

T/SAIAS

上海市人工智能行业协会团体标准

T/SAIAS 018—2025

代替：T/SAIAS 018-2024

具身智能语料库建设导则

Construction Guidelines for Corpus of Embodied Intelligence

2025 - 04 - 30 发布

2025 - 05 - 01 实施

全国团体标准信息平台

目 次

前 言	IV
引 言	VI
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	2
5 具身智能数据	2
5.1 一般原则	2
5.2 数据集	2
5.2.1 数据片段	2
5.2.2 时间切片	3
5.3 数据资源	3
5.3.1 数据资源类型	3
5.3.2 数据信息存储格式	3
6 具身智能数据采集	4
6.1 真实数据采集	4
6.1.1 采集方式	4
6.1.2 采集本体	5
6.2 合成数据采集	5
6.2.1 采集方式	5
6.2.2 合成数据有效性考虑因素	5
6.2.3 合成数据生成方式	5
6.2.4 合成数据用途	6
6.3 采集数据提交	6
6.3.1 数据文件标识	6
6.3.2 数据提交方式	6
6.4 数据采集场景与任务体系	7
6.4.1 场景	7
6.4.2 任务	7
6.4.3 原子技能	7
6.5 数据采集操作规范	7
6.5.1 采集环境要求	7
6.5.2 采集设备选型	7
6.5.3 采集人员培训	7
7 具身智能语料生产	8
7.1 采集数据完整性检查	8
7.2 数据资源清洗过程	8

7.2.1	主要流程	8
7.2.2	数据规范管理	8
7.2.3	数据整理分类	8
7.2.4	数据脱密脱敏	8
7.2.5	数据去噪	9
7.2.6	数据去重	9
7.2.7	数据存储与备份	9
7.3	数据资源标注过程	9
7.3.1	基本要求	9
7.3.2	标注任务制定	9
7.3.3	数据标注检查流程	9
7.3.4	定义坐标系	9
7.3.5	标注方法	9
7.4	数据质量评估	10
7.4.1	基本要求	10
7.4.2	数据准确性	10
7.4.3	数据一致性	10
7.4.4	数据适用性	10
7.5	数据有效性评估	11
7.5.1	数据集选型	11
7.5.2	模型选型	11
7.5.3	测试本体选型	11
7.5.4	测试任务设计	11
7.5.5	测试指标设计	11
7.5.6	模型测试执行	11
8	数据安全	11
8.1	安全性评价	11
8.2	安全管理流程	11
8.3	安全技术	12
8.4	人员管理	12
附录 A	(资料性) 具身智能语料库数据资源类型	13
A.1	视频数据	13
A.2	雷达数据	14
A.3	音频数据	15
A.4	本体数据	15
A.5	轨迹导航数据	16
附录 B	(资料性) 具身智能遥操作模式数据采集本体参考性能参数	17
B.1	性能参数	17
附录 C	(资料性) “生产伴随”式数据采集设备参考性能参数	18
C.1	性能参数	18
C.2	可靠性与稳定性	18
C.3	适配性	18
附录 D	(资料性) 具身智能语料库数据有效性评估数据集要求	19

D.1 数据集准备	19
D.2 数据质量要求	19
参 考 文 献	20

全国团体标准信息平台

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。

本文件代替T/SAIAS 018—2024《具身智能语料库建设导则》，与T/SAIAS 018—2024相比，除结构调整和编辑性改动外，主要技术变化如下：

- a) 更改了“范围”部分，添加了和语料质量评估、有效性评估相关的表述（见第1章，2024年版的第1章）；
- b) 增加了“合成数据”术语（见3.7）；
- c) 更改了“具身智能数据”的“一般原则”（见5.1，2024年版的5.1）；
- d) 更改了数据集中的“时间切片”定义，及其内容类型分类（见5.3.2，2024年版的5.3.2）；
- e) 更改了“数据资源类型”中各类数据的推荐指标、语义分类、语义特征等内容（见5.4.1、附录A，2024年版的5.4.1）；
- f) 增加了“视频数据分类表”中的数据来源（见附录A.1）；
- g) 删除了“雷达数据”的“分辨率”指标（见2024年版的5.4.1.2）；
- h) 增加了“本体数据”中对于采集设备本体的描述（见附录A.4），“数据信息存储格式”的“基本要求”（见5.4.2.1），“传感器标定信息格式”中“数据文档命名规则及存储路径”（见5.4.2.2），“元数据格式”中本体数据的末端工具信息（见5.4.2.3），“真实数据采集”的采集路径（见6.1.1）；
- i) 删除了“采集本体”中“表14：采集本体硬件”（见2024年版的6.2.1）；
- j) 更改了“采集本体”的参考性能参数（见6.1.2、附录B、附录C）；
- k) 增加了“采集本体软件”对于“数据校验与上传”、“本体数据落盘”及“设备状态管理”的功能要求（见6.1.2），“合成数据采集”的采集路径（见6.2.1）；
- l) 更改了为保证数据有效性宜考虑的主要因素中，“随机化控制”的内容（见6.2.2，2024年版的6.3.2）；
- m) 增加了“合成数据的生成方式”中，仿真合成这一步骤（见6.2.3），“数据采集场景与任务体系”（见6.4），“数据采集操作规范”（见6.5）；
- n) 更改了“采集数据完整性检查”的步骤内容（见7.1，2024年版7.1）；
- o) 增加了“数据脱密脱敏”需要符合的相关规定要求（见7.2.4），“数据去噪”包括的情形（见7.2.5），“数据标注”的基本要求（见7.3.1）以及“标注任务制定”中的“任务标注”和“子任务标注”类别（见7.3.2）；
- p) 更改了“定义坐标系”的类型（见7.3.4，2024年版的7.3.3）；
- q) 增加了“数据质量评估”部分（见7.4），“数据有效性评估”部分（见7.5），“具身智能语料库数据有效性评估数据集要求”作为资料性参考（见附录D）。

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这些专利的责任。

本文件由上海市人工智能行业协会提出并归口。

本文件起草单位：上海库帕思科技有限公司、上海市人工智能行业协会、上海人工智能创新中心、人形机器人（上海）有限公司、上海机器人产业技术研究院有限公司、上海智元新创技术有限公司、上海穹彻智能科技有限公司、上海傅利叶智能科技有限公司、上海开普勒机器人有限公司、中国电子科技集团公司第二十一研究所、星海图（苏州）人工智能科技有限公司、工业互联网创新中心（上海）有限公司、上海国评智检机器人有限公司、复旦大学、上海交通大学、同济大学、上海钛米机器人股份有限公司、松应科技有限公司、上海华为技术有限公司、上海擎朗智能科技有限公司、四川天链机器人股份有限公司、上海市先导产业促进中心、上海中创产业创新研究院、国创智造科技（上海）有限公司。

本文件主要起草人：山栋明、钟俊浩、黄海清、庞江淼、姜育刚、卢策吾、何斌、陈春玉、王洪武、郑忠斌、高平、施佳樑、卢恒、张裕珍、邓思文、曹宇、李晨歌、赵春昊、饶雪、汪汗青、江磊、刘宇飞、李泳耀、许凯、朱玉琛、林雪琴、黄媚、王煜、沈荣港、邢琳、姚卯青、闫维新、沈咏剑、周航、

方汇、熊俊杰、胡静萍、陈敬思超、朱建超、顾捷、周斌、蔡宇圣、陶仁智、高宇翔、胡硕、沈林鹏、陈宇涛、李昀佑、许华哲、虞磊、张宇雷、范佳琪、潘晶、沈满、冯瑞、蒋烁、王志鹏、王宇、聂凯旋、张小波、周顺波、张恒、唐旋来、刘斐、黄巍、颜进、郑理莹、宋琼、李琰、方金武、王道洋、蒋龙泉、朱立奇、贺仁龙、郑茂宽。

本文件及其所代替文件的历次版本发布情况为：

——2024年首次发布为T/SAIAS 018—2024；

——本次为第一次修订。

本文件版权归上海市人工智能行业协会所有。未经许可，不得擅自复制、转载、抄袭、改编、汇编、翻译或将本标准用于其他任何商业目的。

引 言

人工智能是新一轮科技革命和产业变革的重要驱动力量，语料数据则是人工智能研究和应用不可或缺的资源，高质量的语料库更是人工智能赋能新质生产力的关键。

在人工智能的浪潮中，具身智能作为前沿科技的代表，正受到国家和上海市的高度重视。《具身智能语料库建设导则》的编纂，正是在这一背景下应运而生，旨在为该领域的发展提供坚实的数据基础和标准化指导。

国家层面，工信部发布的《人形机器人创新发展指导意见》强调了开展人形机器人标准化路线图研究的重要性，并提出建立健全人形机器人产业标准体系。这不仅为具身智能语料库的建设提供了政策支持，也为行业的健康发展指明了方向。

上海市作为科技创新的先行者，率先发布全国首批人形机器人具身智能标准，包括《具身智能智能化等级分级指南》和《人形机器人分类分级应用指南》。这些标准不仅规范了人形机器人的技术语言和发展路径，也为语料库建设提供了明确的技术等级划分依据。上海还计划建设完成多地联动的具身智能训练场，支持超100台异构机器人采集训练，这将为具身智能的语料库建设提供丰富的实践场景和数据资源。

此外，上海市人民政府印发的《上海市进一步推进新型基础设施建设行动方案（2023-2026年）》中提到，将布局智能机器人创新基础设施，建设“大模型+人形机器人”协同创新平台，这为具身智能语料库的建设提供了政策和技术支持。

本文件的编写，紧密结合这些最新的政策导向，为语料库的建设提供科学、系统、标准化的指导。本导则将详细阐述如何收集、标注和利用数据，并对数据的有效性进行评估，以满足具身智能系统在不同场景下的需求，同时确保数据的质量和多样性。通过遵循本导则，我们期待能够推动具身智能技术的创新和应用，为具身智能产业的科学健康发展注入强大动力。

具身智能语料库建设导则

1 范围

本文件提供了建设具身智能模型训练数据内容、数据采集、语料生产（质量评估、数据有效性评估）和数据安全方面的技术指导方法。

本文件适用于具身智能语料库的研究、开发、维护、应用、评估等工作。其它语料库建设也可参照使用。

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 22239-2019 信息安全技术 网络安全等级保护基本要求
- GB/T 35273-2020 信息安全技术 个人信息安全规范
- GB/T 35274-2023 数据安全技术 大数据服务安全能力要求
- GB/T 35295-2017 信息技术 大数据 术语
- GB/T 36073-2018 数据管理能力成熟度评估模型
- GB/T 41479-2022 信息安全技术 网络数据处理安全要求
- GB/T 44588-2024 数据安全技术 互联网平台及产品服务个人信息处理规则
- GB/T 45574-2025 数据安全技术 敏感个人信息处理安全要求
- GB/T 45674-2025 数据安全技术 生成式人工智能数据标注安全规范
- GY/T 353-2021 网络视听节目视频格式命名及参数规范
- GY/T 360-2022 广播电视和网络视听节目内容标识标签规范
- YD/T 4245-2023 电信网和互联网数据脱敏技术要求和测试方法
- T/SAIAS 015-2024 语料库建设导则
- ISO/IEC 5259-4:2024 人工智能-分析和机器学习的数据质量

3 术语和定义

T/SAIAS 015—2024界定的以及下列术语和定义适用于本文件。

3.1

具身智能 Embodied AI

基于物理身体进行感知和行动的智能系统，其通过智能体与环境的交互获取信息、理解问题、做出决策并实现行动，从而产生智能行为和适应性。

注：本文件限定的具身智能本体包括泛人形机器人、仿人机器人、智能机器人等。

3.2

数据资源 data resources

以电子化形式记录和保存的具备原始性、可机器读取、可供社会化再利用的数据集合。

[来源：GB/T 4894-2024，4.7.3.2.3，有修改]

3.3

具身智能本体 Embodied intelligent ontology

在物理环境或者虚拟环境中进行感知和任务执行的机器人。

3.4

数据片段 data clip

以片段形式记录和保存的具身智能模型训练使用的数据集。

注：是按照一定目的和方法进行选择并有序排列的数据汇集。

3.5

时间切片 time slice

在同一时刻中，由各相机视频帧、雷达数据、本体数据、轨迹导航数据等组成的数据集合。

3.6

模拟仿真平台 simulation platform

基于计算机的系统，用于模拟现实世界或想象中的环境、条件或系统。

3.7

合成数据 synthetic data

通过模拟仿真平台等工具生成的数据。

4 缩略语

下列缩略语适用于本文件。

FOV 视场角 (Field Of View)

NSFW 不适于工作场所浏览 (Not Safe For Work, Not Suitable For Work)

SFT 有监督微调 (Supervised Fine-Tuning)

5 具身智能数据

5.1 一般原则

构建具身智能语料库时，其数据资源原则宜包括：

- a) 数据多样性：覆盖多个领域和应用场景，并包括文本、图像、视频等多模态数据；
- b) 数据一致性：同一信息的多模态数据在采集时间范围上保持一致；
- c) 数据规范性：具有统一的数据格式规范；
- d) 数据完整性：覆盖多模态数据的所有元素，对动作、环境、目标等进行详尽标注；
- e) 数据准确性：数据与客观事实一致；
- f) 数据时效性：标注精确时间戳与事件序列，采用分层存储策略区分实时与历史数据。

5.2 数据集

具身智能语料库数据集，以一系列数据片段的形式存储，由一系列数据片段有序或自由组合而成，适用于具身智能模型训练使用。数据集与数据片段关系如图1所示。

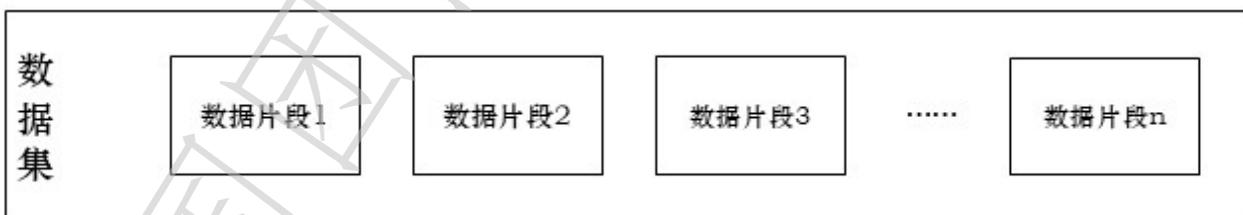


图1 数据集与数据片段关系

5.2.1 数据片段

数据片段由连续的若干个数据时间切片组成。数据片段与数据时间切片的关系如图2所示。



图2 数据片段与数据时间切片的关系

5.2.2 时间切片

时间切片内容包含同一时刻的视频数据、雷达数据、本体数据、导航数据及其他数据，每一个时间切片是时空同步的。

示例：按每秒30帧为例，一个时间切片是在1帧内所有传感器数据的集合，即所有传感器在100毫秒（ms）内产生数据的集合。

时间切片内容根据其来源，包括以下数据类型：

- a) 视频数据：宜包含通过相机等采集设备获得的视觉数据；
- b) 雷达数据：宜包含各种雷达获得的激光点云、距离等数据；
- c) 本体数据：宜包含本体状态数据、控制命令、诊断信息、动态数据、末端设备类型等；
- d) 导航数据：宜包含GPS、地图数据、GNSS数据、IMU数据等；
- e) 其他数据：宜包含音频数据、触觉数据、力觉数据、嗅觉数据、文本数据、传感器标定数据、内外参数数据、场景标签数据（采集场景环境信息、场景任务数据、数据语义特征、格式说明文件等）等。

5.3 数据资源

5.3.1 数据资源类型

5.3.1.1 视频数据

视频数据以一系列序列帧的形式存储，用于视频内容的存储和处理。相关推荐指标、语义分类、语义特征等详见附录A。

5.3.1.2 雷达数据

雷达数据包括激光点云数据和毫米波数据。激光点云数据可发射激光束并接收反射回来的信号，精确测量物理的距离、形状和位置，构建三维环境模型。毫米波数据擅长测量目标的距离和速度，对金属物理比较敏感，在恶劣环境下也能较好工作。雷达数据相关推荐指标详见附录A。

5.3.1.3 音频数据

声音表征的数据，简称音频数据，是以声音波形的形式存储，通常以采样率和位深来表征声音信号。音频数据资源（或可再加工音频数据产品）的推荐指标详见附录A。

5.3.1.4 本体数据

本体数据宜包括执行数据采集的机器人或外骨骼机械臂等设备的本体状态数据、控制命令、诊断信息、本体动态数据、末端设备类型数据，以及传感器标定数据。其中，传感器标定数据在具身智能训练中能够校正误差，统一坐标系，以及确保不同传感器之间的数据关联性，使传感器融合，提高具身智能整体性能。本体数据资源的推荐指标、分类等详见附录A。

5.3.1.5 导航数据

导航数据宜包括定位数据、轨迹数据等。导航数据的推荐指标和资源分类详见附录A。

5.3.2 存储数据信息

5.3.2.1 基本要求

宜支持HDF5、Lerobot、Episode等主流数据格式。

宜包括数据集的标签信息文件、标定参数文件、本体硬件构型（单双臂、末端工具或者URDF）等文件。

5.3.2.2 传感器标定信息

数据采集传感器标定信息应包含：

- a) 传感器设备序列号及名称；
- b) 传感器的类型及数量；
- c) 传感器采集频率；
- d) 传感器内外参数；
- e) 数据文档命名规则及存储路径。

5.3.2.3 元数据

采集的元数据宜包含以下信息：

- a) 通用数据：
 - 1) 时间戳；
 - 2) 唯一 ID。
- b) 本体数据：
 - 1) 类型、厂商及型号；
 - 2) 末端工具信息，如夹爪、灵巧手等；
 - 3) 关节位姿、扭矩等数据。
- c) 视觉传感器数据：
 - 1) 彩色图像数据；
 - 2) 深度图像数据；
 - 3) 采样频率；
 - 4) 图像尺寸。
- d) 雷达传感器数据：
 - 1) 采集范围；
 - 2) 采集精度；
 - 3) 采样频率。
- e) 触觉传感器数据：
 - 1) 灵敏度；
 - 2) 分辨率；
 - 3) 响应时间。
- f) 音频传感器数据；
- g) 力觉传感器数据；
- h) 嗅觉传感器数据；
- i) 轨迹导航数据；
- j) 采集环境信息（包括场景布局、地点、设备/道具布置情况等）；
- k) 采集任务描述信息（包括文字指令和对话信息等）。

5.3.2.4 元数据组帧

元数据组帧宜包含：

- a) 帧索引序列；
- b) 传感器数据序列；
- c) 本体数据序列。

6 具身智能数据采集

6.1 真实数据采集

6.1.1 采集方式

6.1.1.1 遥操作式数据采集

遥操作式数据采集，即人类控制机器设备执行任务，并采集过程数据。通过高精度动作映射与低延迟通信，人类操作员的远程动作被同步至执行终端，同时记录决策逻辑与纠偏过程，具备跨空间操作、人机协同决策、动态环境适应等特点。

6.1.1.2 “生产伴随”式数据采集

生产伴随式数据采集，在真实生产环境中，由真实工作人员穿戴外骨骼等数据采集设备进行实际的生产作业，采集过程由设备上的摄像头以及穿戴的外骨骼数据采集设备等实时记录和存储多模态数据，用于具身智能模型训练，具有进场即采、真实有效等特点。

6.1.2 采集本体

本节规定了采集本体的基本配置，帮助采集方有效改装采集本体，保障采集数据一致性。采集本体的感知相机数量具体宜根据实际采集需要进行加装。根据不同的采集模式分类，其参考性能参数详见附录B和附录C。

采集本体软件功能宜符合下列规定：

- a) 多模态传感器时间同步与空间同步；
- b) 全量满带宽数据存储；
- c) 采集数据本地校验及数据上传；
- d) 包括视觉传感器数据、力觉传感器及执行机构等在内的数据落盘；
- e) 本体的设备状态管理；
- f) 通过OTA进行软件更新。

6.2 合成数据采集

6.2.1 采集方式

数据合成采集，即通过高保真的3D环境构建，构建物理精确的数字场景，实现不同版本的机器人在虚拟场景中的数据合成。同时通过科学高效的仿真软件平台作为数据生成载体，引入视觉和物理随机化处理，增强了仿真数据的多样性，提高模型在不同环境中的泛化能力，实现低成本采集的目标。

6.2.2 合成数据有效性考虑因素

为了确保仿真合成数据的有效性和适用性，宜考虑的主要因素包括：

- g) 高保真度：仿真环境尽量模拟现实世界中的物理和视觉特性，保证数据的真实性和准确性；
- h) 多样性：合成数据包含不同场景、多种任务和多种传感器数据，以确保模型的泛化能力；
- i) 可重复性：仿真数据生成可重复，仿真资产可以重复利用；
- j) 同步性：多传感器采集的数据严格保持时间同步，特别是在多模态数据（包括图像、深度、雷达、IMU）融合时，时间戳精度足够高；
- k) 标准化格式：所有数据符合主流数据格式，以便于存储、共享和分析；
- l) 随机化控制：为了增强数据的鲁棒性，合成数据进行随机化处理，包括视觉随机化（包括光照、材质等）、物理随机化（包括摩擦、重力等）、物体随机化和技能随机化（操作、导航、交互等），以模拟不同环境下的变化。

6.2.3 合成数据生成方式

仿真合成数据生成宜包括以下步骤：

- a) 场景构建：使用3D数字孪生技术对真实场景进行虚拟化，构建高度逼真的仿真场景。场景的视觉和物理特性根据应用场景需求进行细致建模；
- b) 任务设计：机器人在场景中执行多种任务（包括抓取、移动、操作等），通过仿真平台模拟这些任务，并生成对应的动作和感知数据；
- c) 多传感器采集：通过机器人内部和外部的多种传感器（包括RGB相机、深度传感器、激光雷达等）同步采集数据。确保传感器数据的时间同步性，尤其在多模态融合时，数据采集频率需保持一致；

- d) 仿真合成：依托自动化的任务和场景生成机制，同步融入动态随机（物体属性、环境条件）和数据增强技术，并实现时序对齐，最终生成多样、鲁棒、大规模、高保真的多模态合成数据；
- e) 数据随机化：通过视觉和物理属性的随机化，生成多样化的数据集，以增加模型的鲁棒性和泛化能力。视觉随机化涉及材质、光照、视角等，物理随机化处理摩擦、重力等；
- f) 数据处理与存储：采集的数据宜经过适当的预处理，包括去噪、时间同步等步骤，随后以规定格式存储。文件名和目录结构须明确标识场景、时间和数据类型。

6.2.4 合成数据用途

仿真合成数据在具身智能系统中的应用场景主要包括以下几个方面：

- a) 数据泛化，合成数据通过物品泛化和场景泛化来模拟各种现实环境中的变化情况。通过引入不同的物体类型、材质、大小、位置以及各种场景属性（包括光照、天气、地形等），合成数据能够为模型提供更丰富的训练样本，确保模型在实际应用中具备更强的适应性和泛化能力。模型可以在虚拟环境中学会处理不同形状和尺寸的物体，从而在真实场景中也能表现良好；
- b) 极端情况（Corner Case）生成，解决现实世界中极端情况（包括罕见的物体交互、复杂的光照或极端天气等）的采集难题。仿真平台能够生成大量极端情况数据，这些情况在真实环境中难以捕捉，但却是模型泛化和鲁棒性的重要测试场景。通过系统地生成这些难以采集的极端案例，模型能够更好地处理特殊或意外的情境，从而在多变和复杂的环境中表现得更加可靠；
- c) 跨传感器/本体数据迁移，仿真合成数据的另一个重要用途是跨传感器和跨本体的数据迁移。通过充分利用已有数据，模型可以在不同的传感器数据之间进行迁移学习，避免重复采集数据。通过在仿真环境中生成的RGB图像、深度图像、IMU 数据等，模型能够学会在不同传感器输入下进行任务操作。这种迁移机制减少了实际传感器配置下的数据采集成本，并且通过多模态数据的融合提升了模型的性能。

6.3 采集数据提交

6.3.1 数据文件标识

数据（资源）文件应通过文件名称来进行标识：

- a) 文件名称为“文件名+文件扩展名”；
- b) 命名不宜含有中文字符和非法字符等；
- c) 在后续使用过程中，不能对数据集进行重命名，否则会造成数据无法回溯，导致数据丢失。

6.3.2 数据提交方式

数据资源提交方式应符合T/SAIAS 015—2024中的数据资源提交方式。在实体存储介质（如硬盘）、云盘传输、直连在线、数据空间这四种数据资源提交方式中选定一种，来实施具体的数字资源交付。

6.3.2.1 实体存储介质

实体存储介质方式是指将数据资源文件按一定的格式和组织形式（如压缩）存入实体存储介质后进行的数据交换方式。

6.3.2.2 云盘传输

云盘传输方式是指将数据资源文件按一定的格式和组织形式（如压缩）后通过公有或私有云盘转储所实施的数据交换方式。

6.3.2.3 直连在线

直连在线方式是指数据资源供给和接收双方通过光纤专线点对点进行数据传输。这一方式具有较高的安全性和可靠性。若选择该传输方式，所需带宽宜按以下公式（1）进行计算：

$$B = \frac{8 \times 3600 \times S}{t} \dots\dots\dots (1)$$

式中，B—带宽，单位为比特每秒（bit/s）；

S—数据文件大小，单位为兆字节（MB）；

t—拟完成传输的时间，单位为秒（s）。

注：本文件对采用“直连在线”方式所涉及的数据加密方式不作具体规定。

6.3.2.4 数据空间

数据空间是互相信任的合作伙伴之间的数据关系，每一方都对其数据的存储和共享适用相同的规则。在数据空间中，数据不是集中存储，而是存储数据来源，因此只有在必要时才会（通过语义互操作性）共享数据。

数据空间是数据资产化的有效工具体系。

6.4 数据采集场景与任务体系

6.4.1 场景

数据采集场景指在真实物理环境中，通过多模态传感器与实体对象交互，并记录数据的动态过程。如工业领域的码垛、焊接等。

6.4.2 任务

数据采集任务指基于场景目标的闭环连续动作序列，需满足动作连贯性和环境适应性。如家居领域保洁场景的垃圾清运、衣物叠放等。

6.4.3 原子技能

数据采集原子技能指不可再分解的基础能力模块，需具备可复用性和独立评估性。如抓取、放置、旋转、按压等。

6.5 数据采集操作规范

6.5.1 采集环境要求

6.5.1.1 物理环境规范

采集区域宜具有合理的空间布局，确保有足够的空间设备操作，避免相互干扰，同时方便操作人员进行设备的操作和维护；宜为设备提供专门放置区域，不同构型的采集设备分区放置。宜保证光照均匀、稳定，避免出现频闪、阴影、反光等影响采集质量的情况。宜保持采集环境的清洁，定期进行清洁打扫，防止灰尘、毛发等杂物进入采集设备。采集作业宜在0℃至45℃的温度范围，30%至80%的湿度范围内进行。

6.5.1.2 安全要求

对采集作业区域的安全性进行评估，确保无明显的安全隐患，如陡坡、深坑、高压线等，保障人员和设备的安全。

确保设备在采集过程中的安全性，避免设备碰水、受潮、受损等情况的发生。

保证采集人员的安全，进行相关安全意识与防范培训；根据场景需要提供相关防护装备，如头盔、护目镜、手套等，确保其在作业过程中的安全。

6.5.2 采集设备选型

6.5.2.1 遥操作式数据采集设备选型

遥操作终端宜使用VR手柄、动捕服、臂控臂等，要求可以将人类动作准确映射到机器人端，完成遥操作。设备需要支持有线连接，避免无线通信的信号串扰。有线连接状态下，遥操作的端到端延迟不高于200ms。设备如有无线连接模式，宜支持WiFi 6以上或蓝牙连接。遥操作式数据采集设备供参考的性能及技术参数请参见附录B。

6.5.2.2 “生产伴随”式数据采集设备选型

根据适配的目标机器人，生产同构型的“生产伴随式”数据采集设备。设备应能采集视觉、机械臂数据等多种模态数据。生产伴随式数据采集设备供参考的性能及技术参数请参见附录C。

6.5.3 采集人员培训

6.5.3.1 人员筛选规范

数据采集人员宜严谨细致，责任心强，具有良好学习能力。采集人员的身高和体型应与所使用的设备相匹配，确保设备穿戴舒适且不影响操作灵活性。

对于数据采集人员，宜能够根据任务的要求操作采集设备，包括设备的穿戴、操作及动作的准确执行。对于运维人员，宜熟练操作各类设备及相关软件，了解其基本原理和参数设置，能应对常见的设备故障和软件问题。

6.5.3.2 基础培训

宜全面讲解数据采集作业全流程操作规范，确保学员熟悉整个工作流程。针对工作中使用的数据采集设备和软件，进行详细的操作培训。

6.5.3.3 专业技能提升

对于涉及多模态数据采集的工作，宜开展专项培训，讲解不同模态数据采集的特点和技巧。

介绍数据安全和隐私保护的相关法律法规和行业标准，讲解数据采集过程中如何保障数据的安全传输、存储和使用，防止数据泄露和滥用。

7 具身智能语料生产

7.1 采集数据完整性检查

在用于语料生产的所有资源数据存入前，宜按以下过程检查其格式的统一性和数据的完整性：

步骤一：所有资源数据以 5.4.2 中所规定各数据表征模式的文件格式之一的形式存在。如承载资源数据的文件不是对应数据表征模式的规定格式之一的，需采取合适的措施实现资源数据文件格式的转换；

步骤二：验证每个文件的完整性，可以使用对应的解码工具逐一打开文件，确认成功打开后再进行下一步操作；否则，在该文件元数据信息中标记该文件为损坏文件；

步骤三：将资源数据以对象方式存入，并把步骤二中的文件元数据信息存入数据库。

注：资源数据的统一性和完整性检查的重要性在于，能避免因数据解码失败而导致训练崩溃现象的出现。

7.2 数据资源清洗过程

7.2.1 主要流程

具身智能数据资源清洗过程的主要流程包括数据规范管理、数据整理分类、数据脱敏、数据去噪、数据去重、数据存储与备份等。

具身智能语料数据资源包括现实世界采集的真实数据、合成的虚拟数据，以及二者相结合的虚实数据。

7.2.2 数据规范管理

数据规范管理操作宜包含以下步骤：

步骤一：统一命名，所有的资源数据需根据 5.4.2 中所规定的文件标识，进行统一命名；

步骤二：统一格式，所有的资源数据需以 5.4.2 中所规定各数据表征模式的文件格式之一的形式存在。

7.2.3 数据整理分类

数据整理分类主要是对所有的资源数据以 5.4.2 中所规定各数据表征模式进行整理和分类。

7.2.4 数据脱密脱敏

数据脱敏主要是对所有数据资源中的视频、文本、语音数据进行脱密脱敏处理，脱敏过程应符合 YD/T 6225-2024 的相关规定要求，宜包含以下步骤：

步骤一：将对应数据、数据帧发到 NSFW 模型中，模型会返回 0 或 1：

0=NOT EQUAL NSFW

1=NSFW

步骤二：将返回值记录至对应数据帧文本中，若标记为 0 则通过，若标记为 1 则进入人工复核。

7.2.5 数据去噪

数据去噪主要是对所有数据资源进行处理，去除无法使用的数据，包括文件损坏、图像不完整、畸变过大、超出传感器有效值范围且无法修复、数据不符合物理规律与业务需求等的的数据。

7.2.6 数据去重

数据去重主要是对数据进行查重处理，将重复的数据删除。

7.2.7 数据存储与备份

数据存储与备份主要是数据整理完毕后，及时对数据进行存储与备份，遵循一式多份且异地多备。

7.3 数据资源标注过程

7.3.1 基本要求

数据资源标注过程应符合GB/T 42755-2023的规定。

7.3.2 标注任务制定

任务制定主要是制定具身智能数据资源标注，包含任务解决的问题，重点标注的属性信息，以及标注的注意事项等。其中，任务标注是指根据主任务的要求，如目标识别、动作分析等，对数据进行整体标注设计。子任务标注是为实现主任务所分解的具体标注环节，通过多级标注方法的协同，实现从数据标记到模型功能的语义映射。

7.3.3 数据标注检查流程

在具身智能使用数据资源进行标注时，宜按照以下过程检查数据是否遵循基本原则：

步骤一：所有数据资源以时间同步和空间同步进行数据同步，不同传感器（包括摄像头、激光雷达、IMU等）的数据在时间上是同步的，在空间坐标系中是对齐的；

步骤二：目标框宜以矩形框表示；

步骤三：标注类别的分类体系统一；

步骤四：标注属性根据不同的标注类别进行定义，包括物品的颜色等。

7.3.4 定义坐标系

具身智能中坐标系主要定义为四种，分别为本体坐标系、相机坐标系、局部坐标系和传感器坐标系，遵循右手定则，相关定义和作用如表1所示：

表 1 具身智能训练中定义的坐标系类型

序号	坐标系类型	原点位置	X轴方向	Y轴方向	Z轴方向	作用
1	本体坐标系	本体后轴中心或前轴中心	本体前方	本体左侧	本体垂直向上	主要用于获取本体运动状态、相对位置
2	相机坐标系	图像中心或相机镜头中心	图像右侧	图像下方	光轴前向	主要用于2D图像处理、目标检测
3	局部坐标系	环境的固定参考点	设备正前方向	设备左侧方向	设备垂直向上	主要用于全局定位和路径规划，实体在环境中的绝对位置参考
4	传感器坐标系	传感器物理中心	传感器正前方向	传感器左侧方向	传感器垂直方向	主要用于数据采集与标定，包括摄像头、激光雷达、力觉等多模态数据的时空对齐

7.3.5 标注方法

7.3.5.1 关键点标注

关键点标注是指将需要标注的元素按照需求位置进行点位标识，从而实现关键点的识别，并由若干关键点的集合形成具体应用所需的语义功能标识。

7.3.5.2 标注框标注

标注框标注是一种对目标对象进行标注的简单处理方式。在标注框标注的帮助下，模型通过训练，能够检测或识别出所需的目标对象。

框标注包括矩形标注、自由矩形标注、3D框标注，四边形标注和不规则框标注等。

示例：在具身智能训练中，通过标注框将物品的位置确定下来，再进行后续的动作操作。

7.3.5.3 区域标注

区域标注指将图像分成各具特性的区域并提取出感兴趣部分的过程。区域标注包括开区域标注和闭区间标注。

区域标注同时满足均匀性和连通性的条件，其中均匀性指的是该区域中的所有像素点都满足灰度、纹理、彩色等特征的某种相似性准则；连通性是指在该区域内存在链接任意两点的路径。

与矩形框标注相比，其区域标注更加精确，标注边缘可以是多边形甚至是柔性的。

7.3.5.4 属性标注

属性标注就是用一个或多个标签标注目标对象的属性。

注：对具体应用而言，可在标注文件的元数据部分来申明所用的这些标签，这些标签必须是一个封闭的集合。

7.3.5.5 关键帧标注

关键帧标注就是对视频、点云等数据中具有关键动作意义的帧进行属性、区域等内容的文本标注。

7.4 数据质量评估

7.4.1 基本要求

具身智能数据质量评估应符合 ISO/IEC 5259-4:2024 的规定。

7.4.2 数据准确性

数据准确性指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。结合应用场景，设置数据集准确性的评价内容与要求、等级划分与判定依据，评价内容包括但不限于：

- a) 视觉、雷达、触觉、力觉等状态数据的内容正确性与数据格式（如类型、数值范围、数据长度、精度等）的合规性；
- b) 视觉、雷达、触觉、力觉等数据重复率；
- c) 视觉、雷达、触觉、力觉等数据唯一性；
- d) 脏数据的出现率。

7.4.3 数据一致性

数据一致性指数据与其他特定使用的数据之间无矛盾的程度。结合应用场景，设置数据集一致性的评价内容与要求、等级划分与判定依据，评价内容包括但不限于：

- a) 视觉、雷达、触觉、力觉等状态数据的内容正确性与数据格式（如类型、数值范围、数据长度、精度等）的合规性；
- b) 相同数据一致性：同一数据在不同位置存储，或别不同应用或用户使用时的数据一致。数据发生变化时，存储在不同位置的同一数据被同步修改；
- c) 关联数据一致性：根据一致性约束规则，检查关联数据的一致性，数据存在会通过外键、引用或其他关联关系连接在一起，当数据发生变更时，这些关联关系中的数据也要保持一致；
- d) 多模态数据的时间戳一致。

7.4.4 数据适用性

数据适用性指数据符合物理规律与业务需求，能够支撑模型生成可执行的策略。具身智能需确保数据真实反映物理交互约束与任务目标，满足场景使用需求。结合应用场景，设置数据集适用性的评价内容与要求、等级划分与判定依据，评价内容包括但不限于：

- a) 物理规律合理性：数据参数是否满足真实物理规律；
- b) 场景适配性：根据任务指标，数据是否满足该场景定制的数据要求。

7.5 数据有效性评估

7.5.1 数据集选型

待测试的多来源、多场景的具身智能真机和仿真数据，主要包括遥操作采集真机数据、生产伴随式采集真机数据和仿真合成数据等方式。对于数据集的具体要求详见附录D。

7.5.2 模型选型

可选择国内外最新，且具有代表性的模型，基于行业Sota指标，以确保测试结果的全面性和代表性。

7.5.3 测试本体选型

选用行业公认的、具备完善硬件能力的机器人硬件本体。该机器人应能够准确执行模型输出的指令，实现各种复杂操作，为模型在真实场景中的测试提供可靠的硬件支持。

在测试前，对机器人进行全面的校准、调试和维护，确保其硬件性能处于最佳状态，避免因硬件问题影响测试结果的准确性。

7.5.4 测试任务设计

针对每个测试任务，根据任务的复杂性和难度，制定详细的测试用例，明确任务的操作对象、环境背景、初始状态、执行步骤、预期结果以及异常情况处理方式等。

7.5.5 测试指标设计

整体以具身任务完成度进行验证，包括端到端和模块化两种类型的方案评估，评测执行时间、泛化性、端到端任务成功率，或模块化方案中数据集对各模块的支持能力等。

7.5.6 模型测试执行

执行模型测试宜遵循以下步骤：

- a) 初始模型测试：将选定的初始模型（未经过待测具身智能数据集训练）部署到指定机器人硬件本体上，在搭建的各个场景中执行上述测试任务，并按照评测指标进行打分，记录初始模型在各项任务中的表现，作为后续对比的基准；
- b) 模型训练：使用待测具身智能数据集，按照选定的训练方法和参数设置，对选定模型进行训练；
- c) 训练后模型测试：将训练后的模型部署到同一机器人硬件本体上，再次在相同的场景中执行测试任务，并按照相同的评测指标进行打分；
- d) 多轮测试与交叉验证：为提高测试结果的可靠性和准确性，对每个具身智能数据集和模型的组合进行多轮测试，每轮测试后重新训练模型并进行评估。同时，采用交叉验证的方法，在不同的场景和任务组合下进行测试，确保测试结果不受特定场景或任务的影响。

8 数据安全

8.1 安全性评价

数据采集、处理、测试和提供应进行安全性评价，包括但不限于：

- a) 涉及个人信息的，应符合GB/T 45574-2025、GB/T 44588-2024、GB/T 35273-2020、YD/T 4245-2023的规定；
- b) 数据处理过程应符合GB/T 45674-2025、GB/T 35274-2023、GB/T 41479-2022的规定。

8.2 安全管理流程

数据采集、处理、测试和提供建立相关流程，宜包括但不限于：

- a) 数据管理：覆盖上述语料生产的全过程，针对数据准备、数据处理、数据交付等不同阶段的数据存储空间和访问控制，并载明不定期组织内部审查的方案；
- b) 访问控制及权限管理：实施物理的身份验证和授权机制，建立监控和审计机制。

8.3 安全技术

数据采集、存储、处理、测试和提供的软硬件，宜包括但不限于：

- a) 数据网关；
- b) 数据加密与备份装置；
- c) 安全防范和监控系统；
- d) 私域数据存储与传输系统；
- e) 数据资产管理工具，如数据库、数据湖等。

8.4 人员管理

参与数据采集、处理、测试和提供的人员应签署与其职责相对应的保密协议。

附录 A
(资料性)
具身智能语料库数据资源类型

A.1 视频数据

具身智能视频类型如表A.1所示。

表 A.1 视频数据的指标

序号	指标	说明
1	视频类型	见表4
2	数据资源内容	视频及对应文字说明或视频介绍，宜包括拍摄地点、时间、设备、照明条件、天气条件、环境条件、任务描述、对应的开始和结束帧数等信息
3	视频分辨率	宜1080p（1920x1080像素）及以上
4	视频帧率	宜25帧/s及以上
5	单一视频时间长度	宜10s及以上
6	视频文件格式	宜mp4/avi/yuv/h264等
7	视频验收标准	除清晰外，对视频素材的随机抽样中，宜有80%包含主体（不含主体的视频素材示例，包括但不限于航拍、延时风景摄影等）

视频数据宜来自本体头部、胸口、左右手臂末端等采集设备本体部位，以及第三方视角摄像头的视频序列，用于物体检测、环境检测、场景设施识别、语义分割、视觉导航等。摄像头视频数据分类及说明如表A.2所示。

表 A.2 视频数据分类表

序号	视频数据来源	功能	位置	说明
1	采集设备本体	检测识别	头部视频	主要指头部摄像头采集的视频数据
2			胸口视频	主要指胸口摄像头采集的视频数据
3			左右手臂末端视频	主要指左右手臂末端摄像头采集的操作目标近景视频数据
4		视觉感知	肩背部视频	主要指设备肩背部摄像头采集的视频数据
5			腰部视频	主要指设备腰部摄像头采集的视频数据
7			底座视频	主要指设备底座摄像头采集的视频数据
8			双足视频	主要指双足摄像头采集的视频数据
9		第三方视角	检测识别	红外视频
10	立体视频			主要指采用两个或以上的摄像头从不同角度拍摄，用于深度感知和3D建模

视频数据还可能包含天气条件、照明条件和场景类型等数据语义特征，具身智能数据语义特征如表A.3所示。

表 A.3 视频数据语义特征表

序号	类型	说明
1	天气	主要包括晴朗、多云、阴天、雨天、雾天、雪天、沙尘暴、雷电
2	时间	主要包括白天、夜晚、黎明、黄昏
3	光照	主要包括光线较好、光线较弱、黑暗、过曝/眩光、大面积阴影、反射虚影
4	场景	主要包括工厂、家居、办公室、医院、学校、商超、开放道路等

A.2 雷达数据

激光点云数据的指标如表A.4所示。

表 A.4 激光点云数据的指标

序号	指标	说明
1	精度	在±2cm以内
2	点云密度	宜每m ² 90个点及以上, 均匀分布
3	检测范围	宜30m及以上
4	帧率	宜10Hz及以上
5	动态范围	宜10cm-100m
6	数据资源说明	点云数据及对应文字说明或介绍
7	噪声水平	宜距离误差±3cm以内, 角度误差±1° 以内, 时间同步误差±10ms以内
8	单一点云数据大小	宜50MB及以上
9	文件格式	pcd/las/xyz等
10	验收标准	除清晰外, 对素材的随机抽样中, 点数足够, 分布均匀, 且精度误差小, 噪声水平低

相关激光点云数据宜满足以上指标, 若出现以下情形, 则不适合作为激光点云数据, 不适合作为激光点云数据如表A.5所示。

表 A.5 不适合作为点云数据资源的情形

序号	情形	说明
1	数据不完整	主要指缺少关键区域的点, 包括目标物体、环境设备
2	分辨率过低	主要是指点云过于稀疏, 无法捕捉物体细节
3	精度不高	主要是指点的位置信息误差大
4	噪声过多或同步性差	主要指包含大量错误、异常值或者多传感器数据无法准确同步
5	动态范围小	主要指无法覆盖远近距离的物体
6	反射强度不一致	主要指反射强度信息不稳定或不可靠
7	光照影响大	主要指强光或逆光条件下产生阴影或反射
8	天气影响严重	主要指雨、雪、雾等恶劣天气条件下的数据
9	遮挡严重	主要指其他物体遮挡无法获取完整目标物体视角
10	数据过时	主要指与当前环境变化较大, 无法反映当前场景
11	法律和隐私问题	主要是指包含敏感信息, 包括车牌号、人脸等
12	冗余数据	主要指大量重复或非常相似的数据

毫米波数据资源的指标如表A.6所示。

表 A.6 毫米波数据的指标

序号	指标	说明
1	频率范围	宜76GHz及以上
2	检测范围	宜30m及以上
3	角分辨率	±0.5° (长距离) / ±1° (中距离)
4	距离分辨率	宜0.1m及以上
5	速度分辨率	宜0.1m/s
6	更新率	宜20Hz及以上
7	调制方式	FMCW
8	检测精度	宜距离: ±0.05m, 角度: ±0.5°, 速度: ±0.1 m/s
9	文件格式	支持CAN/Ethernet等通信协议
10	验收标准	除清晰外, 对素材的随机抽样中, 分辨率达标, 宜有80%包含主体。

A.3 音频数据

音频数据资源（或可再加工音频数据产品）的指标如表A.7所示。

表 A.7 音频数据的指标

序号	指标	说明
1	语种	汉语（普通话、粤语、沪语等）、英语（英国、美国）、阿拉伯语、俄语、日语等
2	主题领域	参照《广播电视和网络视听节目内容标识标签规范》（GY/T 360-2022）中的“内容类内容特征子类别规范词”
3	数据资源内容	音频及对应文字说明或音频介绍
4	音频采样率	宜不小于44.1KHz
5	通道数	双声道/单声道（由原始资料特性决定）
6	单一音频时间长度	宜60s及以上
7	量化精度	宜不低于16位
8	音频文件格式	宜wav等
9	音频验收标准	对音频素材的随机抽样中，宜有95%不存在表10中所示情况

相关的音频技术指标宜符合GY/T 353-2021的规定。

音频数据资源中包含了资源特征性描述的，相关的标识和（或）标签则宜符合GY/T 360-2022的规定。

音频数据需满足以上指标，若出现以下情形，则不适合作为音频数据，如表A.8所示。

表 A.8 不适合作为音频数据资源的情形

序号	情形	说明
1	文件噪音不合格	整段音频伴有严重的噪音
2	文件内容不合格	音频文件不包含目标类别的音频或含有误导性信息的音频
3	其它不合格	回音很大、过载严重

A.4 本体数据

本体数据资源的指标如表A.9所示。

表 A.9 本体资源指标

序号	指标	说明
1	类别	详见表12
2	数据资源内容	本体数据及对应文字说明
3	格式	JSON或数据库
4	验收标准	本体数据需准确，包含时间数据，且与视频数据保持时空一致性

本体信息数据资源分类及说明如表A.10所示。

表 A.10 本体信息数据资源分类表

序号	类型	说明
1	状态数据	主要包括启动/关闭、传感器健康状态
2	控制命令	主要包括转向、加速、制动、运动
3	诊断信息	主要包括本体故障代码和诊断信息
4	动态数据	主要包括速度、加速度、转向角度、制动力度、横向加速度、纵向加速度、轮速、本体位置、倾角、本体各关节电机功率
5	末端设备类型	主要包括末端工具类型、信息及末端工具等

传感器标定数据资源指标如表A. 11所示。

表 A. 11 传感器标定数据指标

序号	指标	说明
1	常见信息	相机内参、相机外参、激光外参、毫米波外参、INS的内参、INS的外参等
2	参数规定	参数宜与对应本体一一绑定，且定期重新标定进行维护
3	数据资源内容	传感器标定数据及对应文字说明
4	格式	JSON或采集数据包中
5	验收标准	根据各传感器参数，通过采集数据进行投影，根据重叠效果判断标定准确度

A. 5 轨迹导航数据

轨迹导航数据宜包括定位数据、轨迹数据等。轨迹导航数据的指标如表A. 12所示。

表 A. 12 轨迹导航数据资源指标

序号	指标	说明
1	类别	详见表15
2	数据资源内容	轨迹导航数据及对应文字说明
3	格式	JSON或采集数据包
4	验收标准	定位精度符合传感器精度要求，包含时间数据，且与视频数据保持时空一致性

轨迹导航数据资源分类如表A. 13所示。

表 A. 13 轨迹导航数据资源分类

序号	类型		说明
1	定位数据	GPS	主要包括定位精度、定位连续性、时间戳同步
2		IMU	主要包括角速度零偏稳定性、加速度零偏稳定性、时间戳同步
3	轨迹数据	轨迹点	主要包括位置精度、速度精度、时间戳同步
4		轨迹预测	主要包括预测时间范围、预测频率、预测精度（包括minADE、minFDE）

附录 B

(资料性)

具身智能遥操作模式数据采集本体参考性能参数

B.1 性能参数

采集本体的感知相机数量具体宜根据实际采集需要进行加装。其组件结构，及相关臂展、自由度、负载等推荐参数详见表B.1:

表 B.1 遥操作数据采集本体推荐性能参数

序号	硬件组件	关键参数
1	单臂展	宜不小于600 mm
2	单臂自由度	宜不小于6个
3	单臂负载	单臂负载宜不低于3kg
4	感知相机	宜至少包含1个RGB-D相机
5	头部	宜支持左右旋转、俯仰
6	遥操作形式	宜支持VR手柄、动捕手套等
7	末端	宜支持配置夹爪或灵巧手
8	控制模式	宜支持双臂/单臂
9	存储系统	宜支持全量满带宽数据存储，每台配备外接硬盘接口

附录 C

(资料性)

“生产伴随”式数据采集设备参考性能参数

C.1 性能参数

图像采集帧率宜达到15帧/秒以上；如使用外骨骼，关节数据采集率宜在15Hz或更高。为确保数据实时传输和处理，宜采用高速USB接口或以太网接口，满足大量数据的快速传输需求。宜确保在无外部电源的情况下能够连续稳定工作。

C.2 可靠性与稳定性

设备宜具备良好的抗干扰能力和防护性能，适应不同环境条件。设备宜与采集系统的其他硬件和软件兼容，且易于维护和维修，有可更换的零部件和方便的升级接口。

C.3 适配性

C.3.1 机器人设备适配性

设备宜采用关节模块化设计，具备良好的扩展性，支持不同末端工具的接入，如二指开合机构、人手等。

C.3.2 采集场景适配性

需满足室内环境中使用，选择体积较小、便于安装的设备，如使用小型化的相机传感器；宜使用坚固耐用、适应恶劣环境的材质制造。

对于复杂场景，如多物体交互、多人协作场景，宜选择具有高分辨率、多视角、多模态融合能力的设备。

C.3.3 数据处理系统适配性

设备采集的数据格式需与数据处理系统支持的格式一致，数据量宜与数据处理系统的处理能力相匹配。采集设备上的多个相机传感器之间需能够实现同步和协同工作。

附录 D

(资料性)

具身智能语料库数据有效性评估数据集要求

D.1 数据集准备

待测试的多来源、多场景的具身智能真机和仿真数据，主要包括遥操作采集真机数据、生产伴随式采集真机数据和仿真合成数据等方式。

面向数据集的最小可训练数据定义，是一个连续动作单元数据，包含若干段自由运动（接近或者远离物体等），若干个数据动作（抓、放、插、按等技能动作），以及若干模态的数据（同一时间戳下的全局视觉数据、力觉、关节等，传感器帧率不少于15Hz-30Hz）。

面向训练数据集量级准备要求：短序任务中数据量级500-1000条，建议采集5-10小时；长序任务中数据量级1200-3600条，建议采集10-30小时，包含3-5个步骤组成的中等复杂任务，并区分训练数据与测试数据。

D.2 数据质量要求

企业在进行模型训练时，应完成采集数据预处理操作，满足质量保障。常见数据处理方案如下：

- a) 数据格式统一：宜支持HDF5、Lerobot、Episode等主流数据格式。需包括数据集的标签信息文件、标定参数文件、机器人硬件构型（单双臂、末端工具或者URDF）等文件；
- b) 数据采集要求：多模态传感器校准、坐标系选型统一；
- c) 数据清洗要求：去除重复数据、静止数据、黑绿图数据；
- d) 数据准确性要求：传感器精度，如硬件误差（传感器畸变、关节角度传感器偏差、传感器噪声）是否在可接受范围内；通常传感器精度验证指标有线性度误差（Linearity Error）、重复性误差（Repeatability Error）等。硬件误差中的关节角度传感器偏差，即关节编码器测量值与实际旋转角度的偏差，建议在 $\pm 0.5\text{mm}$ 到 $\pm 2\text{mm}$ ；
- e) 数据一致性要求：多模态对齐，如时间同步性（视觉、力觉、关节角度等传感器数据的时间戳、对齐坐标系统一）和语义一致性（不同模态数据对同一事件的描述是否冲突）；时间同步性误差的评估方法可通过时间戳对齐验证各传感器数据的同步性，计算最大时间偏差和平均偏差。建议最大时间偏差在20ms以内。语义一致性可通过自然语言指令与实际动作的逻辑匹配度评估。如BLEU分数、动作执行准确率等。BLEU的取值范围为0至1，得分越高表明语义一致性越高。
- f) 针对仿真数据，额外需要包括视觉真实性（图像质量是否清晰且接近真实世界的视觉效果；物体的纹理、材质和光照效果是否逼真，能否模拟真实世界的光影变化）、物理精准性（数据中物体的运动是否符合物理规律；本体与周围环境的交互是否与真实世界一致）的数据质量验证方案。视觉真实性评价指标通常使用结构相似性SSIM、峰值信噪比PSNR及视觉信息保真度VIF。其中SSIM越接近1表示越接近真实图像；PSNR则通过计算均方误差而后转为分贝值，数值越高则表示失真越小；VIF的数值越大，表明图像质量越好。物理精确性评价指标包含动力学误差（Dynamic Error）与轨迹对齐误差（Trajectory Alignment Error）等。

对于数据处理平台功能要求应包括：全流程覆盖数据采集、数据管理（切片、清洗、标注）、数据训练（预训练、精调训练、模型/镜像管理）、数据评测（数据质量评价、模型质量评价）、数据仿真软件（人工采集、算法合成）等。

参 考 文 献

- [1]GB/T 4894-2024 信息与文献 基础和术语
 - [2]SF/T 0153-2023 图片真实性鉴定技术规范
 - [3]GA/T 1717.3-2020 信息安全技术 网络安全事件通报预警 第3部分：数据分类编码与标记标签 技术体系技术规范
 - [4]YD/T 3470-2019 面向公有云服务的文件数据安全标记规范
 - [5]中国分类主题词表（第二版）
 - [6]中国图书馆分类法
 - [7]中国档案分类法
-