

· 科学技术与社会 ·

算法治理逻辑预设中的悖论与范式变革

Paradoxes and Paradigm Shifts in the Logical Presuppositions of Algorithmic Governance

张海柱 /ZHANG Haizhu 李志新 /LI Zhixin

(东北大学文法学院, 辽宁沈阳, 110169)
(School of Humanities and Law, Northeastern University, Shenyang, Liaoning, 110169)

摘要: 基于人工智能算法技术应用的政府治理实践形成了“算法治理”这一新的治理形态。研究发现, 算法理性与算法统治共同构成了算法治理的基本逻辑, 其运行又依赖于认识论、方法论与价值方面的许多前提预设。然而, 算法治理逻辑预设中存在不确定性、价值权衡、自证预言与去政治化等多重悖论, 由此形成的悖论结构成为算法风险生成的深层根源, 最终会消解算法治理的合理性基础。因此需要推动算法治理逻辑预设的重构, 进而实现算法治理范式变革。为此要基于不确定性背景来对算法治理的理性基础进行重构, 并实现算法治理的“再政治化”与政治控制, 此外还要以公共性为导向进行算法设计。

关键词: 算法治理 算法风险 逻辑预设 悖论

Abstract: The application of artificial intelligence algorithm technology in government governance practices has led to the emergence of a new governance form known as “algorithmic governance”. This study argues that algorithmic governance rests on “algorithmic rationality” and “algotocracy”, yet faces paradoxes of uncertainty, value trade-offs, self-fulfilling prophecy, and de-politicization. These paradoxes generate deep algorithmic risks and undermine its legitimacy. To address these challenges, the paper calls for a reconstruction of its logical presuppositions and a paradigm shift: reestablishing its rational basis in the context of uncertainty, advancing the re-politicization and political control of algorithmic systems, and orienting algorithm design toward the public good.

Key Words: Algorithmic governance; Algorithmic risks; Logical presuppositions; Paradoxes

中图分类号: TP301.6; N031 DOI: 10.15994/j.1000-0763.2026.07.010 CSTR: 32281.14.jdn.2026.07.010

引言

随着人工智能技术的快速发展, 算法正被日益深入地应用于政府公共治理实践之中, 从而产生了一种以自动化决策为特征的“算法治

理”(algorithmic governance)模式。算法常被塑造为一种超越人类局限的“理性权威”, 被认为能够以价值中立、计算精准和效率最优的方式为政府治理提供新的决策依据。然而, 算法在公共领域的应用也伴随着算法霸权、侵犯公众隐私、算法偏见、社会歧视等多重社会风险

基金项目: 国家社会科学基金一般项目“政府算法治理中的不确定性风险及其应对机制研究”(项目编号: 24BZZ050)。

收稿日期: 2025年9月27日

作者简介: 张海柱(1985-)男, 河北盐山人, 东北大学文法学院教授, 研究方向为政策过程、风险治理与数字治理。Email: haizhu1213@sina.com

李志新(2001-)女, 内蒙古赤峰人, 东北大学文法学院博士研究生, 研究方向为数字公共治理。Email: xinneu2023@163.com

的产生。在此背景下,算法风险的有效治理已成为算法治理实践良性发展面临的重要挑战。

政府公共部门对算法技术的采用所催生的新型治理形态问题已经引发了学界的持续关注。纵观现有文献,算法治理相关研究或从宏观视角考察算法在国家治理^[1]与社会转型^[2]中的理论创新与耦合逻辑,或基于中观视角来考察算法应用推动下的公共组织变革与治理流程重塑^[3]以及决策权力重新配置^[4]等问题。类似研究为理解算法技术产生的社会政治影响提供了有益探索。然而,既有研究主要聚焦于“算法如何影响治理”的外向性探讨,在相当程度上遮蔽了一个更为内在且根本性的问题,即“治理中的算法本身为何”,由此普遍忽略了对算法内在技术逻辑及其运行所依赖的理性前提等基础性问题的深入探讨。这种对算法微观逻辑基础的忽视,导致现有研究无法充分揭示算法风险的深层次成因,从而在理论解释力和实践指导性上均显现出明显不足。因此当前亟待引入新的分析视角与理论资源,以填补算法治理研究在微观基础层面上的欠缺。

有鉴于此,本文将引入科学技术哲学、科技与社会研究等学科理论视角,对算法治理的逻辑预设问题进行一种批判性审视,进而深入考察算法治理逻辑预设中的各种内在悖论或困境问题。本文认为它们将在很大程度上挑战甚至消解算法治理的正当性基础。基于此,本文最后将讨论算法治理的范式变革问题,以期为算法治理实践的改革完善提供启发借鉴。

一、算法治理的逻辑预设

现实中,尽管算法治理在具体实践模式上形态各异,但其均依赖于具有共性的运行逻辑与前提预设条件,由此导致不同的算法治理实践形态却会产生相似的社会政治影响。基于此,下文将分别对算法治理运行的基本逻辑与前提预设进行考察。

1. 算法治理的基本逻辑

当前算法治理以机器学习特别是深度学习算法技术的应用为主流趋势,典型地体现出“数

据-计算-分类-预测”的运行逻辑。从政府治理决策的角度,上述逻辑又可以进一步提炼为治理决策所需的知识“如何生产”与“如何应用”两个方面。在算法治理中,“算法理性”和“算法统治”(algoracry)分别对两个问题进行了回答,由此共同构成了算法治理的基本逻辑。

算法理性在理论上解决了算法治理决策所需理性知识的来源问题,其实现涉及“大数据”和“计算”两个构成要素或支撑条件。其中,数据作为算法训练的基础,训练数据规模与算法结果的有效性和准确性呈正相关,特别是在算法训练初期更是如此。大数据则意味着数据占有量上的极限突破,它颠覆了经典统计学的“随机抽样-推断总体”逻辑,而是以“全体数据”作为直接的分析对象,认为从中可以识别出人类认知难以发现的、稳定的相关性关系来进行预测,从而替代对因果关系的追寻。在此情况下,数据最大化原则成为算法理性实现的重要基础。在获取大数据的基础上,更重要的对数据进行处理分析或“计算”。算法即是直接用于计算数据的计算机程序、模型或代码。就此而言,算法理性的核心是一种计算思维。随着人工智能技术的发展,在充足的算力支持下,算法常被认为能够以远超人类的速度进行“无限的场景仿真和试错”,最终达成一种“确定性”。^[5]这种“确定性”正是现代理性主义所追求的最终知识形态。也正因如此,深受西方理性主义传统影响的许多人已经“将算法思维的某些形式和假设深深地视为现代性的不可或缺的方面,并将其他形式(如机器学习)提升至崇高地位”。^[6]

算法统治则回答了算法治理的运行需要解决相关知识如何应用的问题,涉及由谁来应用知识(掌握治理决策权威)以及知识如何转化为行动(做出怎样的治理决策)两方面。所谓“算法统治”,是指“基于计算机编程的算法进行组织和结构化的一种特殊治理系统”,^[7]意味着治理决策权威由传统的官僚组织转移给了算法系统,“机器”(代码)取代“人”成了掌握真实决策权力的主体。无论算法自身掌控决

策权的程度如何,决策权力主体的“非人化”倾向正是算法治理的重要特征。在算法统治的具体运作方式上,算法系统主要基于分类与预测等功能做出治理决策。其中,分类可以视为一种“认知机制”,通过辨别并标识包括人在内的各种治理对象的属性,进而完成类别划分,使算法可以针对不同对象自动配置差异化的政策安排,进而分配不同的利益或惩罚,因此被视为算法权力的重要来源。^[8]相比之下,预测功能在治理决策中更为关键。算法通过对大规模数据的处理与相关性识别,便会使用概率统计知识来预测个体的行为或计算将来的事实,从而将算法治理实践指向的对象从已经发生的事实转变为某种事件发生的可能性。在此基础上,算法治理主体会对特定对象进行方向性引导或干预,以助推或强制方式促使其采取特定行动或改变原有行为,从而实现一种控制目的。这种“预测-控制”导向事实上也是算法治理运行逻辑的重要构成。

2. 算法治理的前提预设

算法治理的基本逻辑得以有效运行需要各种前提假设或预设条件,它们是算法治理更为基础性的支撑要素,否则算法理性与算法统治将难以实现。算法治理是一个知识生产与应用的过程,同时知识应用的结果需要体现公共部门的价值导向。由此,认识论、方法论与价值方面的前提预设共同构成了算法治理运行的支撑条件。

首先,认识论主要探究知识的本质及其来源问题。算法治理能够有效运行所依赖的一个重要的认识论预设是算法理性的有效性或者算法生产知识的确定性。就数据而言,一个基本假设是一切包括人与物在内的社会对象都可以通过测量的方式提取出关于其基本特征或属性方面的数据,进而通过这些数据来观察社会对象。也即,数据被视为对社会世界的客观呈现或表征,基于数据的治理往往持有一种“数据普遍主义”假设,认为数据作为信息的客观载体可以超脱不同社会背景或文化差异的影响。^[9]此外如前所述,基于大数据的算法分析得出的是一种相关性知识,这种相关性知识被认为

是对社会现象客观规律的体现,因而可以替代因果知识。就算法模型的计算而言,对包括人在内的各种社会对象进行数字化或数据化改造是算法应用的前提,治理对象的可量化或可计算成为算法治理的预设条件。

此外,算法治理有效运行的另一个必要的认识论前提是算法模型与被感知对象之间的“同构关系”,即“客观世界的结构可以与人们构建的理论模型的结构进行类比而被认识”,^[10]由此确保算法知识是对客观世界的准确呈现。

其次,方法论主要回答“如何做”的问题。算法治理的方法论预设主要涉及相关主体在应用算法知识进行治理决策的过程中所预设的前提条件。对于分类来说,它的潜在假设是所有社会对象及其物理和社会属性均可以被识别出来并归之于特定的类别范畴内,并可通过对特定分类数据的提取和组合准确地“还原”某一社会对象。例如,作为算法治理对象的居民不再以“个体”身份呈现,而是被分割为无数的对应居民的特定身份属性的“分体”,经数据化后成为“数据主体”。在不同的治理场景下,这些“数据主体”的呈现样貌存在差异,但它们均应当能够真实准确地反映居民个体的现实身份和属性特征,这是算法治理的重要前提预设。而对于算法的预测功能而言,算法要基于过去的数据进行训练和分析,因此算法预测中的一个重要的方法论预设是“过去的模式会自我重复”。^[11]

最后,价值预设是指特定主体在应用算法技术时所服务于或追求的价值目标,公共部门的算法治理应当以公共利益的实现或公共价值的创造为根本目标和原则。而且从实然角度来说,算法治理的支持者也确信,算法技术的合理应用能够更好地促进公共利益或公共价值的实现。这种信念建立的前提即算法的“价值中立”预设。具体来说,算法常被视为纯粹的技术工具,其风险主要源于算法使用者不当的价值意图或使用方式。正因如此,私人企业基于私人利益追求所开发的算法系统,被认为完全可以应用到政府公共部门,用于解决公共问题并追求公共利益。这是当前许多政府部门通过

政府购买方式引进算法技术并与科技公司合作进行治理改革的前提预设。此外,由于算法被赋予“科学、价值中立”的形象,基于算法的治理决策被认为可以摆脱利益、权力或意识形态等政治因素对理性决策的干扰。因此,当人们担忧人类决策者由于自身的主观因素会造成偏见与歧视时,公众更倾向于采纳或服从算法自动化决策的结果,从而产生自动化偏见的现象。

二、算法治理逻辑预设中的内在悖论

风险社会理论指出,风险社会的产生并不是源于工业现代化的失败,而恰恰是工业现代化的胜利或成功。^[12]相应地,算法治理中的算法偏见、歧视、操纵等算法风险的产生并非源于算法技术的“失败”,而是由于算法技术的“成功”应用。算法治理的基本逻辑预设运行得越成功、越彻底,其内在矛盾就展现得越充分,风险积累也越显著,从而呈现为一种悖论状态。具体考察发现,算法治理的逻辑预设中存在不确定性悖论、价值权衡悖论、自证预言悖论、去政治化悖论等多重悖论。

1. 不确定性悖论

不确定性悖论主要存在于算法治理的认识论预设中。基于大数据分析的算法理性或算法式知识生产被认为可以提供治理所需的“确定性”知识,但从数据到算法模型及其输出结果,算法治理过程中将面临许多知识的不确定性问题。

首先,“大数据=全体数据”在现实中难以成立。数据只是对人或物的数字抽象记录,而无法等同于事物本身。现实中,许多人、物、价值或目标因无法数字化或转化为算法可以处理的数据而遭到排斥,从而造成“非存在的社会存在”。^[13]

其次,算法模型实质上是对现实世界真实状况的不充分呈现或近似模拟,在模型拟合度上存在不确定性。即便是一些包含数百万亿个参数的超大模型算法,其学习和推理过程也依赖于各种简化假设,由此导致了算法建模和应用中的不确定性。

此外,算法输出结果所呈现的统计相关性体现的是一种“可能性”而非逻辑演绎的“必然性”,基于统计推理的算法预测在应用于个案时往往产生统计上的不确定性问题。正是由于上述多重不确定性的存在,基于大数据分析的算法式知识生产将面临“有数据的无知”。科学研究的推进不仅无法根除“无知”或不确定性问题,反而可能因“知识悖论”产生更多未知,暴露算法理性的根本局限,也意味着许多算法技术支持者对算法治理模式的高度推崇事实上缺乏必要的确定性基础,将会出现“理性试图用自身的有限性来把握对象世界的无限性之间的张力”。^[14]

2. 价值权衡悖论

在公共部门算法治理的实践中,效率、准确性、社会公平、公开透明、政治合法性等多元价值目标常常存在着内在的张力与冲突,甚至形成难以化解的悖论难题。例如,算法训练有效性的提升需要海量数据支撑,然而大规模数据收集则可能损害公众隐私权利,从而构成了算法技术的两难困境。

算法技术最主要的技术特性是处理大数据上的高效率 and 准确性,然而当算法应用在经济社会结构不平等的社会环境时,此时算法输入数据中也会隐含这些不平等因素,那么高效准确的算法会“让已经存在歧视的系统运行得更顺畅”,^[15]从而产生“算法越高效,社会不公平越严重”的悖论。此外,算法技术的升级和性能的快速提升带来许多新功能或新突破“涌现”,但随着模型在扩大规模时提高性能,它们也可能增加不可预测现象的可能性,包括那些可能导致偏见或伤害的现象。^[16]鉴于算法应用可能对公共价值造成的损害,许多人呼吁进行“价值对齐工程”,让算法的运行受到人类价值观的约束。然而,基于价值对齐理念开发的算法技术不仅大幅度失去了输出结果的多样性,而且其智能表现也会受到不同程度的削弱。^[17]更根本的是,机器学习中的神经网络技术的真实性和有效性恰恰是建立在无法解释性或无法知晓性的基础之上,^[18]“不透明”是算法模型设计的一项规则。若完全公开算法逻

辑又可能引发策略性应对行为，从而形成“透明即失效”的悖论。

3. 自证预言悖论

预测是算法最为重要的功能之一。算法要基于过去的数据进行预测，该功能有效发挥的前提是过去的数据要能够准确代表未来的情况。然而社会现象以及人的行为总是处于持续发展变化之中，从而导致算法的预测往往失灵。更为复杂的是，机器学习算法尽管无法基于过去的数据来准确地预测未来的可能性，但是它却有力量构建一个“虚拟世界”，^[19]并将其等同于“未来”。也即，当算法预测的输出结果被视为一种权威有效的知识后，它会真实地影响治理决策，从而对“未来”进行干预，并制造出原本可能并不存在的“未来”。在此过程中，由于“未来”本身就是算法的参与或主导下建构的，因此会形成“自证预言”或“自我实现的预言”悖论问题。

现实算法治理实践中应用的预测类算法的典型案例分析“预测警务”算法。预测警务算法会基于过去的犯罪逮捕数据来绘制城市的“犯罪热点地图”并进行警力配置。如果该算法错误地将某个低风险区域标注为“高犯罪风险”，警方会投入更多警力在该区域监视或巡逻，从而可能发现更多犯罪行为并带来更高的逮捕率。这一结果看似“证明”了算法预测的准确性，但实际上只是算法干预下的“自证预言”。^[20]同时，前述结果作为新的训练数据也会被用来强化原有的算法模型，从而导致算法预测出现难以发现和纠正的“负反馈循环”问题，不当的警力配置得以延续，此时算法预测结果的真实准确性将难以判断。

4. 去政治化悖论

算法理性以及算法统治的非人化特征为算法治理赋予了“客观中立”的形象，基于算法的自动化决策常被认为可以摆脱特殊利益、意识形态和权力主体的干预或操纵，从而呈现为一种去政治化的治理实践。然而，算法技术不仅会在设计与应用过程中被人为地嵌入政治意图，其本身也具有内在政治性，会呈现出特定政治倾向进而产生社会政治影响。

在算法治理的实践过程中，科技企业及其背后的资本力量通过主导算法代码与程序的设计与部署，实现了对算法的实质性控制，甚至支配政府的日常行为。这一过程催生了算法权力，并随之带来了以私人利益置换公共利益的治理风险。在此情况下，算法设计的主导者却可以基于“算法中立论”来规避自身责任，算法的直接使用者则成为主要问责对象。由此来看，算法中立论本身就是一种“技术政治性”的体现，它产生了影响责任分配的政治后果。进一步而言，算法的内在政治性主要强调算法技术自身的能动性或意向性，其问题更加具有隐蔽性和抽象性而容易被忽视。即便没有特殊利益或价值意图的人为嵌入，算法应用也会导致利益再分配或差异化赋能等各种社会政治后果，在优先满足特定群体利益与价值的同时，对其他群体造成伤害。总之，去政治化本身就是一个政治过程，由此形成一种去政治化悖论。现实中，各种算法歧视、排斥、压迫等不公平现象可能在“去政治化”的表象下被自然化或合理化，这些正是需要进行批判性审视的重要问题。

三、算法治理的范式变革

如前所述，在算法治理中，意在追求确定性而应用的算法技术却可能引发更多的不确定性，意在实现公共价值却可能造成公共价值的损害，意在预测未来却可能导致过去的延续，意在祛除政治影响却可能导致更多的政治化。这些悖论难题构成了算法治理的根本挑战，也成为算法风险产生的深层根源，因此亟须对“算法治理”范式本身进行变革，也即算法治理的“再治理”。首先，针对算法理性面临的不确定性悖论来讨论算法治理的理性基础重构问题；其次，基于对算法统治中的“价值中立”误区以及去政治化悖论的认识，讨论对算法治理实践进行“再政治化”与政治控制的必要性；最后，基于算法技术的内在政治性，从算法设计角度探讨重塑算法技术发展导向以解决价值权衡、自证预言等悖论问题的可能性。

1. 不确定性背景下算法治理的理性基础重构

如前所述,机器学习算法从数据到模型以及输出结果等方面存在多重不确定性或“无知”问题,这表明基于大数据分析的“算法理性”事实上存在自身的局限性,而无法为算法治理决策提供充分的确定性知识。因此,政府治理主体在运用大数据与算法分析工具时,要清晰地认识到其自身难以克服的局限性,从而谨慎地确定它们的应用领域、范围和方式。例如,对“大数据”的强调并不代表它可以取代“小数据”即“抽样数据”。所谓的大数据事实上可能忽视或排斥了许多难以数字化的人、物或价值,此时大数据的代表性反而不如传统的抽样数据。而且大数据分析所暗含的“数据普遍主义”假设将数据抽离于其社会文化背景,因此会造成大数据的“肤浅性”,此时诸如“深度数据”“厚数据”等其他数据类型的作用不可或缺。^[21]同样,算法分析所得出的相关性知识也无法取代因果理论,后者相较于前者而言具有更为显著的稳定性和确定性。

鉴于不确定性背景下算法理性的局限性,除了上文所述对“小数据”等多种数据类型以及因果理论的重视外,许多学者基于科学理性的局限性而主张的“社会理性”对于算法治理来说同样具有借鉴价值。基于科学理性与社会理性的结合,贝克(Ulrich Beck)提出了“风险理性”的概念,它是指“在一种开放的、允许充分弹性的新思维模式下,全方位地认识风险的各种可能来源与可能后果”。^[22]这种基于风险认知的理性思维对于当前的算法治理改革尤为重要。这是由于算法技术应用具有显著的隐蔽性特征,社会公众多在“技术无意识”状态下与算法发生遭遇。^[23]

因此,针对算法治理的风险理性需要唤起包括普通公众在内社会主体的算法风险意识,针对算法技术相关的不确定性风险进行持续的集体反思、沟通与学习。就此而言,风险理性也是一种“学习型理性”,它无法直接解决不确定性与无知问题,但是经由不断的集体学习可以完善对算法技术属性与应用后果的认知,从而有助于在技术创新与安全以及公众接受之

间做出更为审慎的选择。

2. 算法治理的“再政治化”与政治控制

如前所述,在基于算法理性的算法治理中,由算法系统(模型或代码)掌握决策权力的“算法统治”成为一种理想化的实践模式,其逻辑前提则是算法的“价值中立”预设。在当前的算法治理实践中,必须正视算法治理中不平等的权力结构特别是算法自身蕴含的政治属性问题,从而对算法以及算法治理实践进行“再政治化”,其核心是重新塑造算法技术与专家、政府、企业和公众等主体之间的关系,进而谋求对算法治理过程施加有效的政治控制,确保政府治理场景中算法应用过程与结果的合法性和正当性。

具体而言,“再政治化”要求打破“政治-非政治”的二元划分。对于那些能够影响社会资源或利益分配的算法而言,算法治理从设计、配置到应用的各个环节所涉及的决策在实质上都属于政治决策,而非纯粹的技术决策。这意味着,对算法各环节的规制不能仅停留在伦理道德层面的规范或约束,而有必要进行更为有效的政治控制。在实践模式上,“再政治化”强调要包容多方参与者进行审议性实践。与此相应,对算法治理进行政治控制的核心则是以公共监督和社会参与为核心的民主控制,它强调要将算法治理置于受到算法影响的各方主体共同参与讨论、批判性反思和学习的氛围中,以此形塑算法应用的恰当范围和方式。然而包容性审议模式需要思考和解决如何确保普通民众参与能力、搭建有效对话渠道,以及“算法无意识”下弱势群体话语权缺失等问题。

除此之外,当前亟须重视社会层面越来越多针对算法风险的社会论争和抵制行动。越来越多的社交媒体用户开始意识到自身正成为算法“算计”的对象,并主动采取包括数据隐匿、选择性互动、平台迁移等策略进行规避与抵抗,类似算法抵制实践被称为“算法行动主义”。^[24]随着这种算法行动主义范围和影响力的扩大,逐渐显现出贝克所揭示的“亚政治”特征,由此构成算法治理改革的社会压力或动力。就此而言,通过有效方式激发算法用户尤其是普通

公众的算法意识特别是算法风险意识，成为对算法治理进行政治控制的重要基础。

3. 以公共性为导向的算法设计

在技术政治理论看来，技术并非中立工具，而是发挥着一种“中介”作用，会“框定着我们该做什么以及我们如何体验世界”，“技术不仅仅提供手段，而且也促进新目的的形成”。^[25]另外，技术自身包含着某种形式的意向结构或价值倾向性，会优先服务于特定的利益或价值诉求，并排斥其他利益或价值。鉴于算法的这种政治性或能动性，应通过算法的重新设计来改变其“中介”作用或价值倾向，从而使算法技术能够更好地服务于公共利益或公共价值的实现。这要求科技专家在进行技术设计时，不仅要关注所设计的对象，而且要考察自身在设计过程中做出的价值判断或意图、诠释、权力和责任。例如当前机器学习算法因其高效率而备受推崇，但是我们不能忽视在该技术的设计中，效率价值的实现是以其他社会价值的受损为代价的。而如果想要优先实现其他价值，则需要在算法设计过程中做出不同的价值权衡选择。

公共部门的算法治理应坚持公共价值导向，算法设计同样应以公共性为导向，即要求算法技术自身能够体现出对公共性的倾向性，或者算法能够更好地扮演促进公共性价值实现的中介角色。在具体实现方式上，首先要改变算法对数据的无度提取，尽可能降低对公众隐私权的侵犯，需要拒绝“不断收集人类数据是一种组织人类生活的理想方式”。^[26]联合国大数据和数据科学专家委员会在2023年初发布了《隐私增强技术指南》，旨在从技术设计层面更好地保护公众隐私，这种技术发展导向在很大程度上契合了算法公共性理念。其次，亦可倡导“带有小数据主义色彩的绿色人工智能技术”，它是一种“对现有的人类价值体系扰动较小的技术”。^[27]

最后，针对算法只能揭示数据的相关性关系从而导致的难以解释、自证预言悖论等问题，一些科学家正探索如何设计出“可以提供自我解释的机器学习算法”^[28]等新路径。另有学者提出了一种“诠释学”导向的数字技术或算

法设计思路，它旨在改变传统数字技术对公众信息的同质化处理。这种新的设计将算法从纯粹的计算工具转变为对个人或集体解释的支持，为各种异议和争论提供了场所，因此被视为具有了“公共空间”的特征。^[29]类似这些以公共性为导向的算法设计思路值得引起人们的关注。

结 语

在人工智能与政府治理深度融合的背景下，以机器学习为代表的算法技术凭借其显著的效率优势与数据分析能力而被广泛应用于政府治理各领域。然而，诸如算法错误、偏见、歧视或压迫等各类算法风险问题的不断显现则对算法治理的转型发展构成了严峻挑战。本文认为，算法风险的产生不仅源于使用者的不当应用，还有着更为深层次的根源。为了深入探究算法治理中算法风险的根源，本文对算法治理的逻辑基础特别是认识论、方法论和价值方面的前提假设或预设条件进行了微观考察和批判性反思，揭示了其中存在的各种内在悖论。本文认为，算法风险的不断产生实质上正是算法治理结构性基础中内在悖论的客观呈现。不同于一般意义上的危害或负面问题，“悖论”作为算法治理主体所面对的矛盾困境，很难通过简单直接的应对策略予以化解。对此本文指出有必要推动算法治理的范式变革，通过理性基础的重构进而谋求重塑算法的应用模式以及技术发展方向，以期算法治理能够更好地服务于公共性价值的实现。

需要说明的是，本文的主要意图在于弥补政府算法治理改革的现有研究在微观逻辑基础方面的缺失，而非针对特定算法技术应用模式的实践考察。在现实治理场景中，算法治理往往受到不同治理场景的深刻影响，算法技术的嵌入方式、治理主体间的互动结构及其权力分配均呈现出高度的情境依赖性，其具体运作可能会呈现出比单纯理论推演更为复杂的逻辑。因此，对算法治理运行逻辑的深入理解还需要结合对具体实践过程的细致观察和剖析。就此

而言,本文最后针对算法治理范式变革的讨论更多体现为一种方向性引导或理念构想,如何将前述理论构想转化为现实实践方案仍需治理模式、工具与评估标准等可操作层面上的具体考量,这是未来研究需要进一步探索的方向。

[参考文献]

- [1] 张爱军、孙玉寻. 算法权力及其国家能力形塑的主体透视[J]. 学术月刊, 2021, 53(12): 96-105.
- [2] 梁玉成、政光景. 算法社会转型理论探析[J]. 社会发展研究, 2021, 8(3): 21-43.
- [3] 张敏. 算法治理: 21世纪的公共管理现代化与范式变革[J]. 政治学研究, 2022, (4): 50-62.
- [4] Alon-Barkat, S., Busuioc, M. 'Human-AI Interactions in Public Sector Decision Making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice'[J]. *Journal of Public Administration Research and Theory*, 2023, 33(1): 153-169.
- [5] 宋冰. 智能与智慧: 人工智能遇见中国哲学家[M]. 北京: 中信出版社, 2020, 165.
- [6] Morgan, A., Mazzotti, M. *Algorithmic Modernity: Mechanizing Thought and Action, 1500-2000*[M]. New York: Oxford University Press, 2023, 15.
- [7] Danaher, J. 'The Threat of Algocracy: Reality, Resistance and Accommodation'[J]. *Philosophy & Technology*, 2016, 29(3): 245-268.
- [8] 刘河庆、梁玉成. 透视算法黑箱: 数字平台的算法规制与信息推送异质性[J]. 社会学研究, 2023, 38(2): 49-71.
- [9] 莉娜·丹席克等. 数据正义[M]. 向秦译, 上海: 上海人民出版社, 2023, 45.
- [10] 董春雨. 从机器认识的不透明性看人工智能的本质及其限度[J]. 中国社会科学, 2023, (5): 148-166.
- [11] O'Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*[M]. New York: Penguin, 2016, 38.
- [12] 乌尔里希·贝克、安东尼·吉登斯、斯科特·拉什. 自反性现代化: 现代社会秩序中的政治、传统与美学[M]. 赵文书译, 北京: 商务印书馆, 2001, 5-6.
- [13] 林文源. AI的能与不能: 以医疗与照护为例发挥人社想象力[J]. 人文与社会科学简讯, 2019, 20(2): 99-103.
- [14] 肖瑛. 从“理性vs非(反)理性”到“反思vs自反”社会理论中现代性诊断范式的流变[J]. 社会, 2005, (2): 1-24.
- [15] 本·格林. 足够智慧的城市: 恰当科技与城市未来[M]. 李丽梅译, 上海: 上海交通大学出版社, 2020, 114.
- [16] 人工智能学家. AI大模型出现了人们不可预测的能力[EB/OL], 凤凰网, <https://tech.ifeng.com/c/8Ooc7ZyLiCW>. 2023-04-08.
- [17] 吴冠军. 大语言模型的技术政治学评析[J]. 中国社会科学评价, 2023, (4): 27-37.
- [18] Beer, D. *The Tensions of Algorithmic Thinking: Automation, Intelligence and the Politics of Knowing*[M]. Bristol: Bristol University Press, 2023, 100.
- [19] 吴畏. 算法的认识论逻辑[J]. 哲学动态, 2023, (3): 117-126.
- [20] Busuioc M. 'Accountable Artificial Intelligence: Holding Algorithms to Account'[J]. *Public Administration Review*, 2021, 81(5): 825-836.
- [21] 狄波拉·勒普顿. 数字社会学[M]. 王明玉译, 上海: 上海人民出版社, 2022, 121.
- [22] 顾忠华. 第二现代: 风险社会的出路?[M]. 台北: 巨流图书公司, 2001, 26.
- [23] Beer, D. 'The Social Power of Algorithms'[J]. *Information, Communication & Society*, 2017, 20(1): 1-13.
- [24] Velkova, J., Kaun, A. 'Algorithmic Resistance: Media Practices and the Politics of Repair'[J]. *Information, Communication & Society*, 2021, 24(4): 523-540.
- [25] 彼得·保罗·维贝克. 将技术道德化: 理解与设计物的道德[M]. 闫宏秀等译, 上海: 上海交通大学出版社, 2016, 1, 58.
- [26] 戴林杰. 大数据、殖民主义与资本主义——评《连接的代价: 数据如何殖民人类生活并为资本主义所用》[J]. 智能社会研究, 2023, 2(3): 199-207.
- [27] 徐英瑾. 人工智能哲学十五讲[M]. 北京: 北京大学出版社, 2021, 154.
- [28] Rudin, C. 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead'[J]. *Nature Machine Intelligence*, 2019, 1(5): 206-215.
- [29] Alombert, A. 'From Computer Science to "Hermeneutic Web": Towards a Contributory Design for Digital Technologies'[J]. *Theory, Culture & Society*, 2022, 39(7-8): 35-48.

[责任编辑 李斌]