

T

ICS 35.240.50

CCS L 67

# 团体标准

T/CWDPA XXX—2026

## 工业智能体跨平台协同评价规范

Specification for cross-platform collaborative evaluation of industrial intelligent agents

2026-X-XX 发布

2026-X-XX 实施

中国西部开发促进会 发布

## 目 次

|                      |    |
|----------------------|----|
| 前言 .....             | 11 |
| 工业智能体跨平台协同评价规范 ..... | 1  |
| 1 范围 .....           | 1  |
| 2 规范性引用文件 .....      | 1  |
| 3 术语和定义 .....        | 1  |
| 4 评价原则 .....         | 2  |
| 5 评价指标体系 .....       | 2  |
| 6 评价方法 .....         | 5  |
| 7 评价流程 .....         | 7  |
| 8 评价报告 .....         | 9  |
| 9 持续改进 .....         | 9  |

## 前言

本文件按照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》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国西部开发促进会提出并归口。

本文件起草单位：南京财经大学。

本文件主要起草人：。

本文件为首次发布。

# 工业智能体跨平台协同评价规范

## 1 范围

本文件规定了工业智能体跨平台协同的评价原则、评价指标体系、评价方法、评价流程以及评价报告的要求。

本文件适用于工业智能体跨平台协同系统的规划、设计、开发、部署、运维和升级过程中的质量评价、效能评估与持续改进。本文件不适用于军用工业智能体跨平台协同系统的评价，也不适用于未接入工业互联网平台的孤立工业智能体的评价。

## 2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

- GB/T 7713.3-2014 科技报告编写规则
- GB/T 19001-2016 质量管理体系 要求
- GB/T 22239-2019 信息技术 网络安全等级保护基本要求 GB/T 25000.10-2016 系统与软件工程 系统与软件质量要求和评价（SQuaRE） 第10部分：系统与软件质量模型
- GB/T 35273-2020 信息安全技术 个人信息安全规范
- GB/T 36461-2018 物联网标识体系 OID应用指南
- GB/T 37973-2019 信息安全技术 大数据安全管理指南
- GB/T 38624.1-2020 物联网 网关 第1部分：面向感知设备接入的网关技术要求
- GB/T 39116-2020 智能制造能力成熟度模型

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

#### 工业智能体 *industrial intelligent agent*

具备自主感知、决策、执行能力，能够在工业场景中自主完成特定任务，并可与其他智能体或系统进行交互协同的智能化实体。其核心构成包括感知模块、决策模块、执行模块及通信模块，广泛应用于生产制造、设备运维等工业环节。

### 3.2

#### 跨平台协同 *cross-platform collaboration*

不同硬件平台、操作系统、开发框架或归属主体的工业智能体，通过统一的通信协议、数据交互标准及协同机制，协同完成复杂工业任务的过程。其核心特征包括互操作性、数据一致性及任务协同性。

### 3.3

#### 协同能力 *collaborative capability*

工业智能体之间实现跨平台信息交互、任务分工、资源共享及冲突协调的能力，是衡量跨平台协同效果的核心基础指标。

### 3.4

#### 协同效能 *collaborative efficiency*

工业智能体跨平台协同完成任务过程中，在效率、成本、质量等方面呈现的综合效果，反映协同模式的实际应用价值。

## 3.5

**互操作性 interoperability**

不同平台的工业智能体之间交换信息并正确理解和使用所交换信息的能力，包括语法互操作、语义互操作及语用互操作三个层面。

## 3.6

**评价指标体系 evaluation index system**

由一系列相互关联、相互补充的评价指标构成的有机整体，用于全面、系统地反映工业智能体跨平台协同的综合水平。

**4 评价原则****4.1 科学性原则**

评价指标选取与评价方法确定应基于工业智能体跨平台协同的技术原理与产业实际，符合客观规律，指标体系需层次清晰、逻辑严谨，准确反映协同核心特征与关键环节，避免主观臆断。

**4.2 客观性原则**

评价过程需基于可验证的数据、事实及客观证据，规避人为因素干扰，数据采集遵循统一标准，评价方法具备可重复性，确保不同评价主体在相同条件下得出一致评价结果。

**4.3 系统性原则**

评价应覆盖工业智能体跨平台协同全生命周期，含规划、设计、开发、部署、运维等环节，指标体系需涵盖协同能力、效能、安全性、兼容性等多维度，全面反映协同系统综合水平，避免片面性。

**4.4 可操作性原则**

评价指标需明确具体，数据易于采集和量化，评价方法简便易行，规避过度复杂的计算或难以获取的参数，评价流程清晰规范，便于企业、科研机构、第三方检测机构等各类评价主体实际应用。

**4.5 动态性原则**

结合工业智能体技术发展趋势与产业应用需求变化，评价指标体系应具备可调整性，定期复核指标适用性与科学性，及时补充或修订相关指标，确保标准时效性与前瞻性。

**4.6 安全性原则**

评价过程需充分考虑工业智能体跨平台协同过程中的数据安全、设备安全及生产安全风险，将安全性指标作为核心评价内容，确保协同系统稳定可靠运行，符合相关安全标准要求。

**5 评价指标体系****5.1 构建依据**

5.1.1 技术依据：工业智能体、分布式系统、工业互联网、信息物理系统（CPS）等相关领域的核心技术原理与性能标准，参照 GB/T 39116—2020、GB/T 36461—2018 等标准的相关要求。

5.1.2 产业依据：离散制造、流程工业等典型场景下对跨平台协同的实际需求、最佳实践与痛点问题。

5.1.3 政策与标准依据：国家智能制造、工业互联网创新发展等相关政策规划，以及信息安全、数据治理、系统互操作性等领域的国内外已有标准。

5.1.4 专家共识：通过德尔菲法、层次分析法等方法，综合领域内技术专家、企业代表及学者的意见与判断。

**5.2 指标选取原则**

指标体系的构建应符合GB/T 25000.10—2016中关于软件质量模型的通用要求，并结合工业智能体跨平台协同的特殊性进行调整，原则性要求如下：

a) 具体性（Specific）：指标应定义清晰，无歧义。

- b) 可测量性 (Measurable)：指标可通过定量或定性方法进行度量与评价。
- c) 可实现性 (Attainable)：指标要求应符合当前技术发展水平与产业现实。
- d) 相关性 (Relevant)：指标应与“跨平台协同”的核心目标紧密相关。
- e) 时效性 (Time-bound)：指标应能反映一定时期内的协同水平，并适应发展。

### 5.3 权重确定方法

各级指标权重通过德尔菲法（专家问卷调查）与层次分析法相结合的方式确定。首先由专家组对各指标间的相对重要性进行多轮背对背打分，达成共识后构建判断矩阵，进而计算并校验得到各级指标权重。权重结果反映了各指标在综合评价中的相对贡献程度。

### 5.4 体系框架

工业智能体跨平台协同评价指标体系采用三级分层递阶结构，包括目标层、准则层和指标层。一级指标（准则层）包括协同能力、协同效能、安全性、兼容性4个核心维度。二级指标在一级指标框架下展开，共12项。三级指标为可具体测量与评价的底层指标，共36项。指标体系总体框架见表1。

表1 指标体系框架

| 一级指标（权重）       | 二级指标（权重）         | 三级指标      | 指标说明概要                | 数据来源概要    | 评价类型  |
|----------------|------------------|-----------|-----------------------|-----------|-------|
| 协同能力<br>(0.35) | 通信互连能力<br>(0.12) | 通信协议兼容性   | 支持标准协议的种类及与主流协议的适配程度。 | 技术手册、实测数据 | 定量+定性 |
|                |                  | 通信延迟      | 跨平台数据传输的平均延迟 (ms)。    | 性能测试报告    | 定量    |
|                |                  | 通信稳定性     | 规定时长内的通信中断次数。         | 长期运行测试记录  | 定量    |
|                | 数据交互能力<br>(0.10) | 数据格式兼容性   | 支持的标准数据格式及跨平台转换成功率。   | 技术手册、转换测试 | 定量+定性 |
|                |                  | 数据传输完整性   | 接收数据量与发送数据量的完整比率(%)。  | 数据传输测试记录  | 定量    |
|                |                  | 数据同步时效性   | 跨平台数据同步的平均时间差 (s)。    | 同步测试报告    | 定量    |
|                | 任务协同能力<br>(0.13) | 任务分解与分配效率 | 分解并分配复杂任务的平均时间 (min)。 | 任务测试记录    | 定量    |
|                |                  | 任务执行协同度   | 子任务实际进度与计划进度偏差的平均值。   | 任务执行监控记录  | 定量    |
|                |                  | 冲突协调能力    | 面对资源或任务冲突时的协调解决效果。    | 冲突模拟测试报告  | 定性    |
| 协同效能<br>(0.30) | 效率提升效果<br>(0.10) | 任务完成时间缩短率 | 协同与非协同模式任务耗时差值的百分比。   | 任务对比测试记录  | 定量    |
|                |                  | 资源利用效率提升率 | 协同与非协同模式资源利用率差值的百分比。  | 资源监控数据    | 定量    |
|                |                  | 响应速度提升率   | 对场景需求响应速度提升的百分比。      | 响应测试报告    | 定量    |
|                | 成本控制效果<br>(0.08) | 协同运维成本降低率 | 协同与独立运维总成本差值的百分比。     | 成本核算数据    | 定量    |
|                |                  | 系统建设成本性价比 | 总建设成本与协同效能提升值的比       | 成本与效能分析报告 | 定量    |

| 一级指标 (权重)  | 二级指标 (权重)     | 三级指标      | 指标说明概要                | 数据来源概要      | 评价类型  |
|------------|---------------|-----------|-----------------------|-------------|-------|
|            |               |           | 值。                    |             |       |
|            |               | 能耗降低率     | 协同与非协同模式系统总能耗差值的百分比。  | 能耗监测记录      | 定量    |
|            | 质量提升效果 (0.12) | 任务执行准确率   | 协同完成任务中结果准确的比率 (%)。   | 任务执行记录      | 定量    |
|            |               | 故障发生率降低率  | 系统故障发生率降低的百分比。        | 故障统计数据      | 定量    |
|            |               | 用户满意度     | 用户对协同效果的评分 (10分制)。    | 满意度调研问卷     | 定量    |
| 安全性 (0.20) | 数据安全 (0.07)   | 数据加密强度    | 传输数据加密算法等级及符合安全要求的程度。 | 安全测试报告、技术文档 | 定性+定量 |
|            |               | 数据访问控制有效性 | 权限管理严格性及防越权访问效果。      | 权限测试报告      | 定性    |
|            |               | 数据脱敏效果    | 敏感数据脱敏处理后的安全性。        | 脱敏测试报告      | 定性    |
|            | 设备安全 (0.06)   | 设备接入认证安全性 | 设备接入身份认证机制的安全性。       | 接入测试报告      | 定性    |
|            |               | 设备运行稳定性   | 协同运行中设备的故障率。          | 设备运行记录      | 定量    |
|            |               | 异常行为识别能力  | 对设备异常行为的识别准确率与响应速度。   | 异常模拟测试报告    | 定量+定性 |
|            | 控制安全 (0.07)   | 控制指令安全性   | 控制指令传输的完整性、真实性及防篡改能力。 | 指令测试报告      | 定性+定量 |
|            |               | 应急处置能力    | 安全事件应急响应时间及处置效果。      | 应急演练报告      | 定量+定性 |
|            |               | 安全审计完整性   | 关键操作审计记录的完整性与可追溯性。    | 审计日志        | 定性    |
| 兼容性 (0.15) | 硬件兼容性 (0.04)  | 硬件平台适配范围  | 可适配的不同品牌、型号硬件设备的数量。   | 技术手册、适配测试报告 | 定量    |
|            |               | 硬件接口兼容性   | 硬件接口与不同设备接口的适配程度。     | 接口测试报告      | 定性+定量 |
|            |               | 硬件扩展兼容性   | 新增设备接入的便捷性及对原系统的影响。   | 扩展测试报告      | 定性    |
|            | 软件兼容性 (0.06)  | 操作系统适配性   | 支持的主流工业操作系统种类及版本。     | 技术手册、适配测试报告 | 定量+定性 |
|            |               | 开发框架兼容性   | 与主流开发框架的适配程度。         | 框架测试报告      | 定性    |
|            |               | 应用软件兼容性   | 与行业应用软件的协同运行效果。       | 软件测试报告      | 定性    |
|            | 协议兼容性 (0.05)  | 工业通信协议适配性 | 支持的工业通信协议种类及适配稳定性。    | 协议测试报告      | 定量+定性 |
|            |               | 互联网协议适配性  | 支持的互联网协议种类及数据传输效果。    | 网络测试报告      | 定量+定性 |

| 一级指标（权重） | 二级指标（权重） | 三级指标    | 指标说明概要             | 数据来源概要 | 评价类型 |
|----------|----------|---------|--------------------|--------|------|
|          |          | 协议转换兼容性 | 不同协议间转换的成功率及信息完整性。 | 转换测试报告 | 定量   |

## 5.5 指标体系的运用与动态管理

5.5.1 在实际评价中，可根据被评价系统的具体行业属性、应用场景及评价目的，在保持框架稳定的前提下，对非强制性指标及权重进行适应性微调，并在评价方案中予以明确说明和备案。

5.5.2 归口单位应定期组织对指标体系进行系统性复审与更新，周期为2-3年为宜。

## 6 评价方法

### 6.1 数据采集方法

#### 6.1.1 定量数据采集

6.1.1.1 定量指标数据宜采用自动化监测与人工记录相结合的方式采集。

6.1.1.2 通信延迟、数据传输完整性、能耗等反映系统实时性能的指标，应通过工业互联网平台数据采集模块或智能传感器等自动化设备进行采集。

6.1.1.3 任务完成时间、成本、资源利用率等过程性指标，应结合系统日志、运行报告与人工核算记录进行采集。

6.1.1.4 数据采集频率应满足指标特性与评价需求：

- f) 实时性指标（如通信延迟）的采集频率不宜低于1次/秒；
- g) 周期性指标（如日均能耗）的采集频率不宜低于1次/天；
- h) 统计类指标（如月度任务完成率）的采集频率不宜低于1次/月。

6.1.1.5 数据采集应符合GB/T 38624.1-2020中关于工业互联网数据采集的通用要求。

#### 6.1.2 定性数据采集

6.1.2.1 定性指标数据宜采用测试模拟、调研访谈、专家评审、文件审查等方式采集。

6.1.2.2 冲突协调能力、安全审计完整性等指标，应通过构建典型或边缘模拟场景进行测试，记录测试过程与结果。

6.1.2.3 用户满意度、操作便捷性等主观感受指标，应采用经过验证的标准化问卷或结构化访谈提纲进行数据收集。

6.1.2.4 技术兼容性、协议适配性等专业技术指标，应组织相关领域专家，结合技术文档、接口协议、测试报告及现场查验结果进行评审打分。

### 6.2 数据验证与处理方法

6.2.1 应对采集到的原始数据进行有效性验证。验证内容包括但不限于数据的完整性、格式规范性、逻辑合理性及时间一致性。

6.2.2 应识别并剔除因设备故障、网络中断、人为误操作等原因导致的异常数据，并记录异常原因。

6.2.3 对于缺失数据，宜根据数据特性和缺失原因，采用插值法、均值法、回归预测等合理统计方法进行补充，并注明补充方法。

6.2.4 当数据质量存疑或关键指标采集失败时，应进行重复测试或补充测试，以确保数据源的准确性和可靠性。

6.2.5 数据安全管理应符合GB/T 37973-2019和GB/T 35273-2020的相关要求。

### 6.3 指标评分计算

#### 6.3.1 定量指标评分

6.3.1.1 各定量指标应根据其类型，按以下公式计算实际值A，并结合预先设定的判定标准进行标准化评分 $S_{quant}$ （满分100分）。

### 6.3.1.2 比率类指标（如数据传输完整性、任务执行准确率）：

$$A = \frac{N_{qualified}}{N_{total}} \times 100\% \quad (1)$$

$$S_{quant} = \min \left( \frac{A}{A_{opt}}, 1 \right) \times 100 \quad (2)$$

式中：

$A$ ——指标实际值；

$N_{qualified}$ ——符合要求的数量；

$N_{total}$ ——总数量；

$A_{opt}$ ——该指标的最优标准值（目标值）；

当  $A \geq A_{opt}$  时， $S_{quant} = 100$ 。

### 6.3.1.3 差值类指标（如通信延迟、任务完成时间）：

$$S_{quant} = [1 - \frac{A - A_{opt}}{A_{worst} - A_{opt}}] \times 100 \quad (3)$$

式中：

$A$ ——指标实际值；

$A_{opt}$ ——该指标的最优标准值（期望达到的最好值）；

$A_{worst}$ ——该指标的最差允许值（评价底线）；

当  $A \leq A_{opt}$  时， $S_{quant} = 100$ ；当  $A \geq A_{worst}$  时， $S_{quant} = 0$ 。

### 6.3.1.4 提升/降低率类指标（如成本降低率、效率提升率）：

$$S_{quant} = \min \left( \frac{A}{A_{target}}, 1 \right) \times 100 \quad (4)$$

式中：

$A$ ——指标实际值；

$V_{non-coop}$ ——非协同模式下的基准值（如成本、耗时）；

$V_{coop}$ ——协同模式下的实际值；

$A_{target}$ ——该指标的预期目标提升/降低率；

当  $A \geq A_{target}$  时， $S_{quant} = 100$ 。

## 6.3.2 定性指标评分

定性指标应采用等级评分法。由评价团队（不少于5人）根据预定标准独立打分，取平均值为最终得分  $S_{qual}$ 。等级与分值对应关系如下：

- i) 优秀:  $100 \geq S_{qual} \geq 80$
- j) 良好:  $79 \geq S_{qual} \geq 60$
- k) 一般:  $59 \geq S_{qual} \geq 40$
- l) 较差:  $39 \geq S_{qual} \geq 0$

## 6.3.3 综合评分计算

系统综合得分  $P$  采用分层加权求和法计算，具体步骤如下：

- a) 计算二级指标得分  $I_{2j}$ :

$$I_{2j} = \sum_{k=1}^m (S_{jk} \times w_{jk}) \quad (1)$$

式中：

$I_{2j}$  —— 第  $j$  个二级指标的得分;

$S_{jk}$  —— 隶属于该二级指标的第  $k$  个三级指标的评分;

$w_{jk}$  —— 第  $k$  个三级指标在该二级指标内的权重, 且  $\sum_{k=1}^m w_{jk} = 1$ ;

$m$  —— 隶属于该二级指标的三级指标总数。

b) 计算一级指标得分  $I_{1i}$ :

式中：

$I_{1i}$  —— 第  $i$  个一级指标的得分;

$I_{2j}$  —— 隶属于该一级指标的第  $j$  个二级指标的得分;

$v_{ij}$  —— 第  $j$  个二级指标在该一级指标内的权重, 且  $\sum_{i=1}^p v_{ij} = 1$ ;

$p$  —— 隶属于该一级指标的二级指标总数。

c) 计算系统综合得分  $P$ :

式中：

*P* —— 系统综合得分, 满分 100 分;

$I_{1i}$  —— 第  $i$  个一级指标的得分;

$u_i$  —— 第  $i$  个一级指标在总评价体系中的权重, 且  $\sum_{i=1}^q u_i = 1$ ;

$q$  —— 一级指标总数。

注：综合得分满分为100分。得分越高，表明跨平台协同水平越高。

## 6.4 评价等级判定

应根据综合得分，按表2的规定判定工业智能体跨平台协同水平的等级。

表 2 得分及等级说明

| 综合得分   | 评价等级 | 等级说明                                    |
|--------|------|---|
| 90分及以上 | 优秀   | 跨平台协同能力强, 效能显著, 安全性与兼容性优异, 可稳定支撑复杂工业任务  |
| 80-89分 | 良好   | 跨平台协同能力较强, 效能较好, 安全性与兼容性良好, 可支撑常规工业任务   |
| 60-79分 | 合格   | 跨平台协同能力基本满足需求, 效能一般, 安全性与兼容性达标, 需优化部分指标 |
| 60分以下  | 不合格  | 跨平台协同能力不足, 效能差, 存在安全性或兼容性风险, 无法满足工业应用需求 |

## 7 评价流程

## 7.1 评价程序

工业智能体跨平台协同评价工作应遵循系统性与闭环管理原则,按顺序开展评价准备、数据采集与验证、指标评分、综合评价、报告编制与归档五个阶段。各阶段应目标明确、衔接有序,确保评价过程规范、结果可信、结论可追溯。

## 7.2 评价准备

7.2.1 应组建评价团队。团队成员应涵盖工业智能体技术、工业互联网、数据安全及标准化等相关专业背景，人数不应少于 5 人。

7.2.2 应明确评价对象与范围，形成书面说明，内容应至少包括协同系统的构成与边界、应用场景、核心任务及本次评价的目的。

7.2.3 应系统收集评价所需的基础资料，包括但不限于系统技术文档、设备参数、运行记录、测试报告及管理制度文件，并对资料的完整性与有效性进行确认。

7.2.4 应制定详细的评价实施方案。方案内容应至少包括：

- a) 评价指标体系与判定标准；
- b) 数据采集计划与方法（参照第 6 章）；
- c) 人员分工；
- d) 时间进度安排及预期成果。

7.2.5 评价方案应经团队内部评审，并与委托方沟通确认。

## 7.3 数据采集与验证

7.3.1 应依据评价方案和第 6 章的要求，开展定量与定性数据采集工作，并规范填写数据采集记录表单。

7.3.2 应对采集到的原始数据进行验证与处理，形成数据验证报告。报告应说明数据质量状况、异常数据处理方式及缺失数据补充方法。

7.3.3 经验证和处理后的数据集，应作为后续指标评分的唯一输入依据。

7.3.4 网络安全等级保护相关数据的采集与验证应参照 GB/T 22239—2019 的相关要求。

## 7.4 指标评分

7.4.1 应依据第 6 章规定的方法，对全部三级指标进行评分，记录各项指标的计算过程、评分依据及佐证材料。

7.4.2 应逐级计算二级指标、一级指标得分及系统综合得分，形成指标得分汇总表。

7.4.3 指标评分与计算结果应由非直接计算人员进行独立复核，确保评分准确、计算无误。

## 7.5 综合评价

7.5.1 应根据综合得分，对照本标准规定的评价等级表，确定最终评价等级。

7.5.2 应系统分析各级指标得分所反映的系统优势、薄弱环节及潜在风险，挖掘问题产生的技术与管理根源。

7.5.3 应基于问题分析，提出具体、可操作、有针对性的改进建议，并形成综合评价分析报告。

## 7.6 报告编制与归档

7.6.1 应依据第 8 章的要求，编制完整的评价报告。

7.6.2 评价报告应依次通过评价团队内部技术审核和委托方确认。对于提出的修改意见，应予以及时处理并记录。

7.6.3 评价工作全部完成后，应将所有过程文件与最终报告进行系统归档。

7.6.4 归档资料应至少包括：

- a) 评价方案；
- b) 原始数据记录；
- c) 数据验证报告；
- d) 指标评分表；
- e) 综合评价分析报告；
- f) 评价报告及其审核记录。

7.6.5 归档资料应同时保存纸质与电子版本，并采取必要的安全与管理措施。

7.6.6 档案保存期限自评价工作结束之日起不应少于 5 年。

## 8 评价报告

### 8.1 报告内容

评价报告应全面、客观、准确地反映评价全过程与核心结论，报告正文应包含但不限于以下章节：

- a) 概述：评价的背景、目的、依据、对象与范围；
- b) 评价过程简述：评价团队、主要工作阶段及实施概况；
- c) 评价结果：各级指标得分详情、综合得分、评价等级，并附关键数据图表；
- d) 分析与建议：对系统协同水平的具体分析、发现的主要问题及改进建议；
- e) 结论：总体评价结论。

### 8.2 报告编制要求

8.2.1 报告应结构清晰、逻辑严谨、表述准确，使用规范的专业术语。

8.2.2 报告中的所有数据、图表和结论均应源于评价过程，并可追溯至原始记录和过程文件。

8.2.3 报告中的图、表应编号连续，并配有明确的图题或表题，其编排宜参照 GB/T 7713.3—2014 编写。

### 8.3 报告的审核、交付与归档

8.3.1 评价报告编制完成后，应由评价团队负责人组织内部技术审核，重点关注内容的真实性、方法的符合性及结论的客观性。

8.3.2 经内部审核定稿的报告应提交委托方进行确认。

8.3.3 最终生效的评价报告应由评价方与委托方授权代表签字或盖章确认。

8.3.4 评价工作宜在完成后 15 个工作日内完成报告交付。

8.3.5 报告的归档管理应符合本标准 7.6 的规定。

## 9 持续改进

### 9.1 评价结果的应用与反馈

9.1.1 评价委托方应将正式评价报告作为系统优化和管理决策的关键输入。

9.1.2 针对报告中提出的问题与改进建议，应制定具体的整改实施方案，明确责任主体、整改措施、资源保障、时间节点及预期目标。

9.1.3 鼓励委托方在完成主要整改措施后，申请进行阶段性复评或专项验证，以客观评估整改成效，形成“评价-改进-再评价”的闭环。

9.1.4 评价实施方应建立评价结果的后效跟踪机制，在委托方同意的前提下，定期收集系统优化后的运行数据，分析评价建议的实际应用效果，并以此作为优化评价方法、基准值的实践依据。

### 9.2 指标体系的复审与更新

9.2.1 应对评价指标体系进行定期复审。复审周期不宜超过 3 年。

9.2.2 当出现重大技术变革、产业政策调整或应用范式迁移时，应及时启动非定期复审。

9.2.3 复审工作应组建专家委员会，通过调研、研讨、公开征求意见等方式，综合评估各项指标的适用性、科学性与前瞻性。重点审视内容包括：

- a) 指标是否仍能有效表征工业智能体跨平台协同的核心特征与关键能力；
- b) 指标评价基准值（如最优标准值、目标值）是否与当前行业先进水平及普遍实践相匹配；
- c) 指标权重分配是否合理反映各维度的重要性变化；
- d) 是否有必要增补新兴技术（如数字孪生、主动元学习）相关指标，或删减已过时的指标。

9.2.4 基于复审结论，归口单位应按标准制修订程序对本文件进行修订、补充或确认延续有效。

### 9.3 评价方法的优化与经验共享

9.3.1 评价机构与人员应在实践中不断总结数据采集、验证、评分计算等方面的有效方法与常见问题，探索更高效、更精确的技术手段。

9.3.2 鼓励评价相关方，包括归口单位、评价机构、科研院所及典型应用企业，建立经验交流与知识共享平台。通过发布典型案例、编制实施指南、组织培训宣贯等方式，推广先进评价实践。

9.3.3 在遵循知识产权与保密要求的前提下，推动评价数据（经脱敏处理）和评价模型的合理共享与研究，促进跨平台协同评价领域方法论的整体提升与创新。

#### 9.4 评价活动的质量管理

9.4.1 评价实施方宜参照 GB/T 19001—2016 的原则，建立并实施覆盖评价全流程的内部质量控制体系，确保评价活动的规范性、一致性和结果的可信度。

9.4.2 宜引入第三方同行评议或能力验证机制，对评价机构的技术能力与评价过程规范性进行监督与评估，以促进评价服务市场的健康发展和评价结论的公信力。

---