

算法影响评估制度的构建机理与中国方案

张 欣*

摘要:作为底层架构的算法,在扩展式嵌入社会经济领域时已逐步超越代码特质,整合并再生为新型“社会权力”。面对这一挑战,算法影响评估制度应运而生。其有助于创制合理的算法透明度并构筑算法信任,助力场景化和精准化算法治理,与数据保护影响评估制度形成有效勾连。美国和加拿大以公共事业为核心领域,以框架式治理和协同治理为实现路径,以“技术架构、影响维度和问责制”为核心支撑,以敏捷治理理念打造智能评估工具,形成了全周期覆盖、闭环联动的评估框架。在构建本土化方案时,应当科学制定算法影响评估制度的构建规划及基本路径,精细化制度设计并强化制度执行实效,合理构建内外兼具、激励相容的协同评估机制。

关键词:算法影响评估 算法治理 人工智能法律规制 算法问责制

一、问题的提出

伴随着数据和人工智能技术的迅猛发展,以商业资本或者政治权力为驱动力,算法已经逐步超越代码这一特质,内嵌在搜索引擎、新闻推荐、风险评估、人脸识别、自动驾驶等多元化公私场景中,成为构建社会秩序的代理决策者。^①作为底层架构的算法,虽然形式客观、运行高效,但其技术特质无法为决策的稳健性、正当性和合理性作出充分证成。实际上,一系列算法偏误已经在社会、经济领域带来不良影响。例如,脸书的算法在就业和住房广告精准投放时产生了性别倾斜和种族歧视。^②数百万非裔美国人因患者分配医疗算法的偏误设计在医疗护理上受到了系统性歧视。^③更令人震惊的是,2019年两架波音飞机因自动驾驶软件程序故障造成空难,致使346名

* 对外经济贸易大学法学院副教授

基金项目:国家社会科学基金项目(17CFX058)、对外经济贸易大学惠园优秀青年学者资助项目(19YQ13)

① See David Beer, *The Social Power of Algorithms*, 20 *Information, Communication and Society*, 4 (2017).

② See *Department of Housing and Urban Development v. Facebook, Inc., Charge of Discrimination*, https://www.hud.gov/sites/dfiles/Main/documents/HUD_v_Facebook.pdf, 2019-03-28.

③ See Heidi Ledford, *Millions of Black People Affected by Racial Bias in Health-Care Algorithms*, 574 *Nature*, 608 (2019).

乘客不幸罹难。^①由此可见,当算法与公私权力主体相结合,整合并再生成为新型“社会权力”时,^②各国立法者亟须以平等、公平和安全为内核构建科学合理的治理框架,弥合算法系统与社会核心价值之间的潜在张力。^③但由于一个简单的算法应用场景至少涉及算法开发者、部署者、使用者等多个主体和多项法益,不同的算法运行场景也意味着不尽一致的算法治理要求,因此,还需建立一种具有动态性、精准性、场景性的算法治理机制,打破单一治理框架与多元应用场景之间、技术复杂性和动态性与治理规范的稳定性和普遍性之间的紧张内核。

面对日益攀升的算法治理压力,算法影响评估制度作为重要的治理方案被提上中外立法议程,成为当下算法治理实践中最为瞩目的焦点。2018年美国纽约市通过颁布《算法问责法》首创自动化决策系统的影响评估制度。^④2019年加拿大政府颁布《自动化决策指令》系统化创建算法影响评估指标。^⑤2020年欧盟《人工智能白皮书》也不约而同地提出应针对人工智能应用建立清晰、易懂且兼顾各方利益的影响评估标准。^⑥在我国,从人工智能治理的顶层设计^⑦到地方人工智能行动方案,算法影响评估制度已作为一项核心制度被明确提出。例如,新近颁布的《关于建设人工智能上海高地 构建一流创新生态的行动方案(2019—2021)》就明确提出建立人工智能风险评估和法治监管体系。^⑧由此可见,算法影响评估制度不仅已经成为各国立法者竞相关注的焦点,而且还成为人工智能未来监管框架的重要起点。但面对日益迫切的制度建构需求,我国学界对于算法影响评估制度的研究却很薄弱。^⑨本文拟聚焦算法影响评估制度,系统梳理其制度优势,对比挖掘域外方案,提炼制度构建的核心机理,为探索建立中国方案奠定有益的理论基础。

① See Muhammad Ali, et al., Discrimination Through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes, 3 Proceedings of the ACM on Human-Computer Interaction, 1(2019).

② 参见胡凌:《超越代码:从赛博空间到物理世界的控制/生产机制》,《华东政法大学学报》2018年第1期。

③ See Ari Ezra Waldman, Power, Process and Automated Decision-Making, 88 Fordham Law Review, 2-4 (2019).

④ Zoë Bernard, The First Bill to Examine Algorithmic Bias in Government Agencies Has Just Passed in New York City, Business Insider, <https://www.businessinsider.com/algorithmic-bias-accountability-bill-passes-in-new-york-city-2017-12>, 2019-06-06.

⑤ See The Government of Canada, Directive on Automated Decision-Making, <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592>, 2020-03-01.

⑥ See European Commission, The White Paper: On Artificial Intelligence: A European Approach to Excellence and Trust, https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en, p.17, 2020-03-01.

⑦ 参见习近平:《推动我国新一代人工智能健康发展》, http://www.xinhuanet.com/politics/leaders/2018-10/31/c_1123643321.htm, 2020-02-10.

⑧ 参见上海市经济和信息化委员会:《上海市经济和信息化委员会关于印发〈关于建设人工智能上海高地 构建一流创新生态的行动方案(2019—2021年)〉的通知》, <http://www.shanghai.gov.cn/nw2/nw2314/nw2319/nw12344/u26aw62725.html>, 2020-06-02.

⑨ 目前,我国尚无对算法影响评估制度进行专门研究的成果,但有些与该制度研究相关的文章。参见沈伟伟:《算法透明原则的迷思——算法规制理论的批判》,《环球法律评论》2019年第6期;张欣:《算法解释权与算法治理路径研究》,《中外法学》2019年第6期;周辉:《算法权力及其规制》,《法制与社会发展》2019年第5期;张凌寒:《算法权力的兴起、异化及法律规制》,《法商研究》2019年第4期;等等。

二、算法影响评估制度的治理优势

算法影响评估制度是指依据系统制定的衡量标准对自动化决策系统的应用流程、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。^①从技术运行机理和算法治理目标看,算法影响评估制度具有下列3项核心优势。

(一) 创制合理的算法透明度并构筑算法信任

在算法对社会不断重建的过程中,个性化和效率性的增加还伴随着多元风险,致使决策透明度和可控性下降,甚至形成“黑箱社会”。^②降低算法决策透明度的根源主要有3个方面:其一,传统的知识产权和商业秘密制度成了算法开发者和部署者对抗算法问责的有力武器;^③其二,算法开发者和设计者因受到商业利益的驱动而具有将算法模型不断复杂化、黑箱化的内在激励;^④其三,现有算法治理框架未能准确捕捉算法运行的内在机理,将算法决策的合法化审查转变为浮于形式的合规化管理。^⑤算法决策因此变得难以审查。

算法影响评估制度采用全周期视角,以中立、专业、可信的评估主体为保证,对算法设计、部署、运行的全部流程予以动态评估,要求算法系统在用于商业以及公共事业场景前就接受独立的社会技术分析。其犹如一把利剑,精准刺破了“黑箱社会”的神秘面纱,创制合理的算法透明度并构筑算法信任。首先,算法影响评估制度从程序和实质两个方面创建了合理的算法透明度,摆脱了算法设计者妄想藏匿于知识产权和商业秘密制度屏障后,从而试图逃避算法责任的治理难题。一方面,算法影响评估制度建立了程序性的评估机制,赋予利益相关主体以程序性保障和参与渠道,从而创建了算法决策运行过程中的信任沟通机制,增加了算法决策的适当性。另一方面,算法影响评估制度从实质层面预先识别和系统跟踪算法内置或者潜在偏误,增加算法运行系统的可追溯性、稳健性和准确性。其次,算法影响评估标准的建立可以促使研发主体以负责任的态度对系统进行开发,在系统设计之初就预估其可能产生的风险,在运行过程中也实时检测系统是否按照预期稳定运行,由此弱化算法开发者和设计者不断将设计复杂化、“黑箱化”的不良激励。最后,算法影响评估制度还将法律和行业规范与各项治理标准有机结合,形成细化可行的评估框架,解决算法问责制流于形式的治理难题,促使各科技公司强化现有控制流程,实现更高效率的风险管理,将算法影响评估制度内嵌于技术开发、流程管理和实际应用的全过程。通过多管齐下,算法影响评估制度既增加了算法决策在多元应用场景中的适当性和可接受性,又有助于营造

^① See Government of Canada, Algorithmic Impact Assessment, <https://canada-ca.github.io/digital-playbook-guide-numerique/views-vues/automated-decision-automatise/en/algorithmic-impact-assessment.html>, 2020-01-18.

^② See Frank Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information*, Harvard University Press, 2016, pp.3-14.

^③ See Frank Pasquale, *Secret Algorithms Threaten the Rule of Law*, MIT Technology Review, <https://www.technologyreview.com/2017/06/01/151447/secret-algorithms-threaten-the-rule-of-law/>, 2020-02-10.

^④ 参见张欣:《算法解释权与算法治理路径研究》,《中外法学》2019年第6期。

^⑤ 有外国学者对算法问责制被重新塑造为算法决策合规性进行了论述。See Ari Ezra Waldman, *Power, Process and Automated Decision-Making*, 88 *Fordham Law Review*, 16(2019).

可信赖的人工智能应用生态,推动“黑箱社会”转变为“显式社会”,从根本上构建算法信任。^①

(二)助力场景化和精准化的算法治理

人工智能时代的治理逻辑已经发生全方位的深度变革,单一框架的治理路径渐被抛弃,一种基于场景化、精细化的治理实践已经出现。例如,2020年欧盟《人工智能白皮书》就明确提出应当在应用场景、部署目的、安全保护、消费者利益和基本权利5个维度建立基于风险的五级监管体系,清晰区分各类人工智能应用的评估标准,实施差异化监管。^② 场景化、精准化的算法治理,意味着各算法应用所涉及的治理目标、治理工具和治理内容有所不同,与具体关涉到的主体、行为、规范、价值具有复杂的匹配关系;其是将技术应用场景中的各方利益和社会影响不断加以明晰的过程,是事实与规范的精准细化和科学组合。^③

正如有的学者所言:“治理活动就是一种场景化行动。一旦某种特定的场景被建构起来,支撑场景化行动的治理技术就有了自主性,成为治理场景的有机组成部分。”^④在算法治理领域,算法影响评估制度是场景化、精准化算法治理框架的核心实现路径,可以帮助开发者、设计者、部署者以及公众等利益相关方清晰理解算法设计与应用程序所应配备的治理和监管要求。算法影响评估制度还有助于在算法权利配置和责任承担方面建立依据具体场景而触发的精细化标准和规则,准确提炼算法应用场景中具有普遍性的治理需求,使复杂多元的算法应用既满足共性化的设计、部署和运行标准,又可以达致精准化的个性化适配。对于社会嵌入性强、预期风险较大的算法应用而言,算法影响评估制度还可以显著增加系统的可控性、透明度和稳健性。

(三)有效联结数据保护影响评估制度

在自动化决策系统运行过程中,算法与数据紧密相连,共同在场景应用中发挥作用。在设计和运行环节,若没有海量、可靠的数据喂养和支撑,则再好的算法设计也如空中楼阁;但仅有海量数据,如果缺乏具有高度解析能力的算法进行高效分析和挖掘,那么海量数据也可能被浪费,难以得到有效应用。因此,在网络治理领域,数据保护影响评估制度与算法影响评估制度相辅相成,互为补充,构成算法问责制的一体两翼。二者制度设计的内在机理存在诸多相同之处,如以项目而非组织为执行基础,具有事先预期性而非回溯性特质,具有动态性和持续性,评估系统和数据使用的影响和风险等,^⑤但二者的制度设计也存在诸多不同之处。这使得仅依靠数据保护影响评估制度难以解决自动化决策系统在运行过程中存在的歧视、偏误、不透明等治理难题。因此,就自动化决策系统而言,算法影响评估制度不仅可以有效联结数据保护影响评估制度,而且还能从以下两个方面弥补其不足。首先,有效弥补了数据保护影响评估制度中协同治理架构设

^① See Frank Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information*, Harvard University Press, 2016, p.189.

^② See European Commission, *White Paper: On Artificial Intelligence: A European Approach to Excellence and Trust*, https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en, 2020-03-01.

^③ 参见吕德文:《治理技术如何适配国家机器——技术治理的运用场景及其限度》,《探索与争鸣》2019年第6期。

^④ 吕德文:《治理技术如何适配国家机器——技术治理的运用场景及其限度》,《探索与争鸣》2019年第6期。

^⑤ See Margot Kaminski, et al., *Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations*, 19-28 University of Colorado Law School Legal Studies Research Paper Series, 22-23(2019).

计的不足。以《通用数据保护条例》为例,数据保护影响评估制度主要立足于数据控制者运行的内部层面。虽然《通用数据保护条例》第35(9)条规定:“在适当情况下控制者在不影响商业利益保护、公共利益保护或者处理操作安全的情况下,应当就将要进行的处理向数据主体或者代表寻求意见”,第36(1)条规定:“若根据本条例第35条的规定进行的数据保护影响评估表明,若控制者不采取措施降低风险,处理将带来高风险,数据控制者应当在处理数据之前与监管机构进行事先协商。但并未强制性地要求外部法律、社会或者数据安全领域的专家实施独立监督,也并未要求评估过程接受公共机构监督。这种制度设计虽然体现了对数据质量和数据安全的关注,但对协同治理实践的关注有所不足,对社会和经济影响的评估亦有所欠缺。^①与数据保护影响评估制度相比,算法影响评估制度不仅要求设计者、部署者和运行者对算法设计和相关数据展开系统评估,而且还要求在评估过程中纳入外部问责和审计力量,全面践行协同治理理念,相比较而言具有更为完善的治理架构。其次,有效解决改进数据保护影响评估制度中公众披露不足的问题。在数据保护影响评估制度中,虽然立法和相关指南推荐数据控制者对评估结果予以公布,但并不强求其公布评估文件和流程信息,可以仅公布核心概要。^②在数据和算法治理实践中,由于一定程度的公众参与和公开披露已被广泛视为构成有效治理的核心要素,^③因此,算法影响评估制度基于差异化场景建立合理透明度,尤其是对公共事业场景中高风险的自动化决策系统建立严格的公开披露制度,可以有效补齐数据保护影响评估制度中公众披露不足的短板。

三、算法影响评估制度的构建机理

在技术领域,算法影响评估是指对自动化决策系统的稳健性、公平性和可解释性等特性进行的评估。其目标是设定系统的运行边界和使用周期,创建系统设计者对运行结果负责的切入点。但在算法治理领域,由于算法架构、应用领域、治理框架和治理逻辑等不同,因此存在多个版本的制度设计和提案。本文拟选取美国和加拿大的算法影响评估制度作为代表性的立法样本,从应用领域、协调机制、评估标准和评估工具4个方面系统梳理其核心构建机理。

(一)以公共事业场景为核心应用领域

现代治理的复杂性和信息科技的广泛应用极大推动了现代“行政国”向“自动化行政国”的全面转化。^④从立法、执法到司法,从资源分配、战略部署到责任认定,一系列公共决策正由算法辅助乃至替代执行。与此伴随而来的是算法歧视、算法霸权等算法适用危机。这些危机共同指向

^① See Margot Kaminski, et al., Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations, 19-28 University of Colorado Law School Legal Studies Research Paper Series, 19 (2019).

^② See Margot Kaminski, et al., Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations, 19-28 University of Colorado Law School Legal Studies Research Paper Series, 21-24 (2019).

^③ 例如,美国电气和电子工程师协会自主和智能伦理全球倡议(The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems)领衔制定的一系列准则和倡议就反映了公众参与和透明度的要求。See IEEE, Ethically Aligned Design of Autonomous and Intelligent Technologies, <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>, 2020-02-26.

^④ See Danielle Keats Citron, Technological Due Process, 85 Washington University Law Review, 1252-1258 (2008).

3个治理难题:首先,公共事业领域的算法应用攸关公共利益。一旦发生决策失误,就可能带来系统性社会风险。其次,算法决策偏误引发的权益损失目前尚无高效、可行的配套救济措施。例如,公民在被算法系统错误标记后,很难通过高效的方式自证清白并迅速恢复信誉。相反,其可能反复遭受歧视或者错误待遇。最后,由于算法模型架构复杂,因此应用于公共事业场景时难以保障相关人知情、参与、异议和救济性权利,外部主体难以在传统制度框架下进行有效的算法问责。针对这些治理痛点,美国和加拿大两国率先行动,首先聚焦于公共事业场景中的算法影响评估制度,希冀以此提升行政领域的算法治理实践。

2018年10月,美国纽约市率先颁布《算法问责法》,揭开了美国在算法治理领域的立法篇章。该法明确规定行政机构以及慈善团体应用自动化决策系统时应当接受自动化决策系统工作组在公正性、可问责性和透明度等方面的评估。^①与纽约市的立法类似,2019年华盛顿州提出的S.B.5527和H.B.1655法案以及加利福尼亚州的S.B.444法案也明确规定政府和公共实体机构在公共事业场景应用自动化决策系统应当进行算法影响评估。^②2019年,加拿大政府颁布《自动化决策指令》,以透明、问责、合法、程序公正等核心行政法原则为指引,系统构建算法影响评估制度。^③该指令聚焦公共事业场景,主要适用于联邦行政机构以及为政府服务的外部主体利用自动化决策系统作出行政决定的情形。虽然美国和加拿大两国在微观层面的制度设计有所不同,但算法影响评估制度均率先在公共事业场景适用。其核心目的都在于通过系统评估和识别自动化决策系统在公共事业场景对个人和团体的潜在影响和风险,为公众和外部主体审查和稽核算法奠定基础,并为探索精准化、场景化的算法治理框架提供决策依据。

(二)以框架式治理和协同治理为实现路径

人工智能技术的创新具有迅速迭代性。算法作为人工智能技术链的源头更是如此。算法技术日益攀升的发展速率和不断复杂多元的应用场景使得立法者只能处于被动追赶的艰难境地。^④观察美国和加拿大立法者对于算法影响评估制度的构建策略可知,虽然微观规则设计各有特色,但均采用框架式治理和协同治理的思路,在充分考虑技术发展速率和应用复杂性的基础上,积极尝试建立一种动态开放、自我演化的算法治理机制。所谓框架式治理,是指在确定性的法律规则之外,以可度量、可标定、可操作、可区分的分级化指标、标准和清单为基础,构建起约束相关主体认知和行为框架的治理方式,可以为分级化、场景化、精确化的算法治理实践提供决策基础。算法治理实践中的协同治理则是指突破传统政府主导的线性管理模式,借助行业代表、外部专家和社会公众等多元力量,对算法应用予以解释、评估、监测和完善的治理策略。可以说,算法影响评估制度预示着人工智能治理实践开始跨越传统孤立式的单线治理模式,向以协同化、立体化、框架化、精准化为特质的多元治理模式的深度转变。

^① See New York City, Algorithmic Accountability Act, <https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0>, 2020-01-03.

^② See Yoon Chae, U.S. AI Regulation Guide: Legislative Overview and Practical Considerations, 3 The Journal of Robotics, Artificial Intelligence & Law, 23(2020).

^③ See Government of Canada, Directive on Automated Decision-Making, <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592>, 2020-02-01.

^④ 参见张欣:《数字经济时代公共话语格局变迁的新图景:平台驱动型参与的兴起、特征与机制》,《中国法律评论》2018年第2期。

在实施路径上,依据《算法问责法》,纽约市成立了自动化决策工作组,由市长办公室主任主持,同时由市长数据分析办公室首席分析官和人权委员会战略计划专员联席负责。工作组的成员由各相关政府机构和办公室、企业和行业、非营利组织、算法和数据领域的专家等共计17名成员组成。^①为增加规则的可操作性,工作组自成立以来就制定了认定某一工具或者系统是否属于行政机构自动化决策系统的自检清单,以清晰界定应当适用的系统类型。清单和标准相结合的方式不仅有助于规则的持续更新,而且在技术上可度量、可标定,在具体适用层面可观测、可区分。在具体表现形式上,该清单以详细问卷的方式对机器学习、数据处理、人工智能技术和公众影响等术语作以引导式规定。参评主体可以通过依次回答问卷所列问题从而得出是否应当适用该法的结论。^②美国联邦《算法问责法案》则采取自我评估和政府评估双轨并行的路径。依据该法案的规定,联邦贸易委员会作为主管行政机构颁布和制定评估标准。对适用该法的实体而言,应当按照法案规定的评估标准展开影响评估。对属于法案中规定的“高风险自动化决策系统”,由联邦贸易委员会在认为必要时与独立的稽核人员、技术专家咨询合作展开评估。^③加拿大《自动化决策指令》同样采用清单式和问卷式评估框架,通过设定60个与自动化决策系统业务流程、数据和设计相关的问题,以计分卡的形式得出初始评估结论并设定细化的协同评估方案。依据评估后对社会和公民的影响程度,自动化决策系统会受到来自国家研究委员会、统计局或者通讯安全机构的专家、高等教育机构的研究人员、具有相关专业知识的第三方合同供应商、国库委员会秘书处数据和自动化咨询委员会以及同行评审的综合性评估。为确保评估规定符合最新算法实践,该指令采用动态治理思路,每隔6个月就会对评估标准和治理框架进行必要更新。^④由此可见,在算法影响评估机制的构建环节,明晰的概念界定和量化可行的评估标准是保证治理实效的重要起点,外部专家和公众力量的多元化实质参与是确保其动态更新、自我演化的关键路径。框架式和协同式治理机制的构建,深度契合人工智能技术的运行机理,有助于科学有效地开展算法影响评估并精准辨识应予适用的监管框架。

(三)以技术架构、影响维度和问责制构建差异化评估框架

以美国和加拿大为代表的算法影响评估方案虽各具特色、各有侧重,但在构建逻辑上不约而同地反映了推动技术创新以及促进技术应用公平、公正、问责和透明的治理共识。通过详细对比以美国联邦《算法问责法案》为代表的5部美国立法和加拿大《自动化决策指令》,^⑤可以发现立

^① See NYC Automated Decision Systems Task Force, <https://www1.nyc.gov/site/adstaskforce/members/members.page>, 2020-02-01.

^② See NYC Automated Decision Systems Task Force, Checklist for Determining Whether a Tool or System is an ADS/Agency ADS as Defined by Local Law 49 (2018), <https://www1.nyc.gov/site/adstaskforce/about/about-ads.page>, 2019-12-25.

^③ See S.1108, Algorithmic Accountability Act of 2019, (1)(2), <https://www.congress.gov/bill/116th-congress/senate-bill/1108/text>, 2019-11-30.

^④ See Government of Canada, Directive on Automated Decision-Making, <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592>, 2020-03-01.

^⑤ 美国在联邦和州层面主要有5部与算法影响评估相关的立法:2018年纽约市《算法问责法》,加利福尼亚州S.B.444法案、华盛顿州S.B. 5527和H.B.1655法案,以及2019年新泽西州《算法问责法》,美国联邦《算法问责法案》。See Yoon Chae, US AI Regulation Guide: Legislative Overview and Practical Considerations, 3 The Journal of Robotics, Artificial Intelligence & Law, 23(2020).

法者以“技术架构、影响维度和问责制”为核心支撑,构成三位一体、互有补充的闭环评估框架,聚焦差异化、场景化和精准化的算法影响评估实践。

首先,以技术架构为基础维度,采用事前、事中、事后的全周期视角设定技术评估指标体系。通过比较发现,以技术架构为评估维度,立法者多以技术设计效率性和应用目的合法性、技术架构复杂性、系统执行任务类型、数据使用特性、技术所有权、技术运行监测机制、系统可解释性、系统精确度、技术性和物力性保障措施等内容构建核心评估指标。例如,纽约市自动化决策工作组的官方指南指出,应当依据系统执行的任务类型、核心技术组件复杂性、对个人可识别信息的使用程度、系统解释能力、部署系统紧迫性、系统预期效率等核心指标对技术架构特性进行评估。^①与这一立法思路类似,美国联邦《算法问责法案》亦从技术架构切入,明确提出“自动化终局决策系统”与“自动化辅助决策系统”的区分,^②并结合技术新颖性、决策应用范围、背景和用途,以及所涉数据和个人信息的敏感程度建立技术维度评估指标。^③

其次,采用场景治理和精准治理的思路,以影响主体、影响范围、影响程度为基础,建立基于影响维度的评估指标体系。在影响维度指标的设定过程中,各国立法者还嵌入了算法治理实践所要维护的社会核心价值。以加拿大《自动化决策指令》为例,由于立法者十分关注自动化决策系统在经济、社会和生态等方面的影响,因此以系统对个人或者团体在基本权利、健康福祉、经济利益以及环境生态方面的影响建立评估指标。为进一步精准匹配影响等级和监管义务,立法者以风险为基础将评估后的影响等级划分为可逆和短暂性影响、可逆和短期性影响、难以逆转且持续不断性影响以及不可逆转且永久性影响4个等级。^④受媒体报道等公共议程的影响,美国立法更多聚焦于自动化决策系统对公平、偏见、歧视、隐私和安全等核心价值的影响。例如,美国联邦《算法问责法案》中自动化决策系统影响评估条款(C)项就规定应当评估系统对消费者个人信息和隐私以及安全构成的风险,同时应评估自动化决策系统可能引致消费者遭遇不公、偏见和歧视性决策的风险。^⑤纽约市则以自动化系统潜在影响的性质和程度、潜在影响的持续时间、任何类型的已知或者潜在风险,包括可能带来的差别性影响或者偏见、对网络安全、隐私或者个人对关键信息的受限访问等为基准建立影响维度的评估指标体系。^⑥

最后,围绕透明度、合法性、公平和公正等核心价值建立基于算法问责制维度的评估指标体系。具体而言,算法问责制维度的评估指标可被进一步细分为防治不良影响的技术措施、权利救济渠道、算法透明度实现机制3个层面。以2019年美国联邦《算法问责法案》第2条为例,第(b)

^① See Automated Decision Systems Task Force, New York City Automated Decision Systems Task Force Report, 17-20(2019).

^② 前者指一项自动化系统在没有人工干预的情形下做出终局性决策、判断或者论断,后者指一项自动化决策系统服务于人类决策,为人类最终决策、判断或者论断提供所需信息。

^③ See S.1108, Algorithmic Accountability Act of 2019, <https://www.congress.gov/bill/116th-congress/senate-bill/1108/text>, 2020-02-03.

^④ See Government of Canada, Directive on Automated Decision-Making, <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592>, 2020-03-01.

^⑤ See S.1108, Algorithmic Accountability Act of 2019, <https://www.congress.gov/bill/116th-congress/senate-bill/1108/text>, 2020-02-03.

^⑥ See Automated Decision Systems Task Force, New York City Automated Decision Systems Task Force Report, 18-21(2019).

(iii)和(iv)项规定应当评估消费者获得自动化决策系统相关信息以及查阅运行结果并对其进行更正或者反对的难易程度;该条(d)款规定:“应当评估所涉主体为减少自动化决策系统对消费者隐私或者安全风险以及可能导致的偏误和歧视风险而采取的技术性和物力性保障措施”。^①由此可见,立法者是遵循技术防治、权利救济的思路来具体设定问责维度的评估指标体系的。加拿大《自动化决策指令》设定的评估指标,重点强调数据质量问题检测流程、系统设计、开发、维护和改进的归责机制和措施以及规范检测数据偏差等技术防治措施。^②加拿大《自动化决策指令》从评估决定的通知和公布、决策后提供解释、访问组件、政府对所有已发布专有组件审计、调查、检查保留权、源代码发布等算法透明度措施加以规定。

综上所述,美国和加拿大的立法者以技术架构、影响维度和算法问责制为核心,构建了三位一体、互有补充、同向发力的算法影响评估框架。虽然具体指标各有侧重,但均以建立差异化、场景化、精准化的算法治理框架为建构宗旨。因此,纽约市自动化决策工作组提出应尽快制定优先性指南,提出关键概念和原则,对最重要的工具和系统按照优先级展开评估。^③加拿大的立法更为系统和精确,依据影响评估后获得的4个风险等级,从同行评审、通知、人际回圈决策、解释要求、测试要求、检测、训练、应急计划和系统运行批准共9个方面,形成了具有精确性、协同性和系统性特征的治理框架,为场景化、精准化算法治理愿景提供了有益的范本。

(四)以敏捷治理思路打造智能评估工具

面对监管资源稀缺、监管对象复杂的挑战,监管者亟须开发和尝试新的思路,合理设定监管阈值,在科技创新与风险控制之间寻求良好的平衡。^④因此,以创建具有柔韧性、流动性、灵活性和适应性监管框架为宗旨,以实现监管目标多元平衡、监管过程动态优化以及监管工具灵活转换的敏捷治理理念应运而生。^⑤在这一理念的引领之下,各国算法影响评估制度重点对监管工具进行了智能化革新,以数字化、实时性、可扩展性和智能化为目标,运用技术手段对评估工具进行深度的改进。^⑥

以加拿大《自动化决策指令》为例,立法者运用动态编程语言创建开源性算法影响评估模型,以系统类别和目的、决策效用类型、决策解释性、系统潜在影响检测机制、公众回应和投诉机制等类别生成评估所需要的系列问题,依据该指令附录C的评估要求按照每类问题的重要性精确设定权重。^⑦例如,评估模型将自动化决策系统对社会、经济和政府影响的权重设定得更高,大于

^① S.1108, Algorithmic Accountability Act of 2019, <https://www.congress.gov/bill/116th-congress/senate-bill/1108/text>, 2020-02-03.

^② See Mathieu Leymay, Understanding Canada's Algorithmic Impact Assessment Tool, <https://towardsdatascience.com/understanding-canadas-algorithmic-impact-assessment-tool-cd0d3e8cafab>, 2020-02-26.

^③ See Automated Decision Systems Task Force, New York City Automated Decision Systems Task Force Report, 26-27(2019).

^④ See Michael Guihot, et al., Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence, 20 Vanderbilt Journal of Entertainment and Technology Law, 420-426(2017).

^⑤ 参见薛澜、赵静:《走向敏捷治理:新兴产业发展与监管模式探究》,《中国行政管理》2019年第8期。

^⑥ 参见张家林:《监管科技发展及应用研究——以智能投顾监管为例》,《金融监管研究》2018年第6期。

^⑦ See Supergovernance, The Government of Canada's Algorithmic Impact Assessment: Take Two, <https://medium.com/@supergovernance/the-government-of-canadas-algorithmic-impact-assessment-take-two-8a22a87acf6f>, 2020-03-17.

其对数据管理实践和程序公平性影响的权重。在每个类别项下,每一具体问题又相应配比风险系数,依据不同的答案选项单独设定评分编码,据此生成5个类别的加总得分。^①参与者仅需花费35分钟即可利用该评估模型完成一项评估,得出参评系统影响风险所在等级的精确指引,既便于监管者和设计者追踪风险产生的具体环节,又可依据影响等级精准化地确定应履行的程序性义务。当自动化决策系统的功能或者应用场景发生改变时,评估模型会依据新的因素及时更新。

通过对核心应用领域、评估实现路径、指标设计逻辑和评估运行工具的深入分析,可以发现不同国家的制度设计方案虽各具特色,但仍然反映了人工智能治理框架和规范构建的核心共识和发展规律。对这些有益实践的科学借鉴,既有助于我国算法治理体系的建立和完善,又为探索更为人性化、场景化、精准化和敏捷化的人工智能治理框架奠定了重要的基础。

四、中国算法影响评估制度的构建方案

全球人工智能发展已经迈入新阶段。为促进新一代人工智能健康发展,2019年6月,我国国家新一代人工智能治理专业委员会发布《新一代人工智能治理原则——发展负责任的人工智能》(以下简称《治理原则》),提出了我国人工智能的治理框架和行动指南。《治理原则》不仅强调了人工智能的可问责性,而且还提出了公平公正、尊重隐私、安全可控、开放协作、敏捷治理等8项重要原则。^②这些原则既体现了在新一轮科技革命中我国作为人工智能发展大国的责任与担当,充分展现了我国科技创新政策体系的建设能力,又为实现创新型国家和世界科技强国愿景提供了坚实的制度基础。目前,人工智能全球治理已经在规则制定层面展开博弈,以美国、加拿大和英国为代表,多国立法者已经在算法影响评估领域率先布局,旨在占据国际标准和治理规则制定议程中的优势地位。2019年8月,美国商务部国家标准和技术研究所发布了美国联邦参与制定技术标准的详细计划,确定了9个人工智能技术标准领域,其中对人工智能安全、风险管理以及可解释性和安全性方面施以特别关注;该计划还特别提到3个“非技术性”的人工智能标准,包括社会和道德相关、治理相关以及隐私相关的标准。^③2020年欧盟《人工智能白皮书》明确提出欧洲尚缺乏共同的人工智能监管框架,应当尽快建立基于风险和场景的应用风险等级评估,构建卓越、可信的人工智能体系框架,重获技术主权优势地位。^④由此可见,以算法影响评估制度为代表的人工智能领域立法具有国内和国际层面的现实紧迫性,我国应予以重点关注。作为人工

^① See Supergovernance, The Government of Canada's Algorithmic Impact Assessment: Take Two, <https://medium.com/@supergovernance/the-government-of-canadas-algorithmic-impact-assessment-take-two-8a22a87acf6f>, 2020-03-17.

^② 参见钟源:《我国新一代人工智能治理原则发布》,《经济参考报》2019年6月18日。

^③ See National Institute of Standards and Technology, U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools (2019), <https://www.nist.gov/news-events/news/2019/08/plan-outlines-priorities-federal-agency-engagement-ai-standards-development>, 2020-03-03.

^④ See European Commission, White Paper: On Artificial Intelligence: A European Approach to Excellence and Trust, https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en, 2020-03-01.

智能的底层技术,算法影响评估制度已经成为人工智能治理议程的源头节点。鉴此,在厘清算法影响评估制度的核心优势以及对域外算法影响评估制度进行审视和剖析的基础上,本文结合我国人工智能产业发展和规制现状、算法技术运行逻辑、制度设计内在机理和协同评估机制,提出构建中国算法影响评估制度方案的3项建议。

(一)科学制定算法影响评估制度的构建规划和基本路径

算法影响评估制度的现有立法样本均体现了阶段性、场景化、精准化、类型化的创制思路。算法影响评估制度作为人工智能立法的先行领域,虽然具有较强的现实紧迫性,但立法者和监管者应审慎冷静,从算法应用现状、核心治理目标出发综合确定优先适用的核心领域和关键系统类型,科学合理地确定构建规划和基本路径。

首先,应当根据我国自动化决策系统应用现状,以公共事业场景为先导,结合风险等级确定优先适用的核心领域。就行业发展规律而言,我国算法技术应用的价值链布局侧重于技术层和应用层,比较偏好技术相对成熟、应用场景清晰的领域。^①具体到公共事业领域,对社会治理创新和治理现代化目标的切实推进促使政府端成为自动化决策系统切入智慧政务和公共安全应用场景的主要平台。例如,深圳市宝安区利用覆盖全区的公共信用信息平台集合所有法人、自然人的信用档案形成信用画像,同步实现自动推送信用信息、自动识别黑红名单、自动提示奖惩依据、自动实施限制处理、自动反馈处理结果等信用联合奖惩智能模式。^②除此以外,算法技术在智能安防、金融风控、城建监管、公共卫生防治、警务预测和司法审判等先导领域也已经得到广泛适用。^③由于在这些领域应用自动化决策系统,既可能会对经济、社会、生态产生重大影响,也可能对个人和特定团体的权利产生实质影响,因此,处于算法应用链条的高风险一端,具有优先评估的现实紧迫性。在进行具体的制度设计时,建议以“自上而下”的思路,系统评估算法应用在我国公私领域的风险特征,以公共事业场景为基础先导,结合风险等级确定优先适用的关键领域。

其次,根据核心目标、技术架构和影响维度并结合已识别的关键领域来确定优先和重点评估的关键系统类型。依据《治理原则》,结合人工智能治理的全球共识,我国应当围绕公平公正、尊重隐私、安全可控、可信负责建立核心目标,探索我国的算法影响评估标准。在以公共事业为先导的布局之下,结合技术架构和影响维度来识别高风险应用场景和系统类型并予以优先评估。具体而言,应当以架构复杂性、部署目的、任务类型、影响主体、影响范围、影响程度、所涉数据和个人信息的敏感程度为基础,探索建立识别高风险自动化决策系统的评估指标体系。仍以社会信用治理实践为例,不同治理环节所涉及的模型系统具有风险差异性:在信用画像环节,可能涉及预测算法、机器学习、推荐算法、相似度计算、分类和聚类计算以及文本挖掘算法等多种算法技术和模型,实现决策执行、决策辅助甚至代理决策等功能;而在信用信息推送、黑红名单识别、奖惩依据自动提示、自动反馈处理结果等执行环节,则普遍应用语义分析、协同过滤等模型,多体现

^① 参见德勤:《中国人工智能产业白皮书(2018)》, <https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/innovation/deloitte-cn-innovation-ai-whitepaper-zz-181126.pdf>, 2020-06-07。

^② 参见印朋、李晓玲:《“政府+市场”推动深圳社会信用体系走上快车道》, http://www.xinhuanet.com/2019-08/29/c_1124937116.htm, 2019-08-29。

^③ 参见《国务院关于印发促进大数据发展行动纲要的通知》, http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm, 2020-05-30。

为纯粹的决策执行功能。后者与前者相比,对经济、社会和个人权利的影响程度有所不同。因此,信用画像环节的自动化决策应用应当被进一步清晰识别为关键领域从而进行优先和重点评估,适用更严格的风险监管要求。除此以外,系统部署的紧迫性、系统技术复杂程度、数据使用特性等重要指标均应容纳在内,作为共同识别关键领域和优先重点评估系统类型的衡量标准。总体而言,除“自上而下”的思路之外,还应“自下而上”地以核心价值目标、技术架构和影响维度为基准,结合已识别的关键领域动态确定优先和重点评估的关键系统类型,科学匹配评估指标和风险管控等级。

(二)精细化制度设计并强化制度执行实效

人工智能时代的技术建模和运行模式以精细化、颗粒度为未来迭代方向。受益于数据深度挖掘技术的日益成熟,人工智能应用更加注重投入产出比,更加注重精准预测和精确分析,具有架构复杂、场景多元、流程繁复等特质。这就要求算法治理领域的规则设计更加精细化,治理秩序的建构更具实效性。^① 因此,有学者提出“个性化法律”理念,以充分考虑规制对象的异质性、规则对场景的反馈性以及规则制定的精细性。^② 具体到我国算法影响评估领域的规则构建,重点应考虑以下两个方面。

首先,应以精准化治理为目标推动算法影响评估制度的精细化设计。算法影响评估制度旨在通过建立科学合理的评估指标对多元场景的算法应用予以系统评估,为实行差异化、精准化监管奠定规则基础。要实现这一目标,尤为需要践行科学化、精细化的制度建构思路,避免因缺乏必要精度或者量化标准而使评估流于形式。目前各个版本的算法影响评估制度不同程度地面临着精细化制度设计的挑战。例如,算法影响评估制度需要对“风险”“重大利益关系”“责任”等关键概念进行微观制度设计。但目前域外立法对于这些关键概念仍然缺乏量化标准和必要精度,这导致评估量表中的诸多关键问题缺乏可操作性,进而影响算法问责制的有效实施。因此,我国立法者应秉持结构细分、精度量化的设计思路,参考 ISO31000 风险管理标准等构建策略,对关键概念进行必要的量化和精准设计。^③

其次,应探索建立稳定的执行保障机制以确保制度执行实效。如前所述,算法影响评估制度服务于人工智能治理创新的关键需求,是算法问责制有效建立的重要基础。但美国纽约市的算法治理实践表明,因缺乏有力的执行保障机制,政府仅通过建立统一的评估框架来对算法应用实践予以引导,此种规制策略的效果不容乐观。例如,在美国纽约市的算法影响评估实践中,由于纽约市自动化决策工作组不具有明确的执法权,法条中也未明晰拒绝执行评估制度的主体应当承担何种法律责任,因此造成制度难以长久产生实效的困境。^④ 以此为鉴,美国联邦的立法通过明确授予美国联邦贸易委员会根据《美国联邦贸易委员会法》第 5 节关于欺骗性和不公平行为的

^① 参见马长山:《智能互联网时代的法律变革》,《法学研究》2018年第4期。

^② See Christoph Busch, Implementing Personalized Law: Personalized Disclosures in Consumer Law and Data Privacy Law, 86 The University of Chicago Law Review, 313-314(2019).

^③ See ISO 31000 Risk Management, <https://www.iso.org/iso-31000-risk-management.html>, 2020-06-01.

^④ See Colin Lecher, NYC's Algorithm Task Force Was "A Waste", <https://www.theverge.com/2019/11/20/20974379/nyc-algorithm-task-force-report-de-blasio>, 2020-06-01.

规定以及授予其通过系统所在州总检察长提起民事诉讼的权力,较好地回应了实效不足的挑战。^① 加拿大《自动化决策指令》第7条也明确规定,如果不遵守该指令,那么可由财政委员会依据《金融管理法》采取适当和可接受的任何措施。与此同时,依据《合规管理框架》,相应的机构和个人应当承担相应的法律后果。^② 借鉴这些做法,我国立法者应当在规则设计时重视执行保障机制的设立,细化未能有效履行评估义务的主体责任,科学借助算法公开制度赋予公众的知情权和监督权,适时设定专门机构或者行政职位,为建立稳定的执行保障机制奠定组织基础。

(三)合理构建内外兼具且激励相容的协同评估机制

如前所述,美国纽约市是建立算法影响评估制度的先行者。推动其立法的一项核心动因在于非营利组织“为了人民”(ProPublica)于2016年5月发布的“机器偏差”调查报告。该报告重在评估和审查,对美国司法领域被告再犯风险预测和评估算法的潜在偏误和种族歧视问题予以系统揭露。^③ 在“为了人民”组织(ProPublica)之外,依托纽约大学运行的人工智能现代研究所作为跨学科机构也对纽约市算法影响评估立法议程和实施机制进行了有力的推动。^④ 因此,纽约市的算法影响评估制度呈现出政府主导、行业参与、团体推动的协同评估特质。这种内外兼具的多元化协同机制既为公私场景中的算法影响评估实践提供了制度灵活性和延展性,又从内在机理上有效贴合算法设计者和运行者的激励机制,有助于知识产权、商业秘密和算法评估制度的兼容与调和。

在我国,《中华人民共和国数据安全法(草案)》第9条明确规定:“国家建立健全数据安全协同治理体系,推动有关部门、行业组织、企业和个人等共同参与数据安全保护工作,形成全社会共同维护数据安全和促进发展的良好环境”。《中华人民共和国个人信息保护法(草案)》[以下简称《个人信息保护法(草案)》]第11条亦明确规定:“国家……推动形成政府、企业、相关行业组织、社会公众共同参与个人信息保护的良好环境。”由此可见,无论是在数据安全领域,还是在个人信息保护领域,探索内外兼具、激励相容的协同治理机制已成为我国在数据、算法等科技领域打造新型治理机制的重要抓手。但在实践层面,这一理念尚未得到充分重视。例如,伴随着2015年国家创建社会信用体系建设示范城市工作的脚步,杭州、苏州、荣成等10多地政府都已推出体现地方信用治理特色的社会信用评分项目。以芝麻信用、腾讯征信、京东征信等为代表的第三方信用评估机构提供的信用分以及以联通和移动构成的电信运营商信用分渐次展开。但是,无论是城市个人信用分,还是第三方机构的商业信用分,在制定、异议、救济和评估环节均未向公众提供参与渠道,第三方非营利组织的外部监督和有效评估更为鲜见。例如,《福州市个人信用积分管理暂行办法》第13条规定:“市信用办应定期对个人信用积分评价规则进行修正、补充与完善,并

^① See S.1108, Algorithmic Accountability Act of 2019, <https://www.congress.gov/bill/116th-congress/senate-bill/1108/text>, 2020-02-03.

^② 例如,在更严重的后果中,可以要求相应机构采取补救措施,对相应的资金加以限制并由内阁直接讨论;在最严重的后果中,可以要求冻结资金,限制一定价值的交易,约束相应的权限,对机构进行推荐性重组,甚至将机构置于被接管状态。参见加拿大《自动化决策指令》第7.2条。

^③ See Lauren Kirchner, New York City Moves to Create Accountability for Algorithms, <https://www.propublica.org/article/new-york-city-moves-to-create-accountability-for-algorithms>, 2020-11-12.

^④ See AI Now Institute, Algorithm Impact Assessments Report, <https://ainowinstitute.org/announcements/algorithmic-impact-assessments-report.html>, 2020-11-12.

向社会公布”。但这一规定仅对公布环节进行了规定,而在指标设定、权重计算、模型评估等设计环节却付之阙如,凸显出协同治理机制的缺失。

因此,以《个人信息保护法(草案)》的制定为契机,我国在立法中应当进一步细化对算法影响评估制度多元化协同机制的具体设计和实施机制。首先,应当以评估优先级为指引,建立政府主导的算法影响评估机制。具体而言,可以由国家互联网信息管理部门作为牵头机构组建集中化、一体化的评估机制,按照前文提出的评估优先级对高风险自动化系统开展定期评估。其次,由国家互联网信息管理部门按照职责权限推进算法影响评估服务体系建设,支持有关机构开展算法影响评估和认证服务。具体的开展模式可以借鉴德勤等开展算法审计业务专门机构的业务规则,支持政府和商业主体在分级分类基础上对算法应用开展不定期的内部和第三方影响评估。^①最后,参考和借鉴美国纽约市自动化决策工作组创建的多方协作机制,定期与高校和科研机构、社会公众以及非政府组织开展协作,探索构建开放型的多元化评估协同机制。

五、结 语

随着数字经济和智能经济的迅猛发展,我国经济和社会正经历着前所未有的深度变革。算法技术在商业和公共事业领域的嵌入式应用虽然蕴含着巨大的发展潜力,但是也对我国自工业社会以来形成的立法模式和治理模式发起不断冲击,引发了诸多治理痛点。与其他数字技术有所不同,作为底层架构的算法,以商业资本或者政治权力为驱动力,在扩展式嵌入社会经济等多元决策场景时逐渐展现出系统性和积累性决策效果。有学者曾因此提出应当通过一定的程序进入到自动化决策系统内部,了解系统的技术细节以及其对于社会秩序的相应潜力,并对其融入组织、程序和决策的过程进行系统问责。^② 算法影响评估制度以确保决策质量和安全为基本目标,采用全周期动态思维,系统评估算法应用可能带来的社会影响和多元风险,同时致力于探寻降低风险的应对措施,为算法问责制在微观规则层面提供有力抓手。通过设定核心评估体系,算法影响评估制度将基础法律原则融入技术设计和行业规范之中,在法律规制与技术创新之间动态构建良性平衡,以此构筑算法信任,助力场景化和精准化的算法治理,与数据保护影响评估制度形成有效勾连。本文以美国和加拿大算法影响评估制度为典型样本,深度剖析了制度构建的共识性机理,并从行业发展规律、构建路径、制度运行逻辑和新型协同机制等方面提出了构建本土化方案的核心策略。须指出的是,由于制度的有效性不仅取决于内生因素,而且还受到社会价值、具体情境以及更为广泛的制度环境的综合影响,因此未来还应当对算法影响评估制度的配套机制、与数据保护影响评估制度的衔接机制等核心议题进行深入研究,以期为算法影响评估制度的有效运行构建良好的运行生态。

责任编辑 王虹霞

^① See Deloitte Insights, Algorithm Auditor, https://www2.deloitte.com/content/dam/insights/us/articles/4767_FoW-in-govt/DI_Algorithm-auditor.pdf, 2020-11-15.

^② See Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, 2016, pp.1-8.