

团体标准《人形机器人数据集数据标注》（征求意见稿）

编制说明

一、工作简况

（一）任务来源

根据《上海市人工智能行业协会关于<人形机器人平均无故障时间测试与评定>等 10 项团体标准立项通知》（上智协标〔2024〕15 号），由上海是人工智能行业协会归口的团体标准《人形机器人数据集数据标准》获批立项。

（二）主要参与起草单位

本文件的主要参与起草单位：上海机器人产业技术研究院有限公司、人形就机器人（上海）有限公司、中国科学院上海微系统与信息技术研究所、启元实验室、深圳市优必选科技股份有限公司、上海库帕思科技有限公司、上汽通用动力科技有限公司.....

（三）主要工作过程与主要起草人所做工作

本标准编制过程到目前主要经历了标准工作组成立、标准调研、立项阶段和标准研制四个阶段。各阶段主要工作总结如下：

1. 标准工作组成立

2024年11月25日，为保证标准编写质量，面向行业征集标准起草单位，成立了标准起草工作组，吸纳了包括人形机器人制造商、用户、科研院所、检测认证机构的专家，共同开展标准研制工作。

2. 标准调研

调研国内外开源机器人数据集现状，梳理存在的问题：通用性方面，现有数据集仅针对特定型号机器人，不同型号的机器人数据无法打通形成合力；可用性方面，不同主体产生的数据集在数据类型、数据格式等方面参差不齐；规范性方面，在数据采集、预处理、标注等环节缺少统一的管理与规定，数据集质量测评基准缺失，数据集安全无法保障。

调研国内外人形机器人数据集标准现状，目前尚未存在人形机器人数据集管理类标准。在国际标准组织，如 ISO、IEC等发布的众多标准中，虽涵盖了机器

人领域的部分内容，但多聚焦于机器人的硬件安全、通信协议、性能测试等方面，未深入到数据集管理这一关键环节。

调研人形机器人数据集建设与应用需求。在建设需求方面，亟需构建标准化、应用场景任务需求多样化的人形机器人数据集，建立标准化的数据采集、预处理、标注等流程。在应用需求方面，要求数据集具备通用性，能够无缝适配不同型号、不同厂家的人形机器人。同时保障数据集的安全使用。此外，还需要对数据集的质量进行准确评估，实现数据集优化。

3. 立项阶段

2024年11月12日，上海市人工智能行业协会发布《上海市人工智能行业协会关于〈人形机器人平均无故障时间测试与评定〉等10项团体标准立项通知》（上智协标〔2024〕15号）。批准同意《人形机器人数据集数据标注》团体标准立项。

4. 标准研制

（1）2024年11月12日，为保证标准编写质量，面向行业征集标准起草单位，成立了标准起草工作组，吸纳了包括人形机器人制造商、用户、科研院所、检测认证机构的专家，共同开展标准研制工作。

（2）2024年12月~2025年1月，标准起草工作组基于前期调研结果，形成标准草案初稿；

（3）2025年1月15日，标准起草工作组召开了第一次工作组会议。讨论确定了标准草案的范围、框架及分工。会上提出明确标准相关方责任划分、针对数据匹配标注关键要求、明确标注工具选取原则、修改数据标注准确性要求、补充标注流程图、标注工具质量控制要求、轨迹、嗅觉等数据标注结果要求。

（4）2025年1月~4月，标准起草工作组共同对标准各章节内容进行了补充完善，整合修改后形成了第二次会议讨论稿。

（5）2025年4月，标准起草工作组召开了第二次工作组会议。会上完善标注任务实施要求，新增常用标注工具示例内容。

（6）2025年4月，标准起草工作组根据会议上讨论确定的修改意见，对标准进行了进一步地完善及全文复核，形成了标准征求意见稿。

二、标准编制原则和确定主要内容的论据及解决的主要问题

（一）原则

本文件编制遵循“统一性、适用性、一致性、规范性”的原则，注重标准的可操作性。本文件的编写符合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》的编制要求。

标准内容充分考虑了人形机器人数据集的特点与相关方应用需求，制定了切实可行、合理有效的人形机器人数据集数据标注要求。

（二）确定主要内容的论据

本文件规定了人形机器人数据集数据标注任务说明文档、数据标注质量特性、标注与质控流程以及标注结果交付与验证要求。

本文件适用于人形机器人数据集数据标注活动。

主要技术内容如下：

——第4章 标注任务说明文件：包括标注任务分类、标注任务描述。提出标注人员、标注工具、标注、输入输出数据要求。该章主要技术内容基于目前人形机器人数据现状，并与相关应用方、第三方讨论确认。

——第5章 数据标注质量特性：包括准确性、一致性、时效性、可访问性、可移植性、保密性、可追溯性。该章的技术内容为参考数据标注行业标准，结合机器人企业建议，讨论确认。

——第6章 标注与质控流程：包括标注流程、标注任务创建、分配、实施、标注结果质量控制与评价、安全管理。该章主要技术内容结合现有人形机器人数据集标注需求，并与相关应用方、第三方讨论确认。

——第7章 标注结果交付与验收：交付内容、典型数据标注结果。该章主要技术内容为总结现有头部人形机器人企业技术经验。

——附录A（资料性）“标注示例”，给出二维图像的实例分割标注任务示例；

——附录B（资料性）“常用标注工具示例”，给出数据标注常用工具建议。

（三）解决的主要问题

通过本文件的制定，旨在解决人形机器人数据集的标注过程中，不同的标注人员因理解差异导致标注不一致以及效率低下、标注工具如何选用、标注任务的质量控制缺失的问题。

人形机器人数据集数据标注标准的制定，能够规范数据标注全流程，确保所有标注人员在标注过程中遵循统一的标准，从而保证标注结果的一致性和可靠性。

三、主要试验情况分析

无。

四、知识产权情况说明

本文件中没有涉及专利和相关知识产权问题。

五、产业化情况、推广应用论证和预期达到的经济效果

本标准的实施将填补国内外人形机器人数据集数据标注标准空白，为人形机器人数据标注提供统一的人员、工具、环境、标注规则、标注结果要求，提高标注效率，降低企业的人力和时间成本，减少因标注不规范导致的数据质量问题，提升大模型的性能，支持人形机器人数据标注细分市场健康发展。

六、转化国际标准和国外先进标准情况

无。

七、与现行相关法律、法规、规章及相关标准的协调性

国内外没有发布关于人形机器人数据集数据标注的相关标准。

国外标准情况：ISO/IEC DIS 5259-4该标准定义的数据质量过程框架包括数据准备、数据处理、数据评估、数据优化等，仅规范了数据标注质量的评估方法和指标。

我国发布的数据标注标准，包括：

1) GB/T 42755-2023 《人工智能 面向机器学习的数据标注规程》规定了人工智能领域面向机器学习的数据标注框架流程，适用于指导人工智能领域面向机器学习的数据标注以及与之相关的研究、开发和应用等。本标准侧重数据标注的流程管理，但没有对具体数据标注要求进行规定；

2) YY/T 1833.3-2022 《人工智能医疗器械 质量要求和评价 第3部分：数据标注通用要求》提出了人工智能医疗器械领域的数据标注通用要求和评价方法，该标准对象为医疗器械，是医疗场景下的图像分割典型应用，可提供参考价值。

山西省地方标准也对人工智能数据标注进行规定，包括DB14/T 2463—2022《人工智能 数据标注总体框架》、DB14/T 2464—2022、《人工智能 数据标注

一般技术要求》、DB14/T 2465—2022 《人工智能 数据标注通用工作规程》，但内容主要为数据标注生产的管理要求，对数据标准的技术流程进行描述，内容较简略，不具备参考价值。黑龙江发布DB23/T 3857-2024《人工智能数据标注总体框架规范》、DB23/T 3479-2023《人工智能数据标注人员培训服务规范》对人工智能数据标注任务总体框架、标注从业人员服务提供要求。可为本标准撰写提供框架参考。

八、重大分歧意见的处理经过和依据

无。

九、标准性质的建议

本标准批准后作为推荐性团体标准使用。

十、贯彻标准的要求和措施建议

建议该标准发布即实施，发布后，建议项目提出单位组织开展标准宣贯，出台相关文件推动该标准的落地应用。

十一、替代或废止现行相关标准的建议

无。

十二、其它应予说明的事项

无。

《人形机器人数据集数据标注》团体标准编制起草组

2025-04-18