《证券法苑》(2021)
第三十五卷,第227~246页

算法权力之规制:算法影响评估制度的生成及展开*

张永亮** 林盛浩*** 张洁玉****

摘要:算法权力是算法在数据处理过程中借助自我学习而产生的技术优势,以算法决策的方式对社会主体产生的影响力和控制力。异化的算法权力在私营平台领域侵犯了公民的平等权、隐私权和自由选择权,在公权力部门造成了公众自由表达、正当程序、信息公开与司法公平的缺失。算法影响评估制度是规制算法权力异化的一种有效路径,应构建内外兼容的协作治理评估体系,根据算法场景化的不同设定不同的评估标准,增设信息公开流程。

关键词:算法权力 算法影响评估 协作治理 算法问责

---

* 本文系司法部项目"人工智能的法律规制研究"(项目编号:18SFB2049)、浙江省哲学社会科学规划重点项目"监管科技应用之规制体系构建研究"(项目编号:20NDJC15Z)、浙江省属高校基本科研业务费专项资金资助"乡村治理数字化法制体系建设研究"(项目编号:2020TD002)阶段性成果。

** 浙江农林大学文法学院教授、法学博士。
*** 浙江农林大学文法学院法律硕士研究生。
**** 伦敦大学学院(University College London)研究生。

伴随人工智能、大数据分析、云储存等技术的迅猛发展,数据与算法被赋予了新的含义。一方面,数据在数字时代同时拥有了自然属性和社会属性,其既是客观对象的真实记录,又承载了记录主体在数字社会中的"行为轨迹"。另一方面,算法通过对数据的收集、处理和输出的全过程参与,逐渐有了自主性和认知特征,已经不再局限于代码的表现形式,其以算法决策的方式摆脱了纯粹的工具性角色而成为"决策者"。算法不仅在招聘、教育、信贷、证券交易等私营部门中主导着决策结果,而且在公权力领域,诸如智慧司法系统、犯罪算法预测机制、算法推荐中也扮演着至关重要的决策角色。[1] 那么,当承载着社会利益的数据不断累积,算法逐渐取代人类成为数据处理的唯一主体时,社会资源分配权力正在悄然让渡于数据与算法。这也成为算法权力在数字时代崛起的契机,因为当特定主体拥有足以支配他人或影响他人资源的能力时,均可成为权力。[2]

算法进入法学视阈,源自1987年的美国华尔街,券商托马斯·彼得非利用分层算法模仿证券交易员进行交易操作的事件。[3] 此后,大数据、人工智能、云计算、区块链等技术对于金融领域的影响不断加强,金融科技监管成为算法在法学界研究的头号阵地。加强算法人工智能技术对于金融监管、社会管理、行政执法以及司法审判的应用成为近年来的国家目标。中共中央办公厅、国务院办公厅于2021年7月6日印发了《关于依法从严打击证券违法活动的意见》(以下简称"意见")指出,"有效运用大数据、人工智能、区块链等技术,建立证券期货市场监测预警体系,构建以科技为支撑的现代化监管执法新模式,提高监管执法效能"。不仅如此,第十三届全国人大第四次会议表决通过的《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》(以下简称"十四五规划")指出,加快数字化发展、建设数字中国;聚焦人工智能关键算法,加快推进基础理论、基础算法的研发突破与迭代

〔1〕 See *Bruno Leprietal*, *Fair*, *Transparent*, *and Accountable Algorithmic Decision – Making Processes*, Philosophy & Technology, Vol. 31, 2018.

〔2〕 参见[德]尤尔根·哈尔贝斯:《作为"意识形态"的技术和科学》,李黎、郭官义译,学林出版社1999年版。

〔3〕 参见张凌寒:《算法规制的迭代与革新》,载《法学论坛》2019年第2期。

应用。由此可见,数字中国建设,既是全球化背景下技术竞争的必然选择,也是证券监管、行政司法升级、互联网经济蓬勃发展的内在要求。

算法权力作为数字时代社会发展的重要推动力,是数字中国建设的关键内容,事关数字中国建设的成败,因此,从法学的视角规范算法开发与使用,防范算法风险,充分发挥算法的积极功效,研究意义重大。在探寻规制算法权力的过程中,算法影响评估机制越来越被各国所重视,美国、加拿大、欧盟等都在近年提出了人工智能的治理框架,而算法影响评估机制的构建成为各国在人工智能领域抢占制定国际标准与规则的高地。我国旨在成为科技强国、创新型国家,尽早构建可信、可靠的人工智能体系显得尤为重要。因此,构建算法影响评估机制无论是对人工智能的国际竞争,还是对提升自身科技水平而言,都显得尤为迫切。

一、算法权力:诞生与风险

(一)算法权力的诞生

算法权力是算法在数据处理过程中借助自我学习而产生的技术优势,以算法决策的方式对社会主体产生的影响力和控制力。在数字时代,数据的社会属性使其成为社会权力的基础,而算法主宰着数据收集、处理和输出的整个生命周期,当一方主体通过占有数据并控制另一方主体获取数据的渠道,其同样可以构成权力的来源。[4] 数据与算法结合的动态过程是算法权力产生的必要条件,算法权力通过算法决策最终得以体现。算法权力是一种技术性权力,在市场经济的竞争压力下,企业对于技术创新的需求十分强烈,算法技术的创新也大都缘起各大私营平台企业。当技术优势不断凸显时,公权力部门也加大算法在社会管理、金融监管、司法裁判等领域的运用,用于推动各项公共管理事业的优化。算法权力的影响力和控制力也在私营平台与公权力部门中产生了不同的表现形式。

在私营平台中,算法权力不仅构建了全新的数字经济投资、发展模

---

〔4〕 See B. H. Raven, *Power and Social Influence*, California University Los Angeles, 1964.

式,而且塑造了无形的资本意识形态。以证券投资为例,智能投顾的出现打破了传统证券投资的投资生态,极大地迎合了投资者迫切希望通过有效的资产管理获取一定收益的需求。[5] 在互联网经济领域,算法推动下的互联网商务已经具备诸多线下市场难以比拟的优越性,其不仅改变了传统的竞争格局,而且营造了一个崭新的市场环境。在数字经济市场环境中,虽然我们还能看到市场竞争的种种特性,但其背后的助推力"无形之手"已经被"数字化的手"取代。[6] 平台成为数字经济时代最重要的社会生产组织,它通过提取的数据流与算法结合,进一步形成用户画像和用户行为预测,再利用平台新闻推荐算法、用户信用评分机制等多种技术手段来塑造用户习惯与价值观,从而主导消费市场。算法权力在构建新的生产方式的同时,也服务于资本逻辑的意识形态。算法权力与资本力量相结合,形成数字资本,新消费主义正是由数字资本所建构的意识形态,数字资本通过变实体消费为电子消费,数据算法直接接管消费世界,消费世界都要按照以数据算法为核心的数字资本逻辑重新规划,数据算法成为一种强劲的意识形态力量,巧妙地实现了对消费者的无意识控制。[7]

在公权力部门中,算法权力嵌入了行政与司法部门,与公权力运行相融合。在行政数字化建设方面,从 2016 年"十三五"时期首次提出"数字中国"概念,到 2019 年党的十九届四中全会首次提出"推进数字政府建设",再到 2021 年"十四五规划"明确提出"加快数字化发展、建设数字中国",要求政府全方位地转型,将算法与数据同政府治理相结合成为近年来国家倡导的趋势。在建的社会信用评估体系、新型冠状病毒肺炎疫情发生以来的疫情轨迹定位防控系统以及政府部门各类人脸识别系统等,都是算法同行政部门深度融合的具体形态。在金融领域,算法应用与公权力合作的脚步也从未停止,从 2010 年科学技术部等五部委发布《关于印发促进科技和金融结合试点实施方案的通知》,到"十四五规划"中提出"构建金融有效支持实体经济的体制机制,提升金融科

〔5〕 参见李文莉、杨玥捷:《智能投顾的法律风险及监管建议》,载《法学》2017 年第 8 期。

〔6〕 参见[英]阿里尔·扎拉奇、[美]莫里斯·E. 斯图克:《算法的陷阱:超级平台、算法垄断与场景欺骗》,余潇译,中信出版社 2018 年版,第 39 页。

〔7〕 参见邓伯军:《数字资本主义的意识形态逻辑批判》,载《社会科学》2020 年第 8 期。

技水平,增强金融普惠性”,再到2021年4月科学技术部发布的《关于加强现代农业科技金融服务创新支撑乡村振兴战略实施的意见》,算法在金融领域的应用愈加受到重视,成为未来金融发展的关键性技术力量。在司法领域,当算法应用于司法部门时,其开始主宰人的自由权利,并引导司法权力的行使。〔8〕美国威斯康星州法院已经使用COMPAS算法对被告埃里克·卢米斯(Eric Loomis)进行了判决。〔9〕在我国,算法与司法系统结合的例子也不胜枚举。例如,最高人民法院与最高人民检察院提出“智慧法院”与“智慧检务”、北京法院的“睿法官”智能研判系统、上海法院的“206”刑事案件智能辅助办案系统。算法权力与各公权力部门的融合正在进一步加深。

算法权力通过在数字空间扩张,以技术力量完成对现实社会的历史性重构。数据与算法将成为影响世界社会完成体系构建的力量来源,以数据算法为核心的社会构筑力正在各大领域形成。算法权力以意识形态力量与公私主体的融合,获得了对世界政治、经济、文化秩序前所未有的操控。

(二)算法权力的异化

在数字时代,算法最大的优势在于其可以迅速整合大量的、任何人类个体都无法精确地压缩和处理的数据,并且算法可以避免人类在决策时的任意、草率与偏见,它是精确且高效的模拟智能化运行机制。但这种精确和高效的算法决策由于缺乏透明度、可预测性与可解释性,在运行过程中会将人类的偏见或其他扭曲决策的因素复制,并造成决策的歧视性或不公正。因为几乎所有算法系统都是私人和秘密开发的运算程序,甚至是那些应用于公共部门的算法系统——公众对其算法的细节知之甚少。〔10〕算法“黑箱效应”的存在使公众无法了解开发者在设计或实

---

〔8〕 参见张凌寒:《权力之治:人工智能时代的算法规制》,上海人民出版社2021年版,第8页。

〔9〕 2016年,一位名为Eric L. Loomis的被告被判处6年有期徒刑,原因是COMPAS认定他为“高风险”。Loomis随即提出审查COMPAS算法的请求,被威斯康星州的州立法院驳回,提交给美国联邦最高法院后也于2017年6月宣告诉讼失败。参见《红星专访美国机器判案法院:机器说你有罪,你果然有罪》,载https://weibo.com/ttarticle/p/show?id=2309404105958811950706,2021年12月10日访问。

〔10〕 See Deirdre K. Mulligan & Kenneth A. Bamberger, *Procurement as Policy: Administrative Process for Machine Learning*, Berkeley Technology Law Journal, Vol. 34, Issue 3, 2019.

施算法系统时所做的具体决策将如何影响其下游结果,公众无法知晓政策目标如何被转化为算法系统的运行逻辑,或者说算法系统如何代表政策权力的发出指令。算法“黑箱”的不可知性,固化了算法权力的排他性。无论是各平台企业所代表的资本权力,还是公权力部门所代表的政治权力,伴随算法权力的嵌入,将造成公正与平等的缺失。权力垄断的异化现象将引发一系列不可规避的风险。

1. 私营平台算法权力使用过程中引发的风险

私营平台作为算法技术发展最大的推动者,也是算法权力最初的控制者。平台利用算法技术优势与架构优势,通过对用户数据的收集来攫取高额的利润,算法权力在商业领域形成了对消费者的掠夺。平台对于数据的掌控也影响着公民隐私信息的安全。不仅如此,基于算法与数据而产生的新型数字产业链在无形之中剥夺了公众的自由选择权。受限于现有法律无法对算法权力膨胀、越界行为施加直接的规制,私营平台所掌控的算法权力带来了严重的法律风险。

算法权力导致平等权受损。算法利用已有的大数据对用户的行为信息和性格特点进行描述,并利用画像数据形成用户画像。用户画像精准的个性化推荐虽然便利了人们的购物需求,但也侵害人们在同货同价购买方面的平等权。算法通过用户画像,对不同群体进行分类与身份构建,从而为消费者制定反映其支付意愿的价格,实施“一人一价”的价格歧视行为。在各大购物平台算法中,会员票价反而比非会员高、使用高档手机购物付费更多、商品搜索频次越多却越贵等算法歧视现象层出不穷。这种歧视非常隐蔽,不易被人们察觉。即使被觉察,也不容易举证。同样,在其他类型的算法中也存在算法歧视。“以 Facebook 泄密门为例,Cambridge Analytica 仅靠‘趣味小测试’就拿到了 32 万名用户的授权,据此推断出 5000 万用户喜好,有针对性地设下桃色陷阱、推送诱导新闻、操纵总统选举等严重侵犯平等权的行为。”[11] 私营平台利用技术能力不对等、信息来源不对称等优势,影响用户在商业活动中最真实的意思表达,使处于弱势地位的消费者无法做出正确的判断,导致不公平

---

〔11〕 郑智航、徐昭曦:《大数据时代算法歧视的法律规制与司法审查——以美国法律实践为例》,载《比较法研究》2019 年第 4 期。

和低效率现象频现。

算法权力导致隐私权受损。数字化的网络世界产出了一个“数字化人格”。所谓数字化人格,即以个人在网络中所留存的数据信息为基础,以数据模型构建起一个借由数字化信息而建立的人格映像。〔12〕数字人格产生的基础是对海量的个人数据的分析处理。个人数据成为算法运行的必要条件,成为互联网经济时代推动经济发展的“原油库”。平台企业的算法,一方面,通过数据的分析为用户提供更为精准化的服务;另一方面,也加剧了数据泄露带来的用户个人隐私的泄露风险。在算法运行过程中可能因本地服务商的不当操作或受非法入侵,而造成大量信息的泄露。同时,算法运行过程中对于数据的获取本身就是对用户隐私的侵害。海量的消费数据被记录在案,用户的偏好和习惯在不经意间就被人知悉。这些数据为平台了解用户、推广产品提供了便利。在这个以商业利益至上的市场经济环境下,平台都极力争取获得用户数据的机会。苹果公司曾表示:“于2021年春季将升级IOS系统,以加强对广告平台访问其数据的限制,其中要求应用程序必须获得用户同意方可追踪用户在其他公司的应用和网站上的活动数据。”此次升级将影响那些依赖苹果个人设备标识符进行数据跟踪的平台。脸书(Face Book)创始人马克·扎克伯格对此表示强烈抗议。原本应由用户自由支配的隐私,在算法权力的发展之下,却掌控在各科技巨头手中。人们无法知悉算法“黑箱”背后的个人数据流向,也无法掌控自己隐私的权属。

算法权力导致自由选择受限。“监视资本”是算法权力膨胀过程中对公民在市场经济体制下自由选择的严重障碍。〔13〕“监视资本”,是指超大规模的互联网技术公司利用全球数字平台实现自动化营销而得到经济增长。此种商业模式以“眼球”而不是收入作为公司的估值基础和投资回报的预测指标。〔14〕监视资本产生利润的主要来源来自对个人信息流量的引导和控制,但是其利润转化过程不完全透明。平台企业可以通过算法和极为复杂的构建,实现对用户数据的挖掘、分析、处理。大量

〔12〕 参见齐爱民:《私法视野下的信息》,重庆大学出版社2012年版,第62页。

〔13〕 参见《网络法学》,中国政法大学出版社2019年版,第269页。

〔14〕 See Shoshana Zuboff, *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization*, Journal of Information Technology, Vol 30, Issue 1,2015.

的元数据经由算法的处理,不断被结构化、同质化、模型化,最终被导入数字经济市场。其借助算法将数据转化为“测试性产品”(用户行为预测),将此类产品出售给各类广告客户,从而产生高额利润。平台用户成为生产链上的重要一环,不仅是数据的产出者,用户同样还以消费者身份对平台企业进行数据反馈。〔15〕此种模式基本可以实现零成本剥削用户“剩余数据”进而完成资本的原始积累。其将用户原有的知情权和自主选择权剥夺,进一步固化了用户的弱势地位。不仅如此,互联网平台从单一业务转向跨界经营,通过打通上下游产业链,打造复杂的“金融超市”生态。〔16〕以已有用户基础为基础,发挥其在数据占有与数据利用上的优势,为广大群众提供“有选择”的金融服务。以金融科技平台为例,其通过大数据结合算法技术的准确计算,对互联网金融服务对象做出准确的信用评分,再据此提供专项的金融服务。以算法技术剥夺了客户的服务选择,同时利用最大的杠杆为金融科技企业带来最大的利润。

2. 公权部门算法权力使用过程中引发的风险

算法权力在公权力部门中的应用主要体现为行政领域与司法领域。算法权力在行政领域的扩张,给传统行政活动的运行规则和基本范式带来了翻天覆地的变化。一方面,算法自动化决策明显地缓解了治理机关与治理对象间信息不对称,进一步增强了行政部门收集、处理、输出信息的能力,提升了行政执法的效率,是行政权力合法和有效运行的催化剂。但另一方面,在算法决策自动化运行过程中,自动化的处理方式剥夺了公众的意志表达,架空了正当程序原则与信息公开原则。在司法领域,一方面,算法自动化决策在预防犯罪、侦查以及司法裁判环节的应用,极大地提升了司法活动的效率;另一方面,算法线性的处理方式无法有效地输出正确的法律结果,将产生司法偏误的现象。针对算法自动化决策在行政与司法领域可能出现的风险,我国现有的法律、法规并未提出系统的规制框架,这将进一步加剧算法权力在公权部门扩张的风险。

算法权力剥夺了行政决策中公众表达意志的自由。在行政决策中,

〔15〕 参见胡凌:《论赛博空间的架构及其法律意蕴》,载《东方法学》2018 年第 3 期。

〔16〕 参见孙方江:《数据垄断视角下金融业和互联网平台的共生发展问题研究》,载《西南金融》2021 年第 3 期。

公众参与的核心理念是“听取公众意见”,这需要通过专门的程序保证公民参与,并保证各方信息的对称性。公众的有效参与,意在通过自己的行为影响某种结果的形成,而不是作为一个消极的客体被动地接受某一结果。但在收编算法权力的行政部门中,“自动化偏见”成为公众意志表达中的障碍,人们对于算法决策的倾向度将大于其自主所做出的决策,其自主性表达被算法权力不断抹杀。〔17〕算法决策系统所使用的“超级推理”能够以微妙而有效的方式塑造用户的感受和行为,破坏个人的独立判断力。〔18〕同时,自动化决策所带来的行政行为自动化运行,造成了行政相对人陈述与申辩环节的缺失。在预测型算法中,算法结果产生的过程都处于算法“黑箱”内部,其完全规避了公众参与、专家论证等信息交换的正当程序控制。在行政相对人面对自动化或半自动化的算法行政决策时,若行政处罚将产生不利于相对人的处罚结果时,相对人囿于算法“黑箱”的不可知性,无法通过有效的陈述申辩来改变行政处罚结果,自动化或半自动化的算法决策相当于否定了行政处罚行程序相对人的陈述申辩权,这将进一步消解公众参与在行政行为中的意志表达。

算法权力架空了行政法的正当程序与信息公开原则。算法的自动化决策为行政执法带来便利的同时,也压缩了原有的行政活动环节,使正当程序原则无法落实。以交通违章为例,算法自动化决策能对交通监控设备采集的数据进行智能识别和分析,并按照输入的技术标准,识别并输出车辆违章数据,经交管部门审核,即可对违法车辆进行处理,最终将违法处理结果通知送至违法者。在此执法过程中,除去人为的审核环节,其各个环节均可通过算法技术在瞬间完成。算法决策的自动化,的确增强了行政权力在运行过程中的效率,但其省略的行政活动环节,将对行政相对人的实体权利造成消减,直接损害其程序性权利。行政程序的设立目的,在于保障相对人能接受客观、公正、合法的行政行为,程序环节的缺失不仅架空了程序正当原则,也架空了信息公开原则。自动化决策的过程本身就无法被行政相对人知悉,加之其压缩了行政行为的必

〔17〕 See Carr N. ,*The Glass Cage*: *Where Automation is Taking Us*, Random House,2015.

〔18〕 See Karen Yeung, *Hypernudge*: *Big Data as a Mode of Regulation by Design*, Information, Communication& Society, Vol. 20, No. 1,2017.

要环节,相对人对于行政信息的了解将进一步被限制。程序正当与信息公开原则,是规制行政权力的重要原则,在算法自动化决策的引入下,两大原则在实践中被架空,这将不仅造成行政权力行使的失范,还将造成算法权力的极度膨胀。

算法权力在司法领域的应用的风险。算法在司法领域的部署、应用都可能因技术与司法实践无法契合而产生对法律权威及当事人合法权利的消极影响。算法技术与司法相融的实质,就是将司法的过程纳入数据与算法的泛在监控之中。以司法大数据与算法技术,打造一套静默化、自动化、可视化的全流程监控系统,构筑"数据铁笼",实现"科技控权"。[19] 算法应用于司法活动的初衷是以人工智能技术来监控司法权力行使的合法性,然而,伴随算法权力影响力的增强,司法活动却被纳入了更隐蔽、更多元、更宽泛的技术权力侵蚀之中,从而丧失司法的独立性。同时,算法权力与司法的结合将进一步削弱法官的主体地位,由算法辅助或主导的审判模式将带来严重的司法风险。一个审判结果的导出,需要大量的法律要素以及对于法律价值的判断,而算法系统无法将法律要素有效地数据化,也无法将法律价值通过数据的形式展现。基于广泛收集数据的充分信息是人工智能有用武之地的基本前提。[20] 缺乏有效的司法数据"喂养",算法系统在司法审判中所做出的决策将存在偏误与不公。

二、算法影响评估制度在规制算法权力中的应用

(一)算法影响评估制度的生成

算法影响评估机制,是指对自动化决策系统的决策流程、数据使用和系统设计等内容进行系统评判,以明确该系统的影响水平和风险等级的一种算法治理实践。算法影响评估机制设立的目的有二:第一,旨在

---

〔19〕 参见王禄生:《司法大数据与人工智能技术应用的风险及伦理规制》,载《法商研究》2019 年第 2 期。

〔20〕 参见左卫民:《关于法律人工智能在中国运用前景的若干思考》,载《清华法学》2018 年第 2 期。

构建一个系统而合理的方法论来审查算法,在算法做出无法纠正的决策前规避风险。第二,创造并提供算法在自我学习过程中所做出的决定及其理由的文件,留档文件既可以便于全面地对决策进行问责,又可以为日后对算法决策进行干预提供有效信息。

算法影响评估机制源自影响评估机制这一监管方式于算法领域的应用。影响评估机制作最早来自环境保护领域,美国1970年通过的《国家环境政策法案》(NEPA)〔21〕首创环境影响评估机制(EIS),后来影响评估机制被不断应用于隐私保护(PIA)、人权保护(HRIA)、数据保护(DPIA)等领域。2016年,欧洲从事人工智能研究的学者联合发表了一份题为《负责性算法的原则和算法社会影响声明》,声明中阐述了算法在运用时的5个高级原则——责任、可解释性、准确性、可审计性和公平性,并针对算法设计阶段、发布前与发布后提出了一套具有探索意义的问题作为评估算法社会影响的方案。〔22〕这成为算法影响评估机制发迹的开端。2016年,欧盟通过的《通用数据保护条例》(GDPR)中提出了"数据保护影响评估"(DPIA),玛戈特·卡明斯基和吉安克劳迪奥·马尔吉里对此提出将"数据保护影响评估"与GDPR中的"协作治理"制度〔23〕有效联结,构建个人数据权利与算法治理相结合的影响评估机制。〔24〕2018年,美国纽约市颁布《算法问责法》,其标志着算法影响评估机制在美国的立法新篇章。2019年,加拿大颁布《自动化决策指令》,将算法影响评估机制广泛地应用于政府公共部门的算法决策过

〔21〕 National Environmental Policy Act, 42 U. S. C. § § 4331 – 4347.

〔22〕 See Nicholas Diakolopous et al., *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms*, Fairness, Accountability, and Transparency in Machine Learning (2018), https://www. fatml. org/resources/principles – for – accountable – algorithms, visited August 16, 2021.

〔23〕 协作治理主要作为工具性治理措施存在,其目的在于减少自动化决策中的错误、偏见与歧视,力图通过公私协作的监管模式对自动化决策系统进行评估监管。

〔24〕 See Margot Kaminski et al., *Algorithmic Impact Assessments under the GDPR: Producing Multi – layered Explanations*, International Data Privacy Law, Vol. 11, Issue 2, 2021.

程之中。[25]美国与加拿大的框架式治理模式与欧盟的协作治理模式对我国算法影响评估机制的构建有一定的参考价值。

(二)算法影响评估制度在防范算法权力风险中的应用

欧盟的算法影响评估机制框架,通过数据保护影响评估(DPIA)的多重作用进行构建。数据影响评估(DPIA)在《通用数据保护条例》(GDPR)中发挥了协作治理和个人数据保护的功能。首先,DPIA 的协作治理功能。DPIA 在协作治理中发挥着"元监管"的作用,即通过寻求外部监管意见,改变决策系统的决策过程与决策启发方式,形成一种自我监控、自我调节的模式。协作治理要求算法决策背景下,算法决策系统要充分考虑算法不公、算法错误、算法偏见与算法歧视等风险,向受算法影响的数据主体、监管机构、内部独立数据保护官[26]以及第三方专家等寻求意见,并在此基础上构建风险应对方案。协作治理的目的在于极力避免算法系统构建时的偏见与偏误,通过多方协作的评估方式校正算法决策系统的合理性。其次,DPIA 的个人数据保护功能。GDPR 中第 13 条、第 14 条和第 15 条,规定了数据收集主体对于个人的通知义务以及个人对于被收集数据的访问权,如此规定赋予了受自动决策系统影响的个人获取决策的运行逻辑以及决策的预期结果的权利。GDPR 第 35 条中还规定,DPIA 必须对系统处理的目的进行描述,给数据主体的权利与自由进行风险评估。GDPR 充分满足了个人数据权利保护的权利要求,通过事先通知、事后了解以及自动决策系统有关主体的主动披露,对个人数据权利进行保护。在算法决策背景下,DPIA 通过将协作治理模式与个人权利相联结,以对个人数据权利的具体描述作为算法决策系统颁布阶段的注意风险点。此种模式的确为各国在算法治理方面提供了很好的监管范本,但其实质上更像一种扩大算法决策系统执行人承

---

〔25〕 Algorithmic Impact Assessment (AIA), Government of Canada (July 28, 2020), https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html, visited July 27, 202 [hereinafter, Canadian AIA].

〔26〕 在数据保护评估程序要求指南中(DPIA Guidelines),欧盟还将设置一个内部数据官员岗位——数据保护官(独立于算法系统负责人),进行算法系统的评估咨询。See *Article* 29 *Data Protection Working Party* (Adopted on 4 April 2017), http://ec.europa.eu/justice/data-protection/index_en.htm.

诺的手段,由执行人自己在风险评估意见反馈之后提供应对方案,监管部门仅提供一个预估风险的流程,并未提出实质性的监管标准。同时,DPIA 最大的不足在于,并未提供强制向公众披露信息的机制。[27] 公众披露机制被认为是影响评估作为监管工具的一个最基本要素。公众披露不仅有利于构筑算法信任,而且能有效地形成公众监督预防算法风险。在 GDPR 中,其对公司的算法、公司的行为与监管机构的能力都很有信心;DPIA 中,要求公司提出如何实现个人数据权利的方案以及如何解决算法不公、算法偏见和歧视的问题,[28] 其通过多方形成的协作治理评估机制对算法决策风险进行预估,从而省略公众披露与公众参与环节。那么,如何确保此种个人权利保护与协作治理的混合系统正朝着公共利益而努力?这是 DPIA 需要解决的问题。

美国联邦《算法问责法案》(以下简称"法案")和加拿大的《自动化决策指令》(以下简称"指令")均采取框架治理[29]与协作治理相结合的模式对算法决策系统进行评估。[30] 纽约市依据"法案"成立了自动化决策工作组,并由其制定自动化决策清单,以问卷调查形式进行算法风险评估;"法案"采取自我评估与政府评估的双轨模式,由联邦委员会制定算法评估标准,并且在面对"高风险自动化决策"时,联邦委员会可在必要时派遣相关人员与专家组进行合作评估。"指令"奉行的是清单式问卷调查型算法影响评估机制,根据"指令"的规定,算法评估标准着重关注经济、社会、生态等方面的影响,评估内容包含了大约 60 个有关业务流程、数据以及系统设计决策问题,并且评估标准和治理框架需每隔 6 个月进行更新。[31] "指令"中还规定使用算法决策的政府机构必须在生

---

〔27〕 See Michael Veale et al., *When Data Protection by Design and Data Subject Rights Clash*, International Data Privacy Law, Vol. 8, Issue 2, 2018.

〔28〕 See Margot Kaminski et al., *Algorithmic Impact Assessments under the GDPR: Producing Multi – layered Explanations*, International Data Privacy Law, Vol. 11, Issue 2, 2021.

〔29〕 在确定性的法律规则之外,以可量化的分级标准,构建起约束相关主体认知和行为框架的治理方式。

〔30〕 参见张欣:《算法影响评估制度的构建机理与中国方案》,载《法商研究》2021 年第 2 期。

〔31〕 See Government of Canada, Algorithmic Impact Assessment (June 3, 2020), https://canada – ca. github. io/aia – eia – js/.

产之前和项目上线前完成算法影响评估机制的相关工作。

无论是美国还是加拿大,都围绕着算法公平、算法透明的核心价值建立问责评估体系,以详细的评估标准遵循技术治理和权利救济的思路。其自上而下封闭式的问卷清单,看似合理但却存在很大的缺陷。奥斯·凯文、杰文·哈森与梅雷迪斯·德宾就以此种封闭式的问卷清单对食品短缺问题设计算法决策系统,他们以各种算法监督框架验证自动化决策结果的公平、透明、无偏见。但在设计评估的封闭式问题的过程中,评估人员必须事先已经对算法决策结果的好坏优劣具有清晰的认识,才能提出正确的评估问题。[32] 此种自上而下的评估模型虽然能让算法决策系统设计师在早期对决策风险进行评估,但算法自动化决策系统在运行时的自主学习,将使自动化决策过程的输出结果处在一个动态更新的状态之中,事先封闭式的问卷评估模式无法做到在早期对算法决策风险的实质防范。

无论是欧盟的 DPIA 模式,还是美国、加拿大的问卷模式,不同的算法影响评估机制模型中都必然存在可取的优势,但如何减少算法决策所带来的风险可能,如何留档以备问责之需,是构建算法影响评估机制时必须着重考虑的问题。

三、算法影响评估机制的法治构建

算法权力的自主性特点给算法权力的规制带来了很大的困境,传统静态的监管体系无法有效地规制算法权力所带来的风险,故应建立动态灵活的监管体系对算法权力进行规制。算法影响评级机制与其他制度的高度适配性,及其协作治理理念将成为算法权力失范规制的有效路径。

(一)构建内外兼容的协作治理评估体系

在欧美的算法评估体系构建经验中,私营平台往往缺乏合规意识,

---

〔32〕 See Os Keyes, Jevan Hutson & Meredith Durbin, *A Mulching Proposal: Analyzing and Improving an Algorithmic System for Turning the Elderly into High – nutrient Slurry*, Chi Conference on Human Factors in Computing Systems (2019).

政府往往缺乏技术与算法流程知识,协作治理能充分发挥政府、行业专家以及社会公众等多元力量,对算法权力失范进行精准化、立体化、框架化的监管。在我国,《个人信息保护法》第11条与《数据安全法》第9条都鼓励推动政府、企业、相关行业组织以及社会公众共同参与数字化治理。因此,构建内外兼容的协作治理评估体系尤为重要。协作治理评估体系不仅要求多元主体共同协作进行算法影响评估,还要求各制度机制之间相互协作配合,对算法的动态变化进行有效监管。首先,从多元主体协作方面出发,在私营平台中,应通过协作治理强化政府正当程序原则对算法决策系统的合规构建,改变系统设计者组织生产模式与算法设计流程,提升算法决策系统的公平性、公正性与合法性。借鉴金融科技领域的监管经验,减少监管惯性,整合来自政府、市场、社会的资源智识,取长补短,相互补充,促进多元主体的合作,利用技术智力构建多层次的层次的治理框架。[33] 通过国家互联网信息管理部门牵头,构建集中化、一体化评估机制,利用统一的算法影响评估标准对自动化决策系统进行定期评估,通过公私协作提升算法影响评估机制的公信力。在公权力部门中,通过协作治理,提升行政、司法部门对算法系统的利用效率,打破公权力部门在技术与算法流程设计方面缺乏经验的窘境。借鉴欧盟DPIA的数据专员评估咨询机制与美国纽约市的自动化决策工作组的多方协作机制,加强公权力部门与平台企业、高校、科研机构以及社会公众的联系,通过多元参与,提升自动化决策系统在公权力部门的合理性、合法性、公平性。通过协作治理模式,充分提高自动化决策系统在公权力部门的透明度,进一步构筑算法信任。在制度协作方面,由于算法影响评估机制不能对整个算法运行周期进行全覆盖,因此需要其他制度进行协作联结。首先,加强数据安全保护制度与算法影响评估机制的协作联结。数据作为算法系统的输入对象,其客观性、无歧视性与否,将影响算法系统的输出结果,充分利用数据安全保护制度能有效地减轻影响评估机制的构建成本,进一步提升影响评估机制的效能。以欧盟的DPIA为例,数据处理评估制度设计并非一种停止数据处理的工具,而是作为一

〔33〕 参见李有星、王琳:《金融科技监管的合作治理路径》,载《浙江大学学报》(人文社会科学版)2019年第1期。

种改进算法处理的数据活动,并为未来追求法律责任提供问责点的方法。根据《数据安全法》第21条的规定,规定各地区、各行业需对目标领域的重要数据进行分类分级保护。分类分级保护制度能有侧重、有效能地对数据安全进行保护,但却缺少了与算法决策公正之间的联结。因此,在算法分类分级保护的基础上,针对算法影响评估机制的对象需要建立数据分类分级登记备案制度,以数据安全类别与级别的不同,制定影响评估方案,实现影响评估机制的场景化应用,并为未来追究法律责任提供数据处理环节的问责点。其次,通过算法问责机制落实算法影响评估阶段的算法责任。利用算法解释权制度,弥补算法影响评估制度在算法运行阶段的不足。通过影响评估机制与各项制度之间的协作配合,进一步加强算法监管,保证算法公正。

(二)根据算法场景化的不同设定不同的评估标准

算法影响评估标准设立的最大困难,源自"自主学习"型算法的无法预测性,不同于训练型算法可以通过对部分数据进行训练得到训练任务所布置的正确答案,并在此基础上自己建立一个模型来解决未来面临的类似数据的相关任务模式,"自主学习"型算法不会"输入"任何正确答案,而是"自由"地破译数据中可能表明正确答案的模式。[34] 虽然算法设计者肯定能影响其负责的算法的选择,但"自主学习"型算法的选择通常是完全不可预见的。[35] 算法设计者可以通过预编程冻结算法在面临新信息时改变结论的情形,但这将失去算法的工具价值。对此,在构建算法影响评估标准时,需要使用不同的合理标准对不同算法类型进行评估。[36] 首先,在训练型算法中,由于训练结果是预设的,算法需要大量训练数据的输入来保证最终的算法结果无限接近于训练结果。在设立影响评估机制时,需要对算法训练数据的合理性、合法性、公正性进

---

〔34〕 See Avigdor Gal, *It's a Feature, Not a Bug: On Learning Algorithms and What They Teach Us*, OECD(June 7, 2017), https://one.oecd.org/document/DAF/COMP/WD(2017)50/en/pdf, *Harry Surden, Machine Learning and the Law*, Washington Law Review, Vol. 89, No. 1, 2014.

〔35〕 See Jason Millar & Ian Kerr, *Delegation, Relinquishment and Responsibility: The Prospect of Expert Robots*, Ryan Calo et al. eds., Robot Law 102 (2016).

〔36〕 See *Karni Chagal – Feferkornt, The Reasonable Algorithm*, University of Illinois Journal of Law, Technology & Policy, Vol. 2018, Issue 1,2018.

行评估,从数据输入环节避免歧视性数据的输入,从而影响算法结果。以证券交易为例,算法可以根据用户数据分析出用户投资行为的关键触点,使模型算法偏离度越来越小,形成不断迭代的运营闭环,从而促进证券投资服务更加精准匹配。因此,在设立训练型算法模型评估标准时,依据不同领域、不同算法使用场景对算法结果准确率的需求,对算法训练结果的准确率进行评估,确保训练型算法在投入使用后的有效运行。其次,针对"自主学习"型算法,在事前影响评估标准设置上,要充分评估算法设计者在设计算法时各项程序的合理性,通过要求算法设计者提供算法神经网络中各层级的运行逻辑、算法决策树模型以及算法决策的可能性结果,在算法自主运行之前排除人为的偏见,从而降低算法失范的风险。在"自主学习"型算法投入运行之后,需要由监管机构进行日常监管,定期对该类型算法进行评估。根据具体领域与场景的不同,检查算法是否存在风险(如预测是否准确、对个人是否存在负面影响、是否存在敏感数据错误),当出现算法风险时,算法设计者与运营人需要提供针对性的解决方案,并进行新一轮评估。

(三)增设信息公开流程以明确被评估主体的强制披露义务

在算法自动化决策中,数据输入与决策结果输出这两个环节的信息公开,将极大地提升算法透明度,促进公众对算法的信任。在数据输入环节的信息公开制度设计上,我国通过《数据安全法》中的数据分级分类制度,以及《个人信息保护法》第 14 条中的"知情同意权",最大限度地对个人数据用途在公私领域进行公开,提升了数据透明度。但在算法决策输出环节上,法律学者一直以算法"黑箱"作为算法不可揭示的理由,而模糊了算法可以被询问与不能实际被询问之间的界限。[37] 人们无法了解算法内部的具体自主学习与决策过程,但可以通过技术细节了解算法的决策结果的产生逻辑。以算法偏见为例,算法偏见的产生可以来自算法自主学习的不同阶段,在算法自主化运行前而产生的"历史偏见"以及在算法运行后基于应用目的而产生的"应用偏见",都将影响算

---

〔37〕 See David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn about Machine Learning*, U. C. Davis Law Review, Vol. 51, Issue 2 (December 2017).

法决策的产出。因此,“黑箱”可以刺破,需要的是技术细节。在部署算法影响评估机制时,平台企业与公权力部门,应公布程序输入和输出的数据类别信息,公布算法所涉及的逻辑,包括算法输入数据占决策的比重,助于了解算法运行逻辑的基本专业知识、算法决策的范围以及程序可能的后果。按照欧盟《通用数据保护条例》背景引言中的披露要求,“数据控制者应当以‘易见、易懂、易读’的方式提供真实、可靠可能产生的影响类型示例,包括可视化技术等方式呈现”。[38] 平台企业与公权力部门在披露算法决策相关信息时,也应遵循最大限度的透明披露标准,以便于公众理解。稍有区别的是,公权力部门有比平台企业更重的算法透明责任与要求。平台企业可以通过商业秘密、知识产权保护的借口,不完全公布设计企业利益的算法内容,但必须以说明报告的形式向公众公布算法决策相关技术细节。

(四)落实算法问责体系加强算法责任追究

算法影响评估机制并非一种独立的机制,其只能成为算法规制体系中的关键部分,算法影响评估机制作用的发挥离不开算法问责在其中的作用。[39] 算法问责的具体要求在公权力部门的体现将更为直接。在公权力部门,算法决策系统无论是作为辅助决策工具,还是独立决策作出者,公权力部门都无法排除其具体行政行为的定性,需要为自动化决策结果负责。而在平台企业,算法问责机制的构建亟须改变“主体—行为—责任”传统理论下的“责任鸿沟”,[40] 从以缺失控制权作为理由主张“无过错则无责任”转变为以主观过错归责原则,刺破被隐藏在自动化技术面纱之下的算法设计者部署者的主观意图。具体实施路径需从事前、事中、事后3个算法问责时间节点进行设置。在算法投入运营的事前阶段,首先,将算法影响评估机制与算法问责机制相结合,通过对用户行为干预程度、社会动员程度以及潜在风险可能等为标准,建立不同的算法风险等级评估审计,详尽地调查算法设计者与部署者是否存在主观过错,以便归责。其次,建立算法备案机制,将平台问责节点进行固

〔38〕 See General Data Protection Regulation, Official Journal of the European Union, L 119/2, 2016.

〔39〕 参见前注〔24〕。

〔40〕 参见张凌寒:《网络平台监管的算法问责制构建》,载《东方法学》2021年第3期。

定,获取平台设计部署时具有潜在危害和风险的算法系统的相关资讯,以固定问责点为今后的监管提供信息基础。算法备案形式可以通过监管机构发布模板,由私营平台企业填写,将算法部署的目的、风险、评估过程与风险控制方案记录在案。在算法运营的事中阶段,由于人工智能系统面临的监管挑战部分来自缺乏留档文件以及文档标准。〔41〕为进一步强化问责体系,化解风险,需要建立算法日志记录制度,通过在算法运行过程中存留周期性日志,建立类似于金融领域的"审计线索",为监管部门提供监管线索,加强对运行阶段的算法责任落实。在算法决策完成的事后阶段,以算法解释制度作为问责的最后环节。若算法决策结果产生损害,则法律问责必然产生,但算法决策的自主性特点使问责流程必须给算法设计者、部署者以解释说明的机会。以证券交易中的因算法做出的虚假申报操作为例,授予金融监管机构针对存在异常交易的市场主体解释其算法交易程序的权利,并责令算法交易者限期修正算法参数瑕疵。〔42〕这不仅有利于问责流程的程序正当,也有利于监管部门对其他相类似算法的合规检查留下数据支撑与技术支持。当平台能够记录并诚实地重放导致特定决策结果的计算时,算法解释制度的作用才能真正发挥。〔43〕算法解释制度无论从监管成本、监管对象还是监管内容角度来看,都应成为算法问责制的独立环节。〔44〕除此之外,建立与前文证券"意见"中提及的证券纠纷代表人诉讼制度相类似的算法纠纷诉讼制度或算法纠纷集体诉讼制度,通过对算法领域相关司法问题的精细化诉讼制度划分,进一步保障因算法决策受损害主体的合法权益。

---

〔41〕 See Ben Hutchinson et al., *Towards Accountability for Machine Learning Datasets: Practices from Software Engineering and Infrastructure*, Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021.

〔42〕 参见夏中宝:《算法交易对虚假申报操纵法律认定逻辑的新挑战》,载《证券市场导报》2017 年第 10 期。

〔43〕 See Philip Adler et al., *Auditing Black – Box Models for Indirect Influence*, Knowledge and Information Systems, Vol. 55.

〔44〕 参见前注〔8〕,第 221 页。

四、结语

在“十四五规划”背景下,算法权力的规制在数字中国建设过程中成为越来越重要的议题,并成为推动数字化建设的重要驱动力。对于算法权力,我们不仅要深刻认识其技术本质,而且要时刻警醒算法权力展现出的异化风险,以发挥算法对于加快数字化发展、建设数字中国的推动作用。对于算法权力的规制,我国应充分汲取域外经验,将欧盟模式、美国与加拿大模式中的协作治理与框架治理理念与现有法律法规体系融合,构建以算法影响评估机制为核心的规制框架。必须指出,制度的有效性不仅取决于制度本身,还受制度实施环境的综合影响,因此,在未来还需要构建与之相配套的算法问责、数据评估值等衔接制度,从而达到对算法权力的一体化规制。

(编辑:谢贵春)