《人工智能社会实验 设计指南》

(征求意见稿) 编制说明

一、工作简况

(一) 任务来源

《人工智能社会实验 实施指南》是根据国家标准化管理委员会关于下达2024年第七批推荐性国家标准计划及相关标准外文版计划制定的，项目计划编号“20242886-T-469”，计划完成时间为2025年。该计划项目由全国智能技术社会应用与评估基础标准化工作组（SWG 35）提出并归口。

(二) 制定背景及必要性

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。2019年以来，按照中央领导同志部署要求，中央和国家有关部门大力推动开展国家智能社会治理实验基地建设，通过组织长周期、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。

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。《人工智能社会实验 设计指南》旨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。

(三) 起草过程

按照国家标准制修订程序的要求,《人工智能社会实验 设计指南》国家标准的编制完成了以下工作:

1. 资料的收集

在编制过程中,起草工作组收集研究了以下资料:

——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)》;

——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。

2. 文件的起草

(1) 2024年9月29日,本标准正式经过国家标准化管理委员会的立项审批,并下达标准制定计划任务。

(2) 2024年11月,项目组召开会议,正式启动《人工智能社会实验 设计指南》国家标准的编制工作,明确了标准编制的基本思路和计划进度安排,组建起草工作组;起草组完成资料收集、翻译及分析研究工作,确立了文件的基本技术内容,形成工作组讨论稿第一稿。

(3) 2024年11月-2025年7月,起草工作组先后于2024年12月、2025年3月、2025年5月、2025年7月召开了4次标准文本讨论会,针对文件中的关键性问题进行专家研讨,根据会议结论及专家提出的意见建议,逐条研究修改,形成征求意见稿。

二、编制原则、主要内容和确定依据

(一) 编制原则

本标准制定过程中,主要遵循以下几个原则:

1.规范性。在标准的结构设计、要素表述以及标准化术语使用等方面,本标准严格遵循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》和GB/T 20001.7-2017《标准编写规则 第7部分:指南标准》和GB/T 20001.8-2023《标准起草规则 第8部分:评价标准》的有关条款。

2.系统性。人工智能社会实验系列标准包括《人工智能社会实验 设计标准》《人工智能社会实验 实施标准》《人工智能社会实验 评价标准》,涵盖人工智能社会实验的设计、实施和评价三个关键环节,从实验目标设定、方案规划到实

验过程管理、结果评估，形成一个完整的标准体系。三项标准之间结构紧密衔接、内容相互呼应、逻辑高度统一，确保实验活动的全流程规范化和标准化，避免各环节之间的脱节和冲突。

3.可操作性。人工智能社会实验系列标准立足我国人工智能技术应用与社会治理的现实场景，深度融合实验基地实践经验及典型案例，这套标准方法体系为提供人工智能社会实验设计的基本方法论准则，而且最终总结形成面向智能社会治理的经验规律和政策制度，能够为后续应用示范案例提供方法层面的支撑，为实验应用主体、研究主体和技术主体提供清晰具体的可操作性指导，强化实验成果向治理政策的转化效能，有效支撑人工智能社会实验的顺利开展和规范实施。

4.前瞻性。人工智能社会实验系列标准在充分调研智能技术发展趋势、国内外先进实践经验和未来应用场景的基础上，精准把握技术演进与社会发展规律，确保标准能够适应人工智能技术快速迭代和社会治理需求升级的双重要求。同时，标准充分考虑技术兼容性与场景扩展性，为人工智能社会实验的设计优化、实施创新和评价升级预留充足的扩展空间，为行业规范化发展提供方向指引，引领人工智能社会实验领域的技术应用与治理实践创新。

(二) 标准的主要内容

本文件提供了人工智能社会实验从实验设计的原则、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。

《人工智能社会实验设计指南》主要包括 4 个组成部分：

1. 人工智能社会实验组织实施的基本原则和实验方法选取

给出了人工智能社会实验组织实施的四项基本原则，包括遵循科研伦理、遵照实验逻辑、执行随机处理和规范设计流程，以及人工智能社会实验的具体开展场景中自然实验、实地实验、调查实验、计算实验方法的选取建议。

2. 人工智能社会实验干预设计

从明确干预变量、开展预实验、科学选取被试、随机化分组、干预效果评估等方面规范人工智能社会实验的干预环节。

3. 人工智能社会实验指标测量设计

从统计类指标、调查类指标、行为痕迹类指标三个维度提供了研究特定人

人工智能应用场景的适配指标选取范围。

4. 人工智能社会实验数据处理设计

从样本描述分析、因果效应分析、因果机制分析等方面规定人工智能社会实验的实验数据处理环节。

(三) 标准主要内容的确定依据

该标准的主要内容形成基于理论研究与实践调研的双重支撑。一方面，标准借鉴了社会实验领域的经典理论与方法框架，结合人工智能时代复杂系统变迁与治理需求的特征，吸纳了对干预设计、指标监测和数据分析等方面的最新进展。另一方面，标准主要内容扎根于国家智能社会治理实验基地的实践基础，系统调研了 16 个省份 24 个基地的多领域人工智能社会实验开展情况，在总结其设计经验与关键难点的基础上提炼形成通用方法，确保标准具有现实适配性和可操作性。

三、试验验证的分析、综述报告，技术经济论证，预期的经济效益、社会效益和生态效益

(一) 试验验证情况

在标准制定过程中，起草组开展了系统性的调研工作：首先，对涵盖城市管理、教育教学、医疗健康、政务服务、环境保护等领域的智能社会治理实验基地进行了全面考察，形成了详实的调研报告；其次，通过组织试验区、实验基地的应用方、技术方、研究方及领域专家的多轮研讨，广泛收集意见建议；确保本标准既能有效反映技术效能与社会价值，又能平衡治理需求与人文关怀，最终形成具有前瞻性和指导性的共识性标准框架。同时，起草组对国内外社会实验方法相关文献进行了系统梳理，为人工智能社会实验的规范化发展提供坚实基础。。基于这些工作，起草组从两个维度构建了标准框架：一是根据实验设计与干预方式的差异，将社会实验划分为自然实验、实地实验、调查实验和计算实验四大基础类型；二是聚焦实验实施过程中的三大关键技术环节——实验干预、指标监测与数据分析，明确了其在实验控制、数据获取和结果解释中的核心作用。为验证理论模型的适用性，起草组还从微观、中观和宏观三个层面精选了 11 个典型的人工智能应用示范案例，汇编形成案例集，既展示了方法体系在多领域场景中的适用性，也明确了其边界条件。这些工作为标准的可实

施性提供了有力支撑，确保标准发布后能够有效落地应用。

(二) 社会效应

本文件的应用将产生社会效益。研制《人工智能社会实验设计指南》国家推荐性标准，在人工智能社会实验设计流程中引入标准化手段，能够构建人工智能社会实验开展程序的话语体系和技术解决方案，为人工智能社会实验提供可参照适用的程序指引，优化实验流程、提升实验效率。符合开展社会实验本身的科学性要求，可以促进智能技术良性发展和广泛应用，推动建设有人文温度的智能社会。本标准将有助于巩固人工智能社会实验的积极效应，提出可复用、可推广的有关智能社会治理的政策建议，形成对技术、产业与社会的反向回馈，可以为人工智能技术发展路径的优化和治理政策的完善提供科学的指导，可以为国家整体的人工智能标准化战略布局提供参考，为标准化相关政策及市场监管机制的建立提供理论依据，在国际社会创建一个新的供讨论的话语体系和话语空间，帮助我国抢占国际智能社会治理话语权和制高点。

四、与国际、国外同类标准水平的对比情况

当前，在国际国外层面，人工智能技术已在各领域实现广泛应用，但专门针对人工智能的社会实验相关研究尚不充分。国际标准化组织（ISO）、国际电工委员会（IEC）、国际电信联盟（ITU）等国际标准组织以及电气与电子工程师协会（IEEE）等专业团体已经开展了一些间接涉及人工智能技术社会影响及其治理方面的标准化工作，包括人工智能安全、伦理一致性、隐私风险管理、决策偏见评估及组织治理等，但尚未形成针对人工智能社会实验的标准化研究与实践。例如 2025 年 7 月，ISO 发布评估人工智能伦理与运营影响的国际标准框架《人工智能系统影响评估》（ISO/IEC 42005:2025），为开发、提供或使用 AI 系统的组织提供 AI 系统影响评估指南。与此同时，电气与电子工程师协会（IEEE）等国际性专业标准组织也在围绕人工智能安全、人工智能伦理影响分析等方面制定标准化文件。例如 IEEE P7000 系列详细论述了人工智能伦理规范，分别对系统设计中伦理问题、自治系统透明度、系统/软件收集个人信息的伦理问题、消除算法负偏差等方面进行规范。上述国际国外以及国内有关人工智能社会治理的政策指向和标准化实践，能够为研制《人工智能社会实验设计指南》国家推荐性标准提供参考和借鉴。

五、以国际标准为基础的起草情况，以及是否合规引用或采用国际国外标准，并说明未采用国际标准的原因

由于尚未发布相应国际标准，因此该文件没有以国际标准化文件为基础起草，且没有引用国际国外标准。

六、与有关的现行法律、法规和标准的配套情况

人工智能社会实验是人工智能技术与社会治理交叉领域的创新实践模式。2022年3月，中共中央、国务院颁布《关于加快建设全国统一大市场的意见》，明确要求深入开展人工智能社会实验，推动制定智能社会治理相关标准。与此同时，人工智能社会实验也已写入《国家中长期科学和技术发展规划纲要（2021—2035）》和《“十四五”国家科技创新规划》。2025年6月，中央网信办秘书局、市场监管总局办公厅近期联合印发《智能社会发展与治理标准化指引（2025版）》，旨在准确把握社会智能化转型趋势，发挥标准化的基础性、引领性作用，深化智能社会发展与治理，其中部署开展人工智能社会实验，推进智能社会发展与治理标准化工作。为落实有关部署要求，充分发挥标准在智能社会发展与治理中的基础性、引领性作用，研制《人工智能社会实验设计指南》《人工智能社会实验实施指南》《人工智能社会实验评价标准》系列标准，作为贯彻实施《关于加快建设全国统一大市场的意见》《智能社会发展与治理标准化指引（2025版）》等的重要技术支撑。

七、重大分歧意见的处理经过和依据

无。

八、涉及专利的有关说明

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。

九、实施国家标准的要求，以及组织措施、技术措施、过渡期和实施日期的建议等措施建议

建议在标准正式实施之前预留3个月的时间，组织开展标准宣贯培训活动，推动标准更加有效地贯彻实施。

十、其他应予说明的事项

无。