

《人形机器人应用环境数据采集规范》

（送审稿）编制说明

一、项目背景

近年来，国家工信部、发改委等部委密集发布机器人相关的政策性文件，其中包括支持机器人技术的研发和应用，支持机器人行业发展的投资和融资政策，以及大力发展机器人新兴产业的政策措施等。如 2024 年 11 月，《5G 规模化应用“扬帆”行动升级方案》中，提出推动基于 5G 的智能机器人、智能移动终端、云设备等研发应用，推进 5G 与智能农机深度融合，提升基于 5G 的农业传感器、控制器、机器人、无人机等智能化装备研发生产水平。推动人形机器人产业向高端化、智能化方向发展。2024 年 1 月，工信部等七部门联合出台了《关于推动未来产业创新发展的实施意见》，重点指出打造标志性产品要求做强未来高端装备，人形机器人要求突破高转矩密度伺服电机、高动态运动规划与控制、仿生感知与认知、智能灵巧手、电子皮肤等核心技术，重点推进智能制造、家庭服务、特殊环境作业等领域产品的研制及应用。可以看出，以人形机器人为代表的智能机器人行业也已经成为深圳市、广东省乃至全国的重点发展领域和重点工作任务。

人形机器人作为具有类人形态、功能的智能机器人，与传统机器人相比具有更强的适应性和通用性。人形机器人是国际公认的机器人技术竞争制高点，国内外的头部科技公司，如特斯拉、亚马逊和小米等已经开始研发产品。预计到 2028 年，全球人形机器人在服务机器人中的普及率将达到 3.5%，市场规模超过 20 亿美元。未来人形机器人将成为重大产业之一，具有广阔的市场和巨大的经济和社会效应。然而，目前还面临着通用智能难、研究支撑平台和数据缺失、软硬件要求高、系统集成难等多方面挑战。

国内缺乏具身智能研发过程中所需数据采集相关的标准，各行业环境/场景数据没有互联互通，因此存在数据可利用率低、更新效率慢等问题。如何聚焦重点、整合优势、补足短板、规划布局，加快人形机器人技术和产业发展、带动整体生态发展成为重要课题。人形机器人应用环境数据采集标准的制定将促进人形

机器人的系统研发，推动整体规划布局走向产品成熟化、产业生态化，赋能机器人行业健康有序地发展。

二、工作简况

1. 任务来源

本标准由深圳市市场监督管理局于 2024 年 7 月正式批准立项，由深圳市人工智能与机器人研究院提出，腾讯科技股份有限公司、深圳国创具身智能机器人有限公司、松应科技（深圳）有限公司、河海大学人工智能与自动化学院、工业和信息化部电子第五研究所、诺亦腾（北京）科技有限公司、广东电网有限责任公司、华为云计算技术有限公司、深圳赛西信息技术有限公司、深圳兆威机电股份有限公司、帕西尼感知科技（深圳）有限公司、深圳市地铁集团有限公司负责起草。

2. 主要工作过程

（1）预研阶段

项目预研阶段，深圳市人工智能与机器人研究院组织成立了标准编制组，落实了标准主要内容、涉及范围、基本原则、时间节点等工作。通过收集、整理相关技术资料，对机器人数据采集、数据处理的国内外相关现状展开调查研究并梳理问题，为标准制定提供技术参考和支撑，并编制标准立项申报材料，申请立项。

（2）立项阶段

2024 年 3 月 6 日，市市场监管局、市工业和信息化局组织专家对《人形机器人应用环境数据采集规范》深圳市地方标准进行立项评审，经专家组评审，该地方标准符合立项条件，批准立项。

（3）起草阶段

2024 年 12 月至 2025 年 6 月，标准编制小组快速推进项目研究、加快项目执行进度，为实现预期目标，由深圳市人工智能与机器人研究院组织，采用线下、线上结合方式召开《人形机器人应用环境数据采集规范》标准启动会及多次标准讨论会，确定标准草案分工，共同对各章节内容进行补充完善。腾讯科技股份有限公司、深圳国创具身智能机器人有限公司、松应科技（深圳）有限公司、河海大学人工智能与自动化学院、工业和信息化部电子第五研究所、诺亦腾（北京）

科技有限公司、广东电网有限责任公司、华为云计算技术有限公司、深圳赛西信息技术有限公司、深圳兆威机电股份有限公司、帕西尼感知科技（深圳）有限公司等单位参与了会议讨论，并对标准提出修改意见。标准编制小组根据各专家提出的意见进行多次整理并修改确定标准讨论稿，在进一步完善以及全文复核后，形成标准征求意见稿。

（4）征求意见阶段

2025年6月至7月，起草单位充分讨论并修改完善初稿，形成《人形机器人应用环境数据采集规范（征求意见稿）》，面向市、各区有关单位公开征集意见。

三、标准编制原则和标准主要内容

1. 编制原则

本标准严格遵照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》的有关规定编写。

2. 标准主要内容

本标准主要结构如下：

1. 范围

本文件规定了人形机器人应用环境数据采集模型、数据类型、数据要求、数据接口要求、数据采集质量控制与安全控制要求。

本文件适用于人形机器人系统研发实际应用环境数据的体系规划、设计和建设。

2. 规范性引用文件

本章列出了本文件所引用的标准。

3. 术语与定义

本标准涉及的相关术语，参考已发布的国家标准和行业标准，并根据其在本标准中特指的含义和内容对其进行定义。

4. 数据采集模型

本章利用模型图明确了人形机器人数据采集方法、处理方式和应用环境。

5. 数据类型

本章对人形机器人采集数据类型按照数据来源进行了划分。

6. 数据要求

本章对人形机器人中传感器数据、执行器数据、应用程序中间结果数据、参数数据和配置数据的内容和格式提出了要求。

7. 数据接口要求

本章明确了采集数据的通信协议、传输带宽和可靠性要求。

8. 数据采集质量控制与安全控制要求

本章规定了全过程中的数据采集质量控制和数据安全控制要求。

9. 参考文献是本标准起草所参考的相关资料。

四、与国内外同类标准现状

国内外没有发布关于人形机器人环境数据采集的相关标准。

国际标准情况：ISO/IEC 8183:2023 Information technology – Artificial intelligence – Data life cycle framework该标准定义了整个AI系统生命周期中数据处理的阶段及相关行动，包括获取、创建、开发、部署、维护和退役。但没有对具体数据采集过程要求进行规定；ISO/IEC 5259-2:2024 Artificial intelligence – Data quality for analytics and machine learning (ML) – Part 2: Data quality measures该标准规定了数据质量模型、数据质量度量以及在分析和机器学习（ML）背景下报告数据质量的指南。

我国发布的数据采集和数据接口相关标准，包括：

（1）GB/T 42127—2022《智能制造 工业数据 采集规范》规定了智能制造工业数据采集的通用规范。适用于离散及流程工业企业数据资源的规划和采集。该标准对象为智能制造系统，是智能工厂场景下的实施应用，可提供参考价值。

（2）GB/T 36625.3—2021《智慧城市 数据融合 第3部分：数据采集规范》规定了智慧城市数据融合过程中数据采集过程、数据采集内容、数据采集技术、数据采集质量控制和数据采集安全控制。适用于智慧城市各系统的规划设计、建设和管理，其他信息化领域的数据采集过程也可参考使用。

（3）GB/T 39401—2020《工业机器人云服务平台数据交换》规定了工业机器人与工业机器人云服务平台数据交换过程中数据集成模型、数据分类、数据描

述规范和模型。适用于在运行生命周期内工业机器人与云服务平台之间信息资源交换体系的规划、设计、建设和实施。

(4) GB/T 30269.702—2016《信息技术 传感器网络 第702部分：传感器接口 数据接口》规定了传感器和传感结点的参数编码，以及数据交互规范。本部分适用于传感器和传感结点产品的设计、生产、测试、系统集成与应用。

同时参考YD/T 4608.1—2023《5G网络测试数据采集统一文件接口技术要求 第1部分：文件基本结构要求》、YD/T 4608.2—2024《5G网络测试数据采集统一文件接口技术要求 第2部分：采集事件要求》、YD/T 4608.3—2024《5G网络测试数据采集统一文件接口技术要求 第3部分：芯片原始帧采集要求》、YD/T 4586—2023《网络安全态势感知 数据采集要求》、YD/T 4848—2024《接入网运维智能化技术要求 数据采集和通用数据模型》、JT/T 1394—2021《航标数据采集规范》、DB31/T 845—2020《新能源汽车公共数据采集技术规范》、DB11/T 2107—2023《供热系统智能化数据采集及通信规范》等，结合人形机器人所应用环境数据采集的实际情况制定本地方标准。

五、主要内容确定依据

1. 数据采集模型

本标准中数据采集类型的设计，主要基于人形机器人实际运行环境及其本体数据生成机制。人形机器人的数据采集行为直接发生于其应用环境中，其核心本体数据源包括各类传感器设备、执行器状态信息以及任务程序运行数据，是获取环境交互与本体状态的关键。这些数据最终将服务于人形机器人模型训练、系统开发、系统调试与验证等核心应用场景。此结构设计符合机器人系统分层架构的技术逻辑，能够满足全生命周期开发对数据的一致性和可用性要求。

2. 数据类型

本标准依据人形机器人系统的核心构成要素（感知、执行、应用）划分数据类型。环境与本体感知数据（摄像头、力传感器等）的分类对应机器人感知外部环境和自身状态的需求，符合系统层级划分原则；执行器工作与反馈数据（电机、减速器等）因其直接关联动作执行精度，需独立采集以满足控制可追溯性要求；应用程序数据（参数、配置、中间结果）源自开发调试全流程的实际需求，参数

定义本体结构，配置设定工作模式，中间结果支撑算法验证。此分类体系完整覆盖“感知－执行－决策”闭环，确保数据有效支撑模型训练、系统调试等核心应用。

3. 数据要求

本标准的数据内容与格式要求设计，遵循人形机器人系统全栈数据治理的核心逻辑。内容需满足数据要素的完整性，通过统一纳入时间戳、设备/模块标识与数据本体字段，确保传感器与执行器数据的时空可追溯性，强化应用程序数据的模块溯源能力；参数和配置数据增设版本号与更新时间字段，以便于兼容系统迭代。格式要求基于工程实践，强制同类数据格式一致性并优先采用通用结构化格式，规范时间戳、参数值等关键字段的表达形式，同时引用 GB/T 30269.701 编码标准保障异构系统互操作性。

4. 数据接口要求

本标准规范数据接口的通信协议、传输带宽和可靠性要求，源于人形机器人多源异构数据采集的实际需求。设计兼容以太网/USB（有线）、Wi-Fi（无线）和 CAN（高速串口）等通用传输方式，通过对齐工业标准保障传感器数据、执行器指令、软件中间结果与硬件状态数据的无缝流转。带宽与可靠性指标的设定便于直接响应高频环境感知、实时运动控制等场景的时效性约束，其技术依据参照 ROS 2.0 DDS、OPC UA 等成熟通信框架的验证结论，确保全链路采集的稳定性与系统集成性。

5. 采集质量控制与安全控制要求

本标准的数据质量控制要求以 GB/T 36344 为基础，通过完整性（识别离群值）、准确性（符合模型训练标准）、一致性（符合国家标准相关规定）、时效性（及时更新）、可访问性（保障安全访问）、可追溯性（数据溯源）和信息标注（转换为结构化数据）等七项原则覆盖采集全流程，确保数据可直接用于模型训练与评估。数据安全控制聚焦生物识别数据（含人体图像、声音、指纹等），依据隐私保护核心逻辑：保密性（防未授权访问与意外丢失）、目的限定性（限定合法用途）、数据最小化（禁止过度收集）和存储限制（超期删除或匿名化），形成符合 GDPR 等国际规范的技术实施方案。

六、标准中涉及专利和知识产权的情况

本标准不涉及专利和知识产权问题。

七、预期达到的社会效益

本标准的实施具有重要的社会意义，通过统一人形机器人应用环境数据标准、提升数据质量和接口要求，能够为企业降低研发门槛与成本，解决数据采集过程中数据质量参差不齐、数据格式不一致等问题，提高数据的可用性和可信度。提升数据管理和信息安全水平，促进数据共享与流通，促进产业链协同，实现数据价值的最大化。从而推动人形机器人产业化落地和在家庭服务、医疗护理、教育辅导等高价值社会服务领域的广泛应用。

八、与我国有关的现行法律、法规和相关强制性标准的关系

本标准与国家现行法律、法规和相关强制性标准相协调配套。

九、重大分歧意见的处理依据和结果

编制组根据起草前确定的编制原则进行了标准起草，标准起草组前期进行了充分的准备和调研，并做了大量调查论证、信息分析工作，标准在主要技术内容上，在行业内取得了较为一致的意见，标准起草过程中未发生重大分歧意见。

十、实施地方标准的措施建议

拟通过开展标准的宣贯、培训和标准实施跟踪检查等方式推动标准实施。

十一、其他需要说明的事项

无。