

《数字人产品功能质量分级评估规范》 标准编制说明

《数字人产品功能质量分级评估规范》起草组

2026 年 7 月

1、标准范围

本文件规定了数字人产品（包括 2D 数字人、3D 数字人）功能质量评估的评估框架、核心维度、评估流程、评分方式、评分规则、动态权重、总分计算方法及等级划分，并明确评分方式采用区间评分（代表分法）、离散评分和布尔评分三类方式。

本文件适用于面向政务、企业及行业应用场景的各类数字人相关产品和解决方案的功能质量评测与分级，可为政企单位、金融机构、电商企业、文旅教育等行业用户采购、选型和验收数字人系统或服务提供依据，也可作为数字人技术提供商开展产品自评、能力对标、等级认证以及行业协会、研究机构制定相关标准或发布评测报告的技术参考。本标准主要评价数字人产品功能质量，不评价模型训练能力、算法先进性及企业经营能力。

2、工作简况

本项目计划名称为“数字人产品功能质量分级评估规范”。本文件由中国互联网协会归口，由百度、联通在线、中关村智用人工智能研究院牵头研制，北京智谱华章科技股份有限公司、小哆智能科技（北京）有限公司、中国传媒大学数字人研究院、北京三快科技有限公司、马上消费金融股份有限公司等单位参与编制，并在多轮专家研讨、评审和意见反馈基础上形成标准终稿。

根据项目建议书，本项目自 2025 年 12 月启动。标准起草组围绕数字人产业应用场景、产品形态、评估需求和安全合规要求等方面开展调研和研讨，并结合专家评审意见和反馈意见对标准文本进行多轮完善，主要工作过程如下：

阶段	主要工作
2025 年 11 月	启动数字人标准预研究和筹备工作，召开多次研讨会对标准方向等进行交流探讨。
2025 年 12 月	在研讨基础上启动项目研究，梳理数字人产业应用场景、产品能力构成和标准化需求，形成项目建议书和标准研制思路。
2026 年 1 月-2 月	开展国内外相关标准、技术资料和行业实践调研，形成标准草案初稿，提出 TCEMS 五维评估框架。
2026 年 3 月-5 月	围绕评估维度、指标体系、评分方法、等级划分、动态权重和安全合规约束开展编制组内部讨论，完善标准草案。
2026 年 6 月	根据评估流程和分级规则要求，进一步细化基础权重、总分计算方法、等级判定方式和评分表结构。
2026 年 7 月-8 月	根据专家评审意见和参编单位反馈，对产品等级划分正文化、评估流程图示化、评分方式三分法、评分项拆分、统一测试数据集与测量规范、区间表达统一和伦理风险预警能力等内容进行修订，形成标准修订稿及本编制说明。

3、标准编制原则和确定标准主要内容的依据

3.1 编制原则

- 规范性原则。标准编制遵循 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》的要求，确保结构完整、表述规范、条款清晰，术语、符号、区间表达和表格格式保持一致。
- 系统性原则。围绕数字人产品技术实现、内容呈现、用户体验、商业落地和安全合规等关键环节构建 TCEMS 多维评估框架，形成覆盖产品功能质量全链条的评价体系。
- 可实施与可复核性原则。标准在评估流程、评分规则、统一测试数据集、测量规范、基础权重、场景权重、等级判定、评估证据留存和评分表方面给出可执行方法，便于评估机构独立实施、复测和复核。
- 科学性与一致性原则。评分档位和阈值综合参考数字人行业发展现状、典型产品能力水平、公开技术资料、行业应用实践以及专家咨询和产品调研结果设置，遵循可区分、可测量、可实施的原则；同类指标采用同类评分方式，区间表达统一为左闭右开区间，相关阈值可按标准修订程序动态更新。
- 安全合规原则。标准将安全合规设为刚性约束维度，强化数据安全、内容合规、知识产权、算法透明度和人机伦理等要求，并对重大违规情形设置一票否决条款，支撑数字人产品可信、可控、可追溯应用。

3.2 主要内容依据

本标准主要依据项目建议书、数字人产品应用需求、专家评审意见、公开技术资料、典型产品能力分布调研结果以及国内外相关标准化成果编制。国际方面，重点参考 ITU-T F.748.15 关于数字人应用系统框架和指标的思路，吸收其在多模态输入输出、交互性能、图像、语音、动画等方面的指标设置理念。国内方面，结合《GB/T 46483—2025 信息技术 客服型虚拟数字人通用技术要求》以及虚拟数字人、人工智能服务、安全合规等相关标准和行业实践，形成适用于产品功能质量分级评价的技术内容。评分档位、代表分和阈值设置综合参考行业发展现状、典型产品能力水平、公开技术资料、行业应用实践以及专家咨询和产品调研结果，并遵循可区分、可测量、可实施的原则。

标准主要内容包括：范围、规范性引用文件、术语和定义、符号和缩略语、评估框架、评估方法、评估维度与指标要求、分级规范和参考文献。其中，评估框架采用 TCEMS 模型，包含技术实力（T）、内容表现（C）、用户体验与运维（E）、商业可用性（M）、安全合规（S）五个维度。正文第 8 章统一规定评分方式说明、评分项权重、测试方法、评分规则、样本规模、统计方法、人工标注要求和证据留存要求。

评估维度	主要指标项
技术实力（T）	多模态生成、渲染与建模、动作驱动、智能交互、模型自适应。
内容表现（C）	形象逼真度、语音质量、思维能力、动作表现、输出质量。
用户体验与运维（E）	平台易用性、响应速度、系统稳定性、客户支持、集成兼容。
商业可用性（M）	成本结构、扩展性、交付效率、落地案例、生态兼容。
安全合规（S）	数据安全、内容合规、知识产权、算法透明度、人机伦理。

标准明确评估流程包括申请受理、评测准备、测试评估、评分汇总、等级判定和结果发布等环节；明确基础权重参考值为技术实力 25%、内容表现 20%、用户体验与运维 20%、商业可用性

15%、安全合规 20%；规定场景修正系数、归一化权重和功能质量总分计算方法；将产品功能质量划分为卓越级（L5）、先进级（L4）、成熟级（L3）、发展级（L2）、起步级（L1）五个等级，并在评分方式上统一采用区间评分（代表分法）、离散评分和布尔评分三类方式。

根据专家意见，标准在正文第 8 章中对评分表进行了细化，将原附录中的等级划分、评分表和相关说明调整至正文，统一采用左闭右开区间表示法，补充评分方式说明表、满分触发条件、统一测试数据集与测量规范、量化指标计算口径、评分档位设置依据、布尔子指标拆分、重大违规一票否决和评估记录示例等内容，使评分规则、测试方法和等级判定形成闭环。

4、主要试验（或验证）的分析、综述报告

本标准属于数字人产品功能质量评估与分级方法类标准，不涉及特定产品的强制性型式试验。标准起草过程中主要通过资料调研、标准对照、场景分析、指标拆解、样本测试、公式复算、专家评审和文本复核等方式开展验证。

验证方式	验证内容
标准对照验证	对照数字人、虚拟人、多模态交互、智能云服务、安全合规等相关标准和技术文件，验证标准框架与现有标准体系的衔接性。
场景适用性验证	结合政务服务、传媒传播、教育培训、金融营销、客服、品牌内容等典型场景，验证评估维度和指标项的覆盖性。
评分可操作性验证	围绕评分方式说明、代表分法、离散评分、布尔评分、统一测试数据集、测量规范、量化指标计算口径、满分触发条件、评估证据留存和总分计算进行逻辑校核，提升评分过程的一致性。
专家意见验证	根据专家意见补充产品等级划分正文内容、评估流程图、基础权重说明、评分方式说明、评分档位设置依据、统一测试数据集与测量规范、总体评价关系和伦理风险预警能力，增强标准可理解性和实施性。

经上述验证，标准提出的 TCEMS 多维评估框架、评分表拆分方式、评分方式三分法和等级判定规则能够覆盖数字人产品功能质量的主要评价要素，可为相关产品评估、分级认证、采购验收、第三方检测 and 行业对标提供可操作、可复核依据。

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

在标准起草过程中，编制组重点讨论了标准名称、正文与附录边界、基础权重与场景权重关系、评分档位逻辑、区间表达、统一测试数据集、量化指标口径、布尔评分拆分、满分触发条件、重大违规一票否决、评估证据管理以及伦理风险预警能力等问题。针对上述问题，编制组按照专家意见进行了修改完善：将产品等级划分、评分表和总分计算示例纳入正文第 8 章；统一采用区间评分（代表分法）、离散评分和布尔评分三类评分方式；补充基础权重、场景修正系数、归一化权重和总分计算公式；增加评分方式说明表、流程图、统一测试数据集与测量规范、量化指标计算口径、评分档位设置依据、评估证据管理要求、术语和表述统一要求，并强化人机伦理风险感知、预警、阻断、复核和追溯要求。

经协调和修改，目前无重大分歧意见，无需要特别说明的重要技术问题。

6、与国外标准的关系：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

经调研，本标准没有完全对应的国际标准或国外先进标准。本标准在编制过程中参考了 ITU-T F.748.15 等国际化成果中关于数字人系统框架、模态覆盖、多模态一致性和交互性能等内容，但未等同采用、修改采用或非等效采用国际标准。

本标准结合我国数字人产业发展、行业应用、产品评估和安全合规需求，突出功能质量分级、量化评分、场景权重调整和安全合规约束，形成面向产品采购、测评、认证和行业对标的分级评估方法，与国外相关框架相比更强调可操作性、等级判定和本土监管适配。

7、修订标准时，与标准前一版本的重大技术变化及废止/代替现行有关标准的建议

本标准为首创制定，不属于修订标准，不涉及与前一版本相比的重大技术变化，也不涉及废止或代替现行标准。

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

本标准属于数字人产品功能质量评估规范类标准，与现有数字人术语、系统框架、通用技术要求、人工智能服务、安全合规和数据治理等相关标准形成互补关系。现有标准多侧重于术语定义、系统能力、通用技术要求或单项能力评价，本标准进一步面向产品功能质量综合评估和等级划分，构建覆盖技术、内容、体验、商业和安全合规的量化评价框架，并通过重大违规一票否决、安全合规约束和评估证据留存等要求与现行法律法规和强制性国家标准保持一致。

本标准不设置强制性条款，不涉及降低现行法律法规、强制性国家标准和行业监管要求的内容。标准中有关数据安全、内容合规、知识产权、算法透明度和人机伦理等要求，与现行法律法规和强制性标准的基本要求保持一致，不存在冲突。

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团体标准发布实施。

10、贯彻实施标准的要求和措施建议：标准发布后对业界可能产生的影响

建议标准发布后，由中国互联网协会组织标准宣贯、培训和试点评估，引导数字人产品提供商、采购单位、测评机构和行业研究机构理解标准内容，推动标准在采购选型、产品验收、能力对标、等级认证和行业报告中的应用，并同步建立统一测试数据集、测量规范和评估记录模板。

建议结合政务服务、金融营销、教育培训、电商直播、文旅传播、智能客服等典型场景开展试点应用，收集评估数据和实施反馈，持续完善场景权重、评分阈值、测试样本、评价证据要求和复核机制。

本标准发布实施后，有助于形成统一、量化、可比的数字人产品功能质量评价体系，提升市场评价透明度，降低采购和验收沟通成本，促进数字人产品能力提升和产业规范发展；同时通过强化安全合规、一票否决和伦理风险要求，可提升数字人产品规模化应用的可信度和社会接受度。

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

本标准未引用专利条款，未发现涉及专利等知识产权问题。标准实施过程中如发现涉及知识产权事项，将按照中国互联网协会团体标准相关管理要求处理。

12、其他应予说明的事项

本文件无其他应予说明的事项。