



中华人民共和国国家标准

GB/TXXXXX—XXXX

人工智能社会实验 设计指南

Artificial Intelligence social experiment - Design guideline

(点击此处添加与国际标准一致性程度的标识)

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 I

引言 II

1 范围 3

2 规范性引用文件 3

3 术语和定义 3

 3.1 人工智能社会实验 3

 3.2 自然实验 3

 3.3 实地实验 3

 3.4 调查实验 3

 3.5 计算实验 4

4 通则 4

 4.1 遵循科研伦理 4

 4.2 遵照实验逻辑 4

 4.3 执行随机处理 4

 4.4 规范设计流程 4

5 实验方法选取 4

 5.1 自然实验 4

 5.2 实地实验 5

 5.3 调查实验 5

 5.4 计算实验 5

6 实验干预设计 5

 6.1 明确干预变量 5

 6.2 开展预实验 5

 6.3 科学选取被试 5

 6.4 随机化分组 5

 6.4 干预效果评估 6

7 实验指标测量设计 6

 7.1 统计类指标 6

 7.2 调查类指标 6

 7.3 行为痕迹类指标 6

8 实验数据处理设计 7

 8.1 样本描述分析设计 7

 8.2 因果效应分析设计 7

 8.3 因果机制分析设计 7

参考文献 8

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和规程的基础上，结合我国人工智能社会实验实践需求，从实验设计的原则、适配场景、实验类型、实验干预、指标测量、数据处理等方面确立实验设计的一般方法，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，确保人工智能社会实验开展有据可依，提高实验设计的科学性、规范性。

人工智能社会实验 设计指南

1 范围

本文件提供了人工智能社会实验设计的基本方法准则，给出了人工智能社会实验方法体系、设计原则、适用场景、操作方案。

本文件适用于人工智能社会实验相关应用主体、研究主体和技术主体等相关方开展人工智能社会实验的设计工作。

2 规范性引用文件

本文件没有规范性引用文件。

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人、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：人工智能社会实验通常包括组织应用、科学测量、综合反馈三个实验阶段。

3.2

自然实验 natural experiment

借助由自然事件、重大政策、偶然事件等外生因素创设的人工智能应用场景，寻找其对于社会群体的近似随机干预效果，进而将样本划分为实验组和对照组，观测和采集不同分组受试者的表现数据，比较分析实验组和对照组的组间差异，判断人工智能应用对被试对象产生影响程度的社会实验方法。

3.3

实地实验 field experiment

由实验者自主构造人工智能应用场景，并随机将其分配给被试对象，确定实验组和对照组，观测和采集不同分组受试者的表现数据，比较分析实验组和对照组的组间差异，判断人工智能应用对被试对象产生影响程度的社会实验方法。

3.4

调查实验 survey experiment

研究者通过将自主构造的人工智能应用场景描述信息嵌入到问卷调查中，并随机分配给被试对象，确定实验组和对照组，观测和采集不同分组受试者应答问卷的表现差异，判断基于描述信息模拟的人工智能应用场景对被试对象产生影响程度的社会实验方法。

3.5

计算实验 computational experiment

通过计算仿真建模，对难以在真实社会中构造或开展追踪观测的人工智能应用场景进行模拟，再现社会系统的基本环境、不同类型主体的行为特征与交互关系，通过对模拟样本动态变化的比较分析，揭示人工智能技术应用影响社会活动演化规律的社会实验方法。

4 通用规则

4.1 遵循科研伦理

宜严格谨慎遵循尊重、不伤害、有利、公正等基本科研伦理，充分尊重被试对象的自主性，确保被试者的知情同意权、数据信息隐私权得到有效保护，降低被试对象需要承担的风险和伤害，充分保障研究的程序公正、回报公正、分配公正，强化对弱势群体的保护，防止引发新的不平等现象。

4.2 遵照实验逻辑

宜严格遵照“控制—对照—比较”的实验推理逻辑，明确区分实验组和对照组。实验组和对照组被试对象之间的关键差异为是否受到待检验人工智能应用场景干预的影响，并在其他变量上尽可能接近。

4.3 执行随机处理

宜通过被试对象的随机化处理，消除实验组和对照组在实验前存在的系统性差异，增强人工智能社会实验内外部效度与信度，具体包括：

- a) 在实验方案设计阶段，宜根据问题类型，综合运用简单随机化、分层随机化、区块随机化等随机化方法，设计随机化处理方案。
- b) 在执行实验设计阶段，宜检查是否存在实验组未接受干预和对照组接受干预等异常情况，并对相关被试对象进行处理。
- c) 对于自然实验等难以实现完全随机化的实验类型，宜关注影响被试对象接受干预的因素，并在数据分析过程中加以控制。

4.4 规范设计流程

宜根据不同场景和研究问题的特点，选择适配的实验方法（见第5章），并有序开展实验干预设计（见第6章）、指标测量设计（见第7章）和数据处理设计（见第8章）。

5 实验方法选取

5.1 自然实验

自然实验适用于智能技术作用在宏观社会系统、中观组织和微观个人层面影响分析。如人工智能应用场景政策规划的社会影响评估、各类人工智能应用平台上线的效果评估、人工智能应用相关突发事件的社会影响评估。

5.2 实地实验

实地实验适用于智能技术作用在中观组织和微观个人层面影响分析。如在机构、行业或特定社会单元部署，对不同人工智能应用功能的效能效果评估；在不同人工智能应用条件中，对用户行为和心理认知的评估研究。

5.3 调查实验

调查实验适用于智能技术作用在微观个人层面的影响分析。如微观个人对于现实世界中尚不存在的假想人工智能应用场景的态度与认知研究；微观个人对于现实世界中大规模推广成本较高或暂未实现大规模推广、仅在小规模试点应用的人工智能应用场景的态度与认知研究。

5.4 计算实验

计算实验适用于智能技术作用在宏观社会系统和中观组织层面影响分析。如人工智能应用场景的社会舆情演化、经济效益模拟、行业生态模拟研究；人工智能应用场景下微观个人心理认知对宏观社会系统和中观组织的影响机制研究。

6 实验干预设计

6.1 明确干预变量

宜根据不同地区、不同领域、不同场景条件，设计相匹配的人工智能应用场景干预手段。干预变量可以为人工智能应用本身或基于人工智能应用衍生的新兴制度、商业模式、政策措施等场景。干预形式可以为被试者在物理空间中接受的真实干预，或借助网页、视频等现代信息手段构建的模拟干预。

6.2 开展预实验

正式开展实验前，宜专门设计覆盖一定规模样本的预实验环节，用以了解样本对实验干预的潜在反应是否符合预期，避免样本因感知实验干预过于困难、过于简单、过于不切实际等问题而出现异常实验反应。

6.3 科学选取被试

宜根据实验的主要目的和内容，结合研究对象的特征和可用资源，科学选取被试。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。

6.4 随机化分组

宜根据研究问题和被试对象的特点，设计适当的随机化分组处理方案，设计要点包括：

a) 宜优先采用完全随机化分组，通过随机数生成等机制，确保每个被试对象被分配到实验组或对照组的概率完全相同，最大限度消除选择偏差。

示例：研究人工智能教学软件对100名学生学习效果的影响，可使用随机数表将学生编号1-100，抽取奇数进入实验组（使用智能教学软件），偶数进入对照组（不使用智能教学软件）。

b) 若不能实现完全随机化分组，宜根据被试对象的某些重要特征进行分层选取，在分层内实现随机化分组。

示例：研究人工智能诊疗系统对医院诊疗效果的影响时，可先按医院等级（三甲/二甲）分层，再在各等级内随机分配医院至实验组或对照组，避免医院水平等级差异干扰评估结果。

6.4 干预效果评估

宜设计专门环节，对实验干预效果进行评估，以确保实验干预的科学性、有效性。针对不同类型的人工智能社会实验方法，实验干预效果评估关注的要点分别包括：

——开展自然实验时，宜重点关注被试对象分组的随机化程度，即是否有混淆因素显著影响了被试接受干预的概率，以及实际中的分组情况与理论上预期的分组情况存在多大程度的差异。

——开展实地实验时，宜重点关注被试对象对于实验干预的遵从程度，即受试者在多大程度上完全遵循实验者的安排。

——开展调查实验时，宜重点关注被试对象在阅读实验干预材料时的专注程度和对实验干预材料的理解程度，确保实验对象准确理解干预材料试图传递的干预信息。

——开展计算实验时，宜重点关注仿真模型基础参数设定的来源、依据及其与真实情况的拟合程度。

7 实验指标测量设计

7.1 统计类指标

统计类指标宜作为研究特定人工智能应用场景对宏观社会系统、中观组织所产生影响的测量指标，也可在研究人工智能应用场景对中观组织和微观个人所产生影响时，用以测量中观组织和微观个人所处的外部环境特征。统计类指标包括但不限于：

a) 宏观社会经济发展指标，如国民经济统计指标、社会发展统计指标、人口普查统计指标等。

b) 中观组织、行业层面的专业数据统计指标，如工商企业注册相关数据统计指标、全国税收相关数据统计指标、上市公司相关数据统计指标、专利相关数据统计指标。

7.2 调查类指标

调查类指标宜作为研究特定人工智能应用场景对中观组织和微观个人所产生影响的测量指标。调查类指标包括但不限于：

a) 被试对象的人口统计特征指标：如被试对象的年龄、性别、地区、职业、政治面貌、婚姻状况等指标。

b) 被试对象的心理认知指标：如被试对象的态度、认知、偏好与观点类指标，宜充分利用现有科学文献已经编制成熟并经过反复检测的量表进行测量。

7.3 行为痕迹类指标

行为痕迹类指标宜作为研究特定人工智能应用场景对中观组织和微观个人的行为所产生影响的直接观测指标。行为痕迹类指标包括但不限于：

a) 结构化的行为痕迹指标，如微观个人被试对象使用智能技术应用的时长、地理空间足迹、社交媒体好友关系，中观组织被试对象的研发投入、纳税额、专利申请等。

b) 非结构化的数字痕迹指标，如被试对象对新闻消息的评论数据、被试对象在社交媒体发布的信息数据等。

8 实验数据处理设计

8.1 样本描述分析设计

对实验数据的处理，宜首先对被试对象的分组规则和基本统计结果进行描述和分析，证明实验组与对照组在控制变量层面无显著差异，并验证实验干预对被试对象的结果变量产生了影响。分析项目包括但不限于：

- a) 观测样本数目
- b) 各组变量数据的均值、中位数、四分位数、最大值、最小值
- c) 各组变量数据的标准差、方差
- d) 各组变量数据分布的偏度、峰度
- e) 实验组和对照组被试对象涉及的不同变量数据的平均值差异比较结果

8.2 因果效应分析设计

对人工智能应用场景干预的因果效应的分析，宜根据实验场景和研究问题的特点，选择适当的数理统计分析模型，如双重差分模型、断点回归模型、工具变量模型、合成控制模型等。

8.3 因果机制分析设计

在分析人工智能应用场景干预的因果效应基础上，宜进一步揭示人工智能应用场景发挥作用的机制，如利用中介效应、调节效应等定量分析方法，分析人工智能应用场景的作用机理，或通过参与式观察、内容分析、深度访谈等定性研究方法，深度挖掘人工智能应用场景的作用机制。

参 考 文 献

- [1] 中央网信办秘书局、市场监管总局办公厅. 《智能社会发展和治理标准化指引（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.