

《AI 算法可解释性与决策透明度通用技术规范》

编制说明

2026 年 4 月

一、工作简况

（一）任务来源

本项目根据中国互联网协会 2025 年团体标准制定计划（标准计划号：143-T/ISC-25），项目名称为“《AI 算法可解释性与决策透明度通用技术规范》”的任务而进行制订。

（二）起草单位及主要起草人

本文件起草单位：中国联通国际有限公司、南京南瑞信息通信科技有限公司、珠海市人民医院、南京航空航天大学、中国信息通信研究院、北京邮电大学、中科数测科技有限公司、华兴中科标准技术（北京）有限公司。

本文件主要起草人：刘书博、朱世顺、魏兴慎、张浩天、王涵、肖雨果、于向荣、张吉、陈文弢、静静、马若龙、邵彦华、刘欣然、董坤、董婧一、成瑾、李华、任国静、丁月。

（三）标准制定目的和意义

人工智能技术是数字时代发展的核心驱动力，AI 算法的可解释性与决策透明度作为人工智能安全可控应用的关键前提，其保障能力直接关系到算法应用的公信力与社会伦理底线。《新一代人工智能发展规划》等政策明确要求“提升人工智能可解释性水平，增强决策透明度”，为 AI 算法技术规范建设提供了根本遵循。

AI 算法可解释性是指算法决策过程和结果能够被人类理解、追踪与验证的技术特性；决策透明度则是在算法设计、部署与应用全流程中，向相关方公开决策逻辑、数据来源及影响因素的机制。AI 算法可解释性与决策透明度体系是通过技术规范与管理流程的融合，将算法解释、决策追溯、风险评估等环节进行标准化，实现 AI 算法应用全生命周期的可管、可控、可追溯。

随着 AI 技术在金融、医疗、司法、公共服务等关键领域的深度应用，算法黑箱导致的决策偏差、歧视性结果等问题频发，对社会公平正义和用户合法权益造成潜在威胁。行业依赖传统经验式算法评估方法，存在解释模糊、追溯困难等问题，且在 AI 算法可解释性与决策透明度环节缺乏统一技术规范，存在显著应用风险，亟需通过标准制定填补技术空白，提升全行业算法治理水平。本标准的制定可以补充完善我国人工智能技术标准体系，推动 AI 算法可解释性技术规范化发展，同时也有助于降低算法应用风险，保障人工智能健康有序发展。

（四）主要工作过程

2025年12月本团体标准由中国互联网协会正式立项，立项名称为：《AI算法可解释性与决策透明度通用技术规范》。

2025年12月5日通过腾讯会议召开了本团体标准的第一次讨论会。本次会议确定了本标准的框架结构，收集了与会各单位对标准草案（初稿）的修改意见或建议，明确工作组各单位的工作分工，以及会后的具体工作安排。

2026年3月24日通过腾讯会议线上召开了本团体标准的第二次讨论会。编制组就本团体标准第一次讨论后的修改情况、意见采纳情况进行了综合阐述。与会专家及代表对该项团体标准的定义内容、技术内容等进行了进一步的深入讨论和修改，编制组认真听取和记录了与会代表提出的问题和调整建议。

二、标准编制原则和依据

（一）编制原则

标准起草小组在编制标准过程中，以国家、行业现有的标准为制订基础，结合我国电子油泵产品的现状，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》的规定及相关要求编制。

（二）标准范围

本文件规定了 AI 算法可解释性与决策透明度的技术要求和评估方法。

本文件适用于各类 AI 系统的全生命周期过程，包括需求分析阶段的可解释性目标锚定、数据采集与预处理阶段的透明度基础构建、模型设计与训练阶段的可解释性嵌入、部署运行阶段的解释生成与透明度公开、监控迭代阶段的可解释性优化及透明度更新、评估审计阶段的合规性验证。

本文件不替代高风险领域专项法规及国际区域性合规要求中关于可解释性与透明度的更严格规定；当行业专项法规或跨境合规要求有更具体要求时，优先遵循专项法规及区域合规要求，同时本文件提供技术适配支撑。

（三）标准主要内容与确定依据

1.算法可解释性技术要求

本章节对 AI 算法可解释性的基本要求、技术方法、动态调整要求、解释结果呈现要求、解释可验证性要求等内容做出了规定。

依据：结合行业需求和企业生产实际。

2.决策透明度技术要求

本章节对决策透明度的数据透明度、算法透明度、决策过程透明度、信息公开范围与权限控制、决策结果反馈与解释获取机制、高风险领域透明度专项要求、透明度信息动态更新要求、算法偏见的透明度披露要求等内容做出了规定。

依据：结合行业需求和企业生产实际。

3.评估方法

本章节对 AI 算法可解释性与决策透明度评估的评估维度与指标、评估流程、评估结果应用等内容做出了规定。

依据：结合行业需求和企业生产实际。

三、国内外情况简要说明

经查，暂无相同类型的国际标准与国外标准，故没有相应的国际标准、国外标准可采用。本标准达到国内先进水平。

四、与有关的现行法律、法规和强制性国家标准的关系

本标准的制定过程、技术要求的选定、试验方法的确定、检验项目设置等符合现行法律法规和强制性国家标准的规定。

五、重大分歧意见的处理经过和依据

本文件在制定过程中未出现重大分歧意见。

六、废止现行有关标准的建议

本标准不涉及对现行标准的废止。

七、其他应予说明的事项

无。

团体标准编制组

2026 年 4 月