



中华人民共和国国家标准

GB/T 46801—2025

人工智能社会实验 设计指南

Artificial intelligence social experiment—Design guidelines

2025-12-02 发布

2025-12-02 实施

国家市场监督管理总局 发布
国家标准化管理委员会

目 次

前言 III

引言 IV

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 2

 4.1 遵循科研伦理 2

 4.2 遵照实验逻辑 2

 4.3 执行随机处理 2

 4.4 规范设计流程 2

5 实验方法选取 2

 5.1 自然实验 2

 5.2 实地实验 2

 5.3 调查实验 3

 5.4 计算实验 3

6 实验干预设计 3

 6.1 明确干预变量 3

 6.2 开展预实验 3

 6.3 科学选取实验对象 3

 6.4 随机化分组 3

 6.5 干预效果评估 3

7 实验指标测量设计 4

 7.1 统计类指标 4

 7.2 调查类指标 4

 7.3 行为痕迹类指标 4

8 实验数据处理设计 4

 8.1 样本描述分析设计 4

 8.2 因果效应分析设计 5

 8.3 因果机制分析设计 5

参考文献..... 6

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。

GB/T 46801《人工智能社会实验 设计指南》与 GB/T 46802《人工智能社会实验 实施指南》和 GB/T 46799《人工智能社会实验 评价指南》共同构成支撑人工智能社会实验开展的国家标准体系。

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/SWG 35)提出并归口。

本文件起草单位：浙江大学、合肥清和智能科技有限公司、清华大学、中共鄂尔多斯市委员会网络安全和信息化委员会办公室、中国标准化研究院、润泽智算科技集团股份有限公司、浙江省人民医院、上海计算机软件技术开发中心、深圳市坪山区政务服务和数据管理局、之行智库(杭州)有限公司、灵心巧手(北京)科技有限公司、贵州大学、枣庄市网络安全保障中心、哈尔滨工业大学(深圳)、青岛通产智能科技股份有限公司。

本文件主要起草人：黄萃、苏竣、周超男、钟军、魏钰明、尚二平、王永杰、张芳、汝鹏、陈俊华、刘洋、金东、李笠、张娴、祝敬、沈诚、曹岗、池浩、王宇、李翰林、吕立远、岑晓腾、杨庆丰、陈敏刚、王旭、郭腾、何道敬、郭小璇。

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路径的关键方法,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,系统识别、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,形成解决方案和规范,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。

我国已经建立了支撑人工智能社会实验开展的国家标准体系。在该标准体系中,GB/T 46801《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、操作建议,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,指导人工智能社会实验相关工作的开展,从而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

本文件确立了人工智能社会实验设计的总体原则和阶段程序,提供了人工智能社会实验方法体系、设计原则、适用场景、操作方案等方面的指导和建议,并给出了示例。

本文件适用于指导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社会活动,形成人工智能应用场景,通过设计实验,将应用场景中的泛意性影响转变为内涵清晰、概念准确的可测度变量,并采用科学的方法进行跟踪和分析微观个体、中观组织、宏观社会系统的发展变化,以全面评估人工智能应用的实际影响和潜在风险,进而提出对策建议的过程。

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征数据,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,采集和观测不同分组受试者的表征数据,比较分析实验组和对照组的差异,评估人工智能应用对被试对象产生影响效应的社会实验方法。

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