

可信互联网智能体能力评估规范

编制说明 (征求意见稿)

标准起草小组
2026年9月

1. 标准“范围”的内容

本文件围绕可信互联网智能体能力评估，结合当前产业发展现状与行业治理实际需求，确立信息真实性、能力真实性、行为可信性、运行安全性四大评估维度，明确各维度的指标要求与管控规则，构建覆盖智能体全生命周期的可信能力评价体系。

本标准适用于互联网智能体的设计开发、部署运营、合规测评与监督管理，可结合不同服务场景、权限等级差异化适用，为企业能力建设、第三方评估认证及行业治理提供统一技术依据。

2. 工作简况，主要包括：任务来源、主要工作过程、各起草单位和起草人及其在起草标准过程中所承担的工作等情况、对标准草案进行会议讨论范围、征求意见的范围、审查的范围

2.1. 立项阶段

根据中国互联网协会团体标准管理相关规定，经立项评审，《可信互联网智能体能力评估规范》团体标准符合立项要求，现予以立项，项目计划号为XXX-T/ISC-XX。

2.2. 起草阶段

标准立项任务下达后，由中国信息通信研究院牵头组建标准起草小组，正式启动编制工作。起草组围绕可信互联网智能体能力评估的核心定位，面向智能体服务运营企业、安全厂商及金融、工业等重点行业用户开展调研，梳理产业可信能力建设现状、治理痛点与实际评估需求，保障标准指标贴合产业发展实际，兼具适用性与前瞻性。

编制过程中，起草组内部组织召开多次专题研讨会，针对评估维度划分、分级管控规则、场景化指标设置等核心内容反复论证，充分吸纳各方技术建议，对标准框架与技术条款持续打磨优化，形成了征求意见稿。

2.3. 征求意见稿阶段

XX

2.4. 送审阶段

XX

2.5. 报批阶段

XX

2.6. 起草单位和起草人

本文件起草单位：中国信息通信研究院。

本文件主要起草人：郭雪、卫斌、马铭洋、景韩愈。

3. 标准编制的意义；标准原则和确定标准主要内容（如技术指标、参数、公式、性能要求、试验方法、检验规则等）的依据（包括试验、统计数据）

标准编制的意义：本标准可为互联网智能体运营企业构建可信能力体系、完善信息披露机制、强化行为管控与运行安全防护提供技术指引，也可作为企业开展可信能力自评、合规自查、风险排查的参照依据，支撑第三方机构开展可信能力评估认证，助力行业建立统一的可信治理标尺。

标准制依据 GB/T 1.1-2020《标准化工作导则 第一部分：标准的结构和编写》的规则编制。

本标准立足互联网智能体产业发展现状与可信治理实际需求，围绕信息真实性、能力真实性、行为可信性、运行安全性四大核心维度构建评估体系，按操作权限与应用场景分级设置指标要求，为企业可信能力建设、行业评估与监督治理提供统一的技术规范与评价参考。

4. 主要试验(或验证)的分析、综述报告

本标准起草过程中，面向多家互联网智能体服务提供商、安全厂商及金融、工业等重点行业用户开展深度调研，系统梳理当前互联网智能体可信能力建设现状、典型风险场景与行业治理痛点，提炼形成核心评估维度与关键管控方向。

结合行业典型实践案例与多方技术意见，对评估指标的合理性、可操作性与场景适配性进行充分论证，优化评估方法与分级管控要求，经多轮完善后形成本标准。

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

在本标准的修订过程中，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

该项目没有完全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，并列出新、旧版本的有关条款(可引用标准前言的内容)；废止/代替现行有关标准的建议

本标准为制定标准，非修订标准。

8. 说明标准与其他标准或文件的关系(可引用标准前言的内容)，特别是与有关的现行法律、法规和强制性国家标准的关系

本标准非系列标准。

9. 标准作为强制性标准或推荐性标准的建议

建议本标准作为推荐性团体标准。

10. 贯彻国家标准的要求和措施建议(包括组织措施、技术措施、过渡办法等内容)；标准发布后，对国内外业界可能产生的影响

建议本标准作为协会团体标准发布实施，为互联网智能体研发运营企业构建可信能力体系、行业用户单位推进智能体选型与可信合规管控、第三方评估机构实施可信能力测评认证，以及行业治理部门履行监督管理职责提供统一的技术依据与评价参考。

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

本标准未涉及。

12. 其他应予说明的事项

本标准未涉及。