

团体标准《人形机器人操作任务数据集数据格式要求》

（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《人形机器人操作任务数据集数据格式要求》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、人形机器人（上海）有限公司、北京小米机器人技术有限公司、中国科学院上海微系统与信息技术研究所、深圳市优必选科技股份有限公司、上汽通用动力科技有限公司、上海库帕思科技有限公司、五八智能科技(杭州)有限公司、同济大学、上海开普勒人形机器人公司、广东中创智家科学仪器有限公司、上海识度数字科技有限公司、威凯检测技术有限公司、节卡机器人股份有限公司、北京他山科技有限公司、诺亦腾机器人科技(北京)有限公司、比亚迪.....

（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，未深入到数据集建设与管理这一关键环节。

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《人形机器人操作任务数据集数据格式要求》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2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2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。会上提出增加数据集描述要求及示例、增加元数据格式单位、时间戳数据格式要求等内容。

（4）2025年2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25日，标准起草工作组召开了第二次工作组会议。会议上完善主数据定义、元数据格式要求。

(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。

统一性原则：确保数据集相关术语、数据格式等内容保持统一，避免不必要的重复或差异。

适用性原则：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，制定了合理有效的人形机器人操作任务数据集数据格式要求，指导数据集建设工作。

一致性原则：标准内容全文保一致，与人形机器人数据集相关其他标准协调一致。

规范性原则：遵循标准化文件的起草规范，提高文件的可操作性和可读性。

(二) 确定主要内容的论据

本文件规定了人形机器人操作任务数据集的一般要求、数据集描述、总体框架、数据格式要求。

本文件适用于人形机器人操作任务数据集。

主要技术内容如下：

——第5章 一般要求：给出人形机器人操作任务数据集基本要求，说明数据集主数据以及构成组件；

——第6章 数据集描述：包括整体描述文件与片段描述文件。该章的技术内容为总结现有头部人形机器人企业技术经验，并与相关应用方、第三方讨论确认；

——第7章 总体框架：给出人形机器人操作任务数据集总体框架。该章主要技术内容结合相关企业建议讨论确认；

——第8章 数据格式要求：包括元数据格式、本体坐标系设置、数据集存储格式、数据集共享格式要求。该章主要技术内容为结合现有人形机器人操作任务数据集数据格式现状，并与相关应用方、第三方讨论确认；

——附录A（资料性）“数据集描述文件示例”，给出整体描述文件与片段描述文件的示例；

——附录B（资料性）“数据存储格式示例”，给出数据存储格式建议；

——附录C（资料性）“数据集转换RLDS规范示例”，给出数据共享格式建议。

（三）解决的主要问题

通过本文件的制定，旨在解决不同主体产生的数据集在数据类型、数据格式方面不一致、数据结构不统一，导致数据集在使用前需要进行额外的格式转换和校验，增加了开发和调试的成本，难以有效进行数据集质量控制的问题。

人形机器人操作任务数据集数据格式要求标准的制定，能够统一数据类型和格式，降低数据处理成本，保障数据的一致性与完整性，打通数据孤岛困局，促进人形机器人数据集共享与交流。

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

目前，人形机器人行业正处于快速发展阶段，国内外众多科研机构和企业都在积极投入相关研发工作。本标准的研制与实施将填补标准空白，为数据集的采集、存储、共享提供数据格式规范，数据集供应商、机器人制造商、算法研发企业等可基于该标准进行数据集开发和优化，降低数据集处理成本，提高数据集利用效率，加速人形机器人产业的成熟和壮大。

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操作任务数据集数据格式的相关标准。

我国关于数据集格式标准针对卫生健康、气象等领域，具有明显的行业特殊性，不适用于人形机器人，包括但不限于：

1)GB/Z 40104.103-2023：规定了太阳能光热发电站设计采用的气象数据集的数据格式。

2)WS/T 305-2009：规定了卫生信息数据集元数据内容框架、卫生信息数据集核心元数据、卫生信息数据集参考元数据、引用信息与代码表。

3)GB/T 39662—2020：规定了证券投资基金注册登记机构向中央数据交换平台报送每日业务数据时所采用的数据交换格式与数据接口。

4)DB35/T 1777-2018《政务数据汇聚 数据集的规范化描述》：内容针对政务数据集的内容与具体格式要求，不适用于人形机器人操作数据集；

5) 20240869-T-469 《信息技术 脑机接口 多模态数据格式》：主要为脑机接口多模态数据，不适用于人形机器人。

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。

《人形机器人操作任务数据集数据格式要求》团体标准编制起草组

2025-04-28