（2016 年 11 月 7 日第十二届全国人民代表大会常务委员会第二十四次会议通过）

[2] 中华人民共和国数据安全法（2021 年 6 月 10 日第十三届全国人民代表大会常务委员会第二十九次会议通过）

[3] 中华人民共和国个人信息保护法（2021 年 8 月 20 日第十三届全国人民代表大会常务委员会第三十次会议通过）

[4] 互联网信息服务算法推荐管理规定（2021 年 12 月 31 日国家互联网信息办公室、工业和信息化部、公安部、国家市场监督管理总局公布）

[5] 互联网信息服务深度合成管理规定（2022 年 11 月 25 日国家互联网信息办公室、工业和信息化部、公安部公布）

[6] 生成式人工智能服务管理暂行办法（2023 年 7 月 10 日国家互联网信息办公室、国家发展和改革委员会、教育部、科学技术部、工业和信息化部、公安部、国家广播电视总局公布）

[7] 人工智能生成合成内容标识办法（2025 年 3 月 7 日国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局公布）

[8] 科技伦理审查办法（试行）（2023 年 9 月 7 日科学技术部、教育部、工业和信息化部、农业农村部、国家卫生健康委员会、中国科学院、中国工程院、中国科学技术协会、中国社会科学院、中央军委科学技术委员会印发）

[9] 具有舆论属性或社会动员能力的互联网信息服务安全评估规定（2018 年 11 月 15 日国家互联网信息办公室、公安部发布）

[10] 网络信息内容生态治理规定（2019 年 12 月 15 日国家互联网信息办公室令第 5 号公布）

[11] 新一代人工智能治理原则——发展负责任的人工智能（2021 年 6 月 17 日国家新一代人工智能治理专业委员会发布）

[12] GB/T 35770-2022 合规管理体系 要求及使用指南

[13] GB/T 38667-2020 信息技术 大数据 数据分类指南

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

[15] GB/T 43697-2024 数据安全技术 数据分类分级规则

[16] 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

[17] Blueprint for an AI Bill of Rights, OSTP, 2022

[18] Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, White House, 2023

[19] Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST, 2023

[20] EU General Data Protection Regulation，2016

[21] Ethics Guidelines for Trustworthy AI, European Commission, 2019

[22] WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust, European Commission, 2020