ICS 点击此处添加 ICS 号
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 实施指南

Artificial intelligence social experiment - Implementation guideline

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

(征求意见稿)

本草案完成时间：2025-7-23

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

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估基础标准化工作组（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、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，从实验场景研究对象、观测变量、组织应用、科学测量、综合反馈等方面确立实验实施的一般流程，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，确保人工智能社会实验开展有据可依，提高实验实施的科学性、规范性和有效性。

人工智能社会实验 实施指南

1 范围

本文件提供了人工智能社会实验的基本原则和阶段程序，包括了组织应用、科学测量、综合反馈等三个阶段的操作和管理方法。

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组织实施工作，也适用于有关主体推进智能技术社会应用的场景实践。

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 评价指南》

3 术语和定义

下列术语和定义适用于本文件。

3.1

人工智能社会实验 artificial intelligence social experiment

将特定人工智能技术引入真实社会，形成人工智能应用场景，采用实验性方法，科学跟踪和分析人工智能应用场景中微观个体、中观组织、宏观社会系统的发展变化，全面研判人工智能应用的实际影响和潜在风险，进而提出促进智能社会发展与治理的政策建议的过程。

注1：人工智能社会实验主要采用自然实验、实地实验、调查实验和计算实验四类方法。

注2：人工智能社会实验的组织实施通常包括组织应用、科学测量和综合反馈等阶段。

[**来源**：GB/T XXXXXX 《人工智能社会实验 设计指南》]

3.2

实验场景 experimental scenario

人工智能社会实验进行的具体背景和环境，包括实验所处的地点、条件和相关的情境因素，涵盖实验的物理空间、设备和材料的配置，以及与实验目的和研究问题相关的其他元素。

3.3

研究对象 experimental subject

人工智能社会实验中被研究者或研究团队选择并进行观察、调查或实验的个体、组织、群体或现象。

3.4

观测变量 observational variable

人工智能社会实验中被观察、测量或记录的特征、属性或量化指标。

3.5

组织应用 application organizing

建立组织实施机制，协调各方分工合作，明确技术与应用场景、参与主体、实验方案，有效防范伦理风险，部署人工智能社会实验的过程。

3.6

科学测量 scientific measurement

选取合理的研究对象和观测变量，采用科学的抽样与分组方法，运用多种数据收集和分析手段，开展人工智能社会实验研究的过程。

3.7

综合反馈 comprehensive feedback

对实验过程和成果进行全面评价，综合技术、社会、文化等因素形成标准规范、政策建议等并反馈给相关方，开展人工智能社会实验评估和反馈的过程。

4 概述

4.1 基本原则

4.1.1 以增进福祉为目标

人工智能社会实验的实施以推动社会和智能技术的协同发展、增进人类福祉为根本导向，在技术应用场景选择、数据采集分析及成果转化中，优先促进社会效益，防范技术应用加剧社会分化，确保发展成果全社会共享。

4.1.2 保障个体基本权益

人工智能社会实验的实施严格保护个体基本权益，以技术进步促进个人发展。在实验中保障生命健康、人身自由、个人隐私、私人财产等权利不受侵犯，充分尊重对人工智能技术接受程度和使用意愿的个体化差异。通过实施知情同意、隐私保护、风险控制等全周期措施，保障参与者知情权、参加研究的自主选择权与数据安全。

4.1.3 确保可溯审慎包容

人工智能社会实验的实施确保人工智能技术安全和系统可控，为人工智能技术的研发、应用与共享提供敏捷包容的实验环境。明确应用主体、研究主体和技术主体等相关方的责任，推进智能社会发展与治理多元参与，推动政府部门、科研院所、教育机构、企业、社会团体、社会公众互动协调。开展伦理审查，建立敏捷的响应与协调机制，及时发现和化解可能引发的风险。构建多方协同的开放式创新生态，推动治理规则与技术发展同步迭代。

4.2 程序阶段

人工智能社会实验主要包括组织应用（见第5章）、科学测量（见第6章）和综合反馈（见第7章）三个程序阶段（见附录A），具体来说，遴选有代表性的人工智能实验场景，随机建立实验组和控制组，通过科学抽样，甄别运转模式、行为轨迹、社会网络和心理动态等泛意性概念，转化为内涵清晰、概念准确的可测量指标，采用科学的方法进行测量和数据处理，分析智能化转型的综合影响，最终形成技术规范、技术标准、政策建议并进行反馈。

本文件以民生诉求处置场景为样例，给出了人工智能社会实验组织实施中各程序阶段主要工作的参考示例（见附录B）。

5 组织应用

5.1 明确技术类型

选取具有典型代表性的人工智能技术，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，观测人工智能技术对微观个人、中观组织或宏观社会系统的综合影响。明确人工智能技术的具体类型，充分考虑技术对于实验的先进性和适配性。

5.2 明确实验场景

具体实验场景可选择已有场景或新建场景；若为新建场景，宜充分评估，减少重复建设。新建的实验场景宜符合相关政策规定，与经济社会高质量发展中长期目标保持一致。场景可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，满足观测人工智能技术对微观个人、中观组织、宏观社会系统等多层面和维度的综合影响的需求。

示例1：人工智能技术作用于微观个人的主要应用场景包括自动驾驶、智能家居、智能客服等。例如利用生物特征识别、行踪轨迹等敏感信息进行身份验证，通过智能传感技术感知周围环境进行辅助驾驶，智能监测个人健康状况为医师诊断和治疗提供辅助参考，基于推荐算法的智能检索和信息推送服务等。

示例2：人工智能技术作用于中观组织与行业的主要应用场景包括智慧安防、智慧教育、智慧医疗等。例如公安机关使用高精图像识别技术实施监测侦察；线上教育平台利用大数据智能技术进行师生画像，医疗机构运用图像识别技术、自然语言处理进行医学影像智能识别并提供辅助诊断参考等。

示例3：人工智能技术作用于宏观社会系统的主要应用场景包括数字政府、智慧城市、数字乡村、智慧应急管理等。如通过建立城市大脑检测调度城市运行情况；利用能源大数据预测用能需求、经济发展等。

5.3 明确参与主体

实验宜有明确的应用主体、研究主体、技术主体，且各主体间分工明确。应用主体具有承载实验具体场景的建设条件，研究主体具有智能社会治理研究能力和参与相关领域政策制定的丰富经验，技术主体具有较强的实验场景搭建水平。

5.4 明确实验方案

实验方案可包含研究问题、实验设计、样本对象、数据测量、分析计划、伦理考量等部分。实验设计的有关考虑见GB/T　XXXXXX 《人工智能社会实验 设计指南》。

5.5 防范伦理风险

实验宜明确伦理风险监测措施，制定应急预案，注重全流程伦理风险防控，开展必要的伦理审查，按要求跟踪监督人工智能社会实验活动全过程，确保实验符合法规政策、技术伦理、社会实验伦理。根据实验不同情形，可采取会议审查、简易审查、应急审查等伦理审查方式。伦理审查流程宜遵循申请与受理、审查、决定、跟踪审查、申述与专家复核等步骤。

示例1：在实验场景的搭建方面，审查内容主要包括：场景对社会、经济、环境等方面的影响情况；实验地点、实验领域的技术风险以及实验场景对研究参与者可能造成的安全风险等。

示例2：在实验方法的选取方面，审查内容主要包括：对研究参与者在心理、情感方面的影响；对实验对象可能造成的差别对待等。

示例3：在研究对象的确定方面，审查内容主要包括：研究对象的信息安全、隐私和数据保护情况；特殊群体、弱势群体的利益权重的平衡性等。

示例4：在研究对象知情同意方面，审查内容主要包括：知情同意书的内容是否包含研究目的、研究内容、收益与风险等；研究者对研究信息的告知是否充分详尽，态度话语是否存在诱导或胁迫；无行为能力、限制行为能力和无法自己做出决定的研究对象时，签署过程是否合理，是否得到其监护人或法定代理人的书面知情同意等。

示例5：在观测变量的设定方面，审查内容主要包括：对研究对象自主性的影响情况；数据来源的合法性，数据权属处理等。

示例6：在实验的组织实施方面，审查内容主要包括：实施前对人工智能的系统稳健性和安全性的风险评估与应急预案情况；实验过程的公开透明、组织流程清晰情况及实验数据的记录、人员操作记录留存情况等。

示例7：在实验数据的分析方面，审查内容主要包括：数据分析、处理是否符合法律法规；数据保护情况；采用数据分析或处理技术时，处理或分析规则是否在平等、非歧视方面具有潜在影响等。

示例8：在实验结果的反馈方面，审查内容主要包括：研究结果对研究对象和社会的影响情况；研究对象对实验结果应用与发布的知悉及认可情况；成果推广对公民合法权利的影响情况等。

6 科学测量

6.1 选取测量要素

围绕实验要解决的社会问题或要达到的治理目标，选取科学合理的测量要素。要素宜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。

示例1：观察人工智能技术在微观个人层面产生的影响，可选取心理状态、行为模式、价值观念、社会关系的变化情况等作为测量要素。例如人脸识别技术应用对个人安全感的影响，自动驾驶对个人出行范围、社交网络和就业的影响等。

示例2：观察人工智能技术在中观组织与行业层面产生的影响，可选取业务流程、组织关系、责任机制和运行模式的变化情况等作为测量要素。例如智能制造技术应用对行业和组织管理流程再造的影响，智能电网应用对供电企业能源分配和管理方式的影响等。

示例3：观察人工智能技术在宏观社会系统层面产生的影响，可选取公共治理方式、社会服务能力、应急响应流程、公众参与模式的变化情况等作为测量要素。例如互联网平台信息传播对于社会舆论的影响，数字金融技术应用对于金融体系和经济安全的影响，智慧应急建设对于风险管理能力的影响等。

6.2 明确观测变量

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的影响等泛意性概念，转化为经济收入、出行方式变迁、机构调整、满意度和接受度等内涵清晰、概念准确的可测量数据指标。

示例1：评估人工智能技术在微观个人层面产生的影响，可选取个体行为和个体主观感受等作为观测变量。例如用户黏性、使用频率、活动轨迹等个体行为，以及风险认知、技术接受、获得感等个体主观感受等。

示例2：评估人工智能技术在中观组织与行业层面产生的影响，可选取经济效益变化情况和产业生态发展情况等作为观测变量。例如通过在线销量数据分析智能推荐技术应用前后对降低库存的影响，以及组织治理或者行业协同水平变化情况，通过收集产业生态发展情况等数据观察人工智能技术应用对战略实施情况、行业协同效能等的影响。

示例3：评估人工智能技术在宏观社会系统层面产生的影响，可选取政府治理与决策的效率效果和公共服务质量的优化改进情况等作为观测变量。例如通过公共服务平台的数据变化分析人工智能技术应用对应急响应与资源调配速率的影响，以及通过终端传感器运行数据或舆情监测数据分析公共服务均衡度与满意度的改善。

6.3 科学抽样与分组

如研究人工智能应用场景对特定地区或特定领域总人口产生影响的社会实验，宜在综合考虑总人口规模、分布等因素基础上，运用科学的抽样方法，选定具有代表性的被试对象。

根据社会实验的主要目的和内容，结合人口特征和可用资源，确定调查样本总体。结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。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，选择合理的实验类型和规范的抽样方法。充分考虑特殊群体、弱势群体、区域差异、数字鸿沟等问题，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，保证实验数据的科学性和代表性。

示例1：评估人工智能技术在民生诉求处置场景中的综合影响时，可选取实验区域全域民生诉求相关主体作为样本总体，结合区域人口规模、民生诉求事件分布特征及治理资源差异，选取部分区域作为实验组，其他区域作为控制组。实验组通过搭建智能平台整合多元民生诉求受理渠道，应用全流程智能办件体系；控制组保留原有传统受理、分拨及处置模式，通过两组在受理效率、分拨精准度、处置成效等维度的差异，量化评估人工智能技术的实际影响。

6.4 数据采集与汇集

依据科学、规范的社会科学测量手段，综合利用观察记录、问卷调查、大数据捕捉与社会计算等方法，进行数据采集与汇集，落实隐私保护、信息安全及公共安全防护措施，建立符合伦理准则和数据安全要求的全生命周期数据管理体系。

7 综合反馈

7.1 结论提取

通过科学测量情况，提取并形成关于人工智能技术在社会治理中的应用效能，以及对提升公共服务质量、优化社会资源配置等产生具体影响的研究结论。相关结论可以论文、报告、专利、数据库（平台）等形式呈现，并确保其具有理论意义和科学价值。

7.2 综合集成

综合考虑技术发展的可能性、社会个体的接受程度以及经济文化环境等因素，结合社会实验过程和成效，形成标准、规范和政策建议等实验成果。

7.3 实验评价

对实验过程和成效进行评价，内容包括实验场景建设情况、组织与保障机制、风险防范与伦理审查、工作计划执行效果、实验成果及转化等。

实验评价的有关考虑见GB/T　XXXXXX 《人工智能社会实验 评价指南》。

7.4 成果反馈

基于全周期的过程评价和多因素观察，将人工智能社会实验的成果进行反馈，将形成的标准、规范、政策建议等，反馈给技术研发者和政府实践应用部门，为技术研发者提供方向指引，为政府实践应用部门提供决策依据，保障相关治理措施的有效跟进和实施。

附 录 A
（资料性）
实验阶段划分

本附录以流程图的形式给出人工智能社会实验组织实施的阶段划分，目的是为理解本标准中人工智能社会实验的组织应用、科学测量、综合反馈等实施环节提供实践参考。

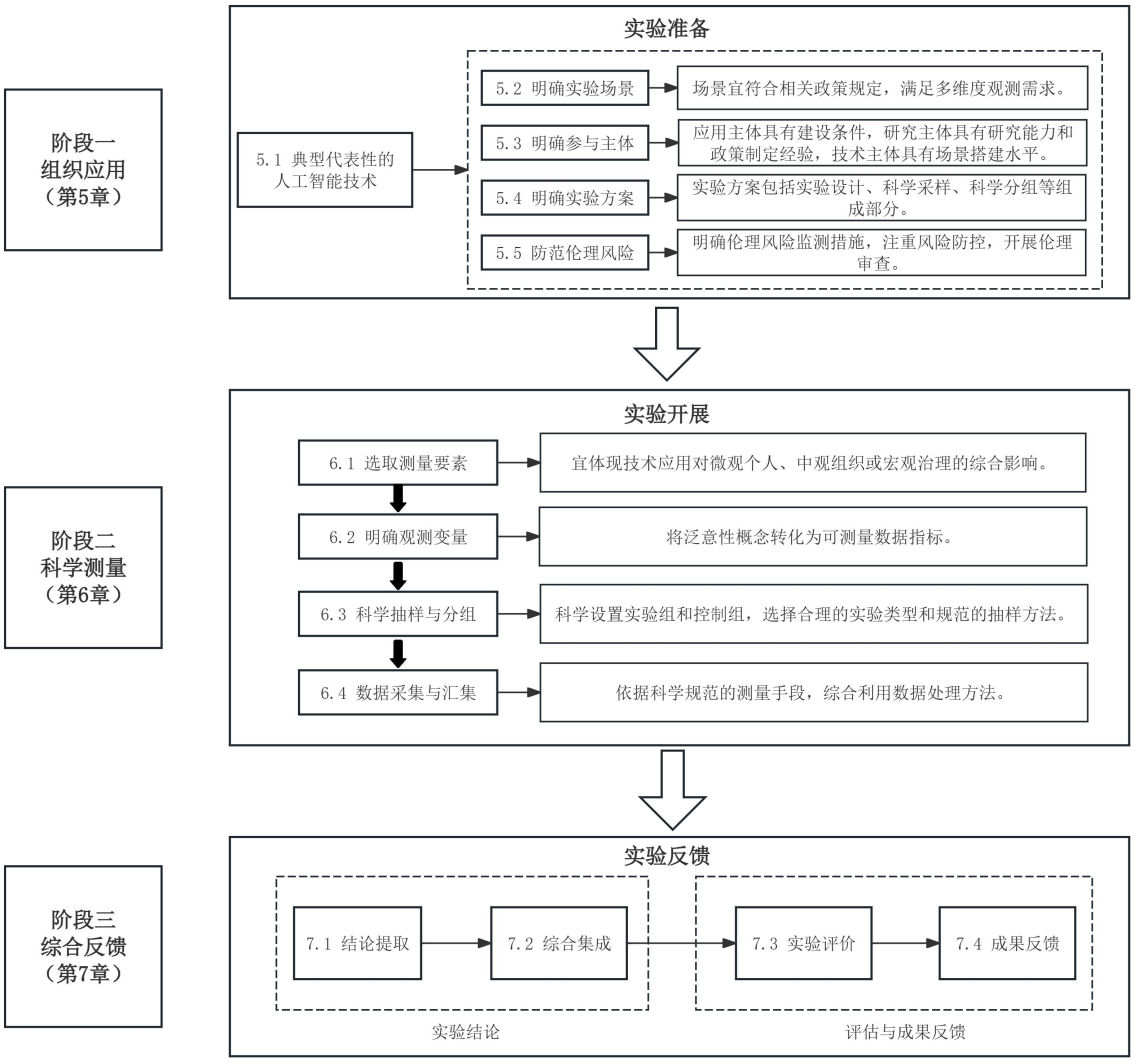

图 A.1 实验阶段划分

附 录 B
（资料性）
实验实施参考示例

本附录以民生诉求处置场景为样例，给出了人工智能社会实验组织实施的参考示例，目的是为理解本标准中人工智能社会实验的组织应用、科学测量、综合反馈等实施环节提供实践参考。示例仅是为了说明相关实施过程，不保证适用于所有人工智能技术应用场景。

B.1 组织应用

在民生诉求处置场景的人工智能社会实验中，明确以人工智能技术革新民生诉求系统为核心目标，聚焦信息收集、处理、反馈等全场景开展智能应用实验。实验参与主体包括技术主体（负责搭建智能平台）、研究主体（联合应用主体和技术主体开展实验并提炼理论结论）及应用主体（负责实际场景落地与协同）。

实验采用对比验证型方法，按照 “小切口、深层次、渐进式” 思路，针对民生诉求受理、分拨、处置、评价等关键环节设计实验方案，选取智能平台、系统应用、业务清单、保障机制等不同服务模式进行对照。同时建立组织实施机制，协调各方分工合作，在实验过程中关注伦理风险防范，重点观测人工智能技术对民生诉求处置全流程的综合影响。

B.2 科学测量

实验从微观、中观、宏观三个层面选取测量要素并明确可测度指标:

微观个人层面：市民提交民生诉求的方式便捷度、对诉求处置的满意度；

中观组织层面：政府部门受理方式规范性、处置协同效率（具体包括分拨效率、处置速度、处置效果、服务态度、反馈质量等）；

宏观社会系统层面：对社会治理 “一网统管” 模式的促进作用。

结合区域人口规模、民生诉求事件分布特征及治理资源差异，选取部分区域作为实验组，其他区域作为控制组。实验组通过搭建智能平台，整合电话、邮箱、社交媒体账号、应用程序等多元民生诉求受理渠道，并纳入“AI + 视频”等智能采集途径实现统一受理分拨；深度融合人工智能算法与语义分析模型，构建全流程智能办件体系及主题分析模型。通过对受理、分拨、处置、督办、评价等环节的数据采集与分析，量化评估技术应用成效。

B.3 综合反馈

实验完成后，结合综合技术应用效果、社会接受度及区域治理环境等因素，对实验过程和成效进行全面评价。实验成果包括智能平台建设规范、民生诉求处置流程优化建议等，已反馈至区域治理相关方，并逐步推广至更大范围。相关经验为人工智能技术赋能社会治理、提升民生服务效率提供了实践参考，获上级部门推广及调研关注，为智能技术支撑民生领域改革提供了可复制的经验模式。

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.