

T/SAIAS

上海市人工智能行业协会团体标准

T/SAIAS XXXX—2025

人形机器人数据集质量评价

Evaluation of humanoid robot dataset quality

（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

上海市人工智能协会 发布

目 次

前言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 评价原则	1
5 评价指标与要求	2
5.1 评价指标	2
5.2 数据集规范性	3
5.3 数据集完整性	5
5.4 数据集准确性	5
5.5 数据集一致性	5
5.6 数据集时效性	6
5.7 数据集可访问性	6
6 评价方法	6
6.1 基本要求	6
6.2 过程质量评价	6
6.3 内容质量评价	6
7 评价组织与流程	9
7.1 评价组织	9
7.2 评价流程	9
附录 A（资料性） 人形机器人数据集评价设置导则	12
附录 B（资料性） 典型场景数据与技术热点匹配度分值	15
附录 C（资料性） 数据集质量提升技术方案示例	17
参考文献	19

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。

本文件由上海市人工智能行业协会提出。

本文件由上海市人工智能行业协会归口。

本文件起草单位：

本文件主要起草人：

本文件为首次发布。

人形机器人数据集质量评价

1 范围

本文件规定了人形机器人数据集质量评价原则、评价指标与要求、评价方法、评价组织与流程。本文件适用于人形机器人数据集质量评价活动。

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 10111 随机数的产生及其再产品质量抽样检验中的应用程序
GB/T 35273 信息安全技术 个人信息安全规范
GB/T 43697 数据安全技术 数据分类分级规则
DB《人形机器人数据集管理规范》
T/SAIAS《人形机器人数据集分类与编码》
T/SAIAS《人形机器人操作任务数据集数据格式要求》
T/SAIAS《人形机器人数据集数据标注》

3 术语和定义

下列术语和定义适用于本文件。

3.1

人形机器人 humanoid robot

具有躯干、头部和部分或全部肢体，外观和动作与人类相似，实现感知、认知、决策、执行等能力的机器人。

[来源：GB/T 12643—202X，4.15.5，有修改]

3.2

数据集 dataset

具有一定主题，可标识并能被计算机系统处理的数据集合。

[来源：ISO 19115-1：2014，4.3]

3.3

元数据 metadata

定义和描述其他数据的数据。

[来源：GB/T 18391.1—2009，3.2.16]

4 评价原则

4.1 应符合 DB《人形机器人数据集管理规范》中 5.3 和 16.1 的要求。

4.2 评价维度包括人形机器人数据集构建过程质量评价与数据集内容质量评价，评价过程应公开透明，评价人员应向委托方详细说明评价指标与要求、评价方法、评价结果等。

4.3 评价指标可根据实际应用与未来发展需求进行扩展，扩展内容不与已有内容冲突。

4.4 数据集质量评价分为符合性评价与分级评价，应在通过符合性评价的基础上，再开展人形机器人数据集分级评价，示例见附录 A：

- a) 符合性评价：评价指标包括数据集过程质量指标与数据集规范性指标，对数据集过程质量指标要求文档的评价，测试前应提供对应文档，文档名称包含评价指标的关键词，并建立索引；
- b) 分级评价：结合应用场景与任务需求选择评价指标，评价指标包括：
- 1) 数据集完整性；
 - 2) 数据集准确性；
 - 3) 数据集一致性；
 - 4) 数据集时效性；
 - 5) 数据集可访问性。

5 评价指标与要求

5.1 评价指标

5.1.1 人形机器人数据集评价指标见表1。

表1 人形机器人数据集评价指标

序号	一级指标	二级指标	评价指标描述
1	过程质量	数据集构建需求与定义	明确定义数据集的目的和预期用途； 清晰描述数据集将要支持的具体应用场景与特定任务； 明确数据集范围边界； 考虑数据集的可扩展性。
2		数据采集	提供数据采集依据的法规、技术文件或其他参考文献； 提供数据采集环境、采集方式（如通过传感器、手动数据输入、外骨骼采集、VR采集、合成数据等）、采集设备/平台/工具链描述、机器人/传感器坐标系描述（如有）等； 提供数据来源的合规性陈述； 提供匿名化或假名化的方法（如适用）； 提供采样的偏差程度信息； 提供数据集操作（合并、切片等）、数据集负责人信息（如有）； 提供数据采集过程中安全审计与风险评估文件。
3		数据预处理	清晰描述数据的录入、排除标准以及数据筛选的方法。 清晰描述数据预处理（如数据清洗、数据增强、数据特征提取、数据分帧、数据对齐与同步校准、数据压缩、数据融合等）的操作方法与参数设置。
4		数据标注	提供数据标注依据的法规、技术标准或其他参考文献； 提供数据标注工具和平台信息； 如具有标注信息，需描述数据标注任务说明、数据验收标准与质控流程。
5		数据存储	提供数据存储信息，包括数据集存储介质、存储方式与存储路径、安全控制、备份、使用权限说明等。
6		数据集使用与流通	提供符合T/SAIAS《人形机器人操作任务数据集数据格式规范》中第6章规定的数据集描述性文件（整体描述文件、片段描述文件）； 提供数据集质量评估结果的记录（如有）； 提供数据集传输与访问的法规、技术标准或其他参考文献； 提供不同安全级别的数据集访问权限说明、访问路径、开放策略（如非涉密数据宜采用通用结构化格式，同步提供标准化元数据，涉密数据应进行加密/匿名化处理等）、获取流程文件（如通过邮箱/组织认证后在线注册；进行自动化/人工审核，学术机构审核时间不超过1天，商业机构审核时间不超过3天等） 提供数据集具备跟踪与改进渠道、数据集优化策略、支撑承诺（如维护计划、持续改进承诺等）信息（如有）
7		数据更新维护	提供数据集版本、更新频率、方式及责任人信息。
		数据集销毁（如有）	提供数据删除、介质报废遵循的法规、技术标准或其他参考文献。
9		合规性审查	提供数据处理活动遵守的相关伦理准则和法律法规； 提供隐私政策同意记录等合规文件； 清晰描述用于保护涉及隐私和个人数据的技术手段，如数据去标识化、数据匿名化等，适当时描述对应规则。

表 1 人形机器人数据集评价指标（续）

序号	一级指标	二级指标	评价指标描述
10	内容质量	数据集规范性	数据集满足安全与隐私要求、可追溯性、数据集标准符合性的程度。
11		数据集完整性	根据数据规则，数据集中数据元素被赋予数值的程度。
12		数据集准确性	数据准确表示其所描述的真实实体（实际对象）真实值的程度。
13		数据集一致性	数据与其他特定使用的数据之间无矛盾的程度。
14		数据集时效性	数据集反映当前技术发展水平、实际应用需求等程度，包括时间覆盖度、技术同步性等。
15		数据集可访问性	数据集能被访问的程度。

5.1.2 应结合应用场景与任务需求，设置数据集内容质量指标的详细评价内容要求。

5.2 数据集规范性

5.2.1 基本要求

安全与隐私、数据集标准符合性、可追溯性的评价内容设置应至少满足5.2.2、5.2.3、5.2.4的要求。

5.2.2 安全与隐私

5.2.2.1 通常数据集不包括以下类型的数据，警用、司法等特定应用场景应另作要求：

- a) 违反价值观的数据：
 - 1) 煽动颠覆国家政权、推翻社会主义制度；
 - 2) 危害国家安全和利益、损害国家形象；
 - 3) 煽动分裂国家、破坏国家统一和社会稳定；
 - 4) 宣扬恐怖主义、极端主义；
 - 5) 宣扬民族仇恨；
 - 6) 宣扬暴力、淫秽色情；
 - 7) 传播虚假有害信息；
 - 8) 其他法律、行政法规禁止的内容。
- b) 歧视类数据：
 - 1) 民族歧视；
 - 2) 信仰歧视；
 - 3) 国别歧视；
 - 4) 地域歧视；
 - 5) 性别歧视；
 - 6) 年龄歧视；
 - 7) 职业歧视；
 - 8) 健康歧视；
 - 9) 其他方面歧视。
- c) 商业违法违规数据：
 - 1) 侵犯他人知识产权；
 - 2) 违反商业道德；
 - 3) 泄露他人商业秘密；
 - 4) 利用算法、数据、平台等优势，实施垄断和不正当竞争行为；
 - 5) 其他商业违法违规行为。
- d) 侵犯他人合法权益：
 - 1) 危害他人身心健康；
 - 2) 侵害他人肖像权；
 - 3) 侵害他人名誉权；
 - 4) 侵害他人荣誉权；

- 5) 侵害他人隐私权;
- 6) 侵害他人个人信息权益;
- 7) 侵犯他人其他合法权益。

5.2.2.2 根据数据集中数据信息的重要性与敏感性，将数据集划分为公开数据集、受限数据集、敏感数据集、涉密数据集4个级别，见表2。根据具体应用场景，应制定对于涉密数据集、敏感数据集或受限数据集的管理类文件。

表2 数据集安全分级

级别	名称	说明
第一级	公开数据集	可对社会公众开放的数据集。 政府、公共机构、研究机构、企业或其他组织发布的可以自由获取和使用的数据集。
第二级	受限数据集	仅限于局部个人和组织使用,达到一定精度(数值、时间、空间精度等)和规模(数据存储量、企业市值、交易量等)的数据集。包括但不限于个人基本资料(如姓名、性别、工作单位名称、职务、联系电话、邮箱等)、组织基本资料(如组织机构名称、代码、简介、电话、邮箱等)。 应确保数据集全生命周期安全。
第三级	敏感数据集	数据泄露后会给社会或个人带来严重危害的数据集,包括个人敏感信息数据和商业敏感信息数据。 应确保数据集访问安全并实时监控。
第四级	涉密数据集	涉及保密数据,不准许公开使用。

5.2.2.3 涉密数据集包含国家政务、安全、科技、军事等领域的绝密文件及保密设施的信息及内容等,经有关部门评估确定的数据集。

5.2.2.4 依据 GB/T 35273、GB/T 43697 的要求,敏感数据集是指包含应受保护的信息,其泄露、修改、破坏或丢失会对人或事产生可预知的损害,详见表3,敏感数据宜采取 C.5 的方式进行脱敏处理。

表3 敏感信息数据

类别	示例数据	
个人敏感信息数据	个人身份信息	身份证、军官证、护照、驾驶证、工作证等
	个人生物识别信息	个人指纹、声纹、掌纹、耳廓、虹膜、面部识别特征等
	网络身份标识信息	个人信息主题账号、IP地址、个人数字证书等
	个人财产信息	银行账户,鉴别信息(口令),存教信息(包括资金数量、支付收款记录等)、房产信息、信贷记录、征信信息,交易和消费记录,流水记录等,以及虚拟货币,虚拟交易、游戏类兑换码等虚报财产信息
	个人健康信息	个人因生病医治等产生的相关记录,如病症,住院志、医单、检验报告,手术及麻醉记录,护理记录、用药记录,药物食物过敏信息,生育信息、既往病史、诊治情况、家族病史,现病史、传染病史等
	个人通信信息	通信记录和内容、短信、即时通信消息、电子邮件,以及描述个人通信的数据等
	联系人信息	通讯录、好友列表,电子邮件地址列表等
	个人网络信息	网站浏览记录、网络日志等
	个人位置信息	包括行踪轨迹、精准定位信息、住宿信息等
商业敏感信息数据	/	符合组织所属行业定义或组织自身定义的敏感数据。包括不限于业务数据(研发设计、产品数据、合同协议等)、经营管理数据(如财务数据、经营战略等)、系统运维数据(网络设备/信息系统配置数据、日志数据等)。
注:个人敏感信息是指一旦泄露、非法提供或滥用可能危害人身和财产安全,极易导致个人名誉、身心健康受到损害或歧视性待遇等的个人信息。通常情况下,14岁以下(含)儿童的个人信息和涉及自然人隐私的信息属于个人敏感信息。		

5.2.3 数据标准符合性

5.2.3.1 数据集分类与编码应符合 T/SAIAS 《人形机器人数据集分类与编码》的规定。

5.2.3.2 数据集描述、数据格式要求应符合 T/SAIAS 《人形机器人操作任务数据集数据格式要求》的规定。

5.2.4 可追溯性

5.2.4.1 数据集可追溯性指追踪数据集来源、处理过程和使用的完整路径的能力，数据集可追溯性应在数据采集阶段、预处理与标注阶段、使用与流通阶段、销毁阶段满足以下要求：

- a) 采集阶段：宜通过数字签名等形式确保其未被私下篡改；
- b) 预处理与标注阶段：
 - 1) 每个数据单元进行处理和标注时应记录操作日志，记录操作者身份、操作时间、具体操作内容；
 - 2) 标注和处理过程中争议数据的处理应保留争议仲裁记录；
 - 3) 数据集每次迭代宜独立存档，应明确标注版本号、变更说明及对应责任方，
- c) 使用与流通阶段：
 - 1) 共享或传输时，应附带相应协议的电子标签，包含使用范围、期限和接收方信息；
 - 2) 数据集访问的终端记录应通过日志记录，包括设备信息、用户账户及访问内容，留存时间宜不少于 180 天，
- d) 销毁阶段：
 - 1) 每个数据单元删除操作需留存删除证明（如可信第三方审计日志）；
 - 2) 关键数据实行逻辑删除与物理删除双阶段机制。

5.3 数据集完整性

应基于应用场景、任务需求，选取与设置数据集完整性的评价内容、评价要求、等级划分与判定依据，评价内容包括但不限于：

- a) 数据文件的完整性：
 - 1) 对于视觉数据：数据文件（图像、视频）均应在存储目录中存在、数据丢帧缺失的占比情况；
 - 2) 对于听觉数据，数据文件均应在存储目录中存在、完整录制指令声音的数据占比情况；
 - 3) 对于触觉数据，数据文件均应在存储目录中存在、数据丢帧缺失的占比情况；
 - 4) 对于嗅觉数据，数据文件均应在存储目录中存在、数据丢帧缺失的占比情况；
 - 5) 对于运动学状态数据，对应的数据文件均应在存储目录中存在、数据丢帧缺失的占比情况；
 - 6) 对于动力学状态数据，对应的数据文件均应在存储目录中存在、数据丢帧缺失的占比情况。
- b) 数据记录的完整性：数据集中非空数据记录占比情况；
- c) 数据模态完整性：数据集中融合多传感器数据情况（至少包括视觉数据与运动学状态数据）；
- d) 标注完整性：提供完整的动作标签、环境状态、任务目标等元数据的情况。

5.4 数据集准确性

应基于应用场景、任务需求，选取与设置数据集准确性的评价内容、评价要求、等级划分与判定依据，评价内容包括但不限于：

- a) 元数据准确性：元数据提供基于来自利益相关方的数据需求说明所要求的信息、以及包含说明文档来正确解释数据的含义和性质的程度；
- b) 数据集不确定性风险：异常值个数指出数据集中数据不准确性的风险；

示例：同一现象有以下值分布(100、105、120、80、75、60、130、2000)，最后一个情况(2000)就表示一个异常值。

- c) 数据集中损坏数据(如空白值、NaN 等)占比情况。

5.5 数据集一致性

应基于应用场景、任务需求，选取与设置数据集一致性的评价内容、评价要求、等级划分与判定依据，评价内容包括：

- a) 多模态数据时间戳一致性：数据集中多模态数据时间戳一致的程度；
- b) 语义一致性：不同模态数据对同一任务/事件描述一致的程度；

- c) 空间一致性：不同模态数据在空间上的对齐程度。

5.6 数据集时效性

数据集应能反映当前技术发展水平、实际应用需求，评价内容包括但不限于：

- a) 时间覆盖度：至少包含近 3 年内的数据样本，且更新频率不低于每半年一次，保留具有长期价值的基础数据（如人体运动学模型）；
- b) 技术同步性：包含技术最新发展的新型硬件、主流算法所需的信息，如支持新型传感器（3D 力触觉传感器、激光雷达等）的数据格式与采样频率，提供适合当前主流模型的标注形式（如时间序列标注、多模态对齐标签等）。

5.7 数据集可访问性

数据集可访问性评价内容包括但不限于：

- a) 数据文件可访问：数据文件均能正常打开与解析，不存在格式损坏或无法解码等情况；
- b) 访问权限控制：数据集应进行分级管理（5.2.2.2），访问权限控制应通过基于角色的访问系统实现：
 - 1) 公开数据集：支持无条件访问；
 - 2) 受限数据集与敏感数据集：访问通过机构资质审核；
 - 3) 涉密数据集：应符合数据安全相关法律法规、强制性标准等文件要求。数据应存储于本地安全环境，原始数据不应离域，使用时应经脱敏处理和技术审查，数据传输应采用国密算法加密等。

6 评价方法

6.1 基本要求

在实际评价中，应根据本文件第5章制定评价细则，确定评价规则、等级划分与判定依据，数据集评价细则的设置可参考附录A。

开展数据集指标评价方法包括但不限于：

- a) 文件审查：对提供满足评价指标描述的文件材料的真实性和有效性进行审查；
- b) 可视化工具或平台测评：通过算法/模型/系统对数据集进行分析和评估；
- c) 人工审查：可根据 GB/T 10111 确定随机抽样规则，人工检查数据样本的有效性。

6.2 过程质量评价

采用文件审查的方式，对提供满足评价指标描述的文档材料的真实性和有效性进行审查。

6.3 内容质量评价

6.3.1 数据集规范性

6.3.1.1 使用可视化工具或平台、人工审查的方式进行评价：

- a) 不应出现 5.2.2.1 中定义关键词的数据；
- b) 不应包括 5.2.2.3 中定义的涉密数据；

6.3.1.2 采用文件审查的方式，对提供的文档材料内容是否包含涉密数据、敏感数据、受限数据的管理进行审查。对于包含 5.2.2.4 定义的敏感数据的数据集，对提供的相关主体授权同意材料、数据脱敏操作说明与审计日志文件有效性进行审查。

6.3.1.3 数据标准符合性使用可视化工具或平台进行测试，应满足 5.2.3 的要求。

6.3.1.4 数据集可追溯性验证方式包括：

- a) 可视化工具或平台测评：通过工具（如数据血缘分析系统）验证数据链条的完整性，对数据集中的数据按照 GB/T 10111 规定的方法进行抽取，验证元数据与操作日志的一致性；
- b) 文件审计：检查数据集变更的相关全文内容是否保持一致，方法包括 MD5 等。

6.3.2 数据集完整性

6.3.2.1 数据文件完整性

使用可视化工具或平台进行测评，数据文件完整性评价结果计算方法应符合式（1）要求：

$$X = \frac{A}{B} \times 100 \dots\dots\dots (1)$$

式中：

X ——评价得分；

A ——特定类型数据需具备符合要求描述（5.3中a）的数据个数；

B ——数据集中特定类型的数据个数。

6.3.2.2 数据记录完整性

使用可视化工具或平台进行测评，数据记录完整性评价结果计算方法应符合式（2）要求：

$$X = \frac{A}{B} \times 100 \dots\dots\dots (2)$$

式中：

X ——评价得分；

A ——数据集中非空数据记录数量；

B ——数据集中至少存在一个数据记录的数据记录总数。

6.3.2.3 数据模态完整性

使用可视化工具或平台进行测评，数据模态完整性评价结果计算方法应符合式（3）要求：

$$X = \frac{A}{B} \times 100 \dots\dots\dots (3)$$

式中：

X ——评价得分；

A ——数据集中融合多传感器数据（5.3中c）的数量；

B ——数据集数据个数。

6.3.2.4 标注完整性

使用可视化工具或平台进行测评，标注完整性评价结果计算方法应符合式（4）要求：

$$X = \frac{A}{B} \times 100 \dots\dots\dots (4)$$

式中：

X ——评价得分；

A ——数据集中满足标注完整性（5.3中d）数据数量；

B ——数据集数据个数。

6.3.3 数据集准确性

6.3.3.1 元数据准确性

使用可视化工具或平台、人工审查的方式进行测评，元数据准确性评价结果计算方法应符合式（5）要求：

$$X = \frac{A}{B} \times 100 \dots\dots\dots (5)$$

式中：

X ——评价得分；

A ——提供满足描述信息的元数据数量；

B ——数据集中元数据的数量。

6.3.3.2 数据集不准确性风险

使用可视化工具或平台进行测评，数据集不准确性风险计算方法应符合式（6）要求：

$$X = \frac{A}{B} \times 100 \dots\dots\dots (6)$$

式中：

- X ——评价得分；
- A ——异常值的数据数量；
- B ——数据集数据数量。

6.3.3.3 损坏数据占比

使用可视化工具或平台进行测评，损坏数据占比评价结果计算方法应符合式（7）要求：

$$X = \frac{A}{B} \times 100 \dots\dots\dots (7)$$

式中：

- X ——评价得分；
- A ——损坏数据的数量；
- B ——数据集数据数量。

6.3.4 数据集一致性

6.3.4.1 多模态数据时间戳一致性

使用可视化工具或平台进行测评，多模态数据时间戳一致性评价结果计算方法应符合式（8）要求：

$$X = \frac{A}{B} \times 100 \dots\dots\dots (8)$$

式中：

- X ——评价得分；
- A ——满足时间戳一致（5.5中a）的数据数量；
- B ——数据集数据数量。

6.3.4.2 语义一致性

使用人工审查的方式进行测评，语义一致性评价结果计算方法应符合式（9）要求：

$$X = \frac{A}{B} \times 100 \dots\dots\dots (9)$$

式中：

- X ——评价得分；
- A ——满足语义一致性（5.5中b）的数据数量；
- B ——数据集数据数量。

6.3.4.3 空间一致性

使用人工审查的方式进行测评，空间一致性评价结果计算方法应符合式（10）要求

$$X = \frac{A}{B} \times 100 \dots\dots\dots (10)$$

式中：

- A ——满足空间一致性（5.5中c）数据数量；
- B ——数据集数据数量。

6.3.5 数据集时效性

6.3.5.1 使用可视化工具或平台进行测评，技术覆盖率计算方法应符合公式（11）：

$$TCR = \frac{A}{B} \times 100\% \dots\dots\dots (11)$$

式中：

- TCR ——技术覆盖率；
- A ——当前节点覆盖的技术数量；
- B ——当前节点行业技术总数。

6.3.5.2 使用可视化工具或平台计算时间衰减指数，量化数据的时间价值衰减与技术相关性，高指数代表数据新鲜、技术匹配，适合当前模型训练；低指数代表数据过时、技术脱节，需紧急更新或淘汰。计算数据集中各样本的时间权重与价值衰减程度，符合式（12）要求：

$$TDI = \frac{\sum_{i=1}^N w(t_i) \cdot Relevance(d_i)}{\sum_{i=1}^N w(t_i)} \dots\dots\dots (12)$$

式中：

TDI ——时间衰减指数；

$w(t_i)$ ——时间衰减函数，通常采用指数衰减（如 $w(t_i) = e^{-\lambda t}$ ），；

t ——数据年龄（当前时间 - 数据采集时间），单位为年；

λ ——衰减率，反映技术迭代速度（如工业机器人领域常取 $\lambda = 0.1 \sim 0.3$ ）

$Relevance(d_i)$ ——数据与技术热点的匹配度（通过关键词匹配或专家评分），如数据相关性评分（0~10），衡量数据与当前技术热点的匹配度。

注：附录B给出了典型场景/技术维度，数据与技术热点的匹配度分值建议。

6.3.6 数据集可访问性

6.3.6.1 使用可视化工具或平台进行测评，数据文件可访问性的评价结果计算方法应符合式（13）要求

$$X = \frac{A}{B} \times 100 \dots\dots\dots (13)$$

式中：

X ——评价得分；

A ——能正常打开并解析，不存在格式损坏或无法解码的数据；

B ——数据集数据数量。

6.3.6.2 针对非公开数据集（5.2.2.2），创建不同角色，并为每个角色分配相应的权限（如公众/学者/企业/管理员等），使用这些账户进行登录，每个角色尝试访问其他角色的权限，记录所有尝试访问的行为，结果分为越权成功和越权失败。评价结果计算方法应符合式（14）要求：

$$X = \frac{A}{B} \times 100 \dots\dots\dots (14)$$

式中：

X ——评价得分；

A ——越权失败次数；

B ——所有尝试访问次数。

7 评价组织与流程

7.1 评价组织

评价组织宜具备下列要求：

- a) 评价团队由3人及以上单数构成；
- b) 评价人员应通过能力测评获取相应证书，熟悉相关文件与标准要求，掌握机器人数据集建设工作的专业知识，遵纪守法、诚实正直、坚持原则、实事求是、科学公正，定期开展技能培训。
- c) 具备保障评价活动的规范性文件。

7.2 评价流程

7.2.1 概述

人形机器人数据集数据质量评价流程划分为4个阶段，见图1：

- a) 评价准备；
- b) 评价实施；
- c) 结果评估；

d) 维护改进。

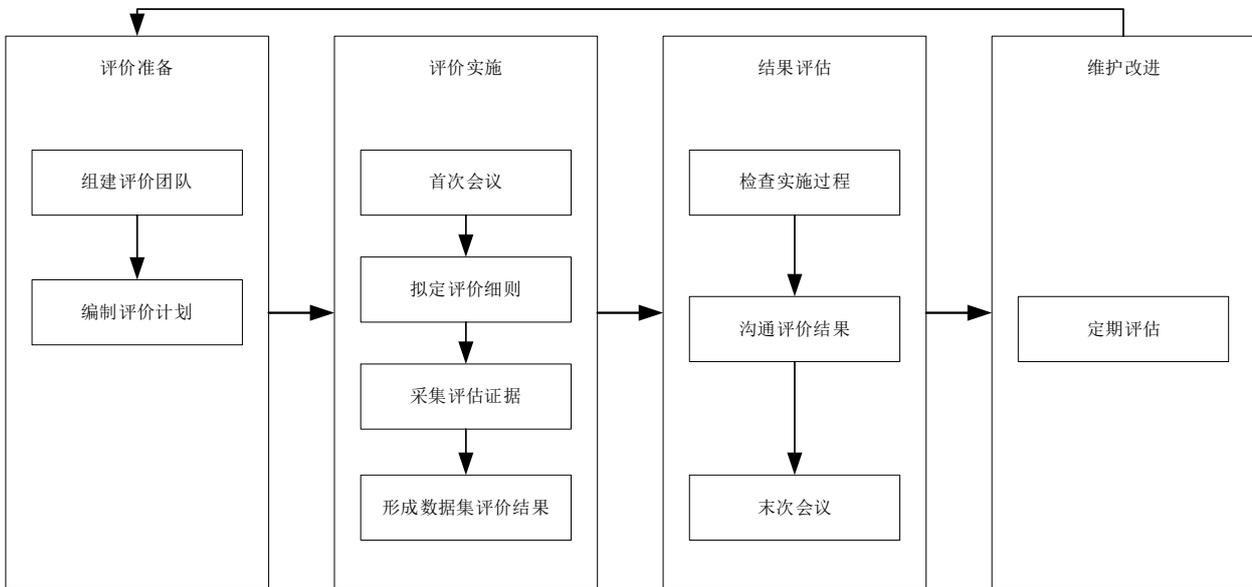


图1 评价流程图

7.2.2 评价准备

7.2.2.1 组建评价团队包括：

- a) 组建 1 个具备评估能力的评估团队实施现场评估活动，见 7.1；
- b) 确认 1 名评估组长及多名评估组员；
- c) 提前将评审时间告知评审团队。

7.2.2.2 编制评价计划包括：

- a) 评价团队讨论确定评价指标及方法，见第 5 章、6.1；
- b) 评价团队编制评估计划，并与受评估方确认。评估计划至少包括评价目的、评价任务、评价时间、评价人员、评价日程安排等。

7.2.3 评价实施

7.2.3.1 首次会议要求包括：

- a) 确认相关方对评价计划的安排达成一致；
- b) 介绍评价团队；
- c) 确保策划的质量评价活动可执行；
- d) 会议说明评价目的、介绍评价方法、确定评价日程以及明确其他需要提前沟通的事项。

7.2.3.2 评价细则拟定满足 6.1 要求。

7.2.3.3 采集评估证据包括：

- a) 实施评价过程中，通过适当的方法收集并验证与评价目标、评价指标有关的证据；
- b) 采集的证据予以记录，采集方式主要为包括访谈、文件评审等。

7.2.3.4 形成数据集评价结果包括：

- a) 对照评价细则，将采集的证据与其进行对比形成评价结果，包括具有证据支持的符合事项、改进方向以及弱项；
- b) 完成测评后出具结果报告，结果报告包括但不限于：
 - 1) 测评报告；
 - 2) 等级证书。
- c) 评价团队对评价结果达成一致意见，必要时进行组内评审。

7.2.4 结果评估

7.2.4.1 检查实施过程包括：

- a) 核查采集证据；
- b) 核查评价过程记录；
- c) 核查评价结果。

7.2.4.2 评价团队将评价结果与受评估方代表进行通报，给予受评估方再次论证的机会，并由评价团队确定最终结果。

7.2.4.3 末次会议要求包括：

- a) 总结评价过程；
- b) 发布评价发现和评价结果；
- c) 会议内容包括评价总结、评价结果、评估强项、评估弱项、改进方向以及后续相关活动介绍等。

7.2.5 维护改进

7.2.5.1 受评估方基于评价结果、改进方向，开展数据集提升活动，附录 C 给出质量提升技术方案示例。

7.2.5.2 定期评估要求包括：

- a) 定期评估数据集内容质量和过程质量，检查其是否符合实际应用需求更新；
- b) 核查意见和问题，调整数据集建设。

附录 A
(资料性)
人形机器人数据集评价设置导则

A.1 人形机器人数据集评价设置导则见表 A.1。

表A.1 人形机器人数据集评价设置导则

序号	一级指标	二级指标	测评项目	评价指标描述	可流通的数据集基本要求
1	过程质量	数据集构建需求与定义	/	明确定义数据集的目的和预期用途，清晰描述数据集将要支持的具体应用场景与特定任务。	提供满足要求的文件。
2		数据采集	/	提供数据采集依据的法规、技术文件或其他参考文献； 提供数据采集环境、采集方式（如通过传感器、手动数据输入、外骨骼采集、VR采集、合成数据等）、采集设备/平台/工具链描述、机器人/传感器坐标系描述（如有）等； 提供数据来源的合规性陈述； 提供匿名化或假名化的方法（如适用）； 提供采样的偏差程度信息； 提供数据集操作（合并、切片等）、数据集负责人信息（如有）； 提供数据采集过程中安全审计与风险评估文件。	提供满足要求的文件。
3		数据预处理	/	清晰描述数据的录入、排除标准以及数据筛选的方法。 清晰描述数据预处理（如数据清洗、数据增强、数据特征提取、数据分帧、数据对齐与同步校准、数据压缩、数据融合等）的操作方法与参数设置。	提供相应证明文件。
4		数据标注	/	提供数据标注依据的法规、技术标准或其他参考文献； 提供数据标注工具和平台信息； 如具有标注信息，需描述数据标注任务说明、数据验收标准与质控流程。	提供相应证明文件。
5		数据存储	/	清晰描述数据存储信息，包括数据集存储方式与存储路径、安全控制、备份、访问路径、使用权限说明等。	提供满足要求的文件。

表 A.1 人形机器人数据集评价设置导则（续）

序号	一级指标	二级指标	测评项目	评价指标描述	可流通的数据集基本要求
6	过程质量	数据集使用与流通	/	提供符合T/SAIAS《人形机器人操作任务数据集数据格式规范》中第6章规定的数据集描述性文件（整体描述文件、片段描述文件）； 提供数据集质量评估结果的记录（如有）； 提供数据集传输与访问的法规、技术标准或其他参考文献； 提供不同安全级别的数据集访问权限说明、访问路径、开放策略（如非涉密数据宜采用通用结构化格式，同步提供标准化元数据，涉密数据应进行加密/匿名化处理等）、获取流程文件（如通过邮箱/组织认证后在线注册；进行自动化/人工审核，学术机构审核时间不超过1天，商业机构审核时间不超过3天等） 提供数据集具备跟踪与改进渠道、数据集优化策略、支撑承诺（如维护计划、持续改进承诺等）信息（如有）	提供满足要求的文件。
7		数据更新维护	/	提供数据集版本、更新频率、方式及责任人信息。	提供相应证明文件。
8		数据集销毁（如有）	/	提供数据删除、介质报废遵循的法规、技术标准或其他参考文献。	提供满足要求的文件。
9		合规性审查	/	提供数据处理活动遵守的相关伦理准则和法律法规；（一票否决条件） 提供隐私政策同意记录等合规文件；（一票否决条件） 清晰描述用于保护涉及隐私和个人数据的技术手段，如数据去标识化、数据匿名化等，适当时描述对应规则。	提供满足要求的文件。
10	内容质量	数据集规范性	安全与隐私	提供相关数据分级说明与管理文件； 检查数据集中是否出现5.2.2中定义的关键词；	提供满足要求的文件。
11			数据标准符合性	提供数据集描述文件，检查是否符合5.2.3的要求。	提供满足要求的文件。
12			可追溯性	提供相关文件，检查是否符合5.2.4的要求。	提供满足要求的文件。
13		数据集完整性	数据文件完整性	从低到高设置5级指标取值要求。	至少满足最低等级指标要求。
14			数据记录完整性	从低到高设置5级指标取值要求。	至少满足最低等级指标要求。
15			数据模态完整性	从低到高设置5级指标取值要求。	至少满足最低等级指标要求。
16			标注完整性	从低到高设置5级指标取值要求。	至少满足最低等级指标要求。

17	数据集准确性	元数据准确性	从低到高设置5级指标取值要求。	至少满足最低等级指标要求。
18		数据集不确定性 风险	从低到高设置5级指标取值要求。	至少满足最低等级指标要求。
19		缺失数据、损坏 数据占比情况	从低到高设置5级指标取值要求。	至少满足最低等级指标要求。
20	数据集一致性	多模态数据时间 戳一致	从低到高设置5级指标取值要求。	至少满足最低等级指标要求。
21		语义一致性	从低到高设置5级指标取值要求。	至少满足最低等级指标要求。
22		空间一致性	从低到高设置5级指标取值要求。	至少满足最低等级指标要求。
23	数据集时效性	时间覆盖度	从低到高设置5级指标取值要求。	至少满足最低等级指标要求。
24		技术同步性	从低到高设置5级指标取值要求。	至少满足最低等级指标要求。
25	数据集可访问性	数据文件可访问	从低到高设置5级指标取值要求。	至少满足最低等级指标要求。
26		访问权限控制	从低到高设置5级指标取值要求。	至少满足最低等级指标要求。

A.2 结合应用场景与任务需求设置二级指标及其测评项目具体权重，计算数据集二级指标评价结果得分，最终得到数据集二级指标对应等级。指标评价结果计算公式应符合式（A.1）要求：

$$Y = \sum_{i=1}^n w_i \times X_i \dots\dots\dots (A.1)$$

式中：

Y ——二级评价指标得分；

n ——二级评价指标覆盖的测评子项目；

w_i ——二级评价指标的测评子项目*i*的权重；

X_i ——二级评价指标的测评子项目*i*的得分。

附录 B
(资料性)
典型场景数据与技术热点匹配度分值

表B. 1~表B. 3给出了家政、商超、工厂3类场景数据与技术热点匹配度的分值建议。

表B. 1 家政场景数据与技术热点匹配度分值

场景/技术维度		数据与技术热点的匹配度
图像/视频	目标检测	5
	图像分割	7
	图像分类	5
音频	语音识别	5
	语音合成	3
	语音对话	9
	指令转换	3
	声源定位	5
	声纹识别	5
嗅觉	气体成分识别	7
	气体浓度	7
触觉	触觉感知	8
	力反馈控制	9
	材质识别	7
	触觉数据融合	8
味觉	味觉成分识别	7
运动学	基于深度学习逆运动学	9
	基于强化学习轨迹预测	9
	MPC全身运动学控制	9
	模仿学习	9
动力学	基于深度学习动力学建模	9
	MPC全身动力学控制	9
	基于强化学习动力学控制	9

表B. 2 商超场景数据与技术热点匹配度分值

场景/技术维度		数据与技术热点的匹配度
图像/视频	目标检测	5
	图像分割	7
	图像分类	5
音频	语音识别	5
	语音合成	3
	语音对话	9
	指令转换	3
	声源定位	5
	声纹识别	5
嗅觉	气体成分识别	7
	气体浓度	7
触觉	触觉感知	8
	力反馈控制	9
	材质识别	7
	触觉数据融合	8
味觉	味觉成分识别	7
运动学	基于深度学习逆运动学	9

表 B.2 商超场景数据与技术热点匹配度分值（续）

场景/技术维度		数据与技术热点的匹配度
运动学	基于强化学习轨迹预测	9
	MPC全身运动学控制	9
	模仿学习	9
动力学	基于深度学习动力学建模	9
	MPC全身动力学控制	9
	基于强化学习动力学控制	9

表B.3 工厂场景数据与技术热点匹配度分值

场景/技术维度		数据与技术热点的匹配度
图像/视频	目标检测	5
	图像分割	7
	图像分类	5
音频	语音识别	5
	语音合成	3
	语音对话	9
	指令转换	3
	声源定位	5
	声纹识别	5
嗅觉	气体成分识别	7
	气体浓度	7
触觉	触觉感知	8
	力反馈控制	9
	材质识别	7
	触觉数据融合	8
味觉	味觉成分识别	7
运动学	基于深度学习逆运动学	9
	基于强化学习轨迹预测	9
	MPC全身运动学控制	9
	模仿学习	9
动力学	基于深度学习动力学建模	9
	MPC全身动力学控制	9
	基于强化学习动力学控制	9

附录 C (资料性) 数据集质量提升技术方案示例

C.1 概述

根据数据集质量评价结果，识别数据集质量不符合项，数据集质量提升环节包括但不限于：

- 数据采集；
- 数据预处理：包括数据清洗、数据规范化与标准化、数据增强、数据分片等；
- 数据标注；
- 数据脱敏。

C.2 数据采集

进行多模态传感器校准，包括硬件同步、坐标系统一致性。

C.3 数据预处理

C.3.1 数据清洗

C.3.1.1 宜使用数据清洗工具：

- a) 传感器数据清洗：使用 ROS bag_tools 过滤异常帧（如点云离群点、图像过曝）等；
- b) 多模态对齐：通过时间戳插值对齐异步数据流（如视觉-力觉延迟补偿）等；
- c) 通用数据清洗工具：如 OpenRefine、Data-Juicer、Trifacta Wrangler 等

C.3.1.2 数据清洗涉及纠正或删除不完整、不正确或不相关的数据。数据清洗方法可以包括：

- 删除重复记录，
- 删除或更正带有错误数据项的数据记录
- 删除或填充空数据项；

C.3.2 数据规范化与标准化

数据规范化与标准化方法包括但不限于：

- a) 归一化处理，通过将值缩放为单位规范（即介于 0 和 1 之间）来在整个数据集上建立共同的区间，如图像数据像素值除以 255 实现归一化；
- b) 抗异常值标准化：采用 Robust Scalar 等；
- c) 缺失值处理：使用迭代填充器（MICE）等。

C.3.3 数据增强

C.3.3.1 视觉数据增强方法包括但不限于：

- 几何变换：缩放、裁剪、旋转等
- 噪声注入；
- 遮挡模拟；
- 物体替换。

C.3.3.2 音频数据增强方法包括但不限于：

- 添加噪声；
- 语速干扰；
- 混响。

C.3.3.3 运动学数据增强方法包括但不限于：

- 轨迹增强；
- 时序伸缩。

C.3.4 数据分片

数据分片方法包括但不限于：

- a) 按任务阶段的分片：包含多模态数据（视觉、触觉、运动学）及元数据（时间戳、场景 ID）等；
- b) 按事件触发的分片：如碰撞事件、关节超限、传感器失效等；
- c) 按时间窗口分片。

C.4 数据标注

数据标注工具符合T/SAIAS《人形机器人数据集数据标注》的要求。

C.5 数据脱敏

除特殊任务需求，个人身份信息的训练、验证和测试数据应去标识化。去标识化是指从获取的数据中删除、更改或限制对个人身份信息的访问，使其无法与一个或更多人相关联。去标识化的方法包括但不限于：

- 匿名化；
- 伪匿名化；
- 解除链接；
- 聚合；
- 差分隐私。

参 考 文 献

- [1] GB/T 12643 机器人与机器人装备 词汇
 - [2] GB/T 18391.1 信息技术 实现元数据注册系统（MDR）内容一致性的规程 第1部分：数据元
 - [3] GB/T 36344 信息技术 数据质量评价指标
 - [4] GB/T 38667 信息技术 大数据 数据分类指南
 - [5] GB/T 40660 信息安全技术 生物特征识别信息保护基本要求
 - [6] GB/T 41352 再制造 机械产品质量评价通则
 - [7] GB/T 41479 信息安全技术 网络数据处理安全要求
 - [8] GB/T 41773 信息安全技术 步态识别数据安全要求
 - [9] GB/T 41807 信息安全技术 声纹识别数据安全要求
 - [10] GB/T 41819 信息安全技术 人脸识别数据安全要求
 - [11] GB/Z 42007 跨境电子商务交易服务质量评价
 - [12] CY/T 103 数据库出版物质量评价规范
 - [13] YY/T 1833.2 人工智能医疗器械 质量要求和评价第2部分：数据集通用要求
 - [14] DB35/T 1952 公共信息资源开发 数据质量评价规范
 - [15] ISO 19115-1: 2014 Geographic information — Metadata Part 1: Fundamentals
-