

# 人工智能嵌入社会治理的风险及其规避

张 铤

(浙江工商大学 公共管理学院,浙江 杭州 310018)

**摘 要:**人工智能嵌入社会治理对创新社会治理方式,提升社会治理能力具有重要的作用。与此同时,人工智能在社会治理中的应用带来了治理技术的安全隐患、治理主体的替代危机、治理行为的价值缺失和治理责任的模糊不清等风险。人工智能的技术复杂性与不确定性、赋能治理与侵犯权利的双重效应以及兼具治理工具与治理对象的自反是其风险根源。人工智能嵌入社会治理的风险规避是一个“系统工程”,应致力于构建一个综合性公共政策框架。概而言之,要以数据与算法治理为基础,加强人工智能风险的根源治理;以优化与完善社会政策为保障,彰显技术治理行为的价值理性;以风险评估与社会实验为手段,开展人工智能风险的监测预警;以社会合作与共治为架构,推进人工智能风险的协同治理。

**关键词:**人工智能;社会治理;风险;公共政策

**中图分类号:**F124.6 **文献标志码:**A **文章编号:**1009-1505(2022)03-0120-07

**DOI:**10.14134/j.cnki.cn33-1337/c.2022.03.012

人工智能的发展和应用正深刻改变社会生产和人们生活方式,引发社会治理领域的变革。作为一种治理技术,人工智能嵌入社会治理对创新社会治理方式、提升社会治理能力具有重要的作用。然而,技术是一把“双刃剑”。人工智能在造福人类、促进经济和社会发展的同时亦带来了不确定性和风险隐患。2018年,习近平总书记在中共中央第九次集体学习时强调,“要加强人工智能发展的潜在风险研判和防范,维护人民利益和国家安全,确保人工智能安全、可靠、可控”<sup>[1]</sup>。那么,人工智能嵌入社会治理究竟存在哪些风险?其风险是怎么形成的?该如何系统性规避其风险?为此,本文拟针对以上问题深入分析和探讨人工智能嵌入社会治理的风险形态、形成根源及其规避之策,以期有效防范人工智能的应用风险,促进人工智能技术的可持续发展,构建人工智能技术与人、社会的和谐生态。

## 一、人工智能嵌入社会治理的风险形态分析

随着人工智能在社会治理中的广泛和深度应用,算法日益渗透社会生活,深刻影响人们的工作、学习和生活。人工智能嵌入社会治理构建起一个基于数据和算法的复合治理网络,形成了一种新的社

**收稿日期:**2022-03-25

**基金项目:**全国教育信息技术研究规划课题“信息技术环境下大学生网络政治参与规范机制研究”(156242525);浙江省哲学社会科学规划课题“抗疫彰显的中国制度优势研究”(21GXSZ019YB)

**作者简介:**张铤,男,浙江工商大学公共管理学院副教授,法学博士,主要从事政府治理、公共政策研究。

会治理范式。作为人工智能的技术架构,算法有其自身的运行逻辑。人工智能嵌入社会治理之所以引发社会治理范式的转换,主要归因于算法权力。算法以其内化、泛在的权力关系影响个人与群体的行为,改变传统的科层制治理结构,重构社会治理的规则。正是在此过程中,人工智能的潜在风险逐渐暴露。简言之,人工智能嵌入社会治理带来了治理技术的安全隐患、治理主体的替代危机、治理行为的价值缺失和治理责任的模糊不清等风险,须引起我们的警惕和重视。

### (一) 治理技术的安全隐患

作为一种具有颠覆性、革命性影响的新兴技术,当前人工智能技术发展尚处于初级阶段,技术自身存在安全隐患。其一,在人为干扰下,人工智能技术依赖的算法可能“失算”,如某些黑客会利用程序漏洞躲避防火装置让智能失效,或通过修改数据源使图像和语音识别系统失去甄别能力,等等。其二,人工智能的算法运行是以数据为基础的。人工智能嵌入社会治理中,智能终端和机器依赖数据分析深度学习。在此过程中,数据源的广泛使用,可能涉及大量个人敏感数据,加剧个人信息泄露、隐私被侵犯的风险。其三,人工智能开发具有很强的开放性。互联网技术的深入发展,开源运动已经成为互联网发展的重要驱动力,为人工智能技术的传播提供了条件。互联网企业或程序员下载已有的开源人工智能代码都能开发人工智能程序。然而,这种“不造轮子直接造车子”的人工智能生产模式在促进技术传播的同时,也扩散了技术风险。例如,恐怖主义分子容易通过下载无人机的操控程序以及人脸识别和攻击代码从而组装攻击性无人机。其他的类似隐私收集、网络攻击等更容易实现。总之,当前人工智能技术远未成熟,其技术自身存在安全风险。由于人工智能技术的复杂性和不确定性,人类很难全面了解和掌握人工智能的潜在风险。

### (二) 治理主体的替代危机

除了技术自身安全隐患,人工智能嵌入社会治理亦可能导致治理主体的角色错位,造成治理主体的替代危机。社会治理场域中,人工智能作为技术治理的工具和手段,短期内确实有助于提升社会治理效率,促进社会治理的精细化、专业化和智能化。长远来看,对人工智能应用功效的“迷信”可能造成社会治理对技术工具和手段的过度依赖,导致社会治理创新驱动动力不足,陷入“内卷化”。例如,有的部门投入大量资源到智能技术系统的建设和升级中。然而,技术资源投入大幅增加的同时,在治理制度、机制等方面缺乏创新。再如,新冠肺炎疫情期间,有的地区因技术防疫系统失灵,防疫工作秩序一度陷入混乱,反映出城市应急管理机制的不足和治理韧性的欠缺。人工智能在社会治理中以其强大的数据处理能力和算法推荐能力,为人们的生活带来快捷和便利,在提升社会治理效率的同时,亦会导致社会治理主体对智能技术形成依赖而产生唯技术论,奉行技术至上。唯技术主义会使社会治理主体对人工智能技术的工具价值产生迷思和崇拜,过于依赖人工智能技术的分析判断。可以预见,长期过度依赖人工智能技术工具的技术治理模式势必会逐渐消解治理主体的主动意识和创新精神,导致人工智能的算法“决策”代替治理主体的自主决策。在算法、算力和大数据等要素的驱动下,随着人工智能技术的升级及其深度嵌入社会治理,社会运行呈现高度“算法化”,社会系统可能被人工智能技术“绑定”,导致治理主体被智能机器取代。

### (三) 治理行为的价值缺失

人工智能嵌入社会治理中,一方面,算法催生出隐性权力。互联网时代的来临带来了人类社会的诸多变革,权力的性质和实现形式也发生了变化<sup>[2]</sup>。算法“黑箱”为权力的隐性运行提供了条件。人工智能具有“算法主观性”特征。算法必然体现设计者和开发者的主观意志和价值取向,将其主观偏见带入算法系统形成算法偏见,导致种族、性别、年龄、阶层或区域等多种类型的算法歧视。例如,相关研究证明,美国芝加哥法院使用的犯罪风险评估算法 COMPAS 对黑人犯罪嫌疑人造成了系统性歧视,其具有高犯罪风险的概率往往被错误地评估为两倍于白人。算法偏见、算法歧视等现象使得技术治理缺失

公共价值和人文关怀,如弗吉尼亚·尤班克斯(Virginia Eubanks)指出,大数据与算法应用于穷人和工人阶层,“强化和延续了近代的济贫措施中的道德主义倾向和惩罚性后果”<sup>[3]</sup>。另一方面,人工智能技术通过解构社会分工模式,颠覆传统劳动关系,造成大量人员结构性失业。智能技术与资本的联姻,会加剧数字鸿沟、贫富差距和社会分化<sup>[4]</sup>。这不仅危及社会公平和正义,甚至可能影响社会和谐稳定。虽然人工智能的发展和应用会创造新的就业岗位和机会,但是大量的技术“弱势群体”恐难以跨越技术鸿沟,其劳动价值可能被剥夺,并逐渐被社会边缘化,甚至被抛弃。

#### (四) 治理责任的模糊不清

人工智能技术算法运作的“黑箱化”不仅对政府有效监管造成困难,而且因为其运行过程的不透明性和算法规则的不易解释,厘定和明晰其侵权行为责任是一个难题。英国《卫报》专栏作家 Ben Goldacre 将算法比喻为一系列的“黑盒子”。面对这样一个知之甚少的数字世界,“人类并不清楚什么是安全的、什么是有风险的”<sup>[5]</sup>。进入人工智能时代,传统的技术责任伦理和制度规范已无法有效应对人工智能技术应用导致的侵权行为责任界定和追究。基于传统的技术责任伦理和制度规范,智能机器的设计者应承担技术风险的主要责任。在人工智能时代,智能机器的设计者是单一的原子责任主体。如果将智能机器的侵权行为完全归责设计者,在法理上和伦理上并不具有很强的说服力。倘若归责人工智能机器本身,显然无法问责一个机器。由此,长久以来基于行为和后果之因果关系的技术责任伦理和制度规范在人工智能时代受到严重的挑战。总之,面对人工智能的快速发展和应用,引入新的技术责任伦理和制度规范,对加强人工智能风险的有效治理,促进人工智能技术的可持续发展,显得必要且紧迫。

## 二、人工智能嵌入社会治理的风险形成根源

人工智能嵌入社会治理的风险生成有其内在逻辑,揭示其风险形成根源有助于找到“症因”,为有效防范其风险提供科学依据。具体而言,人工智能的技术复杂性与不确定性、赋能治理与侵犯权利的双重效应以及兼具治理工具与治理对象的自反是其风险形成的根源所在。

### (一) 人工智能的技术复杂性与不确定性

人工智能技术的长处在于从大数据的复杂关系中寻找出规律性或法则,以此进行预测或判断。人工智能技术依赖算法的运作,但是人工智能技术的算法背后隐藏了“技术黑箱”。一方面,人们很难理解、预测和评估人工智能技术背后的决策逻辑。当人工智能训练完毕,是按照其归纳出来的规则而执行的,人们很难完全掌控人工智能的行为。近年来,人工智能技术与大数据、物联网、云计算和脑科学等技术深度结合后,其算法的运行逻辑更为复杂,导致人工智能技术系统的潜在风险扩大。另一方面,人工智能技术背后的算法并非完全独立运行。人工智能技术的算法具有高度复杂性,其算法需与其他系统结合完成深度学习,才能发挥应有的作用。换言之,人工智能技术的算法不是一个独立的运作体系,其依赖代码和数据样本的机器学习具有较大的不确定性。多次重复使用的样本数据容易误导智能机器,导致智能机器在社会治理应用中做出错误的决策。需要补充的是,如果人工智能的算法模型所采用的数据本身就带有某种偏见和歧视,那么机器学习算法输出的结果会携带这些数据中包含的“偏见基因”,由此引发人工智能的应用风险。

### (二) 人工智能赋能治理与侵犯权利的双重效应

社会治理应用中,人工智能技术是一种集“赋能”与“约束”的力量。人工智能在呈现赋能社会治理正面效应的同时,亦暴露出技术“侵犯性”的“另一面”。例如,有的部门以公共安全为由,应用人工智能技术过度收集和使用公民信息,因疏于管理导致公民隐私泄露。可以说,因技术治理的工具理性



的泛滥,人工智能的社会应用对个人隐私安全带来巨大的挑战,导致技术嵌入与公民隐私权利保护之间的失衡,如有学者指出:“技术治理对秩序唯美主义和全景式监控的过度追求使个人变得更加透明,导致隐私的边界不断收缩。”<sup>[6]</sup>不仅如此,应用人工智能技术的算法推荐会导致“信息茧房”问题,侵犯用户的知情权。无论是基于内容的算法还是协同过滤的算法,都是基于数据化的用户信息偏好。然而,基于用户兴趣的数据挖掘和内容推送,不仅会强化用户的既有偏好,而且会导致他们同质化地获取与自己思想观念和兴趣爱好趋同的信息,窄化信息渠道,陷入“信息茧房”。简言之,人工智能的社会应用如果过多关注技术工具理性而忽视其价值理性和人文关怀。那么,人工智能嵌入社会治理势必会侵犯公民隐私等权利。总之,人工智能发展和应用须坚持以人为本,重视人文关怀,平衡技术治理的工具理性与价值理性,在技术伦理和制度规范的框架内推进技术的可持续发展。

### (三) 人工智能兼具治理工具与治理对象的自反

在科学与技术加速整合的深度科技化时代,人工智能技术的“双刃”效应越发明显。因此,社会治理广泛和深度应用人工智能技术的同时,必须对技术本身进行治理。换言之,社会治理场域中,人工智能技术既是治理的工具又是治理的对象。然而,“以技术治理技术”难免使治理陷入“维护”与“批判”的自反性“悖论”中,扩大技术治理的自身风险。需要指出的是,囿于现实条件,人们对人工智能技术风险的研判和防范能力有一定的局限。人工智能技术背后的算法运行极为复杂,有效解决人工智能技术的不确定性问题可谓困难重重。近年,人工智能在提升社会治理效率、增进社会福祉的同时,其技术应用引发的一系列问题,已对人类的法律制度、伦理规则和技术规范形成挑战。例如,人工智能机器的侵权行为责任如何界定?算法偏见、算法歧视等如何治理?技术鸿沟该如何弥合?等等。总之,人工智能嵌入社会治理的风险已成为社会关注的热点,人工智能的应用风险治理势在必行。有效规避人工智能的应用风险亟待采取可行性对策。

## 三、人工智能嵌入社会治理的风险规避之策

近些年,为了应对人工智能兴起的风险和挑战,世界主要大国颁布了相应的发展规划和政策文件。例如,2016年,美国颁布了《国家人工智能研究与发展战略规划》《为人工智能的未来做好准备》等,英国发布了《人工智能:未来决策制定的机遇与影响》。2017年,法国发布了《人工智能战略》。当前,人工智能风险监管主要有两种模式:“无需批准式”(permissionless)监管和“审慎监管”(precautionary)。美国倾向于“无需批准式”监管,强调除非有确凿证据和案例证明其危害性,否则新技术、新商业模式都是被允许的。相较于美国的“无需批准式”监管,英国、法国倾向于“审慎监管”。例如,英国强调对人工智能制定出的决策采用问责<sup>[7]</sup>,法国强调对新技术的“共同调控”<sup>[8]</sup>。应当说,这两种监管模式都有其可取之处,但局限于“创新”与“安全”的二维选择不免失之偏颇。人工智能嵌入社会治理的风险规避是一个“系统工程”,应致力于构建一个综合性公共政策框架。借此,可为有效防范其风险,促进人工智能技术的可持续发展建立制度基石。

### (一) 以数据与算法治理为基础,加强人工智能风险的根源治理

基于大数据的复杂算法是人工智能技术的核心。加强数据与算法治理是人工智能时代公共政策的应有内容。就数据治理而言,由于数据的收集、利用、共享与公民隐私权利保护等存在一定的张力。建立和完善基于公民授权的数据分类治理机制能有效回应公民对隐私问题的关切。因此,人工智能发展相关公共政策应体现对个人隐私的尊重和保护,在规范数据的收集、利用和共享等环节的基础上,实现数据价值、隐私保护和数据安全等多元政策目标的平衡。具体而言,在数据的采集阶段,要最大限度消除数据隐藏的偏差。因为如果基础数据的产生过程中存在有利或不利于某类人群的偏差,那么用

这些数据揭示的一些统计关系就会使那种偏差固化。在数据共享和数据安全方面,要积极推动智能技术系统和平台的共建共享,打破数据孤岛,畅通数据交流和共享渠道。同时,要注重数据安全,保障个人信息安全,避免数据泄露侵犯个人隐私权利。就算法治理而言,核心议题是要规范和监管算法的运作,增强算法的透明度和可解释性,规制算法权力,彰显算法正义。一方面,要加强算法问责。构建有效的算法问责机制是算法治理的重要内容。对算法应用引发的风险设置一定的责任结构,嵌入责任机制,是加强算法规制必要的制度设计。算法问责的基础是提升算法的透明度和可解释性。因此,要求算法运用履行告知义务、参数报备、参数公开和规定算法可解释权等治理手段应纳入算法治理的公共政策。简言之,要完善人工智能风险治理的法律法规,在人工智能的研发、设计和应用中,有效提升算法透明度和可解释性,将其置于相关监管机构、伦理委员会、社会组织和公众等监督之下,确保人工智能的算法可理解、可解释、可预测,提高人工智能算法的可信度。另一方面,要贯彻算法公正原则。针对人工智能嵌入社会治理存在的算法偏见、算法歧视等问题,要坚持算法公正原则。在具体操作中,算法的开发者、运营者等都应遵守算法透明、算法解释、算法问责、算法包容以及算法审查等相关规定。需要指出的是,算法治理涉及面较广,因此可根据算法应用的成熟程度、影响范围等多种因素,分领域、分优先级的制定相关公共政策<sup>[9]</sup>。总之,数据与算法治理应及时回应公众对人工智能技术引发的伦理、责任和个人隐私权利保护等社会关切。以数据与算法治理为基础,构建算法和数据治理体系,有助于人工智能应用风险的根源治理。

## (二) 以优化与完善社会政策为保障,彰显技术治理行为的价值理性

在技术治理的实践中,必须重视对人的尊严、自由和社会民主、平等、公正等重要价值的维护<sup>[10]</sup>。人工智能技术应用于社会治理,应有助于实质性推动社会整体“赋权增能”和公共性建构,而那些“智能革命中利益受损的群体,作为共同体成员的基本生存和发展权利需要得到充分保障”<sup>[11]</sup>。人工智能技术的发展将导致就业结构、就业方式的深刻变革。人工智能技术具有颠覆性影响,其带来的技术革命成本因此会更大。未来大量的工作岗位可能被人工智能替代,加剧社会不平等,甚至影响社会稳定。因此,在人工智能发展相关公共政策的考量中,不仅要关注人工智能发展的产业和经济政策,也要关注人工智能发展的社会政策。人工智能的风险善治须遵守人本原则<sup>[12]</sup>。为此,一方面,公民应拥有平等享受人工智能的社会应用带来的成果的权利,尤其是政府部门提供的数字化、智能化基本公共服务,应普惠公众。另一方面,要建立和完善适应人工智能时代发展的终身学习和就业培训体系,完善社会保障体系,应对人工智能发展可能带来的结构性失业等问题。基于此,基本收入制度应提上公共政策议程。“基本收入”是指国家(地区)向所有成员支付一定数额的收入,满足其基本生活的需求。该收入不需要任何条件与资格,只要是该国家(地区)的居民,每个人都可以领取。尽管该制度存在争议,芬兰、瑞士和美国的一些州也已经开始了相关实验<sup>[13]</sup>。总之,面对人工智能技术带来的风险和挑战,要完善社会政策体系,彰显技术治理的价值理性和人文关怀,促进技术与治理体制、社会发展的良性互动、相互增效。

## (三) 以风险评估与社会实验为手段,开展人工智能风险的监测预警

人工智能技术具有较大的不确定性,因此需要运用多种手段和工具,对其风险进行科学评估,开展人工智能风险监测预警。一方面,政府可联合高校、社会组织和互联网企业,邀请技术专家和公众代表参与,对人工智能的社会应用风险进行评估,根据其风险类型、等级有针对性地采取治理对策,有效规避人工智能的应用风险。在社会治理的某些重要领域,应加强对人工智能算法的数据使用、应用场景及其风险状况等安全监测,及时发现算法安全问题,进行必要的技术风险预警。另一方面,要积极开展人工智能社会实验研究。社会实验是一种典型的社会科学研究方法,是以现实情境下的人、组织作为研究对象所开展的研究活动,通过某些社会、政治过程或技术变革导致的近似于理想实验的场景来

研究社会科学的问题。为研究工业化社会带来的影响,美国曾通过组织大规模社会实验,称为“霍桑实验”,对工业文明时代人类的心理和行为问题进行深入的解析。当前,人工智能技术已经融入人类社会和生活,人工智能时代已经到来。人工智能的创新发展深刻影响社会,为开展智能时代的社会实验提供了理想的要素和场景。推进人工智能风险治理,消解人工智能技术的负面效应,应深入了解人工智能时代的社会形态和特征。积极开展人工智能社会实验研究,有助于科学研判人工智能技术的社会风险,“准确识别人工智能对人类社会带来的挑战和冲击,深入理解人工智能的社会影响特征与态势,深刻把握人工智能时代社会演进的规律。”<sup>[14]</sup>需要补充的是,人工智能风险的有效治理必然要求加强人工智能研发人员的责任伦理意识,引导他们将道德算法融入人工智能研发设计的算法模型。因此,面向人工智能的研发人员,加强技术道德与伦理规范宣传和教育,亦应纳入人工智能发展的相关公共政策。

#### (四) 以社会合作与共治为架构,推进人工智能风险的协同治理

人工智能是一项充满挑战的颠覆性技术。由于人工智能技术的专业性、不确定性,其嵌入社会治理带来的风险具有复杂性、多样性和严重性,仅凭政府一方难以有效治理其风险。人工智能技术在社会治理应用中必然涉及多元利益主体,因此对其风险的治理亦要求多元主体参与共治,倡导社会合作。仅有政府和技术专家参与的技术治理是不可取的,其治理框架也是不完整的。因此,相关公共政策应着力构建人工智能风险协同治理机制,实现政府单一治理向多元主体合作治理转变。该协同治理机制能“利用各自优势而逐渐组成技术治理的决策权威联盟”<sup>[15]</sup>,形成合作共治的技术风险治理格局,避免单一主体治理人工智能风险的局限性。具体而言,要打破传统的“命令—控制”式技术风险治理模式,促进政府与技术专家、公众和社会组织的交流和协商,引导多元社会主体和利益相关者参与人工智能技术系统和平台的安全论证、讨论,共同应对人工智能技术带来的风险和挑战,实现社会资源配置帕累托改进和社会运行系统整体功能的提升。需要指出的是,人工智能与物联网、云计算、大数据、互联网等技术的深度融合,人工智能所造成的社会影响将会超越一国的边界。当前,人工智能的应用风险已经成为全球关注的问题,因此对其协同治理亦应突破国界。人工智能风险治理需要全球性协调机制、全球性协作平台以及全球性的标准和规则,以及国际范围内人工智能的道德伦理系统和价值理念的碰撞与讨论。为此,我国要积极主动参与人工智能全球治理机制的建构和讨论,在推动形成国际人工智能治理框架和标准规范中发挥应有的作用,提升中国在人工智能全球治理中的国际规则话语权。

## 四、结 语

综上所述,人工智能嵌入社会治理在推动社会治理方式创新、提升社会治理能力的同时也伴随着治理技术的安全隐患、治理主体的替代危机、治理行为的价值缺失和治理责任的模糊不清等不可忽视的风险。对此,我们要抓住超级智能尚未出现的“窗口期”,统筹兼顾,多维并举,制定和实施有效的公共政策防范人工智能嵌入社会治理的风险,平衡人工智能技术创新与其风险之间的矛盾,为人工智能的社会应用打造开放、有序、安全的环境。防范人工智能的应用风险具有重要的社会意义,同时也是一项长期任务。英国物理学家斯蒂芬·霍金和美国企业家埃隆·马斯克都曾发出对人工智能风险威胁的警告。人工智能的风险之所在,就是人类自我警醒之所在。人工智能技术的发展和应用必须以关怀人类为宗旨,创造以人为本的智能社会。在“总体性社会”到“分化性社会”的深刻转型中,人工智能嵌入社会治理有其必然性和现实必要性。随着数据驱动算法的升级和优化,人工智能技术的发展和应用具有广阔的前景<sup>[16]</sup>。我们要在应对人工智能的风险和挑战中推动社会治理变革和创新,促进社会治理体系与治理能力现代化,增进社会福祉和进步。

**参考文献:**

- [1] 习近平. 加强领导做好规划明确任务夯实基础 推动我国新一代人工智能健康发展[N]. 人民日报, 2018-11-01.
- [2] 陈胤. 权力的隐身术——互联网时代的权力技术隐喻[J]. 福建论坛(人文社会科学版), 2015(12):67-72.
- [3] VIRGINIA E. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor[M]. St. Martin's Press, 2018:37.
- [4] 孙伟平. 人工智能与人的“新异化”[J]. 中国社会科学, 2020(12):120-137.
- [5] GGLDACRE B. When Data Gets Creepy: The Secrets We don't Realise We're Giving Away[N]. The Guardian, 2014-12-05.
- [6] 单勇. 跨越“数字鸿沟”: 技术治理的非均衡性社会参与应对[J]. 中国特色社会主义研究, 2019(5):68-82.
- [7] Artificial Intelligence: Opportunities and Implications for the Future of Decision Making[EB/OL]. (2016-11-18)[2021-07-06]. [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf).
- [8] 周衍冰. 大数据产业在法国的发展及应用[N]. 学习时报, 2014-11-03.
- [9] 贾开. 人工智能与算法治理研究[J]. 中国行政管理, 2019(1):17-22.
- [10] 张铤, 程乐. 技术治理的风险及其化解[J]. 自然辩证法研究, 2020(10):42-46.
- [11] 高奇琦. 智能革命与国家治理现代化初探[J]. 中国社会科学, 2020(7):81-102.
- [12] 唐钧. 人工智能的风险善治研究[J]. 中国行政管理, 2019(4):46-52.
- [13] 贾开, 蒋余浩. 人工智能治理的三个基本问题: 技术逻辑、风险挑战与公共政策选择[J]. 中国行政管理, 2017(10):40-45.
- [14] 苏竣, 黄萃. 探索人工智能社会治理的中国方案[N]. 光明日报, 2019-12-26.
- [15] 谭九生, 杨建武. 人工智能技术的伦理风险及其协同治理[J]. 中国行政管理, 2019(10):44-50.
- [16] 张铤. 人工智能的伦理风险治理探析[J]. 中州学刊, 2022(1):114-118.

**Risk Avoidance of Artificial Intelligence (AI) Embedded in Social Governance**

ZHANG Ting

*(School of Public Administration, Zhejiang Gongshang University, Hangzhou 310018, China)*

**Abstract:** AI embedded in social governance plays an important role in innovating social governance methods and improving social governance ability. At the same time, the application of AI in social governance has brought risks, such as the security risks of governance technology, the substitution crisis of governance subjects, the lack of value of governance behavior and the ambiguity of governance responsibilities. The technical complexity and uncertainty of AI, the dual effects of enabling governance and infringement of rights, and the reflexivity of both governance tools and governance objects are its risk sources. The risk avoidance of AI embedded in social governance is a “system engineering”, which should be committed to building a comprehensive public policy framework. In short, we should strengthen the root cause management of AI risks based on data and algorithm management; optimize and improve social policies as a guarantee to highlight the value rationality of technological governance behavior; take risk assessment and social experiment as means to carry out AI risk monitoring and early warning; take social cooperation and co-governance as the framework to promote the collaborative governance of AI risks.

**Key words:** AI; social governance; risk; public policy



(责任编辑 孙豪)