算法透明的困境及规制路径研究

崔冬, 夏玉配
(东北林业大学 文法学院,黑龙江 哈尔滨 150040)

[摘　要]算法决策技术的兴起推动了社会发展,同时也带来了算法黑箱、算法合谋、算法操纵等技术和社会风险,必须对算法进行规制以控制风险。算法透明是进行算法规制的前提和基础,由于技术和制度上的障碍,完全彻底的算法透明不现实也不合理,在具体场景中追求有限、适当的算法透明才是有意义的。在算法透明的落实和保障上,行政规制具有独特的优势,通过构建刚柔并济的行政规制体系、促进企业算法合规监管、推进算法备案制度、追究算法披露不当行为责任四项具体行政规制措施,监督和保障算法透明制度的运行,实现科技创新与社会安全之间的平衡与协调。

[关键词]算法决策;算法透明;平衡协调;行政规制
doi:10.3969/j.issn.1673-9477.2024.01.011
[中图分类号]D922.16　[文献标识码]A　[文章编号]1673-9477(2024)01-0073-09

随着计算机和大数据技术的发展,算法可对海量数据进行处理,分析自然人的个人特征,并广泛应用于贷款、广告、保险等领域。算法带来了信息化科技革命,同时也引发了算法黑箱、算法合谋、算法操纵等技术和社会风险。在面对魔法一样的算法所引发的问题,人们首先想做的是弄明白它是什么,所以算法透明也是人们所熟知的一项规制措施。算法透明是一个完整的过程,要公开算法代码、数据、决策树等信息,同时还要以通俗易懂的图形、文字等方式解释算法决策是如何做出的。

在理论界,国内众多学者对算法是否要透明和透明的尺度意见不一。反对算法透明的学者认为算法透明作用微乎其微,而支持算法透明度的学者,对算法透明到什么样的程度以实现各方利益的平衡未达成统一认识。在实务中,《电子商务法》和《个人信息保护法》强调了算法透明的要求和决策结果的披露义务。《关于加强互联网信息服务算法综合治理的指导意见》与《网络数据安全管理条例(征求意见稿)》等要求构建算法决策披露制度,以促进算法透明。《互联网信息服务算法推荐管理规定》主要规定算法推荐服务提供者的透明义务和责任。

以上法律法规对算法是否透明作出了回应,但是这些规定有的过于笼统、抽象,没有回应关键问题,有的具体实施起来相互矛盾,给实践带来了很大困扰。本文将探讨算法透明的价值,阐明算法透明的技术与法律困境,在困境中寻求解决之道,也即在场景化中把握算法透明尺度,平衡各方利益,减小算法透明阻力。同时通过行政规制机关的相关规制手段实现对算法透明的监督与问责,保障我国算法透明有限、适当,算法结果公平、公正,以期为我国算法透明立法与实践提供参考。

一、算法透明的价值

(一)算法透明是算法规制的基础

有学者认为即便算法透明可知,算法决策结果也不一定是公平的,比如电子酒精测试仪的算法,在操作过程中,工作人员可能因性别歧视、受贿等原因故意操控探头,探头也可能因破损失灵。算法运行的任何一个环节出现失误,即便是透明的算法也可能出现不公平的裁决。[1]此种观点有一定的道理,算法透明确实难以成为化解算法风险的唯一力量。算法透明能揭示"算法歧视"成因,但无法直接消除歧视;算法透明能显示"算法权力"运作的过程,但对已经形成的权力格局无力打破。然而对算法决策规制的基础是识别算法风险。算法透明应当作为其他规制工具的前提而存在,它客观地显示出算法决策的整个过程,为立法机关创设算法相关制度、执法机关执法提供素材和可行性依据,其他规制手段是在了解算法基础上的进一步工作。

在创设算法相关制度中,立法机关不仅要思考算法风险的化解,还需关注算法控制者的义务负担、知识产权的保护等。算法控制者义务的强度、知识

[投稿日期]2023-11-30
[基金项目]司法部法治建设与法学理论研究部级科研项目(编号 22SFB5010)
[作者简介]崔冬(1978-),男,山东禹城人,博士,副教授,研究方向:行政法学。

产权的保护力度都与算法公开的范围有关。在执法机关进行算法问责时,算法透明是发现算法问题的基础。欧盟议会发布的《算法问责与透明的治理框架》指出算法透明和算法问责同等重要,并且不了解一个系统时,就不能追究它的责任,也无法有效规制。

(二)算法透明是正当程序的要求

算法已经成为可以控制和干涉他人的权力,对于权力必须要进行限制以防止滥用,所以算法必须受到正当程序的限制,算法权力的行使要符合正当程序的要求。[2]现在的问题是正当程序一定要透明吗?公众的关注是对权力的有效制约,正当程序可以区分恣意与法治。黑箱之中容易隐藏不公,阳光之下一切黑暗和罪恶都无处藏匿。在程序正义上,隐秘不透明就是不正义的。

有学者对于目前的算法解释技术没有信心,并且难以平衡算法透明与商业秘密、个人隐私等利益冲突,因此无视算法解释技术的发展,放弃利益权衡,转而主张利用其他技术,如"零知识证明"、无偏私和一致性测试、"中间层",它们可以在不打开算法黑箱的情况下监督和控制算法权力。一些学者主张这些手段符合程序正义的要求,认为正当程序不要求算法透明,正当程序可以在输入和输出端介入,并重视输出端的结果,以结果反推算法决策是否正当。这种结果主义、功利主义的正当程序可以说是一种伪程序正义,它只关注了预防性和结果性的层面,没有解释信息处理的基本逻辑、目的,也未阐明对利害关系方的可能影响,回避了基于过程的程序性要求。纯粹的程序正义没有判断正当结果的标准,它是一种公平的程序,只要恰当地遵守该程序,得到的结果也是公平的。这是由于人们很难拒绝自己参与正义程序所产生的结果,虽然这个结果本身是开放性的。一种适当的程序安排,它的意义在于考量作出决策的过程,而不在于追求某种确定的结果。如果仅仅通过结果来判断程序是否正义,正当程序就有了结果功利主义的倾向,因此算法决策不能回避过程性的正当程序。算法决策牵涉公民诸多基本权利和义务,没有过程性正当程序的规制,算法将会暴露其恶的本性——恣意、偏见、专断。如果人们对算法决策的过程一无所知,算法将无法为人类提供一个有稳定预期的世界[3]。

(三)算法透明是建立算法理性信任的关键

有学者认为算法黑箱与大脑黑箱相似,算法黑箱并非是一个负面评价,人类可以容忍大脑黑箱,也一样可以允许算法黑箱的存在,因此没有必要打破算法黑箱,坚持算法透明。虽然人类允许决策者聆听内心的声音,但并不意味着允许算法模型进行自由决策。毕竟人脑决策的"黑箱"是随机的而非系统的,并且人类具有一定的道德感和反思能力,能够对错误的决定纠错,而算法没有这样的特性。

在不了解算法的情况下,对算法惊艳世人的预测和决策能力盲目崇拜、充满信任,允许黑箱的存在,让机器自由地决定人类的基本权利和义务,人类就相当于放弃了决定自身命运的主权。我们必须警惕人类主宰自己命运法治精神的衰落,以及随之而来的人类主体地位的弱化。算法的不透明产生了不可控性,算法可能会侵害人类的权利与自由,人类也会被别有用心的算法控制者摆布、丧失尊严,这是一种伪信任。面对陌生而神秘的算法,对算法技术的恐惧心理不可取,盲目信任算法更要不得,对算法的理性信任必须建立在充分了解算法的基础上[4],对算法扬长避短,让算法造福人类。算法透明有助于人类了解算法的运行逻辑,消除技术恐惧心理,同时避免迷信盲从,达到对算法的理性信任。

虽然算法透明所面临的困境会使人们对它的必要性产生怀疑,但上述重要价值足以认定算法透明是必要的,反对者的声音会使算法透明的支持者审慎思考算法透明遇到的障碍,仔细考量算法透明的尺度。

二、算法透明的困境

(一)技术困境:算法透明的技术难题

1. 原始代码在不断变化且难以理解

算法工程师不断更改算法设计思路和方案,很难按部就班地记录设计过程。经过历任程序员团队的设计与更新,原始代码在不断变化且逻辑混乱不清。因此,公开的静态代码并不一定用在特定决策中,并且静态代码在无拘束情况下的运行可能不符合算法控制者设定的目标。更棘手的是,只有专业人员才能编写和阅读代码,普通公众或者算法使用者不能正确理解公开的静态代码。

2. 准确度高的算法难以具体解释

在人工智能时代,算法的准确性和算法的透明度呈负相关,人工智能系统越复杂就越难以理解,透明度也会越低,但准确性度会越高。为了提高算法的性能和竞争优势,企业往往要加强算法的复杂性,较少为了优先考虑算法的可理解性而去牺牲算法准确度。对于稍微复杂的算法如神经网络、深度学习

等算法模型,大多不能被化约为参数权重或者公式[5],靠语义解释和逻辑推理的人类智能活动无法匹配具有高维度特征的数学优化方法,所以很难用普通的语义对算法进行说明。即使深度学习算法的部分模型完全透明,对这部分模型的实际作用,专家也得不出一致的结论。

3. 输入与输出之间的因果关系难以破解

算法决策规则是从被输入的数据中不断变化形成的,除了少量被严格控制的算法外,通过考察原始数据或者代码来推算算法的输出结果几乎不可能。也即从表面上看,算法输入与输出结果的因果关系非常直观,但是根本无法解释这种因果关系。而且这种因果关系是由输入的数据决定的,这是一种统计意义上的因果关系,和规范意义上的因果关系大不相同。更重要的是,对于深度学习算法,随着输入数据的变化和积累,算法输出结果的深层原因更难把握。

(二)商业秘密保护困境:算法透明的法律障碍

算法透明面临着各项利益冲突,其中与商业秘密的冲突尤为明显。TRIPS 协议第 39 条规定了商业秘密的三个要素:秘密性、保密性、实用性。算法符合商业秘密的条件,因此算法普遍获得了商业秘密的法律保护模式。一个具有准确预测性和较高工作效率的算法,必然要历经艰辛且漫长的研发,在投入市场后,有助于提高企业的市场竞争力和商业效率。将算法进行商业秘密保护是合理且必要的,因为其符合劳动价值论的思想。[6]我国法律规定了大量有关商业秘密保护的条款,算法控制者可根据相关规定拒绝披露算法有关信息,并对算法采取合理的保密措施,禁止他人未经授权披露、获取、使用算法相关信息。因此保护企业商业秘密的目标与增强算法透明度及可解释性之间就存在了巨大张力。如果仅简单认为商业秘密保护的是私人利益,算法透明制度保护的是公共利益,那么商业秘密保护就应该让位于算法透明制度。但在实践中,无论是立法层面还是司法层面,商业秘密都占据了上风,以所谓的公共利益抗衡商业秘密,早有众多败绩。

2016 年,美国纽约市提出了一项立法建议,强制要求使用算法决策的政府机关披露该算法的源代码,然而此项立法建议受到了企业的强烈抵制,企业认为他们的核心商业秘密会遭到泄露,最终该项激进的立法主张只成为了一项政策倡导,没有付诸实践。在卢米斯案(State v. Loomis)中①,卢米斯想要了解为何 Northointe 公司的自动算法 COMPAS 会对其作出高风险评价,故而向法院提出公开算法的诉求,法院因商业秘密保护拒绝了卢米斯的诉求。在维亚康姆诉“油管”(Viacom v. YouTube)案②中,平台不公开算法的做法得到了法院的支持,理由同样是商业秘密。立法机关和司法机关为何会作出如此选择?事实上,虽然算法透明身后蕴含着平等自由、公平正义的价值,但是商业秘密背后也隐含着创新发展、社会进步的深层价值,两者都承载了公共利益,无法厚此薄彼。在卢米斯案中,虽然法院维护了科技创新价值,没有要求公开算法的基本逻辑,但算法的公平正义问题没有得到解决。因此该案的判决一度在美国引发了人们对辅助量刑软件使用合法性的争议,尤其是对黑人的量刑歧视。

三、算法透明困境的化解

算法透明面临着技术障碍及与商业秘密保护之间的冲突,须尽力突破技术障碍、调和矛盾,寻求各方之间的平衡与协调。

(一)算法透明技术困境的消解

尽管目前算法透明存在较大的技术困境,但这不应该成为拒绝算法透明的借口。那种认为人类根本不可能破解深度学习算法的运行逻辑及参数权重等相关信息的看法太过悲观。世界上只有尚未认识之物,未有不可知之物。加利福尼亚大学伯克利分校和马克斯普朗克信息学研究所提出了一种可以进行自我解释的算法系统。这种系统能够分析作用于算法决策的数据,说明使用这种算法决策规则的原

①在该案中,COMPAS 是一款用于判例管理和判决辅助的算法系统,由一家私人企业 Northpointe 研发,被美国多个地区的法院采用,COMPAS 出具的评估报告显示刑事被告人卢米斯再次犯罪的风险较高,法庭以此为参考给卢米斯判了两项最高刑罚。对此,卢米斯上诉称,法院参考私人开发的且未充分公开其计算过程(如算法源代码、不同因素在风险计算中的权重等)的风险评估算法对自己进行量刑,侵犯了其正当程序权利。然而法院最终以商业秘密为由驳回了卢米斯的诉求。

②在该案中,Viacom 是一家主营电影、电视剧、音乐唱片的传媒集团,拥有大量影视作品版权。YouTube 是 Google 名下的一个著名视频分享平台。2007 年,Viacom 起诉 YouTube 侵犯其版权,理由是 YouTube 平台上存在有大量的未经其授权的音视频内容,且 YouTube 知晓这些侵权内容,并在信息搜索排名、内容推送等算法设计方面促进这些侵权内容的传播。为举证证明上述主张,原告 Viacom 向法院申请要求被告 Google 公开谷歌搜索程序和 YouTube 平台上“Video ID”程序的源代码。对此,被告 Google 辩称源代码为受法律保护的商业秘密,拒绝提供。最终,法官驳回了原告的诉讼请求。

因,这将有利于帮助人类理解算法的决策过程。谷歌网络团队最近的研究成果表明他们实现了神经网络算法的可视化,虽然这项成果没有涉及算法决策程序。算法的可视化还有许多未完成的工作,但已经是算法可解释性的一大进步。我国互联网平台也在积极研发模型,优化算法设计、进行数据筛选,尽力解决包括算法黑箱在内的诸多算法风险。为解决算法的可追溯性和透明度,百度和阿里公司研究出了新的算法可视化工具和解释方法,2018年字节跳动公司向公众公布了有关算法推荐原理的信息。综上可知:目前的算法解释技术决定着算法透明的最大边界[7],人类所无法理解的算法机理终将随着算法解释技术的发展慢慢消解。

(二)算法透明与商业秘密冲突的调和

一方面,有的算法获得了商业秘密的形式,如一些"黑箱"算法①、"感知"算法②和"奇异"算法③。在商业秘密和算法透明之间,我们就很难找出一个普适的理由,来使一方凌驾于另一方之上,非黑即白的思路并不可取,双方必须做出适度让步。在两项冲突的利益面前,为避免矛盾激化,可借鉴比例原则的思想。比例原则的本质在于禁止过度,即为保护公众的知情权,可要求适当的算法透明,并将对商业秘密的不良影响控制在最小范围内,同时以缓和的方法实现算法透明的目的。另一方面,算法商业秘密不是绝对的,并非所有的算法都获得了商业秘密的形式。对算法施以商业秘密的法律保护方式为算法控制者获得了极大的竞争优势,但全部采用商业秘密的保护方式在某种程度上是对社会累进创新的抑制。社会上算法的累进创新必须以已有的算法技术为基础进行再创新,但对算法技术进行秘密保护将难以实现社会后续的累进创新。通过专利模式对算法进行保护将有效避免商业秘密的弊端,同时也利于激发社会创新活力。"白箱"算法④和"灰箱"算法⑤以其简单、易于解释和分辨的特性在客观上符合了专利的法定要求,规制机构可以鼓励此种算法的专利模式保护[8],这也将有利于调和算法透明与商业秘密之间的冲突。

(三)有限、适当的算法透明

每一次科技与法律的互动都会引起技术可行性及利益平衡的争论。算法透明并不总是适合通过披露源代码的方式实现,在任何场景都要求披露源代码或公开全部技术细节,不仅在技术上存在困难,也会抑制企业的创新活力。充分彻底的算法透明因技术和制度上的障碍而不具备实施的可行性和现实的合理性。有限、适当的算法透明更有意义,通过合理把握算法透明的尺度,兼顾公众知情权和市场主体技术创新积极性,为算法公平提供依据,为算法的监管和应用提供良好的背景。

何为有限、适当的算法透明?随着算法的迅速迭代和社会生活的广泛渗透,抽象地谈算法应该适当透明和可理解是笼统和不易实施的,不同的主体在不同的场景下对算法模型有不同的要求和认知程度。只有在具体的场景中讨论算法透明尺度,才有可能比照目前的算法解释技术讨论是否可行、是否有泄露商业秘密的巨大风险,以设置相应合理的算法透明方案和目标。本文认为有限、适当的算法透明应当根据不同主体、不同场景采取不同程度的算法透明。

1. 区分不同的算法控制者

对不同算法控制者施予不同的算法公开责任。对公权力机关所使用的算法应当最大限度地公开。公权力机关的算法决策运行体系对公民的基本权利和义务产生着重大影响,加上权力在黑暗之中易滋生腐败,因此其本身必须公开透明。考虑到彻底公开算法会损害算法企业的商业秘密,公权力机关可以考虑使用获得专利保护的算法进行算法决策。对纯商业性且在市场上占有较大份额的企业,使用算法可能会带来歧视、不公影响,这种不良影响的范围面较广,根据企业的不同经营特点,可在不损害商业秘密的情况下要求其披露更多的算法决策信息。对于纯商业性且没有垄断地位的小微企业,对算法公开仅作政策性提倡,原因在于影响面不大,依靠市场调节就可以促进这类企业改进算法,例如消费者更愿意选择评分机制合理且公开透明的网站,此时该网站就会有动力优化自己的算法。

2. 区别风险高度敏感的场景和一般场景

对不同的场景提出不同的算法透明度要求。目前的人工智能技术无法既提高算法模型的预测性,又全面记录算法决策的过程。如深度学习、随机森林、支援向量机等算法模型的预测性较高,算法透明

①"黑箱"算法:算法呈现出突发性,很难甚或不可能预测、解释其特征。

②"感知"算法:算法能够通过图灵测试,已经达到或超过人类智商。

③"奇异"算法:算法能够实现自我递归完善,已经具备"奇异"功能。

④"白箱"算法:算法完全是确定的,即为预先确定的指令集。

⑤"灰箱"算法:算法是不确定的,但易于预测和解释。

度会相对较低,而透明度较高的信念网络或决策树却有着较低的预测性,“预测性”和“透明度”的两难,何尝不是科技与法律的两难。在对公民权利和义务影响不大的算法决策场景中,对算法透明度要求相对不高。但是在一些风险高度敏感的场景中,比如司法、健康医疗等对公共利益和个人利益有重大影响的场景下,对算法透明度提出了高要求,如果不能在原理层面披露算法,必然被人诟病。因此在高风险场景中甚至可以强制要求模型性能低但透明度高的算法。

3. 区别一般用户与监管主体

对不同的主体公开不同的算法信息。按照一般用户的认知取向和理解能力,他们不关心“算法具体的技术细节和复杂的数学解释,他们想要获得的是算法决策的潜在风险、具体用途、基本逻辑、大致权重、主要参数等关键基本信息。因此,针对于一般公众的算法透明要求不在于披露算法的内在逻辑要素,而在于算法控制者以通俗易懂的方式履行算法说明义务,向公众提供算法的一般信息。[9] 而拥有一定技术基础的监管主体会更多关注算法的部分内在细节,寻求“算法为何正当”的相关要素,包括但不限于算法公式,具体参数权重,此时算法控制者算法透明的尺度要达到证明该算法正当合法的地步。

四、算法透明的范式选择

(一)私法规范与救济之局限

1. 引入算法解释权

算法相对人难以理解算法的基本原理,与算法控制者存在信息上的不对称。算法相对人这种相对弱势的地位,导致自身权益极易遭到损害且不自知。为解决算法黑箱带来的负面效应,避免算法权力的滥用,矫正信息不对称的地位,加强算法决策的可责性,专家学者对理论和制度构建进行了详细的探索。在众多方案中,算法解释权是最具代表性的。欧盟《通用数据保护条例》在正文中未明确规定算法解释权,但在序言第 71 条中提及算法解释权,学者结合序言和反自动化决策权的相关规定,对算法解释权进行了归纳,即当算法决策的某项具体决定对算法相对人有法律上或经济上的显著影响时,相对人有权对算法控制者提出异议,要求对前述具体决定进行解释,并有要求更正错误的权利。我国《个人信息保护法》也有与之类似的规定。

算法解释权看似拥有严密的运行逻辑,但作为国外确立的笼统权利,其权利内容、范围、边界都不够具体清晰,处于仅立法而没有运用的尴尬状态。科技的发展总是会带来一些利益冲突,虽然科技有变,但法理有常。一旦遇到利益冲突,不先考虑运用现有法律体系去解决复杂的利益博弈问题,而是径直引入新的法律权利——算法解释权,这种事无巨细的立法主义模式实质上是一种智识上的懒惰。[10] 随便将某些权益命名为权利,会导致现有权利体系的混乱,更会因忽视该权利的具体应用场景和其他应当保护的权益,最终陷入立而不用的尴尬局面。此外,一项域外权利的引入需要充分考虑国内具体情况,审慎考量我国公民和该产业是否特别需要这项新兴权利,以及引入该项权利能否与国内的市场制度、权利框架、法律规范等深度融合,是否与现有权利冲突,经得起实践检验。

2. 现有私法规范的救济

是否赋予算法相对人以算法解释权有待深入论证,但算法解释权所保护的权益可由《民法典》《消费者权益保护法》《电子商务法》《个人信息保护法》等法律规范规定的法律制度涵盖。算法相对人认为权利受到侵害时可根据消费者权利、知情权,以及个人信息处理者的说明义务等相关权利和义务要求算法控制者以通俗易懂的方式说明算法决策的相关信息。算法控制者在接到相对人的请求后,要在规定的时间内,根据算法模型的技术特点决定是否进行说明,并且对于拒绝说明的情况要答复拒绝的理由,由此产生的争议双方可以寻求法院解决。

看似顺畅的现有私法规范救济路径实则困难重重。相对人只有发现自身权利受损或者有受损风险时,才会有意愿主张知情权、消费者权等相关权利,但由于算法相对人本身已陷入被算法普遍统治的环境中,很难对算法的这种侵害有科学的理性认识。即便用户发现权利受到侵害,也一般倾向于息事宁人,不愿意介入法律诉讼。当然也会有很小比例的受害者提起申诉或者诉讼,在理论上受害人可以选择违约或者侵权之诉。违约之诉有着较重的举证责任,救济效果有限,未作为当事人的首要选择。在侵权之诉中,信息的不对称加上司法机关不擅长技术认定,在经过漫长的诉讼和司法论证后,即使受害人得到救济也相当不划算。例如美国的一场诉讼中(Dehoyos v. Allstate)①,双方经过多年的诉讼才就

①在美国的一场诉讼中,保险公司的信用评分标准在法庭上受到质疑,原因是它们对少数族裔的影响各不相同,经过数年的诉讼,双方达成了数百万美元的和解协议。原告称信用评分程序的缺陷导致约 500 万名非裔美国人和西班牙裔客户受到歧视。作为和解协议的一部分,保险公司允许原告的专家对未来的评分模型进行批评和改进。

“信用评分程序的缺陷”达成数百万美元的和解。这种类型的侵权诉讼本身也就具有极强的公益性,很容易出现“搭便车”的情况。

司法系统并不擅长在专业技术上钻研,解释和运用法律才是其优势。技术问题并不是从事实到法律的演绎,它涉及了复杂的数学模型、公式推导、数据统计等。[11]美国的法律系统并不十分满意司法机关对技术问题的应变能力,波斯纳也认为英美法院在医疗事故、药品、专利等技术问题上的回应确实不太理想。从我国的司法诉讼中来看,当涉有技术问题的案件时,诉讼效率较为缓慢,涉及专业化复杂化的技术类官司时,有的案件争议数十年也未有定论。一个比较重要的原因是法院的确不擅长技术认定,不仅当事人双方要聘请专家辅助人,法院也需要寻求外部技术人员协助审案。借助专业技术人员参与司法案件在一定程度上填补了法官在专业技术上的短缺,但这增加了诉讼成本,使诉讼效率变低。

(二)行政规制的比较优势

1. 预防和处理系统性侵权的效率优势

为达到有限、适当算法透明的目的,采取行政规制的过程性控制,还是采用基于司法裁判的事后震慑,应当考量算法黑箱可能发生损害的特点,以及进行事后威慑的效果。在违法行为可能造成系统性且无法挽救的重大损害时,利用事后追责来进行威慑救济效果并不十分显著,利用具有效率和防范风险优势的行政规制是一个行之有效的选择。行政规制系统一直处于运行状态,不需要通过特定损害的发生激活。行政规制机关可以自主研判算法风险并立刻启动风险防范程序,不必等到具体风险的出现,也不需要利害关系人的请求。更关键的是,私法诉讼对每个算法相关案件都需要做到“一事一议”,不会因判决形成一般性规则,存在一定程度的低效率。而行政规制机构可以发挥“规模优势”,对同类问题公布一般性规则,进行整体规制。

2. 专业和信息优势

随着大数据技术的迅速发展,算法透明与商业秘密保护之间的矛盾日益凸显,优化算法透明度更是涉及算法深层次的技术原理问题。规制机构不仅要拥有专业技术,还要能够掌握产业增长趋势,在错综复杂的利益对立中衡量各项规制措施的恰当性,这要求规制机构具备较高的专业能力。由于风险社会的到来,各个国家普遍设置了专业化信息化的行政规制机构,依靠专业化的技术工具进行风险鉴定和防范。这些规制机构的工作人员日常任务就是接近相关的专业技术,因为本身就具备了相关专业技术背景或者实践经验,因此可以与相关领域的技术专家密切联系。另外,行政机关能够设置专门的技术机构和专业技术岗位,并通过聘请技术专家等方式加强技术优势。

(三)行政规制的局限与公私法的调和

行政规制虽然存在诸多优势,但也并非没有短板。与私法救济相比,行政规制也会存在着浪费资源、过度干预、权力寻租等隐患[12]。在有限、适当算法透明的具体实施上,行政规制的瑕疵有两点。一是需要组织机构、人员层面的大量投入,而这些都必须由公共财政予以保障。二是无法妥当地处理棘手个案,并为受害人提供物质上的救济。上述行政规制的短板恰恰是私法救济的优势。首先,私法诉讼救济花费了小部分司法资源;其次可以根据技术变化对个案灵活处理,在侵权诉讼中给予受害人一定的抚慰。

综上所述,由于算法本身的特性,在算法透明度的落实上应该加强行政规制,但并非是要全面代替私法救济体系,而应该立足于两者的制度特性,推动两者取长补短,实现优势互补。行政规制应该聚焦于整体性、高敏感度的算法风险问题,直接推动算法适当透明;私法救济制度应该着眼于个别的、具体的消费者权益、算法知情权隐患,间接促进算法适当透明。

五、算法透明行政规制的具体路径

有限、适当的算法透明的实施,需要关注更多微观问题。因此,以下将提出四个行政规制的具体措施,并结合具体的场景落实算法透明的尺度。

(一)构建刚柔并济的行政规制体系

刚性法律规范是算法规制的重要手段,算法透明的行政规制离不开法律规范支持,否则行政规制效能会被架空,沦为一种倡议性条款。为保障算法透明的妥善实施,必须完善并细化相关法律规范,强化法治力量,明确违法违规责任,在法治的轨道上平衡各方利益,化解冲突,推动构建包括算法备案,算法合规监管、算法问责等制度在内的法律规范体系,逐渐形成算法透明行政规制长效体系,保障算法发展稳定又充满活力。

算法本身是一种技术,对技术规制最直接的手段是制定技术标准,因此技术标准也是算法透明行

政规制的直接依据。算法决策规则存在一定随机性和模糊性,因此促进算法透明以识别算法缺陷是算法规制的重要内容。含有算法透明度要求的算法技术标准能够为算法设计工程师提供明确指引,保证算法系统形成完整可靠的记录,为算法审查和验证提供依据,这样做有利于行政机关进行备案审查。行政机关可以根据差异化、灵活度原则,在标准设计时预留一定的灵活度,根据不同应用场景确定不同的底线和灵活处理空间,在确保伦理价值底线的基础上,形成具有场景差异性的标准体系。

法律规范具有强制力,直接规范人们的行为,塑造算法治理格局。技术标准是典型的柔性规范手段,具有隐形的约束力,依靠人们自愿遵守,能够为企业的探索与创新留下空间。[13]法律规范与算法技术标准缺一不可,需优势互补共同助力算法透明的行政规制。因法律规范和技术标准在效力等级和范围的巨大区别,为保证二者协同共治,需推动算法透明相关法律规范与技术标准的良性互动和有效衔接。首先,要允许多方主体参与算法技术标准的制定。其次,我国未来立法应当明确算法技术规制需求,细化算法透明的技术规范要求。最后,要注重算法技术各标准之间的衔接,还需配套相关算法技术标准评估与实施机制。通过两者的协同共治,构建刚柔并济的算法透明行政规制体系。

(二)促进企业算法合规监管

传统意义上对算法平台的穿透式监管轻视了算法的技术壁垒,并忽视了克服这一难题所需付出的巨大规制成本。这种监管方式意味着政府过多关注企业管理的细枝末节,会造成政府过度干预市场的局面,也将打破政府和企业之间的平衡。规制机关的有限理性和技术无限发展之间的矛盾由来已久,随着算法的迅速发展与广泛开发,规制机关对算法直接的、持续性的监管几乎是不可能的。20 世纪 90 年代,互联网领域开始了隐私保护执法,美国采用灵活且渐进的策略,在算法监管困境的消解上形成了一定经验。根据美国的经验,在企业内部构建起识别、防范、化解风险的合规机制,规制机关对这一合规机制进行监管,实现内部管理型控制,这是一套比较有用的监管制度。[14]

我国可以推行企业算法合规监管,激励和监督企业构建一套算法合规管理体系,实现对算法平台企业的穿透式监管。具体措施包括但不限于企业要建立独立的算法监督委员会并设计自我审查和纠错机制,根据法律法规更新企业的算法披露规则,定期和不定期地对算法运行和决策进行审查监督,并在企业内部开放员工投诉渠道,发现相关算法风险立即采取应对措施,同时还要帮助和支持规制机关的外部审查。规制机关要对网络平台算法披露情况进行实时监督,发现潜在风险或问题时,可要求算法网络平台进行内部自查或者通知平台整改算法披露制度。企业算法合规的履行促进了以算法透明为核心的事前规制路径的具体落实,在降低自身经营风险的同时,维护了公共利益。企业按照合规管理的要求公开算法相关信息时,在一定程度上衔接了算法备案制度。

(三)推进算法备案制度

算法备案是一种特殊的行政备案制度,由各级政府网信办公室统一受理算法控制者有关算法的备案,并负责审查监督,推动实现算法透明的创新制度[15]。行政规制机关可以通过算法备案了解算法的底层运行逻辑,提升规制能力,保证后续的规制活动顺利展开。另外,规制机关通过算法备案为算法相对人固定了一些证据,在事前阶段就明确了算法责任主体,为事后追责提供了保障。目前我国的算法备案制度正在逐步落实,根据《互联网信息服务算法推荐管理规定》,在互联网信息服务领域,算法备案系统已经上线。

我国算法备案制度实施处于初步阶段,尚存在诸多困难,在实施中需要注意以下几个方面。第一,避免算法备案流于形式。虽然我国法律规定了算法备案的大致内容,但其中的具体内容是由算法控制者自主提供,规制机关对算法备案的信息进行形式审查,并不能真正了解算法的重要信息,也不能形成有效的规制。规制机关需要对提供备案的信息按照相关技术标准进行实质审查。第二,实施动态算法备案。算法在不断演进发展,已经备案过的信息早已发生变化,影响规制机关对算法产品的把控。规制机关一方面要提醒算法控制者对更改的参数、程序等信息进行再次备案,另一方面规制机关需主动审查算法模型,发现改变应责令算法控制者进行再次备案。第三,降低备案难度和成本。由于算法种类、形式繁多,算法备案的工作难度较大并需要耗费大量成本,为防止算法控制者判断错误,以及备案算法范围的不当扩大,可以对算法分级分类,建立算法备案“负面清单”,影响力和风险等级都较小的算法无需备案。第四,避免泄露商业秘密。根据前述比例原则的要求,将对商业秘密的不良影响控制在最小范围内,需要为工作人员设置严格的保密义务,避

免工作人员利用职权窃取泄露经过备案的算法信息。

(四)追究算法披露不当行为责任

如果法秩序不设计义务违反的法效果,规范将如同“无牙之虎”,沦为违法者的笑柄。算法控制者按照规定披露算法相关信息存在不当的备案与解释行为、虚假备案、误导性解释、信息重大遗漏、干扰性披露等行为时,需承担相关法律责任。如果侵犯了消费者的合法权益,则算法责任主体应该承担民事责任,进行赔偿或补偿等。如果违反行政规制机关与算法披露的要求,算法责任主体应当承担相关行政责任。如果触发了互联网平台的系统风险,危害了国家利益,构成犯罪的应当承担刑事责任。

需要注意的是算法责任的治理包含了多种手段,要将惩罚性的强制性手段与激励引导手段相结合。一方面对于算法披露的故意违法行为,规制机关需加大责任力度,震慑算法控制者,强制其遵守法律法规。例如规制机关要求企业以通俗易懂的方式披露信息,但有些企业为了逃避监管,披露大量冗余信息混淆真正被需要的信息,使得调查人员不得不花费大量时间和精力在上万份材料里大海捞针,真实的算法逻辑被淹没在海量数据的迷宫里,披露与掩盖的关系可谓是道高一尺,魔高一丈,愈进愈阻,永无止息。互联网算法平台故意披露大量冗余信息的行为,违反了平台的信息披露义务,应当承担与公司法、证券法相类似的虚假陈述责任。另一方面,规制机关必须考虑到目前算法尚处于发展的初期,企业的创新积极性强,还需激励引导算法控制者积极遵守相关算法法律规范。例如算法的某些要素会随着数据的输入慢慢发生变化,算法控制者或是疏忽大意没有发现或是没有能力发现而未更新备案,经过规制机关的要求已更新算法信息,规制机关可适当免除处罚。此举既可以激励算法控制者积极备案,还可减少规制机关的规制成本。

六、结语

虽然有技术限制和商业秘密保护困境,但通过合理把握算法透明尺度及行政规制机构的介入,算法透明已成为算法决策规制中最直接有效且恰当的方式。[16]算法透明在某些情形下有实现“防患于未然”的作用,但是,我们并不能夸大其在规制中的作用,算法透明并不是终极目的,它只能是通向算法可知的一个阶梯。算法透明的实现保障了公众的知情权,但知情同意也并不能满足契约公正的实质要求。如果一方依靠其优势地位牺牲弱小一方利益换取巨大利益,弱小一方的知情同意也不能掩盖契约的不公正,因为契约公正的本质是互惠。也许一个当事人并不在意不正义契约执行所带来的伤害,但由于算法大范围的应用,所积累的负面效应犹如冰川之下的火山,随时有爆发的风险。因此,算法透明还需要结合其他规制手段共同遏制算法的恶,才能发扬算法增益福利、促进人类自由解放的善,构建公正、自由的算法社会。

参考文献

[1]沈伟伟:算法透明原则的迷思——算法规制理论的批判[J].环球法律评论,2019,41(6):20-39.
[2]陈景辉.算法的法律性质:言论、商业秘密还是正当程序?[J].比较法研究,2020(2):120-132.
[3]李晓辉.算法商业秘密与算法正义[J].比较法研究,2021(3):105-121.
[4]安晋城.算法透明层次论[J].法学研究,2023,45(2):52-66.
[5]衣俊霖.数字孪生时代的法律与问责——通过技术标准透视算法黑箱[J].东方法学,2021(4):77-92.
[6]狄晓斐.人工智能算法可专利性探析——从知识生产角度区分抽象概念与具体应用[J].知识产权,2020(6):81-96.
[7]周翔.算法可解释性:一个技术概念的规范研究价值[J].比较法研究,2023(3):188-200.
[8]孙建丽.试论算法的法律保护模式[J].电子知识产权,2019(6):39-47.
[9]吕炳斌.论个人信息处理者的算法说明义务[J].现代法学,2021,43(4):89-101.
[10]贾章范.论算法解释权不是一项法律权利——兼评《个人信息保护法(草案)》第二十五条[J].电子知识产权,2020(12):49-61.
[11]宋华琳.制度能力与司法节制——论对技术标准的司法审查[J].当代法学,2008(1):46-54.
[12]孔祥稳.论个人信息保护的行政规制路径[J].行政法学研究,2022(1):131-145.
[13]胡坚波.多措并举推进我国算法治理[J].人民论坛·学术前沿,2022(10):20-28.
[14]张惠彬,何易平.平台算法监管的困境与出路——基于美国算法监管模式的研究[J/OL].(2023-10-17)[2023-11-20].科学学研究,2023:1-25.
[15]张吉豫.论算法备案制度[J].东方法学,2023(2):86-98.
[16]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.

[责任编辑 李 新]

Research on Dilemma and Regulation Path of Algorithm Transparency

CUI Dong, XIA Yupei

(School of Humanities and Law, Northeast Forestry University, Harbin, Heilongjiang 150040, China)

Abstract: The rise of algorithmic decision-making technology has not only promoted social development, but also brought about such technical and social risks as algorithmic black box algorithmic collusion, algorithmic manipulation, etc. Therefore, algorithms must be regulated to control risk. Algorithm transparency is the premise and foundation of algorithm regulation, however, due to technical and institutional barriers, complete algorithmic transparency is unrealistic and unreasonable. In this sense, it is meaningful to pursue limited and appropriate algorithm transparency in specific scenarios. Administrative regulation has unique advantages in the implementation and guarantee of algorithm transparency. Therefore, the balance and coordination between scientific and technological innovation and social security can be achieved through four specific administrative regulatory measures. The first measure is to establish a rigid and flexible administrative regulation system. The second is to promote the compliance supervision of enterprise algorithms. The third is to promote the algorithm filing system, and the last is to investigate the responsibility of the misconducts in algorithm disclosure.

Key Words: algorithmic decision; algorithm transparency; balance and coordination; administrative regulation