智能社会治理实验的规范遵循与实践模式*

李昊林① 乔天宇**② 李铮②
①对外经济贸易大学法学院
②北京大学社会学系

摘 要：智能社会治理伴生了大量认知需要，有必要通过社会实验方法对其开展研究。智能社会治理实验需在规范与实践层面满足两重基本遵循。规范遵循旨在保障智能社会治理实验的正当性与建设性，包括智能逻辑、社会逻辑、治理逻辑、实验逻辑四个方面的要求。实践模式旨在确保智能社会治理实验能不断生成对人工智能技术与社会治理互动关系的有效认知，它从研究对象、实验媒介、实验路径三个方面，勾勒了智能社会治理实验的基本模式。

关键词：智能社会治理；社会治理实验；人工智能；社会实验
DOI：10.16582/j.cnki.dzzw.2026.02.010

一、引言

当今的人工智能技术已与社会各个领域产生了广泛而深切的互动[1]，智能社会治理正是人工智能技术与社会治理活动交互影响、共生演化的产物[2]。从交通控制到公共安全，从健康诊断到社会信用系统，智能社会治理通过技术手段与人力配合，填补“事前预防”与“事后回应”之间巨大的治理真空，强化“事中动态评估调整”在整个社会治理流程中的功效。[3]可以说，智能社会治理已逐渐成为治理能力和治理体系现代化的重要举措[4]。

2021年，中央网信办等八部门联合公布《国家智能社会治理实验基地入选名单》，中国智能社会治理实验的序幕正式拉开。目前，许多智能社会治理实验基地已经取得一定的建设成果，学界也对其中的机制机理进行了剖析。不过，许多智能社会治理实验基地的建设更加着重实现治理基础设施的数字化迭代与治理方式的数字化再造，即聚焦于更宽泛的数字技术。既有研究在讨论智能社会治理时，也存在将人工智能技术与其他数字技术不相区分的情况。[5]在对实验基地既有建设经验进行分析时，现有研究也多将实验基地所做的各种措施整合理解为“智能社会治理”[6]，但在分析人工智能技术这一智能社会治理的核心要素时，往往较为简略。这种偏重于宏观，不将人工智能技术与其他数字技术相区分的实践与研究现状，导致既有文献中虽已积累了关于数字社会与数字治理的大量研究成果[7]，但却难以回答人工智能技术在社会治理中究竟发挥了多大作用、社会治理因应人工智能技术变革取得了何种效果等问题，难以回应以人工智能技术为核心关切的认知需求。

学界很早便呼吁过要通过社会实验方法揭示人工智能技术与社会治理活动互动的作用机理与特点[8]。迄今为止，现有研究对智能社会治理实验规范遵循的讨论主要集中于社会实验的一般性要求[9]。对于智能社会治理实验自身特殊性的规范要求，现有研究似乎只探讨了人工智能技术自身的风险性这一个侧面[10]。而对于如何在实践层面设计并操作聚焦人工智能技术的智能社会实验，现有研究只做较为零散的探索[11]。

本文旨在弥补这一研究空白，以人工智能技术为核心关切，尝试系统阐述智能社会治理实验的规范遵循与实践模式，论证有效开展智能社会治理实验所需考虑的因素。其中，规范遵循旨在保障智能社会治理实验的开展是具备正当性和建设性的。除现有研究已经充分关注到的实验逻辑外，规范遵循还包括智能逻辑、社会逻

*基金项目：科技部科技创新2030——“新一代人工智能”重大项目“引领赋能型人工智能法治新模式研究”（项目号：2022ZD0120101）；武汉东湖高新区国家智能社会治理实验综合基地项目“智能社会核心特征及治理模式研究”。
**通讯作者 收稿日期：2024—11—04 修回日期：2025—06—13

辑、治理逻辑三方面的要求。实践模式则旨在确保智能社会治理实验能不断生成对人工智能技术与社会治理互动关系的有效认知。它从研究对象、实验媒介、实验路径三个维度，勾勒了智能社会治理实验的基本模式，即以人工智能技术在治理中的应用或应用人工智能技术的社会场景为研究对象，以人工智能技术部署界面或规范为主要媒介，以调节应用、机制、环境等参数为实施路径的社会实验。

二、智能社会治理实验的规范遵循

2021年发布的《关于组织申报国家智能社会治理实验基地的通知》已经从宏观层面详细说明了智能社会治理实验基地的建设目标、建设任务、申报条件。本节所讨论的智能社会治理实验的规范遵循，旨在探讨实验基地在开展具体的智能社会治理实验前及开展过程中，需要考虑和评估哪些要素。

由于智能社会治理实验会对现实社会中的个体与群体产生影响，故对其必须特别注重科学性及伦理性的要求。事实上，从域外实践来看，欧盟和美国部分州对于人工智能技术进入社会治理活动的态度并不十分积极，业已生效的欧盟《人工智能法案》甚至将社会评分、在公共场所为执法目的使用实时远程生物识别系统等人工智能系统列为存在“不可接受的风险”（unacceptable risk）的人工智能系统，禁止其在欧盟范围内投放使用。尽管欧盟对人工智能技术的态度受到欧盟对数字规制整体布局的影响[12]，我国或不必采取相同态度，但开展具体的智能社会治理实验前确有必要评估一系列规范性评价要素，以避免智能社会治理实验单向度地满足有效性评价维度，将社会治理体系扭曲为精巧而陌生的社会控制系统。

除社会实验本身所要求的科学性和伦理性外，智能社会治理实验的规范遵循还需确保即将开展的智能社会治理实验与现有技术条件及治理体系相适应，避免智能社会治理实验沦为成本收益不成比例的浮华工程。鉴于既有研究已经对社会实验的科学性要求做了较为完备的阐释，本节将从智能社会治理实验概念中“智能”“社会”“治理”三个核心要素出发，重点阐释其中所蕴含的智能逻辑、社会逻辑、治理逻辑，从而全面明晰智能社会治理实验的规范遵循，为后续开展人工智能治理实验的建设性与正当性评估提供更为完备的指引。

（一）智能逻辑：技术性、阶段性

人工智能技术是撬动智能社会治理的核心杠杆，因而智能社会治理实验的开展必须观照人工智能技术自身所具有的技术性和阶段性，避免技术盲信与功能错配。所谓技术性，即强调人工智能技术之所以能够为社会治理带来变革，正是因为它具备了自学习、自适应和自组织等智能化能力。这些能力让人工智能在处理复杂数据和执行多样化任务时表现出了高效率和高精确度。智能社会治理实验的成效，首先取决于治理主体能否准确把握智能技术的技术本质，能否将其有效地应用于适当的治理环节。

阶段性则强调智能社会治理实验最终所产出的治理方案必须考虑人工智能将不断发展迭代的现实。人工智能技术不是静止不变的，而是一个持续进化和自我优化的系统。人工智能技术的发展逻辑主要由数据驱动和学习能力的增强所决定。其进化不仅仅表现在算法精度和效率的提升上，更体现在技术的自主性和适应性上。人工智能技术将不断从实践应用中学习，通过算法的优化和模型的迭代，不断提高其智能水平。智能社会治理实验必须考虑到技术的这种自我优化能力，以及这种能力对治理实践的潜在影响。

（二）社会逻辑：现实性、正当性

智能社会治理实验所产生的认识最终会应用于现实社会治理，因而智能社会治理实验还需遵循社会逻辑，即考虑社会现实性与社会正当性。现实性强调智能社会治理实验必须充分考虑现实社会中的各项约束因素，确保治理方案与现实各要素间的相容性与适配性。其一，人工智能技术的部署和应用，必须考虑现有治理技术和治理机制能否支撑和容纳智能技术。其二，人工智能技术的部署和应用，必须充分考虑治理各环节中“人”的特性。智能社会治理实验并非要从社会治理中剔除人的要素，而是在社会治理中更好地构建新的人际互动机制。因而治理主体及治理受体对智能技术的接受能力与驾驭能力，是智能社会治理必须纳入考量的部分。例

如，在卫生医疗领域，人工智能辅助的诊断和病历分析只有在医生和患者能够理解并信任这些新工具的情况下才能发挥作用。同样，在智能养老领域，治理策略需要特别注意老年人对新技术的接受程度，重视易用性和直观性，确保技术对于老年用户的友好性。其三，治理策略还需考虑不同区域和文化背景下的差异，如在技术欠发达地区部署智能技术时需考虑当地居民的技术知识水平和接受态度，制定更符合地方特色的治理策略。

正当性则强调治理策略不仅要在现实社会中是可行的，更要是向善的。人工智能技术的引入确实可能提升治理效能，但由于其技术的复杂性与不透明性，也会不可避免地引入可能降低公众对社会治理信任与认可的要素，甚至放大公众对社会治理的不满。这些要素不仅在于人工智能技术本身，还在于人工智能技术训练及应用所牵涉的数据集及信息系统。以时下当红的生成式人工智能为例，其超大规模、超多参数量、超级易扩展性和多应用场景的技术特性对以算法透明度、算法公平性和算法问责为内核的算法治理体系带来全方位挑战。[13]与此同时，其训练过程中需使用的海量多源数据又带来了多种类型的数据质量和安全风险。[14]因此，智能社会治理实验主体在部署实验时必须要对这些要素给以足够的警惕和关注，分析整合实验各阶段面临的伦理风险，确定不同参与主体在各阶段的职能作用，探索新互动形态下的正当性边界以及操作方案。[15]

（三）治理逻辑：系统性、可靠性

智能社会治理实验同样是一种治理实践，因而需要遵循治理逻辑，尊重治理具有的系统性特性，满足治理强调的可靠性要求。治理逻辑的系统性强调治理不是孤立的环节，而是一个复杂的系统，各环节相互依赖。[16]在一个环节部署人工智能意味着其他环节也需要对应的调整。例如，在“智慧考试”情境下，为了实现人工智能对于测试内容的智能组织，需要构筑人工智能技术嵌入的国家试题库，同时需要逐步构建并普及“手写板+摄像头”“专用考试终端”等新型试题投送方式，从而使得界面可以更好地实现数据收集。[17]正是在这个意义上，智能社会治理实验也可能是一次以人工智能技术引入为契机的治理再造。[18]当然，社会系统本身具有复杂性与动态性[19]，因此，智能社会治理实验必须考虑这些因素的复杂交互，确保实验在取得预期效果的同时不会产生过多的副作用。

智能社会治理涉及大量的治理媒介和治理手段更新调整，由此管控新治理媒介和治理手段自身的风险就成为智能社会治理实验必须重视的课题。新治理媒介和治理手段也可能成为降低治理风险与治理中不确定性的重要工具，但是面对具体而多变的社会现实，治理主体依然不能放松对技术不完备性的警惕。治理逻辑的可靠性一方面要求治理主体充分评估相关技术及所依托数据库的安全性，另一方面要求治理主体对突发状况做好事前的应急预备。对于智能社会治理实验而言，后者显得尤为重要。尽管部署于社会治理环境的人工智能技术必然经过了相当量级素材与案例的训练，但在设计治理流程时，治理主体依然有必要特别预留技术接口记录并提示人工智能系统对极端情况或未知数据的处理，并及时进行人工介入。

三、智能社会治理实验的实践模式

实验法是科学研究的基础，其方法论基本原则在于“控制+干预”，即在一定控制条件下，对实验对象施加有目的的干预，并对干预产生的效果进行考察。与纯粹无介入的观察方法相比，将实验方法应用于社会研究，其优势在于能够通过随机化的方式，控制其他可能的混淆变量带来的影响，进而更容易识别出由干预带来的因果效应。而在过去多数利用非实验数据的社会研究之中，得到可靠的因果推断是极其困难的。

对于智能社会治理而言，人工智能技术在很多社会治理场景中的应用是否真正取得了预期效果，这在科学上属于因果效应推断问题，因此实验法大有用武之地。另外，人工智能技术本身也为实验干预的实施提供了新工具，智能体仿真实验为因果推断新添羽翼[20]，将为社会治理提供新型科学依据。与此同时，在智能社会新型社会关系的产生和人机关系的演化过程中，还有很多与此相关联的新现象、新机制有待认识与厘清，对其中的规律性，尤其是对各种因果关系予以充分了解和科学把握将是后续治理开展的前提，基于这些考虑，开展实验

研究也是必要且亟需的。

智能社会治理实验的实践模式通过界定“控制+干预”中予以重点考量的因素，从而确保实验最终能够产出对智能社会治理规律性的认知。传统上，社会科学研究者喜欢依据干预实施环境对社会实验进行分类，最常提及的实验类型包括实验室实验、实地实验、调查实验等。然而，由于人工智能技术的广泛应用，社会治理情境越发复杂和多样，单一的实验实施环境已无法满足利用实验方法进行智能社会治理研究的需求。本文将从研究对象、实验媒介和实验路径三个维度，对智能社会治理实验中“控制+干预”的重点考量因素重新进行把握，以期形成对智能社会治理实验基本模式的界定。

（一）智能社会治理实验的研究对象

由于智能社会治理有利于保持技术创新优势，重塑社会行动及交互关系的网络化特征[21]，同时还可以促进微观领域整体发展水平的提升[22]。因此以实验为名，推动各治理主体积极探索智能社会治理，本身便构成了智能社会治理实验重要的实践意义。与更偏重应用性的智能社会治理实验基地建设不同，本文讨论的智能社会治理实验更侧重于认识性。学界已经指出，对于智能社会治理，我们最终需要理解智能社会中技术与社会运行的互动机制[23]，提炼智能社会风险识别与形成机理以及智能社会演化路径与内生动力[24]。而开展智能社会治理实验正是期待能通过实验方法产生关于智能社会治理活动的基本知识。从研究对象角度，可将智能社会治理实验分为两类。

第一类智能社会治理实验偏重于研究“智能的社会治理”，其研究对象是人工智能技术在治理中的应用。在这一过程中，人工智能技术被视为提升社会治理效率和效果的工具或手段[25]。对于这类研究对象，智能社会治理实验强调从治理客观效能和治理主观感受两方面，对其进行具体的认知把握。所谓治理客观效能，即实验前后治理目标的实现程度、实现速度以及实现治理目标所消耗的治理资源是否都得到优化提升。以使智能技术识别公共体育场所健康危险事件为例，有研究团队设计了基于视觉的公共体育场所跌倒、溺AI识别，以及基于可穿戴设备的管危险事件AI识别，并将识别系统部署到公共体育场所健危险事件报警系统平台之上。经实验研究测试，AI识别正确率可以达到90%以上，危险事件报警系统的报警准确度达到了95%以上。[26]而所谓治理主观感受，即实验前后社会公众及治理参与者对治理的信任度、满意度等主观感受是否发生变化，以及其变化机制机理。例如，有研究者发现社区智能服务具有增强社区居民的归属感、鼓励居民参与社区事务的作用，并试图对智能服务对居民主观感受产生影响的机制进行挖掘，从而助力长期的社区智能治理。[27]

第二类智能社会治理实验偏重于研究“智能社会的治理”，其研究对象是应用人工智能技术的社会场景，或者也可以说是围绕人工智能技术所形成的社会互动关系。在这一过程中，人工智能技术成为社会治理的环境乃至对象。对于这类研究对象，智能社会治理实验强调对新互动关系的各类影响进行把握，同时探究强化或弱化其中某一影响的方案。目前，我国已经开展的智能社会治理实验较少对这类对象进行研究，对其在认知上存在较大的不足。这种认知不足甚至可能会进一步将“智能社会的治理”简化为对人工智能技术的治理乃至对人工智能技术的限制，从而造成严重的后果。事实上，智能社会的治理在回应人工智能技术所引发的各类问题的同时，更是在积极拥抱人工智能技术发展对人际互动及人机互动关系带来的各类正面变化，探索旧有社会治理难题在新互动关系下可能出现的解法。要想真正在未来实现对智能社会的治理，就必须能够从非工具主义立场对人工智能技术本身及其带来的影响进行认识。

（二）智能社会治理实验的实验媒介

智能社会治理实验的媒介是指实验研究开展所借助的工具。对于偏重于研究“智能的社会治理”实验而言，其实验媒介主要就是智能技术部署界面，即通常所谓的“数字平台”。目前，我国各智能社会治理实验基地普遍运用智能治理平台以链接实际社会治理活动中的诸多方面，实现区域治理效率的提升。

智能社会治理实验与传统的社会实验有一个明显的差异，那就是其实验空间还包括虚拟空间，而这个虚拟空间将同样构筑于平台之上，尤其是在社会治理决策环节进行的智能社会治理实验。由于数字时代大量社会现实已被转化为数据形态，随着机器学习、模拟仿真等技

术的不断成熟，开发生成一整套有人工智能技术支撑的决策支持系统已逐渐成为可行现实。对于许多需要直接与大量个体互动的复杂社会治理场景，治理主体可以借助基于行动者建模（Agent-based modeling，ABM）等计算机仿真建模方式[28,29]，在智能治理决策辅助平台上建立一个能够表征现实社会的“人工世界”，如基于SIR疾病传播模型创造的模拟实验平台，城市规划领域创造的模拟实验平台等。其中，治理主体和治理研究人员可以设置个体的互动规则，进行丰富多样的社会模拟实验，从而在虚拟环境中模拟个体或群体行为的变化，以了解特定政策变动可能产生的集体效应。与基于实证数据检验因素间关系的方法相比，社会仿真研究方法可超越现实环境的诸多限制。研究者可以在模拟的数字虚拟空间中开展现实社会中难以实施或负担成本的场景实验，借此预测特定政策的干预结果，降低政策落地过程中的潜在风险；还可以借由强化学习技术“训练”特定政策模型，寻找更优政策路径。事实上，人工智能对社会治理重塑的关键潜力也正在于它能够提升治理决策的品质和效率。

对于偏重于研究“智能社会的治理”实验，由于平台在议题生产、治理对象认证和治理结构定型等方面都具有重要作用[30]，因而平台依然发挥重要的实验媒介作用。除此以外，研究者还可能借助“规范”这一实验媒介，通过设定或改变规则以实现对新型社会关系中互动行为的干预。这些规则改变也可以以代码形式部署在平台之上。

由此可见，无论是哪一类的智能社会治理实验都高度依赖平台这一实验媒介。因此，构建一个功能良好的智能技术部署平台是有效开展智能社会治理实验的前提，也是具体领域开展智能社会治理实验的真实需求。[31]除平台外，数据也是开展智能社会治理实验不可或缺的重要媒介和基础性资源。尤其对于仿真模拟的计算实验来说，数据的质量是需要考量的重要因素。在建设智能社会治理平台时也应充分考虑其可从多来源汇集高质量数据的功能。目前，不少智能社会治理平台能满足上述要求，例如山西阳泉平定县搭建了“智慧大脑”一体化平台，该平台将全县常住人口、房屋、城市部件等多类型数据信息全部纳入系统，实现了跨部门资源的互联互通、信息共享。[32]

（三）智能社会治理实验的实验路径

智能社会治理实验的具体实践遵循实验方法“控制+干预”的基本方法论原则，而实验中所实施的干预实际上都可以看作是一种调整干预参数的活动。因此，智能社会治理实验的实施需要根据研究对象和具体的研究议题选择适当的干预参数。

围绕人工智能技术在社会活动中的实践，智能社会治理的研究议题可延伸为：①对人工智能技术社会应用的关注；②对人工智能技术的运作方式和在其影响下人类互动机制的关注；③对人工智能技术所处社会环境的关注。三类研究议题对应三种不同的干预参数，即应用参数、机制参数与环境参数。本文将区分“人工智能技术在治理中的应用”和“应用人工智能技术的社会场景”两类研究对象阐释智能社会治理实验的三种不同实施路径（参见表1），并对不同路径下智能社会治理实验的实施方式以及常用的实验情境进行讨论。

1.应用参数

所谓应用指的是治理活动中人工智能算法和模型等具体技术的应用。实验干预中的应用参数可以是：治理活动中是否应用了人工智能技术，在多大程度上应用了人工智能技术，以及应用了何种人工智能技术等。调整应用参数的相关智能社会治理实验可测评不同类型的人工智能技术应用于实际情境的直接效果，比如考察某种人工智能技术在治理实践中的落地应用是否带来了治理效能的提升，这或许是智能社会治理实验最具典型性的实施路径，对于提高治理活动的有效性、科学性也是十分必要的。在实验实施的过程中，可在具体情境中采用实地研究的方式开展实验，如通过将不同人工智能技术在不同时间点上引入特定治理场景作为干预，比较使用不同智能技术时的治理效果。[33]调整应用参数的路径还可用于评估受众对智能技术的接受或信任态度等主观感受，这类实验多采用调查实验的方式进行。比如，在考察公众对人工智能技术参与公共决策的信任问题时，通过在调查问卷中设置不同的情境问题，并将不同的调查问卷随机分配给被试，能够测度社会公众对人工智能技术参与不同类型决策的信任程度[34]。

表1 智能社会治理实验的不同内涵与三种路径区分

		“智能的社会治理”实验	“智能社会的治理”实验
研究对象		人工智能技术在治理中的应用	应用人工智能技术的社会场景
实验媒介		平台或规范	
实验路径	应用参数	考察技术应用，即投入的人工智能算法和模型	
		干预：是否应用智能技术、在多大程度上应用智能技术，应用何种智能技术、如何应用智能技术等	
		实例：如智能技术在环境治理中的应用（实地实验）、人工智能参与公共决策对公众信任的影响（调查实验）	实例：如机器人部署与人类协作任务完成（线上实验室实验）
	机制参数	考察运作方式，如人类的行为方式或互动方式	
		干预：在利用智能技术搭建的人工社会模型上，让某种机制发生或改变其作用方式	干预：在受智能技术影响的环境中或人机互动条件下，让某种机制发生或改变其作用方式
		实例：如在模拟传染病扩散模型中让部分人群接种疫苗（计算仿真实验）	实例：如单面镜实验研究（线上实地实验）
	环境参数	考察物理或社会环境，系约束或影响行动的外部条件	
		干预：在利用智能技术搭建的人工社会模型上，考察不同的环境条件	干预：在应用某种特定的智能技术时，考察不同的环境条件
		实例：如模拟群体踩踏事件发生的风险预警模型中改变空间布局（计算仿真实验）	实例：如道德机器实验（线上调查实验）、数字消费平台的用户如何评价外卖送货员（准实验）

可以认为，上述讨论都属于将人工智能技术在治理中的应用作为研究对象，相关实验偏重研究“智能的社会治理”领域，其重点考察将人工智能技术应用于治理活动时产生的影响。而针对应用人工智能技术的社会场景，调整应用参数的实验也可以用于研究智能技术本身的特征以及由其构成的环境的新特征，为治理应用人工智能后产生的新型社会关系和人机关系提供必要依据，即偏重研究“智能社会的治理”的实验。已有实验研究发现不稳定的人工智能也可以增进人类合作。在一项在线实验室实验中，通过在一个解决颜色协调博弈的小型人类社会网络中引入人工智能机器人，该博弈的集体目标是网络中的每个节点与其所有的邻居具有不同的颜色。实验的结果发现，在网络中以一定方式引入会随机产生噪声的机器人将有助于协调博弈任务的解决。另外，还可进一步讨论这些机器人放置在网络的什么位置上、机器人以何种频率产生噪声时对达成博弈目标的影响。[35]这里的实验干预是引入人工智能机器人以及不同的引入方式，它们均可视为对应用参数的调整。

2.机制参数

所谓机制是指治理活动中人工智能机器的运作方式或者在人工智能机器影响下人类新的互动方式。需要强调的是，与机制参数相关的实验干预将会直接对被试对象的行为产生影响。对于机制参数，同样可以根据对研究对象的宏观分类从两个方面理解。

以人工智能技术在治理活动中的参与为研究对象时，人工智能技术扮演的角色体现在其作为实施实验干预的重要工具上面。应用智能仿真技术搭建人工社会模型，是近年来备受关注的一种范式。在这种情况下，实验干预是在利用人工智能技术搭建的人工社会模型上，让某种机制发生或改变其作用方式，进而观察由该机制导致的可能社会后果，并将这种预测作为进一步治理决策的依据。机制改变在很多公共事务治理实践中都会涉及，治理活动中很多公共政策的实施都可以看作是机制调整。然而，有些潜在的机制改变是不便或不宜在现实世界中开展大规模实验的，或于伦理有悖，或不便操作。但智能技术创造出的仿真社会模型可用于模拟出不同机制实施后的社会结果。这一实验思路被用于设计通过疫苗接种进行疫情防控的可行方案。以某种特定方式为人群接种疫苗就是一种机制改变，因为这改变了病毒的人际传染方式，接种疫苗者由此获得了对抗病毒的免

疫力，其将会进一步阻断病毒在人群中的传播。智能技术设计的行动者仿真模型可整合区域内人口、医院、疾病治疗单位、接触者追踪等数据，利用该模型评估各种不同的疫苗接种策略，从而发现最为经济有效的疫苗接种方案以防止大规模疫情发生。[36,37]在这里，每种疫苗接种策略都可视为一种特定机制，在仿真模型中，实施不同的疫苗接种策略可视为一种调整机制参数的实验干预，对不同疫苗接种策略所取得效果的观察可以通过在模型中调整参数实现。基于行动者仿真模型等社会计算仿真模型提供了一种考察“假设会怎样”的工具，其优势在于可以在一个虚拟的平行世界中先行开展实验以模拟各种可能情况，降低治理活动的信息成本，其结果也可辅助决策，避免治理活动中过度的资源投入。

以应用人工智能技术的社会场景作为研究对象时，可重点研究人工智能技术的应用导致人类产生的新行为方式或互动方式。实验方法可用于探究新行为方式或互动方式带来的可能影响，尤其是特定机制在对主体行为产生影响的过程是如何发生的，从中得到的认识也将作为支撑智能社会治理更有效开展的依据。同样是将机制参数作为研究对象的实验，在这种情况下，实验干预不再是在基于智能技术形成的仿真社会模型上对特定机制进行调整，而是在人工智能技术应用创造的现实环境中，将人工智能技术本身作为重要干预，考察受其影响条件下，某种机制发生或改变而导致的可能后果。“单面镜”实验就是一个借助人工智能技术应用创造的条件来考察新机制引入所导致后果的一个典型实验。[38]实验具体考察线上匿名互动的方式如何改变了线上匹配市场中个体的搜索行为。利用数字平台开展的线上实地实验（或称数字实地实验）不仅能够对匿名互动机制实施操控，也为将实验对象进行随机分组创造了便利条件：实验在一个10万用户的样本中随机选择5万用户，把他们作为干预组，为他们分配匿名浏览的权限，匿名互动就类似构造了一面“单面镜”，被浏览页面的用户并不知道浏览者的到访行为；而其他用户将作为控制组，只能继续进行非匿名浏览。实验干预即是引入了一个匿名互动的机制参数，除用户的浏览机制外，控制组和实验组的用户所在的信息接收环境和浏览到的信息内容均无明显差异。此类通过调整机制参数实施的实验，不但可以验证特殊机制对人类互动产生的实际干预效应，数字实验还具有善于考察干预效应异质性的优势。

3.环境参数

所谓环境指实验中个体或群体面对的外部条件，可以是物理环境，也可以是社会环境，它们对人类或人工智能机器行动主体构成约束和影响。对于不同的研究对象，调整环境参数的实验也有不同的表现。当将人工智能技术作为治理工具用于提升决策力时，实验干预更多还是在利用人工智能技术搭建的人工社会模型上，通过改变人工社会模型内的环境参数观察其可能导致的社会后果，以此作为依据辅助进一步的治理决策。这一实验路径可用于对恶性群体事件的预防。通过还原已发生群体事件的当时场景，综合官方披露的信息以及实地测量的环境和人流等数据，智能社会治理实验有能力构建一个模拟事件发生场景的基于行动者仿真模型。模型能够反映行动主体与物理空间中存在的各主体和其他物品间的相互作用关系，并能对特定情境中发生群体事件的风险作出预测。在相关仿真实验中，可对仿真模型中的环境参数进行调整，发现导致群体事件发生的关键因素。[39]通过改变环境参数，实验能够借助智能技术模拟出现实中无法复现的社会情境，找到最大程度控制风险、保证人群安全的环境参数设置，并为预防类似群体事件的再次发生提供科学依据。

在大模型兴起的背景下，计算仿真实验出现了诸多新形式。斯坦福大学的研究人员率先应用ChatGPT搭建人工智能小镇，观察虚拟行动者之间涌现的复杂社会行为[40]，这种将大模型与仿真实验相结合的研究方式很快就引发了广泛关注[41]。同样以计算仿真实验为基石开展的大型社会模拟器建设，不仅为系统性地刻画更贴近现实的互动场域和社会环境提供了可能，还能促进知识探索与发现，推动基于真实场景的社会治理创新，实现治理优化与效能提升。[42]武汉东湖高新区国家智能社会治理实验综合基地在建设大型社会模拟器方面进行了有益的尝试，目前已发布1.0版本，并在交通治理领域落地应用[43]。

对于以应用智能技术的社会场景为研究对象的智能

社会治理实验，重点是在智能技术应用的社会背景中，通过改变环境条件，使用实验方法深化对人工智能技术与人类行动者互动特征的理解，特别是针对因智能技术应用导致社会情境改变，进而出现有关机器与人类行动者互动关系的新问题。在可能因智能技术引发的潜在道德困境中，可以通过实验方法在线上布置与实际情境相关的实验场景，利用不同的环境参数刻画不同的实验场景，并让被试对象在不同的实验场景中做选择。其中，差异化实验场景中不同的环境参数设置即是实验干预。通过发挥线上实验能够链接来自不同国家、具有不同背景实验对象的优势，实验可根据不同的环境参数下海量被试所做出的选择，发现人类借助人工智能技术开展互动时普遍的道德偏好。[44,45]这一实验路径得到的结论可作为弥补人工智能机器缺陷、讨论通用机器伦理的基石。

在线上渠道搭建的在线实验室，其中对环境参数的调整可以认为是在实验室中完成的。除此之外，这种对环境参数的调整也可以通过实地实验或准实验的方式在非实验室条件下完成。相较于在线实验室的实验方式，实地实验和准实验方式更贴近于实际的社会生活。通过准实验方式可以根据已发生的社会事实，以一种更加自然的方式考察环境参数对人们在数字平台上与其他主体的互动行为的影响。有研究使用某个在线消费配送平台的数据，利用外卖送货员分配的特殊变化识别不同环境条件下用户给出评价的规律性。[46]相似的实验方式和路径还可用于讨论数字平台上不同角色主体间产生矛盾的原因。总之，这种通过改变实验中的环境参数了解不同角色主体间互动行为相关规律的方式，将为完善平台的机制设计，更好地开展有针对性的数字平台的治理提供依据。准实验尽管并非严格实施随机条件下的“控制-干预”，而是利用现有条件形成的自然分组，结合使用统计手段以减少混淆偏差，但它在实施上较严格的“控制-干预”实验更加灵活。更重要的是，将准实验设计广泛用于实施政策评估，对更宏观层面的智能社会治理决策同样具有重要价值。此外，准实验设计帮助识别政策实施的潜在后果，还将有助于减少“一刀切”治理带来的意外风险，从而推动形成更具包容性和适应性的智能社会治理方案。

四、结语

在现有理论探索及实践的基础上，本文尝试系统阐释智能社会治理实验的规范遵循与实践模式，试图为我国深入开展智能社会治理实验提供更为直接有效的理论指引。开展智能社会治理实验必须始终关注其科学性、伦理性、正当性与建设性，因而在开展前以及开展过程中始终有必要充分对照规范遵循对实验设计进行评估。开展智能社会治理实验的目的是对智能社会治理的规律性予以把握，为更好地实现这一目标，需遵循实践模式所勾勒的智能社会治理实验基本路径，即以人工智能技术在治理中的应用或应用人工智能技术的社会场景为研究对象，以人工智能技术部署界面或规范为主要媒介，以调节应用、机制、环境等参数为实施路径的社会实验研究。由此，智能社会治理实验才能在推动人工智能技术与社会治理互动融合的同时，生成关于人工智能技术在治理场景中应用的特征与边界，以及对智能社会新型社会关系特征的微观认知。最终在理论研究者的归纳与总结下，形成对智能社会治理乃至智能社会的宏观系统把握。

本文尽管将实验研究的对象区分为“智能的社会治理”和“智能社会的治理”，但这构成的只是针对实验设计的指导性框架，而非一种绝对的区分。有学者提出有必要开展从“智能社会的治理”到“智能的社会治理”的融合探索[47]。实践中的智能社会治理实验也可能是更加复杂的，在一项具体的实验研究中可能同时涉及这两类研究对象。例如，一个社区智能服务平台上所部署的人工智能技术可能既是治理工具，也塑造了新的社区互动关系及人机关系，平台的搭建也可能是后续社区互动关系建立的基础。区分二者的目的在于指导做出更具针对性的实验设计，同时也助于全面把握实验结论。

还需要指出的是，实验方法尽管在因果推断上面具备其他方法无法替代的优势，但实验方法在外在效度方面的局限也是值得关注的，这一局限同样也会出现在针对智能社会治理议题开展的实验研究之中。另外，有必要对人工智能技术再进行更细致的分类，有些实验路径可能更适合对特定类型的人工智能技术开展研究，对此文中的讨论尚不充分，有待今后做更深入的探讨。随着

人工智能技术与社会治理活动交互程度的进一步加深，人工智能技术本身的行为逻辑，以及智能机器之间的互动及相互影响可能会成为智能社会治理中愈发重要的研究课题。换言之，在智能社会中，我们可能需要将由人工智能驱动的机器也看作是一类重要的社会主体。“机器行为学”[48]的提出倡导了一种全新的跨学科研究领域，智能社会治理实验也将成为这一领域的重要研究工具，未来会大有可为空间，其成果也将会深刻影响人机协同的未来图景。

参考文献：

[1]肖峰. 大模型与智能社会：基于历史唯物主义的探察[J]. 中国社会科学，2024(07): 71–89.

[2]刘毅. 中外智能社会治理研究的文献计量学分析[J]. 理论与改革，2023(04): 147–161.

[3]吕鹏，毕斯鹏，管正青，等. 智能社会协同治理：研究现状与发展趋势[J]. 华南师范大学学报：自然科学版，2023，55(01): 19–35.

[4]乔天宇，向静林. 社会治理数字化转型的底层逻辑[J]. 学术月刊，2022，54(02): 131–139.

[5]舒全峰，杨晓婷，刘璐. 智能社会治理何以产生数字倦怠——基于行政负担理论的分析[J]. 中国行政管理，2025(01): 98–109.

[6]张成岗，阿柔娜. 数字时代社会信任的作用机理与实践路径——基于E市国家智能社会治理实验综合基地的案例探索[J]. 行政管理改革，2025(02): 4–14.

[7]向静林，艾云. 数字社会发展与中国政府治理新模式[J]. 中国社会科学，2023(11): 4–23，204.

[8]苏竣. 开展人工智能社会实验探索智能社会治理中国道路[J]. 中国行政管理，2021(12): 21–22.

[9]苏竣，黄萃. 人工智能社会实验的基础理论与方法——《社会实验理论与方法评介》新著概要[J]. 杭州科技，2022(03): 46–51.

[10]俞鼎，李正风. 生成式人工智能社会实验的伦理问题及治理[J]. 科学学研究，2024，42(01): 3–9.

[11]吕鹏，陈典涵. 社会复杂系统智能模拟：涌现机理与方法路径[J]. 山东大学学报：哲学社会科学版，2024(01): 125–135.

[12]李昊林，王娟，谢子龙，等. 中美欧内部数字治理格局比较研究[J]. 中国科学院院刊，2022，37(10): 1376–1385.

[13]张欣. 生成式人工智能的算法治理挑战与治理型监管[J]. 现代法学，2023，45(03): 108–123.

[14]张欣. 生成式人工智能的数据风险与治理路径[J]. 法律科学（西北政法大学学报），2023，41(05): 42–54.

[15]汝鹏，秦晓阳，苏竣. 风险、原则与责任：基于实验路径的人工智能社会实验伦理规范体系建构探究[J]. 科学学与科学技术管理，2024，45(04): 98–117.

[16]Cosens B, Ruhl J B, Soininen N, et al. Governing complexity: Integrating science, governance, and law to manage accelerating change in the globalized commons[J]. Proceedings of the National Academy of Sciences, 2021, 118(36). https://doi.org/10.1073/pnas.2102798118.

[17]于涵. 打造智慧考试 服务智慧教育[J]. 中国考试，2023(05): 1–10.

[18]朱萌. 流程式赋权：界面治理如何提升政府运行保障效能——兼论数字界面对组织界面的调试与补充[J]. 学术月刊，2023，55(07): 89–101.

[19]乔天宇，邱泽奇. 复杂性研究与拓展社会学边界的机会[J]. 社会学研究，2020，35(02): 25–48.

[20]吕鹏. 智能体仿真模拟：推进行动与结构互构研究[J]. 社会学研究，2024，39(04): 45–68.

[21]李鹏，李雨书. 人工智能社会治理：实验逻辑、建设思路、场景搭建[J]. 信息技术与管理应用，2023，2(01): 13–20.

[22]张强，王家宏. 数字赋能体育场馆智慧化转型发展：突破动因、国外镜鉴与立体化路径[J]. 武汉体育学院学报，2023，57(07): 54–61.

[23]苏竣，魏钰明，黄萃. 社会实验：人工智能社会影响研究的新路径[J]. 中国软科学，2020(09): 132–140.

[24]苏竣. 基于社会实验的循证式智能社会治理研究[J]. 智能社会研究，2022，1(01): 24–29.

[25]雷挺，任宇凡，吴义熔. 人工智能提升政府治理能力的作用机理与组态路径研究——基于欧洲18国的多案例分析[J]. 电子政务，2024(02): 65–78.

[26]刘晓然，刘玉，李健，等. 公共体育场所健危险事件AI识别和报警系统平台构建[J]. 首都体育学院学报，2023，35(03): 267–275.

[27]汝鹏，沈娅云，苏竣. 智慧社区如何影响社区依恋?——基于北京老旧小区智慧化改造的案例研究[J]. 中国软科学，2023(04): 66–75.

[28]夏德龙. 复杂性研究的社会仿真模拟方法述评与展望[J].

华中科技大学学报：社会科学版，2021，35(02): 127–135.

[29]范晓光，刘金龙. 计算社会学的基础问题及未来挑战[J]. 西安交通大学学报：社会科学版，2022，42(01): 38–45.

[30]刘学. 数字平台参与社会治理的三重角色——基于组织的视角[J]. 浙江社会科学，2023(11): 93–101，158.

[31]张生，孙睿，曹榕，等. 基于需求图谱分析的智慧教育督导发展策略研究——以国家智慧教育督导平台建设的实践为例[J]. 中国远程教育，2023，43(06): 39–48.

[32]陶建群，刘哲. 县乡一体智治融合数字赋能县域社会治理的山西平定实践[J]. 国家治理，2023(13): 66–71.

[33]马永喜，辛雅儒，申晨. 人工智能技术应用对城市居民垃圾分类成效的影响——一个实地实验研究[J]. 经营与管理，2022(10): 116–122.

[34]冉龙亚，陈涛，孙宁华. 人工智能参与公共决策对公众信任的影响——一项基于调查实验的实证研究[J]. 科学与社会，2022，12(01): 103–121.

[35]Shirado H, Christakis N A. Locally noisy autonomous agents improve global human coordination in network experiments[J]. Nature, 2017, 545(7654): 370–374.

[36]Waldrop M. Special agents offer modeling upgrade[J]. Proceedings of the National Academy of Sciences, 2017, 114(28): 7176–7179.

[37]Ajelli M, Merler S, Fumanelli L, et al. Spatiotemporal dynamics of the Ebola epidemic in Guinea and implications for vaccination and disease elimination: A computational modeling analysis[J]. BMC Medicine, 2016, 14(01): DOI: 10.1186/s12916–016–0678–3.

[38]Bapna R, Ramaprasad J, Shmueli G, et al. One–way mirrors in online dating: Arandomized field experimentp[J]. Management Science, 2016, 62(11): 3100–3122.

[39]Liu Y Y, Kaneda T. Using agent–based simulation for public space design based on the Shanghai bund waterfront crowd disaster[J]. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 2020, 34(02): 176–190.

[40]Joon S P, Joseph C O, Carrie J C, et al. Generative agents: Interactive simulacra of human behavior[EB/OL]. [2025–06–18]. https://doi.org/10.1145/3586183.3606763.

[41]Guo T C, Chen X Y, Wang Y Q, et al. Large language model based multi–agents: A survey of progress and challenges[EB/OL]. [2025–06–18]. https://arxiv.org/abs/2402.01680.

[42]吕鹏. 大型社会模拟器：社会知识生产的大科学装置[J]. 探索与争鸣，2024 (11): 13–16.

[43]何亮. 我国推出首个通用人工智能大型社会模拟器[N]. 科技日报，2025–03–31(01).

[44]Awad E, Dsouza S, Kim R, et al. The moral machine experiment[J]. Nature, 2018, 563(7729): 59–64.

[45]Maxmen A. Self–driving car dilemmas reveal that moral choices are not universal[J]. Nature, 2018, 562(7728): 469–470.

[46]Kapoor A. Service quality, ratings and digital platforms: An analysis of a medicine delivery platform using a quasi–experiment[EB/OL]. [2025–08–13]. https://ssrn.com/abstract=3120863.

[47]苏竣，魏钰明. 迈向智能社会：现实图景、发展趋向与治理使命[J]. 西北大学学报：哲学社会科学版，2025(01): 78–88.

[48]Rahwan I, Cebrian M, Obradovich N, et al. Machine behaviour[J]. Nature, 2019, 568(7753): 477–486.

作者简介：

李昊林（1997—），男，法学博士，对外经济贸易大学法学院助理教授，研究方向为宪法学与行政法学交叉研究、数字法学。

乔天宇（1988—），男，社会学博士，北京大学社会学系助理教授、研究员，研究方向数字社会发展与治理、计算社会学、组织社会学。

李铮（1998—），男，北京大学社会学系博士研究生，研究方向为技术社会学。