

ICS

CCS 点击此处添加 CCS 号



中华人民共和国国家标准

GB/T XXXXX—XXXX

人工智能社会实验 评价指南

Artificial intelligence social experiment - Evaluation guideline

(征求意见稿)

(本草案完成时间：2025-07-23)

XXXX - XX - XX 发布

XXXX - XX - XX 实施

国家市场监督管理总局
国家标准化管理委员会 发布

目 次

前言 II

引言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 总体原则 1

 4.1 科学性 1

 4.2 客观性 1

 4.3 全面性 1

 4.4 目的性 1

5 评价指标体系 2

 5.1 评价指标体系框架 2

 5.2 评价指标的设立 3

 5.3 组织应用 3

 5.4 科学测量 4

 5.5 结果反馈 5

 5.6 组织保障 6

 5.7 目标成效 7

6 取值规则 7

 6.1 通则 7

 6.2 取值方法 7

7 评价流程 8

 7.1 确定评价目标 8

 7.2 制定评价方案 8

 7.3 设计指标体系 8

 7.4 设置指标权重 8

 7.5 收集数据及支撑材料 8

 7.6 分析论证 8

 7.7 编制评价报告 9

8 评价结果 9

 8.1 评价结果的得出方法 9

 8.2 评价报告 9

参考文献 10

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。

本文件由全国智能技术社会应用与评价基础标准化工作组（SAC/SWG35）提出并归口。

本文件起草单位：……

本文件主要起草人：……

引 言

人类社会正迈向以人工智能为关键技术支撑的智能社会。总结形成智能社会发展与治理的经验规律和理论，超前探索智能社会发展与治理的标准规范，完善适应智能社会发展与治理的体制机制，将有效支撑国家治理体系和治理能力现代化建设。

人工智能社会实验是探索智能社会发展与治理道路的核心方法，也是智能社会发展与治理标准化的基础方法之一。通过开展长周期、宽区域、多学科的人工智能社会实验，识别、界定、观测和评估特定智能技术社会应用场景的影响，分析可能带来的治理挑战，总结应对措施，形成解决方案和标准，对于合理规制技术应用社会风险，促进技术良性发展具有重要意义。

人工智能社会实验系列标准旨在提供人工智能社会实验设计、实施、评价等方面的标准，为开展国家智能社会治理实验基地建设等专项工作，推动智能社会发展与治理提供技术支撑，拟由以下文件构成。

——《人工智能社会实验 设计指南》。目的在于提供人工智能社会实验设计的基本方法准则，给出人工智能社会实验方法体系、设计原则、适用场景、操作方案。

——《人工智能社会实验 实施指南》。目的在于提供人工智能社会实验实施的基本原则、程序阶段，以及各阶段的操作和管理方法。

——《人工智能社会实验 评价指南》。目的在于提供人工智能社会实验评价的总体原则、评价指标体系、评价流程。

本文件在社会实验基本方法的基础上，结合我国人工智能社会实验实践需求，从评价指标体系、取值规则、评价流程、评价结果等方面确立实验评价的程序方法，为从事人工智能社会实验工作的相关技术主体、应用主体、研究主体提供参考，通过客观公正的实验评价，确保实验过程科学规范、实验成果切实有效。

人工智能社会实验 评价指南

1 范围

本文件提供了人工智能社会实验评价的总体原则、评价指标体系、评价指标的取值规则、评价流程和评价结果的指导。

本文件适用于人工智能社会实验应用主体、研究主体和技术主体等相关方开展人工智能社会实验的评价工作，也适用于各有关方面评价人工智能社会实验的综合效能。

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 XXXXXX 《人工智能社会实验 设计指南》

GB/T XXXXXX 《人工智能社会实验 实施指南》

3 术语和定义

GB/T XXXXXX 《人工智能社会实验 设计指南》和GB/T XXXXXX 《人工智能社会实验 实施指南》界定的术语和定义适用于本文件。

4 评价原则

4.1 科学性

评价方根据人工智能社会实验设计和实施的有关要求，基于实验方法、场景适配、实验干预、伦理审查等实验特点，科学确定评价指标，准确确定取值规则，将主观评价与客观测算、静态分析与动态分析、定性分析与定量分析相结合，全面给出评价结果。

4.2 客观性

被评价方如实提供评价要求的各项文件，确保文件的真实性、准确性。评价方客观、准确地开展评价，对各项文件、相关资料进行客观评审分析，真实、准确地报告评价发现与结果。评价过程有完整的文档记录，评价方对支撑评价结果的文件进行归档，确保相关结果可追溯。

4.3 全面性

评价宜分析组织应用、科学测量和综合反馈三个程序阶段（见GB/T XXXXXX 《人工智能社会实验实施指南》4.2）的规范性、有效性，也需要考虑人工智能社会实验组织实施的基本原则，充分关注智能技术在应用过程中对微观个体、中观组织、宏观社会系统产生的影响。

4.4 目的性

评价宜充分考虑实验是否形成了面向智能社会发展与治理的经验规律和政策制度, 是否能有效支撑智能社会发展与治理工作。

5 评价指标体系

5.1 评价指标体系框架

人工智能社会实验评价指标体系设置三级评价指标（见表1）。其中，一级评价指标5个，二级评价指标17个，三级评价指标47个。

一级评价指标	二级评价指标	三级评价指标
组织应用	技术类型选择恰当	技术类型合理
		技术先进性和适配性
		技术可控程度
		技术可解释性
	实验场景搭建得当	符合政策规定和中长期目标
		场景满足观测需求
	参与主体分工明确	组织有效、各主体能力满足实施要求
	实验方案科学合理	实验设计规范
		实验方法与场景适配
		实验方案要素完备
		实验阶段划分合理
	有效防范伦理风险	全过程开展伦理风险监测
伦理风险可逆性		
全过程开展科学必要的伦理审查		
科学测量	测量要素配置科学合理	测量要素与治理目标相匹配
		测量要素体现微观、中观、宏观影响
	观测变量清晰准确	可测度变量清晰准确
	抽样与分组科学合理	干扰变量明确
		科学设置实验组和控制组
		抽样方法规范
		抽样与分组过程符合伦理要求
		实验数据代表性
	数据采集与汇集科学规范	数据采集充分性
		数据采集规范性
		数据采集偏差控制程度
		数据处理规范性
结果反馈	实验结果可复现	实验复现一致性
	形成优质实验成果	实验报告具有理论意义
		产出学术论文、学术著作、专利和发明
		成果转化、形成稳定服务能力
		前沿技术场景得到有效推广
		有效支撑政策制定、标准规范

一级评价指标	二级评价指标	三级评价指标
	社会影响广泛	经验规律被采用推广
		获得宣传报道、奖励奖项
		后续监测效果
		人才培养
		国际交流与合作
组织保障	基础设施保障充分	网络、算力、数据库等设施有效配套
		组织应用所需资源充分
	资金保障合理	经费配置合理
	实验实施安全顺畅	实验过程安全、无重大事故
		数据与个人信息保护安全
目标成效	实验实施安全顺畅	建立有效的管理制度、工作机制
	聚焦技术应用社会影响及重大治理议题	聚焦技术应用社会影响及重大治理议题
	发现治理规律或提出有效治理方案	发现治理规律或提出有效治理方案

表 1 人工智能社会实验评价指标体系

5.2 评价指标的设立

一级评价指标由构成人工智能社会实验实施的三个阶段——组织应用、科学测量、结果反馈，以及实验的组织保障和目标成效构成。人工智能社会实验的实施，见GB/T XXXXXX《人工智能社会实验 实施指南》。

二级评价指标由能表征各一级评价指标的指标构成。实验组织应用阶段的活动情况，通过技术类型选择恰当、实验场景搭建得当、参与主体分工明确、实验方案科学合理、有效防范伦理风险5个二级指标来表征。实验科学测量阶段的活动情况，通过测量要素配置科学合理、观测变量清晰准确、抽样与分组科学合理、数据采集与汇集科学规范4个二级指标来表征。实验综合反馈阶段的活动情况，通过实验结果可复现、形成优质实验成果、社会影响广泛3个二级指标来表征；支撑实验实施的保障措施情况，通过基础设施保障充分、资金保障合理、实验实施安全顺畅3个二级指标来表征；实验目标和成效的实现情况，通过聚焦技术应用社会影响及重大治理议题、发现治理规律或提出有效治理方案2个二级指标来表征。

三级评价指标是直接取值的指标，由能衡量二级评价指标的指标构成，具体取值规则见第6章。

5.3 组织应用

5.3.1 技术类型选择恰当

GB/T XXXXXX《人工智能社会实验 实施指南》中5.1给出了关于确定技术类型的指南。人工智能技术类型选择恰当的评价具体包括：

- 技术类型合理，主要评价实验选取的人工智能技术类型是否明确、合理且具有典型代表性。
- 技术先进性和适配性，主要评价实验选取的人工智能技术对于实验场景是否具有先进性，是否能够聚焦城市大脑、乡村智治、智能交通等智能技术作用于社会的各类应用场景。
- 技术可控程度，主要评价实验选取的人工智能技术可控的情况，以及在突发风险时是否设置了应急响应机制等。

- 技术可解释性，主要评价实验选取的人工智能技术是否能够被清晰地阐释技术的运行逻辑和原理。

5.3.2 实验场景搭建得当

GB/T XXXXXX《人工智能社会实验 实施指南》中5.2给出了关于确定实验场景的指南。人工智能社会实验的实验场景搭建得当的评价具体包括：

- 符合政策规定和中长期目标，主要评价实验场景是否符合相关政策规定，是否与经济社会高质量发展中长期目标保持一致。
- 场景满足观测需求，主要评价实验场景是否能够满足观测人工智能技术对微观个人、中观组织、宏观治理等综合影响的需求，是否有利于调动实验有关的各项资源用于基地场景建设、实验布点、数据收集、风险控制等。

5.3.3 参与主体分工明确

主要评价人工智能社会实验的参与主体间是否有明确的组织架构和分工，各主体是否符合人工智能社会实验要求，包括应用主体是否具有承载实验的具体场景的建设条件，研究主体是否具有智能社会治理研究能力和相关领域政策制定的丰富经验，技术主体是否具有较强的实验场景搭建水平。GB/T XXXXXX《人工智能社会实验 实施指南》中5.3给出了关于确定参与主体的指南。

5.3.4 实验方案科学合理

人工智能社会实验的实验方案科学合理的评价具体包括：

- 实验设计规范，主要评价实验的设计是否符合 GB/T XXXXXX《人工智能社会实验 设计指南》第4章的内容。
- 实验方法与场景适配，主要评价选取的实验方法是否符合 GB/T XXXXXX《人工智能社会实验 设计指南》第5章的内容。
- 实验方案要素完备，主要评价实验方案要素是否符合 GB/T XXXXXX《人工智能社会实验 实施指南》中5.4的内容，要素状态是否适合于实验的开展。
- 实验阶段划分合理，主要评价实验组织应用、科学测量、综合反馈三个程序阶段划分是否合理。

5.3.5 有效防范伦理风险

GB/T XXXXXX《人工智能社会实验 实施指南》中5.5给出了有效防范伦理风险的指南。人工智能社会实验有效防范伦理风险的评价具体包括：

- 全过程开展伦理风险监测，主要评价实验方案中是否包括实验全过程的伦理风险监测，是否制定了应急预案注重实验全过程伦理风险防控。
- 伦理风险可逆性，主要评价实验方案中是否包括实验失败或产生负面后果时的退出机制或补救措施，包括数据清除、影响回溯等。
- 全过程开展科学必要的伦理审查，主要评价实验方案中是否包括实验全过程科学必要的伦理审查机制，是否明确伦理审查主体、依据、程序及责任后果，确保实验符合法规政策、技术伦理、社会实验伦理。

5.4 科学测量

5.4.1 测量要素配置科学合理

GB/T XXXXXX《人工智能社会实验 实施指南》中6.1给出了选取测量要素的指南。人工智能社会实验测量要素选取科学合理的评价具体包括：

- 测量要素与治理目标相匹配，主要评价实验测量要素的选取是否与实验要解决的社会问题或要达到的治理目标相匹配。
- 测量要素体现微观、中观、宏观影响，主要评价实验测量要素的选取是否体现技术应用对微观个人、中观组织或宏观社会系统的综合影响。

5.4.2 观测变量清晰准确

主要评价人工智能社会实验可测度变量是否清晰准确，是否将人工智能技术对组织运转模式、个人行为轨迹、社会网络和心理动态等泛意性概念转变为经济收入、出行方式变迁、机构调整、满意度和接受度等内涵清晰、概念准确的可测量的具体数据指标。GB/T XXXXXX《人工智能社会实验 实施指南》中6.2给出了明确观测变量的指南。

5.4.3 抽样与分组科学合理

GB/T XXXXXX《人工智能社会实验 实施指南》中6.3给出了科学分组的指南。人工智能社会实验抽样与分组科学合理的评价具体包括：

- 干扰变量明确，主要评价是否基于不同地区、不同领域、不同场景的现实条件，设计合适的人工智能应用场景干预手段，并明确实验干预所代表的关键变量。GB/T XXXXXX《人工智能社会实验 设计指南》中 6.1 给出了明确干预变量的指南。
- 科学设置实验组和控制组，主要评价是否根据科学采样步骤中的随机抽样结果科学设置实验组和控制组。
- 抽样方法规范，主要评价是否综合考虑样本总体的人口规模、特征等多种实验因素，评估使用概率抽样或非概率抽样，或根据实验需要结合应用两类抽样方法。
- 抽样与分组过程符合伦理要求，主要评价抽样和分组方法选择是否充分考虑特殊群体、弱势群体、区域差异、数字鸿沟等问题，是否提出具体措施保证抽样与分组过程符合知情同意、尊重隐私、平等保护、比例原则等伦理要求。
- 实验数据代表性，主要评价抽样和分组是否能够确保实验数据的科学性和典型代表性。

5.4.4 数据采集与汇集科学规范

GB/T XXXXXX《人工智能社会实验 实施指南》中6.4给出了数据采集与汇集的指南。数据采集与汇集科学规范的评价具体包括：

- 数据采集充分性，主要评价数据采集是否设置了多样化的测量维度，比如是否有利于微观个人、中观组织或宏观社会系统的影响，以及是否就每个测量维度设置了充分数量的指标。
- 数据采集规范性，主要评价是否设置了完备的数据采集流程，以及数据采集的格式是否标准化，是否可以实现在之后的研究中复用。
- 数据采集偏差控制程度，主要评价对数据采集过程中可能出现的偏差，是否设置了有效识别、量化和纠正措施。
- 数据处理规范性，主要评价对已采集和汇集的数据进行存储、使用、共享和销毁等环节时，是否符合法律法规、伦理准则。

5.5 结果反馈

5.5.1 实验结果可复现

主要评价不同研究者用同样的方法重复实验，实验复现的一致性，是否可以得到可靠的、与原实验结论相似或相同的结果。

5.5.2 形成优质实验成果

形成优质实验成果的评价具体包括：

- 实验报告具有理论意义，主要评价基于实验产出的阶段性实验报告是否具有理论意义。
- 产出学术论文、学术著作、专利和发明，主要评价基于实验是否产出了丰富的人工智能社会实验及智能社会治理相关的学术论文、学术著作、专利和发明。
- 成果转化、形成稳定服务能力，主要评价基于实验是否已有成果的高效应用转化，后续是否可形成稳定服务能力。
- 前沿技术场景得到有效推广，主要评价基于实验的前沿技术是否得到有效推广或具有较大推广价值。
- 有效支撑政策制定、标准规范，主要评价基于实验是否构建智能技术社会应用相关技术规范、评价认证体系和应用守则，形成技术范式、服务范式及政策建议，支撑智能社会治理相关法律法规、政策文件的起草，是否牵头或参与智能社会治理相关国际、国家、行业或团体标准。
- 经验规律被采用推广，主要评价实验成果是否通过政策采用、试点示范等方式实现了经验推广。

5.5.3 社会影响广泛

社会影响广泛的评价具体包括：

- 获得宣传报道、奖励奖项，主要评价实验的社会关注程度，是否获得宣传报道，是否获得奖励奖项等。
- 后续监测效果，主要评价实验结束后一定时限内跟踪监测实验社会效果是否良好。
- 人才培养，主要评价基于实验是否有较好的人才培养。
- 国际交流与合作情况，主要评价基于实验是否开展了国际交流与合作。

5.6 组织保障

5.6.1 基础设施保障充分

基础设施保障是否充分的评价具体包括：

- 网络、算力、数据库等设施有效配套，主要评价是否能够根据实验具体任务需要，配套相关网络、算力和数据库等硬件设施。
- 组织应用所需资源充分，主要评价实验组织应用所需资源是否配套充分等。

5.6.2 资金保障合理

主要评价实验开展的经费保障是否合理等。

5.6.3 实验实施安全顺畅

实验实施安全顺畅的评价具体包括：

- 实验过程安全、无重大事故，主要评价实验全过程是否安全且无重大事故。
- 数据与个人信息保护安全，主要评价实验全过程是否能够保障数据和个人信息安全。

——建立有效的管理制度、工作机制，主要评价实验的开展是否建立有效的实验工作推进机制，包括实验不同主体的工作任务是否确定，以及是否拥有强有力的实验组织领导机制推进实验所需的考核和管理等。

5.7 目标成效

5.7.1 聚焦技术应用社会影响及重大治理议题

主要评价人工智能社会实验是否聚焦技术应用社会影响及重大治理议题。

5.7.2 发现治理规律或提出有效治理方案

主要评价通过人工智能社会实验是否发现了治理规律或提出了有效治理方案。

6 取值规则

6.1 通则

评价方可以根据人工智能社会实验不同的适用场景，选择合适的取值方法。

6.2 取值方法

6.2.1 统计数据法

统计数据法是基于对统计数据的收集、整理、分析、计算得出评价指标取得指标值的方法。使用该方法时，评价标准中的取值规则应规定数据来源、起止时间、单位、计算方法等内容。

6.2.2 试验/测量法

试验/测量法是基于试验方法或测量方法计算、测量、观测得出评价指标取得指标值的方法。使用该方法时，评价标准中的取值规则宜规定所使用的试验/测量方法、试验/测量的结果、与试验/测量的结果相对应的指标值。在规定所使用的试验/测量方法时，如果存在现行适用的标准，那么宜引用这些标准中的试验/测量方法；如果没有，那么宜自行规定试验/测量方法。

6.2.3 证据判断法

证据判断法是基于评价对象所提供的证据，对照判断准则给出评价指标取得指标值的方法。使用该方法时，评价标准中的取值规则宜明确规定指标值及其判断准则、支撑作出判断的证据。作出判断的证据通常为留痕证据，例如记录、录像等。

注：使用证据判断法判定评价指标取得具体指标值时，评价对象所提供的证据能够证明符合判断准则的，即取得对应的指标值。

6.2.4 量表法

量表法是基于等级评价量表计算得出评价指标取得指标值的方法。使用该方法时，评价标准中的取值规则宜规定对应每个评价指标的等级评价量表，以两种相反的状态为两个端点，在两个端点中间按程度顺序排列不同的态度，并用不同的数值来代表某种态度。例如：最高（7）、很高（6）、最高（5）、一般（4）、低（3）、很低（2）、最低（1）；十分满意（5）、比较满意（4）、满意（3）、一般（2）、不满意（1）。

常用等级评价量表一般分为5级评价量表、7级评价量表。

6.2.5 主观赋值法

主观赋值法是基于一定的赋值依据，由评价人员根据经验、证据等情况，在规定的范围内对评价指标赋值的方法。使用该方法时，评价标准中的取值规则宜规定针对每个评价指标的赋值依据、赋值范围。

7 评价流程

7.1 确定评价目标

评价方宜基于客观公正地评价人工智能社会实验的设计、开展、结果反馈情况的目的，参考被评价方制定的人工智能社会实验方案等材料，从科学性、规范性、有效性和先进性等方面明确人工智能社会实验的评价目标，为设计人工智能社会实验综合评价指标体系提供依据。

7.2 制定评价方案

评价人员宜基于评价目标，制定完整的评价方案，包括评价组织架构、工作内容、计划进度、工作界面等相关内容。

7.3 设计指标体系

根据人工智能社会实验的评价目标，参考本文件第5章，设计选取并确定评价指标，构成人工智能社会实验评价指标体系。

人工智能社会实验评价指标体系的设计宜遵循以下原则：

- 指标体系设计符合相关国家、地方标准或规范；
- 指标体系宜包含定性指标和定量指标，两者均可测量，定量指标可计算；
- 在不影响被评价人工智能社会实验的正常开展或由此带来成本增加的情况下，可依据实验特点、所涉行业/地区的相关要求增加自定义指标；
- 所有指标分值保持百分制；
- 指标的数据项尽量保证客观并可取得。

7.4 设置指标权重

人工智能社会实验评价指标权重宜采用分级设置，按照指标分类、确定权重的方法进行设置。

7.5 收集数据及支撑材料

根据指标体系，制定详细的调查提纲，确定拟收集数据的来源、数据采集方法及支撑材料。在该阶段所要收集的数据宜包括：

- 被评价方为开展人工智能社会实验而制定的实验方案、立项及实施资料等；
- 被评价方为记录实验流程各环节数据而准备的相关资料，包括但不限于得出实验结论所依据的数据等；
- 被评价方以多种形式呈现或宣传实验结果或成果的相关资料，包括但不限于阶段性实验报告、学术论文、新闻报告等；
- 被评价方为维持人工智能社会实验正常运行所准备的资金、人员、机制保障等证明材料，包括但不限于实验经费构成说明、专家签约材料等；
- 其他有关资料，包括但不限于满足评价对象具体特点及评价要求的有关资料等。

7.6 分析论证

在充分获得数据及支撑材料的基础上，按照指标体系，对人工智能社会实验设计、开展、结果反馈全过程的科学性、规范性、有效性、先进性等状况做全面的定量与定性分析论证。

7.7 编制评价报告

将分析论证结果与评价目标、评价指标等基准相比较，得出评价结果，编制人工智能社会实验评价报告。

8 评价结果

8.1 评价结果的得出方法

人工智能社会实验评价体系可以采用各级指标加权求和的方式计算得出最终分值，见公式（1）。

$$X = A \times a + B \times b + C \times c + D \times d + E \times e \quad \dots\dots\dots (1)$$

- 式中：
- X——评价结果；
 - A——组织应用评价得分；
 - a——组织应用评价的权重；
 - B——科学测量评价得分；
 - b——科学测量评价的权重；
 - C——结果反馈评价得分；
 - c——结果反馈评价的权重；
 - D——组织保障评价得分；
 - d——组织保障评价的权重；
 - E——目标成效评价得分；
 - e——目标成效评价的权重；

8.2 评价报告

- 通常，评价报告宜包括但不限于以下方面的内容：
- 评价目标；
 - 评价对象；
 - 评价时间、地点、评价人员等；
 - 评价所依据的标准；
 - 评价过程；
 - 评价结果及其计算方法；
 - 评估建议。即针对人工智能社会实验所涉技术的社会影响及社会意义的政策建议。

参 考 文 献

- [1] 中央网信办秘书局、市场监管总局办公厅. 《智能社会发展与治理标准化指引 (2025版)》
- [2] 苏竣、黄萃. 《社会实验理论与方法评介》. 北京: 清华大学出版社, 2023.