Ao

《具身智能实景应用能力要求及评价方法》

标准编制说明

标准编制组

2026 年 11 月

1. 标准范围

本文件规定了具身智能系统实景应用能力的总体要求、场景任务与场景挑战度、实景应用能力要求、实景应用成熟度评价方法，以及评价与结果表示要求。

本文件适用于工业制造、仓储物流、商业服务、医疗康养、家庭服务和特种作业等领域中具身智能系统在真实场景或等效测试场景下的应用能力评价，可用于应用验证、产品选型、项目验收、实景训练评价和第三方评价等活动。

本文件评价对象为在明确软硬件配置、模型版本、工具端、外部依赖和运行边界条件下的完整具身智能系统，不将单一部件、单项算法或脱离规定场景的实验室性能直接等同于系统实景应用能力。

2. 工作简况

2026 年 10 月，本标准编制组正式成立，对该标准的立项达成共识，共同收集相关资料，多次组织专家进行讨论，形成标准草案大纲。

2026 年 10 月 26 日，本标准正式立项。

2026 年 10-11 月，编制组多次召开标准研讨会，针对产学研单位对标准文本提出的意见建议，持续对标准草案进行讨论、迭代，形成征求意见稿。

2026 年 11 月 4 日-2026 年 12 月 3 日，本标准通过中国互联网协会官方网站对外公开征求意见。

3. 标准编制原则和确定标准主要内容

3.1 编制原则

（1）场景任务导向原则。以真实业务任务为评价载体，明确任务目标、起止状态、成功判据、运行边界和场景条件，避免以单项参数替代系统整体应用能力。

（2）系统整体评价原则。面向软硬件、模型、执行机构、感知系统及外部依赖共同构成的完整系统，从任务执行、自主作业、稳定可靠、安全协同、应用效能与经济性五个维度开展评价。

（3）客观可验证原则。指标应具备明确的数据来源、计算口径、样本要求和判定规则，评价过程应可记录、可复核、可追溯。

（4）分级分类原则。区分场景挑战度与应用成熟度，允许不同领域根据任务特点选择适用指标和门槛，同时保持统一的评价逻辑和结果表示方式。

（5）安全底线原则。将人员安全、关键安全功能、危险行为、数据与系统安全等作为评价的重要约束，对不可接受的重大安全风险设置安全否决机制。

3.2 主要技术内容

本标准由范围、规范性引用文件、术语和定义、符号和缩略语、总体要求、场景任务与场景挑战度、实景应用能力要求、实景应用成熟度评价方法、评价与结果表示，以及五个附录构成。主要技术内容如下：

（1）评价总体框架。提出 C-L 场景应用评价模型。其中，C 表示场景挑战度，在测试前依据作业环境、作业对象、动态交互、任务复杂度和异常与扰动等因素确定；L 表示应用成熟度，依据测试结果和适用指标门槛确定。评价结果采用“C 等级—L 等级”的组合方式表示。

（2）场景任务与场景挑战度。规定场景任务的目标、起止状态、作业单元、成功和失败判据、边界条件等基本要素，并将场景挑战度划分为 C1 固定条件级至 C5 开放挑战级五个等级。

（3）实景应用能力要求。从任务执行能力、自主作业能力、稳定可靠能力、安全协同能力、应用效能与经济性五个维度提出要求。任务执行类指标包括任务完成率、作业质量达标率、操作精度、任务完成时间和节拍达标率等；其他维度相应覆盖自主完成与人工介入、故障与恢复、风险事件与协同安全、效率与投入产出等内容。

（4）实景应用成熟度评价方法。规定评价准备、样本与测试实例、数据获取、指标计算、异常数据处理和统计要求，为不同应用场景制定具体评价方案提供统一方法。

（5）等级判定和结果表示。将应用成熟度划分为 L1 演示验证级至 L5 规模化应用级五个等级，并规定必选能力维度、指标门槛、安全否决和综合评分等要求。

（6）附录内容。附录 A 给出场景挑战度确定方法；附录 B 给出典型应用场景参考；附录 C 给出工业制造、仓储物流和商业服务等典型评价方案参考；附录 D 给出综合评分计算规则；附录 E 给出安全否决条件参考。

4. 主要试验（或验证）的分析、综述报告

无。

5. 标准在起草过程中遇到的问题及解决办法；重大分歧意见的处理经过和依据；有无重要技术问题需要说明

在本文件的修订过程中，无重大分歧意见和技术问题。

6. 与国外标准的关系：包括采用国际标准和国外先进标准的程度，与国外标准主要技术内容的差异

该项目没有对应的国际标准或国外先进标准。

7. 修订标准时，说明与标准前一版本的重大技术变化，并列出所涉及的新、旧版本的有关章条；废止或代替现行有关标准的建议

不涉及。

8. 说明标准与其他标准或文件的关系，特别是与有关的现行法律、法规和强制性国家标准的关系

符合现行法律、法规要求。

9. 标准作为强制性标准或推荐性标准的建议

建议本文件作为推荐性标准。

10. 贯彻标准的要求和措施建议；标准发布后，对国内外业界可能产生的影响

建议标准发布后，通过宣贯培训、评价方案示范、典型场景试点和第三方评价实践等方式推动实施。使用单位应结合具体行业和场景，明确评价对象、系统版本、场景挑战度、测试实例、适用指标、门槛值和安全否决条件，完整保存原始记录和评价依据。

标准的实施有助于形成具身智能系统从演示验证到规模化应用的统一评价语言，降低采购选型、应用验证、项目验收和供需沟通中的信息不对称，促进评价活动规范化和产业协同。标准实施过程中应持续收集不同领域的应用数据和问题反馈，适时修订完善指标、等级门槛和典型评价方案。

11. 标准是否涉及知识产权的情况说明；如标准中含有自主知识产权，说明产品研发程度、产业化基础及进程

本文件未涉及。

12. 其他应予说明的事项

无。