

人工智能时代国家治理的算法依赖及其衍生难题

梅立润

(武汉大学马克思主义学院, 湖北武汉, 430072)

摘要: 人工智能时代的到来既丰富了国家治理的技术工具选择, 也促使国家治理产生深刻的算法依赖。算法偏见的存在及其隐蔽性会引发国家治理公平难题, 算法黑箱的存在使得不同主体都可以找到卸责理由从而会引发国家治理责任难题。人工智能企业拥有独特的技术优势, 因而人工智能时代国家治理的算法依赖在某种意义上是国家对人工智能企业的技术依赖。应对算法依赖引发的国家治理难题既需要人工智能企业自律, 也需要国家加强监管。建立硬覆盖算法备案说明制度、不定时算法抽检审查制度、开放式算法结果观测制度, 是推进国家监管的关键行动举措。

关键词: 人工智能; 国家治理; 算法依赖; 国家监管; 技术向善

中图分类号: D035

文献标识码: A

文章编号: 1672-3104(2022)06-0123-09

借助技术的力量提高治理效度与水平业已成为现代国家治理的基本经验。人工智能时代的到来进一步丰富了国家治理的技术工具选择, 算法嵌入国家治理是大势所趋, “借助于海量的大数据和具备强大计算能力的硬件设备, 拥有深度学习算法的人工智能机器可以通过自主学习和强化训练来不断提升自身的能力, 解决很多人类难以有效应对的治理难题”^[1]。与此同时, 人工智能时代的到来也促使国家治理产生深刻的算法依赖。由于算法可能出现偏见以及“算法黑箱”现象的存在^①, 因此深刻依赖于算法的国家治理也可能面临一些难题。识别并有效应对算法依赖的衍生难题, 是人工智能时代国家治理需要直面的重要理论与现实问题。

一、算法依赖: 人工智能时代国家治理的新趋向

不断意识到并努力克服自身局限是人类社会演进的鲜明特征之一, 技术在辅助人类克服特

定局限的过程中扮演至关重要的角色。“从人类文明发展的愿景来看, 人类技术发展多以增强能力、减轻负担为主要目的。人类发展进程中的每一次科技革命都是一次人类能力的飞跃, 机械化延伸、增强的是人的体力, 信息化延伸、增强的是人的感知力。”^[2]在某种意义上也可以认为, 正是由于意识到人类在体力和感知力方面的特定局限以及存在克服这些局限的渴望, 人类社会才发动了多次科技革命, 最终以机械化克服了人类的体力局限, 以信息化克服了人类的感知力局限。不过, 人类对技术的期待与追求不仅仅停留于此。随着人类逐渐意识到在成长和发展过程中所面临的绝不仅仅是体力或感知力方面的局限, 在人类引以为傲的智能方面也存在巨大的突破空间, 于是便又开始探索模仿、拓展或延伸人类智能的技术。时至今日, 已经形成一股席卷全球的智能革命浪潮或者说以人工智能为标志的科技革命浪潮。

随着人工智能相关专项国家战略的密集出台、新创人工智能企业的不断出现、人工智能领

收稿日期: 2022-03-10; 修回日期: 2022-06-14

基金项目: 国家社科基金重大项目“推进全面从严治党重大理论和实践问题研究”(19ZDA009); 中央高校基本科研业务费专项资金资助项目“人工智能时代国家善治的实现机制研究”(413000111)

作者简介: 梅立润, 男, 湖北阳新人, 博士, 武汉大学马克思主义学院讲师, 主要研究方向: 人工智能与国家治理、政党特质, 联系邮箱: meisanbai@163.com

域投融资的持续火热以及人工智能具体应用场景的日益丰富,人类社会的智能化程度正在不断提高,人工智能时代、智能社会已经从概念走向实践。由于算法是人工智能的“灵魂”,是人工智能的技术内核,因此人工智能时代和智能社会的到来即意味着“算法时代”“算法社会”和“算法生活”的到来^[3-5]。众多的事实与现象也已表明,“算法在现代社会愈发重要,通过挖掘分析和创新使数据产生了更大价值,提高了企业的运行效率,优化了产品和服务的质量,提高了现代社会的自动化程度,影响着社会生活的各个方面,显著提高了社会的运转效率和人们智能化生活的水平,甚至通过预测和辅助决策等功能决定着人类社会未来的走向”^[6]。不可否认,当前的人工智能技术应用主要集中于经济领域和社会领域,或用以促进一系列产业的转型与升级,或以为社会大众提供更便利便捷、体验感更好的各种生活服务,而在“算法统治的世界”^[7],算法的作用最终会从经济领域、社会领域逐渐扩散渗透到国家治理领域。一方面,这符合历次科技革命中标志性技术作用扩散的轨迹规律。虽然鉴于治理的广泛关联性影响,国家在决定是否将新技术应用于治理事务中时相对要更审慎,这在一定程度上会延缓新技术的应用进程,但是不会完全封阻这种进程。另一方面,这也契合现代国家治理的回应性特征。毕竟当人工智能带来更佳的生产与生活体验之后,社会大众可能出现“期待转移”,开始期待国家同样利用人工智能提供更佳的治理体验或公共服务体验,而且一旦国家利用人工智能提供更佳的治理体验或公共服务体验之后,就只能按照“百尺竿头更进一步”的逻辑继续努力提供比之前更好的治理体验或公共服务体验。不管比之前好多少,总之就是不能比之前差或者长期保持在同一水平,因为社会大众的心理预期会“水涨船高”。在国家治理越来越注重回应社会期待的环境中,这种“期待转移”在某种程度上会促使或倒逼国家不断扩大人工智能在国家治理领域的应用范围与程度。

随着人工智能嵌入国家治理的范围不断扩大以及程度不断加深,国家治理对算法的依赖程

度也会不断增加。一方面,国家治理日益需要算法辅助进行信息处理。在信息爆炸时代,“一个越来越明显的悖论是,政府的信息收集系统收集了大量的治理信息,却不重视这些信息,更缺少整合、处理这些信息的能力,导致大量的信息沉睡在系统中”^[8]。在人工智能时代,国家可以通过算法辅助按需要快速对各种特定类型的海量信息进行初步归类和整理。比如北大方正技术研究院推出的“方正智思舆情预警辅助决策支持系统”,可以利用算法对海量互联网信息进行自动抓取、自动分类、自动聚类、主题识别与追踪,最终形成简报、报告、图表等分析结果。另一方面,国家治理日益需要算法辅助识别政策问题。在现代社会,人们越来越习惯于遇到问题向网络求解,由此在一些搜索引擎上留下了丰富且真实的“数字足迹”或者说网络行为数据,“基于需求的主动搜索行为反映了人们对某些主题的关注度”^[9]。通过算法对海量“数字足迹”或者说网络行为数据进行挖掘和分析,可以更为全面、客观和清晰地刻画出社会公众关注的议题版图,从而为国家制定公共政策提供更加丰富的议题清单和更加科学的决策依据。再者,国家治理日益需要算法辅助监测或预判社会问题与社会风险。在充满变化和不确定性的现代社会,“未雨绸缪”或者说预见性地制定一些公共政策来应对可能出现的社会问题与风险,越来越成为检验一个国家治理水平的重要维度,而人工智能系统或算法在这方面同样可以“大显身手”。比如北京市怀柔公安分局研发的“犯罪数据分析和趋势预测系统”,基于对当地既往大量犯罪案件数据的分析,能够自动预测未来某段时间、当地某个区域可能发生犯罪的概率以及犯罪的种类。更为关键的是,“这些在过去需要多名专业分析人员花费几天甚至几个月才能做出的分析,如今只需点击几下鼠标就能直观地呈现出来,而且预测结果更加科学、准确”^[10]。尽管算法与国家治理的结合已经日益成为一种潮流,但是鉴于算法本身的技术特性以及国家治理的广泛关联性影响,人工智能时代的国家治理需要特别警惕算法依赖可能引发的衍生治理难题。

二、“隐蔽的不公平”：人工智能时代国家治理的公平难题

人类社会的整体生存与发展不仅需要依持各种物质生产技术的更新进步，还需要依持各种价值理念的确立与修正。公平既是人类社会价值理念系统中的原初构成，也是政治的本质要求和国家治理的核心追求。国家治理公平与否，小至影响社会个体的切身福祉，大至关乎民心向背或者一个政权的兴衰存亡。国家治理越是能充分细致、不偏不倚、一视同仁地考虑各方利益诉求，国家治理的公平性就越高，反之则越低。在前人工智能时代，许多国家主要依恃广泛的代表机制和科层组织以人工采集与分析信息的方式来发现社会公共问题以及识别和处理其中的各种利益诉求。在人工智能时代，“万物皆数”且“万物互联”，许多社会公共问题在生成与爆发之前或多或少会释放出一些“数字信息”，采集与初步分析这些信息的任务主要由相关算法来完成。换言之，国家所发现的社会公共问题以及其中的各种利益诉求是经由相关算法整理与挑选后以“清单”样式呈递在国家面前的。在此过程中，算法扮演着某种类类似于“守门人”的角色，算法通过筛选与过滤决定了何种社会公共问题以及谁的利益诉求可以进入国家的视野或吸引国家更多的注意力。与此同时，算法还可以通过辅助决策或直接决策决定各种社会公共问题以及利益诉求会得到怎样的处理。然而，算法偏见的存在可能使国家治理的公平性难以得到保证。

一方面，作为算法基础的数据难以保证无偏见。由于算法需要数据喂养才能发挥作用，无数数据支撑的算法就是“无根之木、无源之水”，因此国家治理算法依赖的背后其实是数据依赖。而“数据，在本质上，是人类观察世界的表征形式……人类文化是存在偏见的，作为与人类社会同构的大数据，也必然包含着根深蒂固的偏见”^[11]，并且“这些偶然甚至错误的数据从源头改正本就很难，通过大数据分享、加工、提供等方式传播出去之后再想彻底改正更是难上加难”^[12]，这使得以数据为原料喂养出来的算法，

很有可能隐蔽地继承、传导、循环甚至放大其数据母体中含有的偏见。比如，微软2016年3月在Twitter上推出了一个名为Tay的AI聊天机器人，Tay可以通过与网友对话学习怎样交谈，Tay起初还给人一种善解人意、活泼可爱的感觉，然而不到一天时间就被网友“教成”了一个满嘴脏话甚至说出一些极端话语的机器人，随后被微软迅速关闭并下线。“Tay是人工智能映射人类偏见的一个缩影，目前的人工智能算法最本质的规则就是它需要有大量的数据对其进行训练——如果训练的数据本身就带着偏见、错误以及极端的思想，训练的结果就会偏离正常的结果。”^[13]在人工智能时代，国家治理行动的展开所依持的是规模巨大、类别复杂、质量参差不齐的数据，在此基础上训练出来的辅助国家治理的相关算法，也就难以确保无偏见。

另一方面，算法设计过程存在人为操控的空间。在人工智能时代，虽然算法操控着众多事务的运行，但是当算法操控其他事务的时候，其自身也存在被操控的可能，毕竟“算法决策总是需要用到一些目标变量或指标，或者对分析对象进行某种分类。而指标的选取、赋值以及分类方法都是人为设定的。如果这一过程受歧视性观念影响，算法决策产生系统性的歧视后果就不足为奇了”^[12]。正是因为算法设计者可能将自身的偏见嵌入算法中，所以人们才会认为“算法是一面镜子，映射了人类社会许多固有的偏见”^[14]。实际上，在一个充满算法的时代，算法偏见的情形并不少见。比如美国司法部门在罪犯出狱之前会对其进行再犯罪评估，以判断其是否适合出狱或出狱后是否需要采取必要的监视措施。为此，美国司法部门采用了Northpointe公司推出的风险评估系统，而该系统对黑人给出的再犯风险评分是白人的两倍。在洛杉矶市，一名轻微犯罪的黑人妇女被标记为“高风险”，而一名曾经两次武装抢劫的白人男子却被标记为“低风险”。最后的结果是，那名黑人妇女并没有再犯罪，而那名白人男子却再次犯了盗窃罪。国家治理事务相当繁杂，这使得在一些配套性算法解决方案的设计过程中，设计主体嵌入自身偏见的机会更多，空间更大。国家治理具有更加广阔和深刻的影响，当

特定国家治理事务采用的配套性算法解决方案存在偏见时,所带来的负面效应或者说对国家治理公平性的损耗将是巨大的。

更为关键的问题是,算法的隐蔽性和复杂性使得潜藏在其中的偏见很难被发现。有人举了一个非常生动的例子,“如果你是一个穷人,以前买过的东西都是相应品类中单价最低的,那么,在系统画像里一个名为‘价格敏感程度’的维度中,你的分数就会非常高。于是乎,当你搜索一个关键词后,从前到后翻十页,你看到的都是为囊中羞涩的人量身定做的便宜货。如果说LV的店门上挂一个横幅,说‘穷人与狗不得入内’,那肯定会被砸了——这是非常严重的偏见甚至歧视。但是在互联网上,数据和算法带来的偏见却很隐蔽”^[15]。而且在人工智能时代,万物互联会创造出庞大的数据流,“这些数据流在大数据、云计算、人工智能等技术的帮助下形成为‘算法’,届时人类社会的资源配置将在很大程度上被程序化的‘算法’所取代,一切抉择都听‘算法’的,由‘算法’说了算,出现明显的‘算法’崇拜”^[16]。在算法崇拜情结下,人们将更加缺少警惕意识与反思能力,将更加倾向于服从算法或者计算机的判断^[17]。如此一来,即使是有偏见的算法也容易因为人们的信任而“避过检查”,即使是由有偏见的算法导致的不公平结果也容易因为戴上了客观与科学的算法面具而被认为是公平的,即使人们在追查算法偏见导致的不良后果时也不会轻易将线索指向算法。

总的来说,由于数据质量的客观原因或算法设计者的主观原因,算法可能产生偏见。如果国家治理依赖的是有偏见的算法,那么国家治理的公平性就难以得到保证,而且因为算法偏见的隐蔽性以及算法崇拜心理的存在,由算法偏见导致的不公平成为“隐蔽的不公平”,国家治理公平性的流失也由此变得更为隐蔽。

三、“有理由的卸责”:人工智能时代国家治理的责任难题

虽然与权力和权利相关的议题长期以来在政治理论研究以及政治实践中占据显要位置,但

是“在当代民主政治与公共行政实践中,责任越来越成为中心概念”^[18],与责任相关的议题变得越来越耀眼,恰如弗雷德里克·莫舍尔所言,“责任很可能是行政——无论是公共行政还是私人行政——的词汇表中最重要的词语”^[19]。尽管在当代民主政治与公共行政实践中充斥着“责任话语”,然而要理解责任的意蕴却并不容易。总的来说,不管是在汉语语境中,还是在英语语境中,责任概念主要有两种用法:一种是“职责意义上的责任”,即分内应做之事;一种是“过错意义上的责任”,意味着行为主体某种过错导致了不好的后果因而要面临相应的谴责、惩罚、制裁等,即要让行为主体为自身的过错“付出代价”,这背后连带着一个问责或者说责任追究的过程。通过分工明确主体职责并且在特定主体履责不力时进行问责,是一个政治体系健康运行的基础。国家治理是一项颇为烦琐复杂的工程,国家治理的展开既需要各种国家部门履行好各自的职责,也需要对国家部门出现的过错行为进行问责。问责的前提是明确责任的归属,即明确是谁的责任或者说是谁的过错。如果责任归属不清,那么问责就无从谈起,不存在无对象的问责,否则与问责配套的相关制度、规则、标准、程序设计等都会失效或陷入空转。

随着人工智能技术段位的不断提高和应用范围的不断扩大,国家治理的数字化转型和智能化转型进程也不断加快,“循数治国”和“算法治理”成为备受理论与实践双向推崇的治理样态。在此过程中,许多国家部门的行动越来越注重遵循算法的指示,“行政决策中人类形式上的参与无法掩盖算法作出实质决策的真相”^[20]。然而,算法并没有人们想象中的那么靠谱,它也可能产生不可预测的后果。一方面,算法是为解决问题而生的。随着社会问题越来越多且越来越复杂,算法设计的难度也在不断增加,过去由几百行、几千行或几万行代码组成的算法就可以解决问题,而如今想要设计出一种算法来解决某个问题可能需要写几十万行、几百万行甚至是上亿行代码,在此过程中难免会有所疏漏或忽略,从而使算法产生一些不可预测的后果。另一方面,算法毕竟是由人设计的,算法的优劣在很大程度上取

决于算法设计者的认知能力、知识水平。算法设计者的认知能力、知识水平越高,就越有可能设计出优秀的、完善的算法,反之则有可能设计出劣质的、有漏洞的算法。由于人的认知能力、知识水平都是有限的,算法设计者很难掌握与特定问题有关的所有知识,因此在算法设计过程中很有可能出现某些方面的偏差,最终使得算法后果“出乎意料”。再者,虽然针对特定问题设计出来的算法带有一定的稳定性,但是特定问题的发展态势会受各种自然和社会环境因素变化的影响,而算法设计者很难全面了解和控制环境的变化,因而设计出来的算法可能难以解决已经发生了变化的问题,“强行匹配”会使得算法的后果不可预测^[21]。换言之,由于算法本身的复杂性程度日益提高、算法设计者的认知局限以及环境的动态变化,算法可能出现失误,从而造成不可预测的后果或者说不好的、负面的后果。事实上,算法失误已经不是什么新鲜事。比如IBM公司与得克萨斯大学的MD安德森癌症中心在2013年合作开发了“Watson for Oncology”(沃森癌症机器人),旨在辅助临床医生从丰富的患者和研究数据库中有效的诊疗措施。然而,美国新闻媒体STAT News在2018年7月查阅了IBM的内部文件之后发现,沃森癌症机器人有时会给医生提出错误甚至是危险的治疗建议,如建议医生给存在严重出血症状的癌症患者使用会加重出血的药物。关键问题是,由于算法可能出现各种类型的失误,因此依据算法指示而行动的国家部门也可能出现各种类型的失误,从而造成一些影响更加广泛的、不好的或者说负面的后果。

国家治理影响广泛的特性使得国家治理更不容有失,严肃问责可以倒逼国家部门尽职履责,尽可能减少治理失误。问题在于,当治理失误是由算法失误导致时,到底应该向谁问责或者说追究谁的责任呢?虽然从理论上来说应该直接问责于相关国家部门,但是由于一些人工智能企业在数据与算法方面拥有独特的优势,国家在解决社会公共问题时经常需要人工智能企业提供“技术支持”或者说“技术辅助”,使得“政府在理论上仍然是最大的‘公共性’组织,但是实现公共任务的技术路线和工具,将更多依赖于广泛的

协同合作……随着数字革命和人工智能技术的进步,未来可能有更多政府职能将被新技术方案或新技术机构革新或替代”^[22]。因此,当由算法失误导致治理失误时,国家部门可能卸责于算法开发企业或算法设计者。由于算法可以自我学习与升级,因此即使是算法开发企业或算法设计者本身,有时都无法完全清楚地知道算法为何那般决策。也就是说,在数据输入与结果输出之间,人们明确地知道存在一个计算环节,然而人们却无法洞悉计算的逻辑与过程,这就形成了所谓的“算法黑箱”。恰如伊恩·博格斯特所言,算法就像“黑洞”,人们虽然可以清晰感受到它的影响,但是却并不能对其内部一窥究竟^[23]。“算法黑箱”的存在同样使得算法开发企业或算法设计者可以找到卸责的理由。随着卸责链条的拉长,最终导致的局面就是,无法找到合适的对象来为算法失误造成的后果“负责”,问责的大棒高高举起却难以挥下,国家治理陷入问责难题。正如有研究者所言,“即使因算法缺陷导致决策失误,或者公众对算法决策的结果不满意,行政人员往往倾向于把责任诿过算法或算法开发公司,而对算法或算法开发公司又难以司法审查,责任主体模糊往往使得算法决策陷入审查难、问责难的窘境。例如,在英国女性乳腺癌筛查漏检丑闻中,关于‘算法错误’究竟是怎么产生的,国家卫生医疗系统(NHS)、公共卫生局(PHE)以及负责软件维护的日立咨询公司三方互相踢皮球,然而最终结果很有可能是根本法定论”^[24]。强行问责是许多人想到的解决办法,不过结果可能是得不偿失,因为强行问责相关国家部门容易抑制国家部门将新技术带进治理领域的动力,而强行问责算法开发企业或算法设计者又容易消解市场与国家进行技术合作的信心。

总体而言,以特定职责作为各种国家部门行动的方向指引,以问责校准校正各种国家部门的行动选择,是保证国家治理有力有序有效的重要机制支撑。不过,在国家治理走向算法化的时代,当出现治理失误时,“算法黑箱”和卸责链条的存在,使得问责对象较难锁定,本应为治理失误负责的主体很可能找到卸责理由而成功避责,出现问责空转现象,谁的责任、谁来负责或者说问

责于谁将成为一个难题。

四、企业自律+国家监管：人工智能时代国家治理难题的应对

在人工智能时代，国家治理算法化是大势所趋，由深刻的算法依赖衍生出的国家治理公平难题和责任难题在此进程中可能会反复出现。“隐蔽的不公平”和“有理由的卸责”现象在多大程度上能够得以克服或规避，决定了国家治理算法化最终能够走多远。

(一) 企业自律：应对人工智能时代国家治理难题的理想期许

在人类社会漫长的发展演进史中，各种各样的技术既造福过人类社会，也给人类社会带来过麻烦。在技术色彩越来越浓厚的当今时代，技术向善越来越成为一种共识、一种推崇乃至一种信仰。以工具属性为核心属性的技术本身并无善恶，技术向善还是向恶关键在于使用主体的选择。算法从本质上来说不过是一把“技术菜刀”，是无关好坏选择的价值中立的、中性的工具^[25]。人工智能时代的国家治理深刻依赖于算法，表面上来看，国家治理场景中的算法向善还是向恶主要取决于国家的选择，实际上，虽然一部分数据掌握在国家的相关机构手中，但是更多的、每时每刻都在更新的数据掌握在少数网络科技巨头的手中^[26]。由于“现有人类社会大数据的生产主要来自各类应用平台，人们的工作和生活、学习、娱乐等消费记录都通过各类应用平台传递到云端，而这些应用平台的研发者和运行者属于少数互联网企业”^[27]，而且只有少数大型科技公司、互联网企业具备足够的“动机”和“实力”将足够的资源投入算法设计，许多优秀的算法最终掌握在少数大型科技公司和互联网企业(人工智能企业)手中。因此，人工智能时代国家治理的算法依赖在某种意义上是国家对人工智能企业的技术依赖，国家治理场景中的算法向善还是向恶在很大程度上取决于人工智能企业的选择。

如果人工智能企业在算法设计环节对数据的缺陷视而不见、嵌入偏见或者存在谋求私利的倾向，那么依赖于该算法的特定国家治理行动就

必然难以达致公平的结果。如果人工智能企业意识到“卸责链条”的存在而不以追求尽善尽美的态度对待算法设计，那么依赖于不严谨算法或者说有漏洞算法而展开的特定国家治理行动就很难达到预期目标，甚至有可能造成不可挽回的损失。这也表明，如果人工智能企业能够以技术向善“自律”，将技术向善内化为一种企业精神，将技术向善视为一种道德律令般的力量，在设计辅助国家治理的算法时尽可能对所用数据进行细致的检查，尽可能避免掺杂私利与偏见，并且尽可能以负责任的态度增加算法设计的严谨性和道德性，那么实现国家治理场景中的算法向善就要容易得多，或者说国家治理出现偏差的概率就要低得多。其实，不少国家已经意识到人工智能时代企业自律的重要性。比如，中国国家新一代人工智能治理专业委员会2021年9月25日发布的《新一代人工智能伦理规范》就强调相关主体在研发环节要强化自律意识，加强人工智能研发相关活动的自我约束，主动将人工智能伦理道德融入技术研发各环节，自觉开展自我审查，加强自我管理，不从事违背伦理道德的人工智能研发。毫无疑问，当国家治理深刻依赖于人工智能企业的算法设计时，人工智能企业自律是实现国家治理场景中算法向善或防止国家治理出现偏差的基础，是最为节约成本的方式，也是最为理想的途径。

(二) 国家监管：应对人工智能时代国家治理难题的现实选择

在人工智能时代，期待人工智能企业自律以实现国家治理场景中的算法向善或避免国家治理偏差并不稳当，正如曾经的图灵奖得主、加拿大蒙特利尔大学计算机科学教授Yoshua Bengio所言，“自我监管没有用。你认为自愿交税有效吗？没有。遵守道德准则的公司相对于不遵守道德准则的公司会处于不利地位”^[28]。或许可以这么形象地打比方，如果将国家治理场景中的算法向善视为一只飞鸟，人工智能企业自律仅仅是其中一只翅膀，依靠这一只翅膀，国家治理场景中算法向善这只飞鸟难以飞起来。那么，什么才是支持算法向善这只飞鸟起飞的另一只翅膀呢？答案是国家监管。而推进国家监管，制度设计是

关键。

其一，建立硬覆盖算法备案说明制度。凡是需要依靠人工智能企业提供算法支持的国家治理事务，相关人工智能企业都必须向国家提供算法设计理念与思路说明以及相关代码内容，以备检查。硬覆盖算法备案说明制度可以对试图主动嵌入偏见的算法设计行为形成震慑，也可以为日后算法失误的责任认定提供线索与证据，让人为失误与不可控制的技术失误得以辨识，促使算法设计者以更负责任的态度开展工作。这在一定程度上将原本可以不断卸脱的责任巧妙地前移至算法设计环节。或许有人还担心，人工智能企业会以算法是商业秘密为理由拒绝公开算法设计思路或提供相关代码内容。此时就需要国家出场说明，任何意欲与国家合作或者说为国家治理提供算法支持的人工智能企业都需要以公开算法设计思路或提供相关代码内容作为最基本的资格条件，否则即无最基本的合作资格。毕竟算法一旦参与到国家治理事务中，就涉及千千万万人的福祉，在某种意义上就具有公共性，“商业秘密对于效率和创新有一定的激励作用，但政府的合法性和公共利益拥有压倒性考量”^[29]。实际上，算法备案说明制已经在一些领域得到践行，比如中国人民银行、银保监会、证监会与外汇局2018年3月联合发布的《关于规范金融机构资产管理业务的指导意见》就明确指出，金融机构运用人工智能技术开展资产管理业务时应当向金融监督管理部门报备人工智能模型的主要参数。当然，人工智能企业就特定算法设计理念与思路向国家作出说明时需要详细到何种程度以及向国家提供代码内容时需要完整到何种程度，应由国家视具体的治理场景与事务相机而定。

其二，建立不定时算法抽检审查制度。针对为国家治理提供技术辅助的各种人工智能企业以及各种旨在解决特定问题的算法，国家需要不定时进行算法抽检审查，防止人工智能企业“备用分离”，即备案时是一套算法，实际应用时又是另外一套算法，既防止人工智能企业单方面擅自调整算法，也防止算法在自我学习过程中“变坏”而无人知晓。不定时、突击式的算法抽检审查可以让国家更有可能看到处于真实运行状态

中的算法或者说看到算法的真面目，可以压缩人工智能企业机会主义式进行算法设计的空间，使有偏见、有缺陷的算法无法长存，将“变坏”的算法扼杀于摇篮或萌芽之际，从而有助于避免后续的责任推诿。由于国家相关组织与部门在进行抽检审查时可能成为“技术门外汉”，而人工智能企业则拥有更加专业的知识，因此可以考虑从多个人工智能企业中“借调”技术人才成立专门的行业性算法抽检审查机构，以此作为国家推进算法抽检审查的组织基础。行业性的算法抽检审查机构主要通过“过滤机制”发挥监督制约作用，即由算法抽检审查机构对为国家治理提供算法支持的人工智能企业进行算法抽检审查，“过滤”掉有问题的算法。这种方式可以对人工智能企业起到一种威慑作用，最起码可以威慑它们不敢刻意设计存有问题的算法。或许有人存在这样的担心，即有些人工智能企业根本就不在乎自己所设计的算法能否通过抽检审查，那么“过滤机制”就无法起到威慑作用。确实，这种情况是可能出现的，解决的办法就是国家建立“红黑榜”名单，通过抽检审查的人工智能企业暂时进入“红榜”，未通过的进入“黑榜”，对于进入“黑榜”的人工智能企业，国家要压缩其以后为国家治理提供技术支持的空间和机会，让其付出“高昂的代价”。相反，对于进入“红榜”的人工智能企业，国家可以考虑增加与其合作的范围和频率。实际上，算法抽检审查机构、“过滤机制”和“红黑榜”的存在还可以间接发挥促进人工智能企业之间正向竞争的作用，而人工智能企业之间的正向竞争在某种意义上也是一种互相监督制约。因为，如果有人工智能企业想获得为国家治理提供算法支持的空间和机会，所提供的算法设计要经得起检验，就需要尽可能地推出更优秀、更具公平性、更负责任的算法。哪一家企业更愿意对算法结果负责，更愿意增加算法的公开透明度，更愿意解释算法的设计精神，也就更有可能获得国家的青睐。由于获得国家的青睐类似于获得一种名声认证，而好的名声对于人工智能企业自身发展的价值是不言而喻的，因此人工智能企业要想在竞争中胜出，就得确保不被别人发现漏洞或偏差，就要在算法设计过程中精益求精。

其三,建立开放式算法结果观测制度。当人工智能企业将解决特定国家治理问题而设计的算法投入使用之后,国家需要持续系统地观测其输出结果,并根据输出结果决定是否需要调整、修补、优化或废置该算法。毕竟在“偏见进,偏见出”的逻辑下,算法输出结果是推断算法本身是否有问题的最直接与最有效的依据。我国发布的《新一代人工智能伦理规范》就在多处强调“强化人工智能产品与服务的质量监测和使用评估”和“及时监测人工智能系统”,由美国参议院两位议员于2019年4月提出的《算法问责法(草案)》也主张对算法进行系统的影响性评估,这两者都含有加强算法结果观测的意味。推进算法结果观测,国家可以充分借助“技术的力量”,针对特定操作算法再设计一套监督算法,及时识别、纠正或终止特定操作算法在运行过程中出现的偏差^[30]。更为重要的是,国家要充分借助“社会的力量”,形成一支开放式的灵活观测队伍。由于社会大众是治理智能平台、智能软件或智能程序的使用者,或者说社会大众是治理算法的直接受众,因此社会大众对算法到底是在为善还是在作恶最有发言权。如果社会大众对算法作恶睁一只眼闭一只眼或者说熟视无睹,那么算法设计者就会更加有恃无恐,算法向善就少一点阻力;而如果社会大众对算法作恶深恶痛绝,那么算法设计者就会有所顾忌,算法向善就多一点希望。实际上,由社会大众发现算法作恶并最终促使相关责任方承诺整改或者导致相关责任方遭受国家力量惩戒的例子颇多。这表明,通过建立开放式算法结果观测制度,让广大社会民众愿意及时向国家或行业机构反馈算法作恶的信息,而不是盲目相信一切算法和算法的一切判断,国家或行业机构就能依据更迅捷以及更清晰的信号采取行动。这对试图误用和滥用算法的人工智能企业就会形成巨大的威慑,而对算法结果的重视也有利于抑制人工智能企业在算法设计环节嵌入偏见或降低要求的侥幸心理。

注释:

① 需要说明的是,除了算法偏见和“算法黑箱”可能使依

赖算法的国家治理面临一些难题,其他算法问题也可能引发特定的国家治理难题。由于算法偏见和“算法黑箱”更为典型与常见,因此本文主要聚焦于此。

参考文献:

- [1] 陈鹏. 算法的权力: 应用与规制[J]. 浙江社会科学, 2019(4): 52-58, 157.
- [2] 刘俊祥. 智能化时代中国政治学科学学术话语的创新[J]. 学术界, 2020(3): 56-65.
- [3] DANAHER J, HOGAN M J, NOONE C, etc. Algorithmic governance: Developing a research agenda through the power of collective intelligence[J]. Big Data & Society, 2017, 4(2): 1-21.
- [4] BALKIN J M. Free speech in the algorithmic society: Big data, private governance and new school speech regulation[J]. UC Davis Law Review, 2018, 51(3): 1149-1210.
- [5] VINCENT J D C, HOUSE-PETERS L, CRAMPTON J W, etc. The social life of Robots: The politics of algorithms, governance and sovereignty[J]. Antipode, 2020, 52(3): 605-618.
- [6] 李秦梓, 张春飞, 姜涵, 等. 新技术新监管背景下的算法治理研究[J]. 信息通信技术与政策, 2019(4): 30-34.
- [7] 徐格, 李沁. 算法统治世界: 智能经济的隐形秩序[M]. 北京:清华大学出版社, 2017.
- [8] 钱坤. 从“治理信息”到“信息治理”: 国家治理的信息逻辑[J]. 情报理论与实践, 2020, 43(7): 48-53.
- [9] 孟天广, 赵娟. 大数据时代网络搜索行为与公共关注度: 基于2011—2017年百度指数的动态分析[J]. 学海, 2019(3): 41-48.
- [10] 陈潭. 大数据时代的国家治理[M]. 北京: 中国社会科学出版社, 2015: 206.
- [11] 张玉宏, 秦志光, 肖乐. 大数据算法的歧视本质[J]. 自然辩证法研究, 2017, 33(5): 81-86.
- [12] 杨成越, 罗先觉. 算法歧视的综合治理初探[J]. 科学与社会, 2018, 8(4): 1-12, 64.
- [13] 佚名. 人工智能还是人工智障? ——大型算法翻车现场[EB/OL]. (2020-01-17)[2022-06-13]. https://www.sohu.com/a/367510273_99904820.
- [14] 佚名. 算法偏见: 看不见的“裁决者”[EB/OL]. (2020-04-08)[2022-06-13]. <https://zhuanlan.zhihu.com/p/127040870>
- [15] 周涛. 数据时代的伦理困境[EB/OL]. (2017-06-21)[2022-06-13]. https://www.sohu.com/a/150798308_300763.

- [16] 姜太碧. 万物互联时代的“算法”崇拜与资源配置变革[J]. 民族学刊, 2018, 9(6): 72-79, 121-122.
- [17] NICHOLAS C. The glass cage: Where automation is taking us[M]. London: Bodley Head, 2015.
- [18] 李军鹏. 责任政府与政府问责制[M]. 北京: 人民出版社, 2009: 2.
- [19] MOSHER F C. Democracy and the public service[M]. New York: Oxford University Press, 1968.
- [20] 张凌寒. 算法权力的兴起、异化及法律规制[J]. 法商研究, 2019, 36(4): 63-75.
- [21] 郭林生. 论算法伦理[J]. 华中科技大学学报(社会科学版), 2018, 32(2): 40-45.
- [22] 樊鹏. 利维坦遭遇独角兽: 新技术的政治影响[J]. 文化纵横, 2018(4): 134-141.
- [23] IAN B. The new aesthetic needs to get weirder[EB/OL]. (2012-04-13)[2022-06-13]. <http://www.theatlantic.com/Technology/Archive/2012/04/the-new-Aesthetic-needs-to-get-Weirder/255838>.
- [24] 汝绪华. 算法政治: 风险、发生逻辑与治理[J]. 厦门大学学报(哲学社会科学版), 2018(6): 27-38.
- [25] HAUER T. Society and the second age of machines: Algorithms versus ethics[J]. Society, 2018, 55(2): 100-106.
- [26] 王绍光. 新技术革命与国家理论[J]. 中央社会主义学院学报, 2019(5): 93-100.
- [27] 陈鹏. 智能治理时代的政府: 风险防范和能力提升[J]. 宁夏社会科学, 2019(1): 74-78.
- [28] 佚名. AI 与机器人的 42 个大问题[EB/OL]. (2019-10-18)[2022-06-13]. https://www.sohu.com/a/345581103_100048981.
- [29] 汪庆华. 算法透明的多重维度和算法问责[J]. 比较法研究, 2020(6): 163-173.
- [30] ETZIONI A, ETZIONI O. Designing AI systems that obey our laws and values[J]. Communications of the ACM, 2016, 59(9): 29-31.

Algorithm dependence and its derivative problems of state governance in the era of artificial intelligence

MEI Lirun

(School of Marxism, Wuhan University, Wuhan 430072, China)

Abstract: The advent of the era of artificial intelligence not only enriches the choice of technology tools for state governance, but also promotes the profound algorithm dependence of state governance. The existence and concealment of algorithm bias will cause the problem of fairness in state governance. The existence of algorithm black box makes all different subjects able to find the reasons to discharge the responsibility, which will cause the problem of state governance responsibility. Artificial intelligence enterprises have unique technological advantages. Therefore, in the era of artificial intelligence, the algorithm dependence of state governance is the technology dependence of the state on AI enterprises. To deal with the state governance problems caused by algorithm dependence requires both the self-discipline of AI enterprises and the strengthening of state supervision. The establishment of hard coverage algorithm filing and explanation system, irregular algorithm sampling inspection and review system, and open algorithm result observation system are the key action measures to promote state supervision.

Key Words: artificial intelligence; state governance; algorithm dependence; state supervision; technology for the good

[编辑: 苏慧]