《人工智能数据智能体 总体要求》编制说明

一、工作简况

(一)任务来源

本项目根据中国互联网协会团体标准立项及相关工作安排开展制定，项目名称为《人工智能数据智能体 总体要求》。

(二)起草单位及主要起草人

本文件牵头起草单位：中国信息通信研究院；

本文件主要起草人：施建锋、刘远超、xxx。

(三)标准制定目的和意义

随着人工智能技术从模型能力提升向智能体应用拓展，数据智能体作为人工智能技术应用于数据领域的重要形态，能够围绕数据任务开展任务理解、任务规划、数据处理、工具调用及结果校验等活动，在数据采集、标注、治理、发现、传输、交付、加工、分析、运维及安全合规等业务场景逐步落地应用，为提升数据处理效率、促进数据资源开发利用、推动数据业务智能化转型提供新的技术支撑。

当前，行业在数据智能体的设计开发、系统集成、数据处理、任务执行、安全保障及运行管理方面尚未形成统一通用技术要求。现有通用人工智能、通用智能体、数据质量相关标准难以完整覆盖数据智能体在**数据任务执行与结果质量、运行稳定性与异常处理、结果可信与过程追溯、安全合规、自主行为管控、跨系统互操作协同**等方面的专项要求，产品能力参差不齐，缺少研发、选型、测试、运行的统一参考依据，制约数据智能体技术规范化落地推广。

本标准旨在建立数据智能体通用技术框架，明确总体原则、数据任务有效性与结果质量、性能与可靠性、结果可信与过程可追溯、安全与合规、自主行为与人工干预、互操作与协同性、测试与验证等方面技术要求，为数据智能体的设计开发、部署运行、产品选型、解决方案建设、测试验证提供统一依据，引导产业规范建设，提升数据智能体执行任务的有效性、可靠性、可信性与可控性，促进人工智能技术和数据业务深度融合，助力数据资源高效开发利用。

(四)主要工作过程

2026 年 10 月 21 日，召开本团体标准立项评估会，确定标准整体框架结构，收集参会单位对标准草案初稿修改意见，明确工作分工与后续编制安排。

2026 年 10 月 26 日，本团体标准由中国互联网协会正式立项，立项名称为：《人工智能数据智能体 总体要求》。

2026 年 11 月，对《人工智能数据智能体 总体要求》（征求意见稿）面向行

业公开征求意见。

二、标准编制原则和依据

(一)编制原则

1.协调性。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》开展编制，标准内容与现行法律法规、数据安全、个人信息保护、人工智能领域现有国家标准、指导性技术文件协调一致，与GB/T 45288.1- 2025《人工智能 大模型 第1部分：通用要求》、GB/Z 242- 2026《人工智能 智能体技术要求》等形成互补衔接，不重复已有标准内容。

2.科学性。结合数据智能体技术特征、数据业务场景特点，围绕任务全生命周期设置分层分类技术要求；不强制限定特定模型架构、特定实现技术路线，重点规定能力目标，而非固化技术实现路径。

3.实用性。以数据智能体完整业务主线（任务理解- 规划- 执行- 校验- 交付）为核心主线，覆盖开发、部署、运行、测试验证全流程，提出可落地、可验证的技术要求；性能考评参数仅定义统计口径，不设置统一强制阈值，具体阈值由供需双方结合业务场景确定，适配不同行业、不同规模业务的差异化需求。

4.审慎可控性。将数据安全、个人信息保护、操作风险管控、过程可追溯、人工干预机制作为底线要求；针对批量修改、删除、对外传输等高风险数据操作，强化事前确认、人工接管、操作回退相关约束，防范自主执行带来业务与数据风险。

(二)标准主要内容与确定依据

一、范围

本标准适用于企业、政府机构、科研院所等各类组织开展数据智能体的设计、开发、部署、运行及管理，也可作为数据智能体产品、解决方案、应用系统研发、建设、选型和实施的参考依据。标准聚焦数据智能体面向各类数据任务的通用能力要求，不直接规定底层大模型训练开发的技术指标。

二、规范性引用文件

列出本标准引用的基础标准，包含GB/T 45288.1- 2025 人工智能大模型通用要求、GB/Z 242- 2026 人工智能智能体技术要求、GB/T 35273 个人信息安全规范、GB/T 43697- 2024 数据安全技术 数据分类分级规则等，引用文件为标准条款提供基础支撑。

三、术语和定义

对元数据、数据血缘两个核心术语给出定义，明确标准语境下关键概念；其余通用人工智能、智能体相关术语直接引用现有国家标准与指导性技术文件，避免重复定义。

四、符号和缩略语

明确 API 等缩略语含义，便于标准文本理解与执行。

五、总体框架

数据智能体总体要求由总体原则、数据任务执行要求、执行保障与全过程控制要求构成。

六、总体原则

构建由目标原则层、数据任务执行主线、执行保障、全过程约束与控制组成的总体框架；确立有效性、可靠性、可信性、安全性、可控性、协同性六项总体原则，作为后续全部章节要求的顶层指导，明确数据智能体应当达成的基础目标。

七、数据任务有效性与结果质量要求

围绕完整数据任务链路，规定数据任务理解、任务规划与资源选择、数据活动执行、结果校验与质量、结果交付与任务反馈的相关要求。要求智能体能够识别任务歧义、能力边界，合理规划资源，识别处理数据异常，开展结果校验，如实反馈任务完成状态（完成、部分完成、未完成），避免将未通过校验结果直接交付使用。本章节解决数据智能体“能不能正确完成数据任务”的核心问题。

八、性能与可靠性要求

规定任务执行性能、运行稳定性、异常识别与容错处理、任务恢复与执行一致性、资源管理与运行监测；设置时延、吞吐量、任务完成率、服务可用率、故障恢复等考评参数，仅明确统计口径，不做统一数值强制约束。重点解决智能体并发运行、故障中断、重试恢复场景下，避免重复操作、状态错乱等可靠性风险。

九、结果可信与过程可追溯要求

规定数据来源与处理依据、结果依据核查、不确定性与能力边界说明、执行过程追溯、记录与审计支持。要求区分事实与推断结论，对不确定结果作出提示；关键处理步骤、工具调用应当可追溯，留存审计记录，支撑问题定位、结果复核。

十、安全与合规要求

包含数据安全与个人信息保护、身份认证访问控制、模型工具执行环境安全、输入输出安全与攻击防护、安全监测事件处置、合规管理。落实数据分类分级、最小必要原则，防范提示注入、越权访问、敏感数据泄露等风险，将合规约束贯穿任务全流程。

十一、自主行为与人工干预要求

明确任务授权与自主执行边界、操作风险分级与事前确认、执行过程监督、人工接管与任务中止、状态恢复与操作回退。核心约束智能体自主执行的范围，高风险操作设置人工确认机制；支持任务暂停、人工接管、操作回退，避免无约束自主操作带来数据破坏风险。

十二、互操作与协同性要求

规定数据平台与工具互操作、接口调用信息交换、数据语义一致性、多智能体任务协同、人机协作与任务交接。要求智能体能够对接外部平台工具，处理跨系统语义差异，支持多智能体分工协同，支持人机任务交接，打通跨系统协同壁垒。

十三、参考文献

列出本标准编写参考的外部标准与技术文档，作为资料性参考。

三、国内外情况简要说明

国内外已经在人工智能治理、AI 系统生命周期、数据质量、风险管理、可信 AI 方面形成部分标准与实践，ISO/IEC 系列国际标准对 AI 概念、系统生命周期、数据质量、风险管理、可信 AI 给出框架性指导；A2A、MCP 等开放技术规范对智能体协作、工具调用提供技术参考。国内已发布大模型、通用智能体、数据管理能力成熟度相关国家标准/指导性文件。

但是现有标准大多面向通用人工智能系统或者通用智能体，缺少专门针对**数据智能体**在数据任务全流程的完整通用要求，对任务规划、数据活动执行、异常恢复一致性、自主操作风险管控、多智能体数据协同等针对性要求不足。

本标准立足国内数据业务实际场景，参考借鉴国内外标准理念，自主编制形成面向数据智能体的完整通用技术框架。编制过程参考协调的国内标准包括：GB/T 45288.1- 2025、GB/Z 242- 2026、GB/T 35273- 2020、GB/T 43697- 2024、GB/T36073- 2025、GB/T45907- 2025 等。相关法律法规按照现行有效版本执行。

四、与有关的现行法律、法规和强制性国家标准的关系

本标准遵守现行法律法规、强制性国家标准，与数据安全、个人信息保护、网络安全、人工智能服务管理相关规定协调一致。本标准属于团体推荐性要求，不替代数据智能体产品提供方应当依法承担的数据安全、产品质量法律责任，也不降低相关主体法定合规义务。本标准提出的各项能力要求，仅适用于标准声明的运行环境、版本配置条件，不替代业务人员、数据管理人员的专业判断。

五、重大分歧意见的处理经过和依据

本阶段未形成需要提交专门协调程序处理的重大分歧意见。专家反馈意见均逐项研究并形成处理说明；对于标准框架章节编排、风险分级的落地表述、考评参数口径、与通用智能体标准边界划分、规范性引用文件设置等需要和归口及关联项目协调的内容，本文件予以部分采纳，并记录后续进一步优化完善的工作安排。

六、废止现行有关标准的建议

本标准为全新制定的团体标准，不涉及废止现行有关标准。

七、其他应予说明的事项

无。

《人工智能数据智能体 总体要求》团体标准编制组

2026 年 11 月