**《具身智能语料库建设导则》**

**(征求意见稿)**

**编制说明**

标准编写工作组 2024 年11 月

**《****具身智能语料库建设导则》**

**(征求意见稿)**

**编制说明**

一、工作简况

**1、任务来源**

制定《具身智能语料库建设导则》至关重要，一是能够为具身智能领域提供一个统一的标准和规范，确保语料库的建设和使用过程中的数据质量、一致性和可比性；二是有助于提高语料库的可用性和可靠性，减少因标准不一而导致的研究误差和资源浪费，为学术研究和工业应用提供坚实的基础；三是通过提供高质量的语料库，可以加速算法的迭代和优化，促进新技术的诞生和应用，从而推动整个行业的进步；四是有助于不同领域专家之间的合作，促进知识的共享和交流。本项目由上海库帕思科技有限公司提出，上海人工智能协会批准立项。《具身智能语料库建设导则》主要起草单位：上海库帕思科技有限公司，国家地方共建人形机器人创新中心，上海智元新创技术有限公司，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。计划应完成时间为 2024年 11月。

**2、主要工作过程**

**预研阶段**：2024 年 8月，牵头起草单位组织专家组对具身智能语料库的建设现状进行国内外调研，包括相关政策制度、技术路径、市场现状进行深入研究，梳理团体标准相关法律法规在标准审查中的符合性条文，开展具身智能语料库建设的技术要求研究，为后续标准制定提供理论基础和政策指导。

**起草阶段：**2024 年 9 月-11月，对具身智能语料库建设需求进行资料收集、查阅、分析和整理，对团体标准政策制度进行梳理，查找标准技术审查关键问题，确定标准主要技术内容和标准框架结构。对照《标准化法》、《团体标准管理规定》等标准，编写标准草案。

2024 年9 月-11月，面向社会征集标准编写参与单位，组成标准工作组。 为提高标准的普适性和可操作性，标准工作组通过召开研讨会以及书面反馈意见等 多种形式，对标准文本进行修改完善，处理内部意见 100 余条，形成标 准征求意见稿。

**3、主要参加单位和工作组成员及其所作的工作等**

本标准由上海库帕思科技有限公司、上海库帕思科技有限公司，国家地方共建人形机器人创新中心，上海智元新创技术有限公司，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。联合编制。

主要成员：山栋明、黄海清、施佳樑、张裕珍、陈春玉，沈林鹏，陈宇涛、许彬、江磊、刘宇飞、黄媚、刘传厚、陈曦、沈荣港、王洪武、李昀佶、闫维新 姚卯青 朱建超等 40余人。

主要分工：山栋明为工作组组长，主持全面协调工作。施佳樑为本标准主要持 笔人，负责本标准的起草、编写。张裕珍、王洪武、李昀佶等为组员负责对具身智能语料库建设的现状进行全面调研、研究分析和资料汇总整理等工作，提炼总结各模块要点和要求，并进行了反复验证、修改完善。

二、标准编制原则和主要内容

1 、标准编制原则

开展《具身智能语料库建设导则》编撰工作，应把握以下原则：包括符合性原则、先进性原则和协调性原则、规范性原则。

（1）符合性原则，团体标准制修订应符合国家有关团体标准的法律法规、行政 规章的要求；

主要依据：《团体标准管理规定》第十一条明确提出“团体标准应当符合相关法 律法规的要求，不得与国家有关产业政策相抵触。”《中国标准化协会标准管理办法》 中第三条明确提出“标协标准制修订工作应当遵循以下原则：（一）遵守国家有关的 法律、法规；（二）符合强制性标准的要求。”

（2）先进性原则，团体标准的主要技术指标不得低于强制性标准的技术要求， 并应优于国家标准、行业标准；

主要依据：《标准化法》第二十一条明确提出“推荐性国家标准、行业标准、地 方标准、团体标准、企业标准的技术要求不得低于强制性国家标准的相关技术要求。 国家鼓励社会团体、企业制定高于推荐性标准相关技术要求的团体标准、企业标准。” 《团体标准管理规定》第十三条明确提出：“制定团体标准应当以满足市场和创新需 要为目标,聚焦新技术,新产业,新业态和新模式,填补标准空白”。《中国标准化协会标 准管理办法》中第三条明确提出“标协标准制修订工作应当遵循以下原则：（三）优 先支持符合经济发展方向，促进科学技术进步，提高产品质量和满足市场需求的项 目；（四）积极采用国际标准；（五）协调融合、有序优化、技术先进、经济合理；

（3）协调性原则，团标标准文本应与现行标准协调一致；

主要依据：《团体标准管理规定》第十一条明确提出:“对于术语、分类、量值、 符号等基础通用方面的内容应当遵守国家标准,行业标准,地方标准, 团体标准一般不 予另行规定。”第十五条明确规定：“技术审查原则上应当协商一致。如需表决,不少 于出席会议代表人数的 3/4 同意方为通过。”

（4）规范性原则，团体标准编写应规范，符合 GB/T 1.1 的要求。

主要依据：《团体标管理规定》第十五条 团体标准的编写参照 GB/T1.1《标准化 工作导则第 1 部分:标准化文件的结构和起草规则》的规定执行。 GB/T 20004.1-2016 《团体标准化 第 1 部分：良好行为指南》第 7 章 7.2 条编写原则提出：“团体宜按 照 GB/T 1.1 制定统一的标准编写规则，包括团体标准的结构、起草表述方法、格式 等内容，以提高团体标准的适用性 ”。

2、标准主要内容

本文件提出了具身智能语料库建设需要具备的核心能力及方法，包括数据要求、数据采集、数据生产、数据安全等四个方面。

本文件适用于具身智能语料库建设的采集、标注、加工等方面的工作。

三、主要试验（或验证）情况

本标准不涉及试验问题。

四、标准中涉及专利情况 本标准不涉及专利问题。

五、预期达到的社会效益、对产业发展的作用等情况

《具身智能语料库建设导则》团体标准的制定，将为人形机器人具身智能领域的发展提供标准化和规范化的指导，使行业的各环节有章可循、有据可查、有准可依。这有助于引导企业和科研机构更加有序地开展技术研发和应用，避免非良性竞争和资源浪费，为人形机器人具身智能的健康发展奠定坚实的基础。

通过《具身智能语料库建设导则》的实施，可以加速人工智能技术的发展，尤其是在自然语言处理、机器学习和人工智能领域。这将推动传统行业的转型升级，指引新兴行业的加速生长，为产业的科学健康发展注入强大动力。

六、与国际、国外对比情况

本标准没有采用国际标准。

本标准制定过程中未查到同类国际、国外标准。 本标准水平为国内先进水平。

七、在标准体系中的位置，与现行相关法律、法规、规章及标准，特别是强制 性标准的协调性

符合国家有关法律法规、政策制度的要求。

八、重大分歧意见的处理经过和依据 无。

九、标准性质的建议说明

建议本标准的性质为团体标准。

十、贯彻标准的要求和措施建议

建议本标准批准发布 3 个月后实施。

建议本标准由上海库帕思科技有限公司宣贯实施。

十一、废止现行相关标准的建议 无。

十二、其他应予说明的事项 无。