doi:10.13582/j.cnki.1672-7835.2020.06.012

算法“黑箱”的成因、风险及其治理

谭九生,范晓韵
(湘潭大学 公共管理学院,湖南 湘潭 411105)

摘　要:算法作为一种前沿技术,极大地提高了社会运行效率,不断推进了社会发展模式革新。然而,由于算法技术的复杂性、相关法律政策的匮乏、算法素养的限制、公开算法动力不足以及算法安全维护等原因,形成了算法“黑箱”。“黑箱”的存在使得算法在公共领域的应用中,极易衍生出私人资本支配公权力、算法监管的政府失灵、政府信任危机等问题。对此,必须对算法“黑箱”加以治理,治理的关键在于实现“有意义的算法透明度”,具体可遵循构建政府责任义务体系、推动技术公司算法公开、发挥社会组织监督作用等路径。

关键词:算法;算法黑箱;风险;算法治理
中图分类号:D630;TP18　　**文献标志码**:A　　**文章编号**:1672-7835(2020)06-0092-08

大数据、算法和算力的紧密结合给社会运行带来了颠覆性的影响,使我们日益生活在一个“算法社会”①之中。搜索引擎对于用户的信息匹配算法、医疗行业对于病症的诊断算法、公共应急部门对于灾情疫情的估测算法、警用系统对于民众的犯罪风险评估算法,都是算法被广泛应用并改变社会运行与发展模式的典型案例。然而,算法技术的复杂性、相关法律政策的缺乏、算法素养的限制、技术公司公开算法动力不足以及算法安全维护等因素耦合,形成了算法“黑箱”。随着算法在公共领域中的深度融合应用,由算法“黑箱”而衍生的一系列风险,比如私人资本支配公权力、算法监管的政府失灵、政府信任危机等,也引发了社会各界对算法的普遍忧虑。因此,研究算法“黑箱”的成因、风险及其治理,这对发挥算法的技术价值、加强政府依托算法进行治理的正当性具有重要的理论和现实意义。

一　算法“黑箱”形成的原因

算法(algorithm)一词是由拉丁语 algorism 演变过来的,是公元 9 世纪波斯数学家花拉子密(Al-Khwarizmi)的音转,他在其著作《代数学》中引入印度-阿拉伯数字系统及配套算法来解决数学问题,因此,算法意为“花拉子密”的运算法则,成为了任何运算方法的统称。但现在所使用的算法概念已远远超出最初的定义了,不同的学科从不同维度对算法展开研究。概览各学科对算法概念内涵与外延的界定,整体上可分为两个层面:一是关注算法技术本身,表现为数学结构或计算机代码,如数学领域认为算法是通过数字符号、图表等数学工具来解决某一问题的操作程序;计算机科学领域将算法视为使用计算机执行计算或解决问题时的一系列指令②。二是凸显算法技术运用,表现为用计算机语言呈现的影响社会运行规则与过程的技术程序,如公共管理学领域认为算法是一种增强或取代人类分析和决策的特殊决策技术③;或将算法视为建构社会秩序的理性模型,是以计算机代码为载体的“社会运行的基础规则”④。本文从公共管理学科视角出发,聚焦算法

收稿日期:2020-03-15
基金项目:湖南省教育厅重点项目(18A058);湖南省研究生科研创新重点项目(XDCX2020B023)
作者简介:谭九生(1973—),男,湖南茶陵人,博士,教授,博士生导师,主要从事公共行政理论研究。

①Jack Balkin. “Free Speech in the Algorithmic Society: Big Data, Private Governance and New School Speech Regulation”, *U.C. Davis Law Review*, 2017(51):1149—1210.

②Great Britain. *Select Committee on Artificial Intelligence*. AI in the UK: Ready, Willing and Able? .HL Paper100.

③Brent Mittelstadt, et al.,“The Ethics of Algorithms: Mapping the Debate”,*Big Data&Society*,2016, 3(2):3.

④贾开:《人工智能与算法治理研究》,《中国行政管理》2019 年第 1 期。

技术在公共领域中的应用，将算法视作解决社会问题的方法或推动社会运行的技术程序。

（一）算法“黑箱”是什么

黑箱理论源于控制论，指不分析系统内部结构，仅从输入端和输出端分析系统规律的理论方法，这里的“黑箱”是一种隐喻，指的是“为人的不知的、那些既不能打开，又不能从外部直接观察其内部状态的系统”①。而算法“黑箱”与理论上作为系统的“黑箱”又有所区别，算法“黑箱”本质上归属于技术“黑箱”，技术“黑箱”特指作为知识的人工制造品，“其特点是部分人知道，另一部分人不一定知道”②。在这个意义上，算法“黑箱”指的是算法运行的某个阶段“所涉及的技术繁杂”且部分人“无法了解或得到解释”③。

算法“黑箱”与机器学习技术的发展相随，本文根据算法“黑箱”的发展程度将其分为三种常见情形：第一种情形属于算法“黑箱”的初级形态④，与监督式机器学习技术相对应，算法在人为预先确定的结构化数据及规则的基础上运行，输入端的训练数据和输出端的算法目标都是已知信息，那么这里的“黑箱”是指在输入端和输出端之间存在的不能被观察的“隐层”⑤，其特点在于大部分人已预先了解算法信息。第二种情形属于算法“黑箱”的中间形态⑥，对应半监督式机器学习技术，通常情况下，算法的输入端是无人干预的，数据挖掘、数据收集等程序是自动运行的，具有不透明性，但经过算法运行后的输出端是人为预先确定的，因此这里的“黑箱”指的是存在于算法输出之前我们无法洞悉的算法流程，其特点在于部分人了解算法，但另一部分人不一定了解。第三种情形属于“算法‘黑箱’的进阶形态”⑦，对应无监督式机器学习，算法基于训练数据构建程序模型，来模拟人类学习行为等智能活动，并不断利用经验数据来完善已有程序模型、改善自身性能，这一过程可以在没有人为干预的情况下进行。具体而言，在输入端，算法凭借自动学习能力能够自动对数据进行挖掘和收集；在输出端，“学习型算法”凭借高级认知能力自动生成和改善程序模型，因此该情形下的“黑箱”是指包含输入端与输出端的一个全流程闭环，其特点在于仅有机器了解算法，而人不了解。

纵观算法“黑箱”的三种情形，本文主要针对算法“黑箱”的中间形态进行风险治理探析。一方面，就目前发展阶段与实际应用而言，监督式机器学习技术在当下是主要情形，大部分算法仍需在人为监督下运行，这意味着在数据、模型与目标都已预先确定的前提下，初级形态的算法“黑箱”带来的问题是不足为惧的。另一方面，随着人工智能技术更为广泛的应用，人工智能系统的自主性将使人为决策逐步被自动化决策所替代，无监督式机器学习技术将逐渐取代监督式机器学习技术，这意味着整个算法流程都脱离人的监督与干预，而在现阶段，要治理进阶形态的算法“黑箱”是不太现实的。

（二）算法“黑箱”的形成

算法“黑箱”是一个重要的技术价值问题，因为“黑箱”的不透明与难以解释关涉人们的知情利益以及直接影响人们对算法的信任感与认同度。中间形态的算法“黑箱”之所以形成，既包括算法技术本身这一普遍原因，也包括算法技术应用的社会环境这一特殊原因。

算法“黑箱”形成的技术原因，在于算法本身具有的高度技术复杂性和专业性。从算法的内涵构造来说，包括技术基础层（输入端）、技术程序层以及技术结果层（输出端）三部分，从算法的运行流程来说，整个过程涉及庞杂的数据材料和繁复的算法方法，并以计算机代码的形式呈现，而不是能够被大多数人所理解的自然语言。这意味着除了少数算法研发人员之外，更多的外部人员并不清楚算法的设计理念与目标，也无从获悉数据

①陶迎春：《技术中的知识问题——技术黑箱》，《科协论坛（下半月）》2008 年第 7 期。

②陶迎春：《技术中的知识问题——技术黑箱》，《科协论坛（下半月）》2008 年第 7 期。

③仇筠茜，陈昌凤：《基于人工智能与算法新闻透明度的“黑箱”打开方式选择》，《郑州大学学报（哲学社会科学版）》2018 年第 5 期。

④Diakopoulos N. “Algorithmic Accountability: Journalistic investigation of computational power structures”，*Digital Journalism*，2015，3（3）：398-415.

⑤许可：《人工智能的算法黑箱与数据正义》，《社会科学报》2018 年 3 月 29 日第 6 版。

⑥Diakopoulos N. “Algorithmic Accountability: Journalistic investigation of computational power structures”，*Digital Journalism*，2015，3（3）：398-415.

⑦张淑玲：《破解黑箱：智媒时代的算法权力规制与透明实现机制》，《中国出版》2018 年第 7 期。

的挖掘方式与确证情况,更谈不上确定算法责任归属问题及监督评估算法,因此算法对于大多数人而言,是一个难以理解的"黑箱"。

算法技术在应用过程中也会形成"黑箱",中间形态算法"黑箱"的特点在应用过程中的具体表现是,私人技术公司因负责算法设计与运行而掌控算法,政府公共部门仅仅明确算法输出端的信息即算法目标或算法结果,而社会公众则几乎完全被排除在算法"黑箱"之外。因此,可以从算法技术应用的社会环境来分析中间形态算法"黑箱"的形成原因,主要有以下四个方面:其一,相关法律政策的匮乏与模糊。一方面,我国明确要求算法公开的相关法律政策比较缺乏,只在少数政策文件中对算法提出了要求。当以算法为核心的新一代人工智能技术进入研发部署与实践应用之中时,相应的法律政策却难以适用于现实要求,换言之,相关法律政策的匮乏使得算法处于"黑箱"之中。另一方面,已有的法律政策大多存在模糊不清等问题,仅根据人工智能发展的战略态势对算法技术提出了总体要求,却没有构建相应的配套机制,这给算法公开的落地造成了困难,在这个意义上也形成了算法"黑箱"。例如,《新一代人工智能发展规划》(2017年)明确提出了要"实现对人工智能算法设计等程序的全流程监管",然而,想要在未明确要求算法公开或算法公开标准模糊的基础上对算法进行全流程监管,不过是坐而论道罢了。其二,算法素养的限制。这里的算法"黑箱"主要指的是由于算法素养的限制与鸿沟而产生的一种认知上的"黑箱",因为算法作为一种存在技术门槛的复杂性技术和专业性知识,暂时只能被算法技术人员等少数人掌握,而受到自身算法素养限制的社会公众,由于无法理解算法代码、运行逻辑等信息而被排除在算法"黑箱"之外。另外,算法素养的鸿沟不只是形成在算法设计者与算法消费者之间,更加存在于人与机器之间,人工智能技术凭借巨大的机器优势在人与机器之间形成算法鸿沟。由此可见,算法对于大多数人而言是个无法解读的"黑箱"。其三,技术公司公开算法的内生动力不足。技术公司会因以下几种情境面临法律追责困境而不愿公开算法:公开数据挖掘及数据收集详情,可能会因侵犯用户隐私而被索赔;公开算法运行程序或方法,可能会因程序漏洞、方法不当等技术问题而遭受指控;公开算法运行输出结果,可能会因出现算法歧视、结果不确定乃至错误等问题而招致诉讼纠纷。其四,算法安全的维护与算法保密的需要。由于一些算法信息关涉国家秘密、政府机密以及商业秘密,出于防止产生算法泄露风险的考虑,不仅不能盲目地要求算法彻底公开,还需要对相关算法信息进行保密处理。算法的运行需要训练数据、模型参数等,一旦这些算法信息因公开披露的要求而被泄露,该算法则极有可能被其他技术公司或算法设计者恶意复制与修改。若发生上述情况,国家安全、社会稳定以及市场秩序都将面临威胁,个人数据的安全性也将很难保证。也就是说,这种情境下出现的算法"黑箱"主要是人为构设的,是基于算法安全与算法保密需要而产生的一种"故意不透明"①。

二　算法"黑箱"的衍生风险

算法作为一种技术工具,在应然层面理应是客观中立的,但因算法自身的"黑箱"化特征以及技术应用不当,在实然层面又难免陷入偏私与歧视。私人技术公司知道,而政府公共部门与社会公众不一定知道的中间形态算法"黑箱",其存在可能导致算法与资本或权力相连接②,成为"损害个人权益和社会福利的工具"③,影响社会公众对算法技术价值的认知,削弱政府运用技术进行治理的正当性,阻滞政府治理现代化的进程。

(一)私人资本支配公权力风险

在数字化时代,算法技术在提高政府服务效率、降低成本、提升决策精准度等方面发挥着巨大的潜能,它产生了新的强大力量,但算法"黑箱"是造成力量失衡的根源④。这可以运用信息社会论者贝尔(Bell)提出的知识价值说来进行解释,信息技术作为数字化时代社会经济资源的核心,

①Jenna Burrell. "How the Machine Thinks: Understanding Opacity in Machine Learning Algorithms", *Big Data &Society*, 2016(1-2):1.

②肖唐镖:《中国技术型治理的形成及其风险》,《学海》2020年第2期。

③杜小奇:《多元协作框架下算法的规制》,《河北法学》2019年第12期。

④Kubler, Kyle."The Black Box Society: the secret algorithms that control money and information", *Information, Communication & Society*, 2016(2):1-2.

现有的权力分配结构有可能被信息技术的归属所打破、社会秩序也有可能被信息技术的分布所左右。在实际中，政府公共部门迫于技术手段的限制和技术人员的缺乏，不得不将基础数据的所有权和控制权让与私人技术公司，与技术公司合作实施算法流程的开发，并将算法设计、运行和分析交给技术公司来操作。由此可见，因掌握关键信息技术而操纵算法的私人技术公司占据着优势地位，但政府在“算法社会”中却被边缘化，逐渐失去了对算法关键数据的所有权和控制权，面临着去中心化的挑战。例如，智慧城市的领导者——巴塞罗那，其数字技术负责人 David Meyer 认为，以“黑箱”操作系统告终的城市本身将失去对关键信息和数据的控制权，而这些信息和数据本可以被用于做出更好的决策①。诸多案例显示，控制算法数据及算法分析的技术公司占据了“智慧城市”运动的指挥中心，而“民主负责的政府公共部门则转移到了外围”②，这意味着算法“黑箱”的存在直接导致了公共利益被私人利益俘获、资本支配公权力等风险的产生。正如学者 Aneesh A 所指出，“公共领域算法治理的特点是公共权力屈服于技术公司和其他开发人员的私人控制”③。

（二）算法监管的政府失灵

算法作为一种前沿技术，以其强大的数据处理和分析能力在推动社会从数字化到智能化再到智慧化的过程中发挥着独特优势，对于创新政府治理模式、加速国家治理能力现代化进程都发挥着积极作用。然而任何技术都具有两面性，“黑箱”的存在容易隐藏算法缺陷并触发某些风险，因此，政府应在治理算法方面担负重要职责，以推动算法技术良性健康应用。具体而言，政府治理算法至少要包含以下内容：掌握技术公司生成的关于算法运行目标的记录、审查技术公司披露的与算法设计相关的信息、监管算法运行的各流程，并对整个算法过程中产生的问题进行问责、要求技术公司及时纠偏等。然而，“黑箱”的存在让政府治理算法的内容与手段均受到限制，进而使政府陷入监管算法失灵的困境。算法“黑箱”的存在使得算法漏洞、算法缺陷得以隐藏，一旦算法在数据、模型、方法等方面出现问题，算法应用结果将会与应用目标相悖。政府公共部门作为算法技术应用的推动者，理应对算法进行治理以保证算法技术的良性健康应用，但“黑箱”让政府公共部门难以审查算法信息，几乎无法识别和判断产生问题的病灶所在，以致难以及时针对算法问题进行问责和纠偏。更有甚者，技术公司或算法设计者开始利用机器与人之间形成的“黑箱”，以算法自动化运行为由逃避算法责任，将引发风险的责任归咎于算法技术本身，产生技术外包的责任归属问题，使得政府更加难以问责技术公司和治理算法问题。

（三）引发政府信任危机

鉴于“算法过于依赖规则和数字，而无法对规则和数字之外的因素加以充分考量”④，因此政府在“算法社会”中扮演的角色是我们最为关注的问题。公众期待政府公共部门参与算法、代理算法以及治理算法，构建、维护和巩固算法社会秩序，以实现政府对社会公众负责的目标。但算法“黑箱”让诸多算法流程呈现出不透明的状态，使得公众不知道涉及自身利益的算法设计意图是什么、其数据来源是否正当、算法是如何运行的以及运行结果是否公平，这意味着公民的知情权在一定程度上被遮蔽。在此情境下，易引发社会公众对于算法合理性的担忧，进而引发对政府的信任危机。除了由算法信任危机引发政府信任危机之外，还有以下两个原因：其一，由政府监管算法失灵引发政府信任危机。“黑箱”的存在让政府治理算法的内容与手段均受到限制，加之由于技术限制、资金紧张、人才缺乏等原因，政府只能采取与技术公司合作的方式推动算法工具的开发与应用，但算法“黑箱”的存在使公权力屈服于技术公司和算法设计者的私人控制。另外，根据 Robert 和 Ellen 的调研发现，政府根本就没有太多承包商生成的关于算法创建和实施的记录，包括数据选择、模型设计选择、验证设计记录、拟解决的问

①David Meyer. How One European Smart City is Giving Back Power to its citizens, https://www.alphr.com/technology/1006261.

②Kitchin, Rob. “The real-time city? Big data and smart urbanism”, *Geojournal*, 2014, 79(1): 1-14.

③Aneesh A. Technologically Coded Authority: The Post-Industrial Decline in Bureaucratic Hierarchies, http://pdfs.semanticscholar.org/9455/244cad543ea65e7d8089f73446024be9b2fa.pdf.

④宋华琳，孟李冕：《人工智能在行政治理中的作用及其法律控制》，《湖南科技大学学报（社会科学版）》2018 年第 6 期。

题以及成功的度量标准等①。不难发现,政府对由技术公司所操纵的算法了解甚少,更遑论对其进行审查与治理了,政府治理算法失灵是引发政府信任危机的重要原因。其二,由公共领域算法风险引发政府信任危机。算法应用的目标、政策判断、公平与否都被"黑箱"遮蔽,政府无法确定算法运行的结果是否为"善",实际情况也证明在很多方面应用算法来操作公共事务是具有风险的,通过算法运行做出的预测和决定有可能是不公平的,甚至是错误的。例如美国法院使用商业公司设计的 COMPAS 算法分析工具进行犯罪风险评估,被证明存在种族偏见,该算法系统对黑人造成了歧视②。对于在公共领域应用的算法,公平性的考虑要比算法的性能重要得多,但这并不属于私人技术公司和算法设计者的职权范围,只有政府公共部门的介入可以选择公平性而非客观性。然而,政府在此过程中进入了一个自身无法观察、难以理解、不能解释的算法"黑箱",政府职责的有效性和公平性便变得无法评估,无力对社会公众负责。

三　算法"黑箱"的治理之道

算法既能实现政府治理精准化,切实维护公共利益,同时,算法作为带有明显的风险的信息技术③,也可能因"黑箱"的存在而引发诸多社会风险。换言之,算法"黑箱"的治理,其实质在于如何平衡收益与风险的问题。为了实现此目标,治理算法"黑箱"亟待解决以下几个问题:一是如何确保算法的正当性、可责性与公平性;二是如何实现政企之间基于信任基础上的互动与合作;三是如何消除公众的误解与不信任。对于中间形态算法"黑箱"的治理,我们应当设置一个合理的框架,至少应包括追求"有意义的算法透明度"④、构建政府责任义务体系、推动技术公司算法公开、发挥社会组织监督作用等方面。

(一)有意义的算法透明度

"黑箱"的关键问题就在于无法观察和难以理解,治理算法"黑箱"首先要求的就是打开算法"黑箱"、推进算法透明,但过度追求算法披露,要求算法彻底公开,可能会带来消极后果,甚至适得其反。一方面,不能被多数人理解的公开披露的海量数据,不过是"无意义的数据废墟"⑤;另一方面,人们会简单地将算法彻底公开等同于算法正当,忽视了算法隐含的价值观念以及对公共事务产生的影响。相比较而言,追求"有意义的算法透明度"在实践中更具可操作性。"有意义"的主要含义包括以下几个方面:其一是对政府而言有意义,算法透明能够便于政府审查算法信息、监管算法流程、治理算法问题,以确保算法正当、公平、负责任;其二是对技术公司而言有意义,能够在保护商业机密的基础上,取得社会各界对算法技术的信任;其三是对公众而言有意义,能够实现普遍意义上的算法可理解和算法可监督。具体而言,政府通过制定和完善相关法律法规,对算法公开内容及公开程度做出明确规定;要求技术公司在算法设计阶段嵌入算法伦理;对算法信息进行审查校验、对算法流程进行监管。技术公司应向政府提供算法设计的足够信息、公开披露算法运行说明或原理;在政府要求公开披露的算法信息类型及内容之外,允许存在因保护商业机密而豁免公开算法的有限例外。政府与技术公司应通过合作向公众提供特定的算法访问渠道,便于公众了解算法信息并理解算法运行逻辑,维护其知情权,实施其监督权。总而言之,追求有意义的算法透明度,关键在于解开以上几个层次的透明性,确定透明的程度与形式,而非追求从元数据到算法运行结果的彻底公开。

(二)构建政府责任义务体系

现代社会的组织形式和组织活动中,存在着建立在权力关系、法律关系以及伦理关系基础上的三种责任义务,分别是法律的、行政的以及道德的责任义务⑥。三种责任义务共同构建了政府治理算法"黑箱"责任的法律、行政与道德维度,各

①Robert Brauneis, Ellen P. *Goodman. Algorithmic Transparency for the Smart City*, Social Science Electronic Publishing, 2017.

②https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

③颜昌武,杨郑媛:《什么是技术治理?》,《广西师范大学学报(哲学社会科学版)》2020 年第 2 期。

④Robert Brauneis, Ellen P. Goodman. *Algorithmic Transparency for the Smart City*, Social Science Electronic Publishing, 2017.

⑤Stohl C, Stohl M, Leonardi P M."Managing Opacity: Information Visibility and the Paradox of Transparency in the Digital Age", *International Journal of Communication Systems*, 2016(4):123-137.

⑥张康之:《论社会治理中责任义务的实现》,《浙江学刊》2004 年第 2 期。

维度从同一个完整的责任义务体系中分解出来，相互独立又相辅相成。本文借鉴这一理论框架作为分析的起点，从法律责任、行政责任、道德责任三个方面对治理算法“黑箱”的政府责任体系进行证成。

在法律责任层面，着眼于政府制定、完善相关法律法规，在多元主体之间建立起治理算法“黑箱”的共同准则。准则需要对以下几个问题进行回应：算法的哪些内容应该公开披露，算法内容的公开程度如何，如何把握算法公开和算法安全之间的平衡，如何保障公民的算法知情权。其一，政府应当为技术公司设置公开算法说明的强制性义务，对于算法“黑箱”产生的难以发现的问题进行治理。但就算法公开程度而言，并非是追求算法的彻底公开，而应追求有意义的算法透明度。因此相关法律法规应明确规定，算法公开披露的内容不是算法代码或算法元数据而应当是算法说明，具体包括算法的设计者、设计理念、数据来源、运行参数及变量、假设、逻辑、功能、影响、风险、重大变化等①。其二，为了防止一些关涉国家秘密与政府机密的算法信息泄露，不能盲目地要求算法彻底公开，需要在法律层面上维护算法安全。另外，为了实现政府与技术公司在治理算法“黑箱”问题上的有效协同，需要营造互利互惠基础。政府在要求技术公司公开披露算法内容的同时，也要尊重他们保护商业机密的权利，允许技术公司存在豁免公开算法的有限例外。因此，相关法律法规应明确指出豁免算法公开的前提条件、具体内容以及备案方式等。其三，政府要在法律层面上保障公民的知情权与获得解释的权利，如欧盟通过的《数据保护条例》（2016 年），其引言第 71 条指出了具体信息数据主体有权利获得解释。在这个意义上，相关法律法规应明确要求技术公司为公众设置特定的算法信息访问渠道，例如 Facebook 于 2019 年 5 月推出的“why am I seeing this post”功能，旨在帮助用户理解他们看到该文章或信息的原因、了解算法排序的影响因素，并为用户提供信息首选项或隐私键等快捷方式来控制算法结果。

在行政责任层面，着眼于行政审查、行政监管与行政问责，要求技术公司操作的算法符合法律法规要求、符合公共利益，削弱或消除算法“黑箱”造成的负面影响。其一，在政府与技术公司之间存在着由算法信息不对称而形成的“黑箱”，可以从组织结构着手治理这种不透明问题，政府应设立专门机构、建立算法行政责任制度、组织和培训专业人员，依法对算法进行审查、监管与问责，例如美国“通过组建专门机构和人员构成问责主体的方式建立算法问责制”②。其二，政府应对算法应用进行全流程监管，以确保算法正确并符合道德伦理规范。在 2017 年国务院发布的《新一代人工智能发展规划》③中，明确要求政府应在安全防范、市场监管等方面发挥重要作用，并建立人工智能安全监管和评估体系，实行设计问责和应用监督并重的双层监管结构，实现对算法设计、产品开发和成果应用等的全流程监管。其三，政府应建立和完善算法问责机制。算法问责指政府必须要求相关责任主体对算法“黑箱”所引致的风险进行处理，这意味着技术公司有对算法运行结果负责任的义务，为了在实践中实现这一目标，需要算法问责、算法纠偏等过程的确立和实施。

在道德责任层面，着眼于算法伦理的嵌入，政府应该要求算法设计符合社会道德观念、价值与规范，在一定程度上规避算法“黑箱”带来的算法信任危机与政府信任危机。除非使算法变得足够透明，否则我们将不知道算法运行结果是否符合政府对于公平等民主价值的承诺，但算法“黑箱”是客观存在的，这意味着政府也不知道能否将算法运行结果与自己的行政实践相结合。因此处于“算法社会”中心的政府需要思考几个关键性问题：如何部署及治理算法以促进而非阻滞诸如透明、公平、秩序等民主价值，如何规避或消除元数据中映射的社会固有观念或偏见。算法伦理规范的嵌入和实施对于保证算法开发的公正和道德是非常必要的，而技术公司自觉遵循算法伦理规范是需要明确的，这就要求政府为作为算法设计者的技术公司设置嵌入算法伦理的义务。这已经成为全球各界所探讨的共同问题，如电器和电子工程师协会（IEEE）于 2015 年 12 月发起了全球倡

①徐凤：《人工智能算法黑箱的法律规制——以智能投顾为例展开》，《东方法学》2019 年第 6 期。

②张欣：《从算法危机到算法信任：算法治理的多元方案和本土化路径》，《华东政法大学学报》2019 年第 6 期。

③国务院关于印发新一代人工智能发展规划的通知，http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm。

议,旨在确保设计人员在开发自主与智能系统时优先考虑伦理要求;“Google成立的‘AI伦理委员会’致力于建立算法应用的伦理准则”①;欧盟也于2018年成立了“欧洲科学和新技术伦理小组(EGE)”;我国通过的《人工智能标准化白皮书》(2018年)也指出人工智能设计的目标应与人类的利益和伦理道德一致。因此,政府要在明确算法开发的公共目的基础上,要求技术公司平衡算法的工具理性和价值理性。算法以其精准匹配、自动化运作等独特的优势发挥着重要的工具性价值,不断满足社会对效率、效益的追求,而其他民主价值也是不容忽视的。政府应要求算法设计者树立兼顾效率和公共利益的理念,要求技术公司在设计算法、数据挖掘、数据筛选等阶段均应遵循相应的算法伦理规范,使算法兼具正当性和公平性,以降低算法“黑箱”带来的消极影响,提升社会各界对算法技术的认可度。

(三)推动技术公司算法公开

治理算法“黑箱”不仅是政府的责任,作为算法设计开发和运行主体的技术公司,也有向公众解释算法的义务,本文将从自觉公开算法及披露算法说明两个方面来探讨如何推动技术公司算法公开。

自觉公开算法,要求技术公司要有自愿进行算法公开的意识,主动创建、提供、披露算法记录。开放算法实践可能是使技术公司设计与运行的算法透明化的最佳方法,也就是说,技术公司应从一开始就自觉记录算法模型、算法数据来源与结构、算法运行过程、算法输出以及企业风险自查结果等信息,便于让政府对其进行审查和监管,并主动向政府和公众披露算法的设计理念、运行原理与逻辑等。另外,技术公司也可以通过举办论坛、搭建算法访问平台等方式来实践算法的自觉公开,例如在2018年1月,今日头条通过开展算法论坛的形式,面向行业公开其公司的算法原理,期望借此消除公众对算法的误解,使更多的人理解并信任算法,让算法能够为社会创造更大价值。技术公司自觉公开算法可以看作是一种“推”式透明方法,能够减轻政府“拉”式获取算法记录的负担,也更加能够取得社会各界的信任。

披露算法说明,是基于要求技术公司进行有意义的算法公开而提出的治理算法“黑箱”的关键路径。技术公司可以在治理因算法本身的复杂性以及因算法素养的限制与鸿沟而产生的“黑箱”问题方面做出重要的贡献,需要使从算法输入端至算法输出端的整个流程成为社会各界可观察、可理解的“白箱”,这就要求技术公司公开披露的算法内容是简明易懂的算法说明而非庞大复杂的算法代码。有学者指出,算法说明应包括“算法的逻辑、算法的种类、算法的功能、算法的设计者、算法的风险、算法的重大变化等”②,据此,技术公司应将重点放在披露算法“黑箱”中隐藏的算法模型与公式、算法逻辑与规则、算法功能与风险等。关于这一点,在国际上,欧盟于2016年通过的《全面数据保护法》就规定了技术公司应向用户解释算法隐含的逻辑推理过程③。总而言之,技术公司应在树立自觉公开算法的理念基础上,主动公开算法说明,一方面是为了便于政府相关机构审查与监管,及时发现并纠正算法偏误④;另一方面是为了让更多的人理解算法、参与算法、信任算法,目的在于改善算法,让算法更好地服务社会。

(四)发挥社会组织监督作用

由于算法素养的缺乏与限制,公众几乎在整个算法运行过程中都是处于“黑箱”之中的算法文盲,在治理算法“黑箱”的问题上,公众可以通过政府与技术公司合作搭建的算法公开平台来了解、理解算法,也可以以个体形式寻求算法知情权的保障,甚至可以通过技术公司提供的有限渠道来影响算法运行。但公众由于算法常识薄弱,他们并不清楚算法的目标和意图,难以清楚认识到算法的负外部性问题,更难以对算法运行进行监督和评估。相较之下,社会组织在治理算法“黑箱”方面所能起到的作用更大,大部分公众深谙自身力量较为弱小,因此公众可以自发组成社会组织以期集中力量打开“黑箱”。国外在这个方面已经有了一些典型案例,例如,德国成立了以技

①杜小奇:《多元协作框架下算法的规制》,《河北法学》2019年第12期。

②徐凤:《人工智能算法黑箱的法律规制——以智能投顾为例展开》,《东方法学》2019年第6期。

③张淑玲:《破解黑箱:智媒时代的算法权力规制与透明实现机制》,《中国出版》2018年第7期。

④张恩典:《大数据时代的算法解释权:背景、逻辑与构造》,《法学论坛》2019年第4期。

术专家和资深媒体人为主的以评估监控用于公共领域算法的非营利组织①;美国纽约州颁布的《算法问责法案》要求将公民组织代表纳入监督自动化决策算法的工作组,以确保算法公开与透明。社会组织作为第三方机构,其监督算法的功能是对政府算法责任和技术公司公开义务的补充,能够促进算法透明。在某种意义上,算法透明性与民主价值是一致的,它可以使社会公众了解算法的目的是什么,是否可以提高公共利益,而坚持第三方监督是民主价值实现的题中之义。因此,要充分发挥社会组织监督评估算法的作用,它可以在“黑箱”程度日趋深化的情形下,在一定程度上平衡算法信息的不对称性,促进算法“黑箱”打开。

结语

如果算法本身是完美的,那么算法“黑箱”也就不构成问题。然而事与愿违,实际情况表明算法难免会陷于“黑箱”之中,算法“黑箱”的存在导致算法可能受设计者、使用者操纵而带来诸多风险,因而需要对算法“黑箱”的潜在风险进行规避与治理。在劳伦斯·莱斯格(Lawrence Lessig)看来,网络空间的规管方法“由法律、社会规范、市场机制以及程序架构组成”②,这意味着“算法社会”秩序的重塑、算法“黑箱”的治理需要立法者、行政机关、技术公司以及社会公民的通力协作,本文正是在这个意义上提出了算法“黑箱”的治理之道,致力于防治算法“黑箱”的系列风险,促进算法公开透明,以实现智慧政府治理的发展目标。但我们需要注意的是,人工智能技术的快速发展使得技术人员的作用逐渐间接化,算法能够在没有人为干预的情况下利用数据进行自主运行和迭代训练,自动生成运行结果并不断完善算法指令。如果说由于算法本身的技术复杂性而产生的“黑箱”是建立在算法设计者与算法消费者之间的认知阻隔,那么因深度学习算法的应用而生成的“黑箱”则是横亘在人与机器之间的阻隔,即便作为设计者的算法程序员也难以完全知晓和理解该算法信息。不难发现,这种由深度学习算法生成的“黑箱”是更加难以洞悉和治理的,但由于目前以技术治理技术、以算法监测算法已成为现实,伴随着深度学习算法“黑箱”而来的多种风险便可以通过开发“监督算法运行的算法”来进行治理。

On the Cause, Risk and Its Governance of Algorithm “Black Box”

TAN Jiu-sheng & FAN Xiao-yun

(School of Public Administration, Xiangtan University, Xiangtan 411105, China)

Abstract: As a cutting-edge technology, algorithm has greatly improved the social operation efficiency and constantly innovates the social development mode. However, the algorithm “black box” has been formed, due to the complexity of algorithm technology, the lack of relevant laws and policies, the limitation of algorithm literacy, the lack of open algorithm power and the maintenance of algorithm security. The existence of the “black box” makes the application of the algorithm in the public domain easily lead to the problems, such as the domination of public power by private capital, the government failure of algorithm supervision, and the crisis of government trust. Therefore, the algorithm “black box” must be governed. The key of governance is to achieve the “meaningful algorithmic transparency”, which can be achieved by building the system of government responsibility and obligation, promoting the algorithm openness of technology companies, and playing the role of social organization supervision.

Key words: algorithm; algorithm black box; risk; algorithm governance

(责任校对　莫秀珍)

①张淑玲:《破解黑箱:智媒时代的算法权力规制与透明实现机制》,《中国出版》2018 年第 7 期。

②刘曙光:《劳伦斯·莱斯格:代码:塑造网络空间的法律》,《网络法律评论》2007 年第 8 期。