武漢大學學報(哲学社会科学版)
WUHAN UNIVERSITY JOURNAL (Philosophy & Social Science)
第74卷第5期 2021年9月
Vol. 74 No. 5 Sept. 2021 029~040
DOI:10.14086/j.cnki.wujss.2021.05.003

论人工智能算法的法律属性与治理进路

王德夫

摘 要 人工智能算法专注于解决“如何实现智能”这一特定的技术问题,是一种特殊的技术方案,相关的法律治理也围绕这一基本属性而展开。在人工智能算法设计和自我完善过程中所产生的新知识与相关利益,可以通过知识产权制度进行确认和分配;而对于如何平衡投资利益与社会公众利益,以及如何使法律监管突破黑盒障碍等问题,则应设计专门制度予以解决。通过创新性的“算法可理解+数据可信+参数可解释”治理架构,结合算法识别、数据可信以及算法可理解等基础性规则,可以突破算法解释、平台责任等现有治理手段的局限,确保技术理性与社会发展的协调与相互促进。

关键词 人工智能技术;算法治理;知识产权;算法可理解;数据要素市场;数据安全;数据可信

中图分类号 D913 **文献标识码** A **文章编号** 1672-7320(2021)05-0029-12

基金项目 中国法学会2019年度部级法学研究课题(CLS-2019-Y03);武汉大学“人工智能问题”融通研究专项课题(2020AI013)

随着现代信息技术的发展,科幻意味浓厚的机器人或是更能引发人们畅想的人工智能,逐渐从文艺作品和科研机构走进社会公众的日常生活。然而,人工智能的技术边界仍显模糊,其内在的技术对象仍显神秘,相关应用也处于发展变化之中,远未成熟[1](P3-9)。在可预见的未来,是否真的会出现某种足以达到人类智力水平的“强人工智能”[2](P56)甚至“超人工智能”,这一问题仍然充满了悬念[3](P417-457)。但是,对与人工智能相关的市场和社会风险进行识别与应对,已经是迫切的现实需求。当下,我国仅通过《电子商务法》中零散的规定,对各类智能产品、服务底层的算法予以规制:该法第40条规定了搜索引擎以及竞价排名算法的行为规则,第18条规定了与智能推送算法相关联的消费者权益保障义务。事实上,这种做法的应急色彩较浓而系统性不足,也难以适用于电子商务领域之外的人工智能应用活动。对此,在理论层面,相关的治理活动既要充分地应对人工智能算法与生俱来的技术底层性和不透明性,也要展现出足够的前瞻性。在现实层面,则应将人工智能算法治理从宏观讨论层面进一步扩展到治理框架与重要规则设计的层面,探寻符合相关技术条件的治理规则。

一、人工智能算法的基本范畴与法律表达

从概念上,算法并非一个新奇的事物,它与现代信息技术同步产生,距今已有超过半个世纪的历史。人工智能算法作为广义上计算机软件算法的一种特殊类型,也属于为了解决某个特定问题或者达到某个特定目的所要采取的一系列步骤[4](P398-415)。但是,作为相关领域最前沿的对象之一,人工智能算法与传统计算机软件算法之间的差异,才是其自身成为独立技术对象和法律对象的前提。

(一)人工智能算法与传统计算机软件算法的差异

现实中,无论是实现固定功能的传统计算机程序还是自主化运作的人工智能系统,都以各自的算法

作为运行的基本逻辑。正确、合理的步骤设计是相关系统最终实现设计功能的必备前提。但是,极具技术前沿性的人工智能算法与传统计算机软件算法间的差异性仍然十分明显。

一方面,人工智能算法专注于模拟基础智能,不因具体应用中的不同功能而变动,并以此区分于传统的应用软件算法。因此,在技术语境下,人工智能算法往往也被称为通用算法。这意味着,虽然同样名为算法,但人工智能算法与传统的计算机软件算法之间更多地呈现出一种总分关系:前者为基础性的"智能"部分,而后者则视不同的应用场景和方式具象化为不同的信息产品或者服务算法。这也喻示了二者在制度层面的差异:现行法律规范可以较好地延伸至各式各样的信息产品或者服务之中,以解决利用算法工具实施不当行为的问题,但作用于更底层的通用算法则因为识别和针对性方面的不足而显得力不从心。

另一方面,在运行逻辑上,人工智能算法采取了与静态的传统编程逻辑迥然不同的技术路径:它颠倒了传统的计算机程序运行顺序,系统输入的是数据和预期获得的结果,输出的则是另一个更新后的算法。旧算法到新算法的进化即为"机器学习",并与传统的"人力设计规则"相对应[5](P277-288)。人们尝试以不同的方式实践这种"认知可计算化"[6](P99-108),并通过相关系统的重复运行,实现人工智能算法的自主调整与更新。也就是说,人工智能算法的运行是以迭代后的新算法自动更新程序本身,获得学习与进化的能力,进而找寻通用的、足以令计算机具备人类智能的理想方案[7](P58-65)。这种自我编程、自我进化的能力,是人工智能脱胎于静态的计算机程序而展现智能的主要方式。

(二)人工智能算法的技术特征

其一,人工智能算法具备底层通用性:同一个人工智能系统可以运行于不同的场景之中,产生截然不同的技术与应用效果。这是真正意义上的人工智能与商业宣传中人工智能噱头的根本区别。在认知层面上,人工智能必须具备这种灵活应对不同问题的能力,突破摆脱特定应用领域的限制①。在法律层面上,通用算法的运行方式、应用领域和潜在的市场风险乃至伦理道德风险,也不同于虚拟财产或者其他抽象的智力劳动成果,是一种前所未有的新型对象,对法律制度的应对与完善提出了新的要求。

其二,人工智能算法的迭代与进化高度依赖外部的数据输入。这一技术特性会带来两方面的影响,值得引起相关研究的注意。一方面,人工智能的分析能力和决策能力以大数据技术为基础[8](P136-148),并且这是一个动态的过程,始终依赖数据的持续更新,不存在单机或者脱网运行的技术基础[9](P33)。另一方面,人工智能算法在数据规模和更新频率方面的需求,也极大地抬高了相关市场的进入门槛,为相关的市场行为监管、竞争状态分析乃至安全评估等活动带来了挑战。

第三,人工智能算法的输出具有不确定性。这种不确定性体现为人工智能算法单次运行所输出的结果并非确定的数值,而是一个新的决策方法或者过程。也即是说,人工智能算法运行的结果是另一个算法,而且应当是一个比当前运行版本更优化、可以解决更多问题或者更符合人们需求的新算法。这一过程与生物学的进化相类似,在现代信息技术的支撑下,可以在有限的时间内实施生物无法企及的庞大次数,使相对原始、简陋的初始算法在外部数据供给下,通过反复学习,进化到足以实现某种程度的智能的水平。这也意味着,人工智能算法始终是一个不断变化的中间状态,同一名目下不同版本的算法,却有可能在基本功能、关键性能以及潜在应用领域等方面具备截然不同的能力。

(三)人工智能算法的主要分类

必要的类型化分析有助于厘清相关治理活动的对象与边界。在人工智能算法技术因素之外,可以依照不同的标准,将其分为以下三种类别:

其一,按照人工智能技术本身的先进程度,或者相关应用领域的敏感程度,人工智能算法可以被划

① 如市场上常见的智能电饭煲、智能洗衣机等产品,虽被冠以"智能"之名,也在一定程度上具备比传统产品更加复杂的功能,但应用领域十分受限、运行流程单一,与真正意义上的人类智能相去甚远。

分为管制类算法和非管制类算法。对于管制类的人工智能算法，需要通过专门法予以规定，以明确受管制的对象范围和行为规则。在缺乏专门立法的情况下，我国对于管制类算法的识别是通过具体领域或者应用活动的其他规范性文件进行附带规定的。如2020年8月28日，我国商务部、科技部发布《中国禁止出口限制出口技术目录》，新增“基于数据分析的个性化信息推送服务技术”和“人工智能交互界面技术”两项，将智能推送算法纳入“维护国家安全基础上扩大对外技术交流”的范畴，从国家安全关切的角度对相关算法的对外技术交流活动实施管制。而对于非管制类的人工智能算法，相关的权利和义务承担则更多地遵循各专门法或者一般的法律原则，具有较大的自主性与灵活性。

其二，按照相关算法的技术细节是否向控制者之外的其他人公开，人工智能算法可以分为保密类与开放类算法。人工智能算法凝聚了相关研发人员聪明才智和相关机构的投资，算法质量也与其持续的经营状况密切相关，具备重要的经济和社会价值。因此，无论是出于技术原因还是相关主体的主观意愿，相关的算法往往难以表露于产品或者服务外部，也更难以被用户或者社会公众所接触，有着很高的保密性。但是，在当前市场环境下，部分人工智能算法也呈现出一种主动开放的趋势。从技术角度观察，这是因为机器学习对外部数据供给的依赖超过了对算法本身设计精妙程度的依赖。此时，开放算法非但不会导致其失去对自己技术秘密的控制，或者使自己的人工智能系统被他人夺走，反而会因为这种公开吸引到更多的二次开发和社会应用，支撑人工智能系统的不断迭代与进化。并且，在相关法律文件（如开源许可证等）的辅助下，开放算法换取数据回馈的做法更有助于相关主体将数据优势转化为算法优势，继而转化为产品或者服务的性能优势，最终获取市场竞争优势。这种新型的研发或者经营模式，也应当引起治理关注。

其三，按照相关算法的外在表达形式，人工智能算法可以被划分为作品类算法与技术方案类算法。在不同的应用场景下，算法本身可以表现为文字、图案、工程设计图、产品设计图、示意图、计算机软件等多种形式，既有可能直接来源于相关系统的自主运行，也有可能来源于设计、编程人员的创作活动，因此也有可能被纳入著作权保护客体范畴[①]，成为作品类算法。但是，从计算机程序设计、编程的角度看，人工智能算法更多的是一种解决“如何通过技术手段模拟人类智能”问题的技术措施，而非满足相关人员“如何表达内心世界”的愿望。这种技术方案类算法，更多地会与知识产权制度中的专利或者商业秘密产生关联。在确定治理目标与手段时，相关制度设计会也更多地倾向于确保技术信息的安全、有效与开放。

（四）人工智能算法的法律表达

综合人工智能算法的技术原理、特性与分类，笔者所讨论的人工智能算法聚焦于通用算法，即：处于人工智能技术底层的，兼具保密与开放性的非管制类技术方案，具有特定的技术功能和物理边界。其本质是一种技术信息，可以作为民、商事权利客体和公法视野下的监管对象。

在外部形式方面，人工智能算法具有相对固定的形态，可以实现法律上的识别与流转。虽然计算机相关产业中，算法的书写方式并无固定形式，但在设计、编程人员专业习惯的作用下，往往表现为计算机程序或者伪代码[②]，而非可以被计算机运行的程序。人工智能算法也不例外。对此，有两方面的理解：一方面，它包含有人工智能系统运行所需要的完整的技术信息和必要步骤，而不仅仅是抽象的思路或者想法。这意味着，它可以作为一个独立的对象存在于人工智能系统之中，可以被单独地表达出来或者被阅读，也可以被相关领域普通的技术人员所理解。它可以被用来解决特定的技术问题，具有使用和交换的价值。另一方面，它具有跨语言（编程语言）的通用性——不同应用领域、技术背景的使用者（编程人

① 由“机器自主学习”自动生成的新算法，具有作品的外在形式，却不是由自然人创作而来。对于此类对象是否属于著作权意义上的作品、是否应该纳入到著作权保护范畴，有诸多学者进行了热烈的讨论，各方观点碰撞激烈。但相关讨论与本文内容无直接联系，故不作详细论述。

② 一种非正式的程序设计语言，它可以比较容易地被改写为源代码。编程人员在设计算法阶段选择将“算法”直接写成计算机程序，十分耗费精力，也缺乏现实的必要性，故多采用“伪代码”。

员),只要具备基本的专业知识或者能力,即可以根据人工智能算法相关文件,将其与自己所期望实现的功能相结合,形成新的人工智能应用。这意味着,人工智能算法具有技术上的独立性和物理上的可分割性,可以在现实应用中被识别、使用和流转,是法律层面上合格的对象。这两方面的特性,使得将人工智能算法纳入法律治理的范畴,具备了现实的可操作性。

在内在实质方面,人工智能算法是一种特殊的技术方案,应当被纳入处理技术方案相关问题的专门制度体系中。由此也决定了,虽然人工智能算法与知识产权制度存在着天然的密切联系,但也不能简单地将其视作一种新型的知识产权客体。一方面,它表明,人工智能算法是智力劳动的成果,只应用于解决特定的技术问题,而与起到区分作用的标识以及文学艺术或者表达作用的作品相区别。与此同时,它专注于解决如何实现设计人员所理解的智能这一特定技术问题,不是抽象意义上的智力活动的规则和方法,也不被用于解决其他的技术问题。另一方面,强调人工智能算法作为一种技术方案的特殊性①,可以将其与人工智能程序、人工智能数据输入、人工智能数据存储、人工智能硬件等技术层面的对象间划定清晰的边界。这也使它与工程设计图、产品设计图以及计算机程序等技术类著作权客体相区别。

二、人工智能算法治理的基本规定

从治理的本义出发,人工智能算法治理至少应包含两方面内容,即通过制度对人工智能算法相关的各种利益进行确认,以及对相关对象、行为进行必要的定义与规制。

(一)人工智能算法治理的对象

人工智能算法治理的对象,应该是明确、清晰的人工智能算法,而非平台、某种具体应用或者某些特定信息之类的关联对象。当下,我国对人工智能的关注大多体现于相关的纲领性文件之中,更多的是一种系统性和战略性描述:既强调科技伦理,也注重从技术攻关和产业发展规划的角度描述相关基础理论、技术对象和应用场景。对于人工智能的治理问题,在相关学术研究中,既有基于风险视角[10](P128-136)的宏观思考,也有对算法治理路径的归纳与论证:以"个体赋权、外部问责和平台义务"[11](P17-30)的方式来确定人工智能算法背后的权利义务分配规则。这些思考和治理策略具有相当的前瞻性与科学性,但是,这些做法所指向的对象仍然略显抽象。而在立法实践中,更缺乏对人工智能关键性技术对象的准确描述和专门安排。对此,应把相关治理活动的对象明确地指向人工智能算法本身,才能符合技术发展要求和社会现实。

现实中,人工智能算法已经从产品或者服务的技术底层走向了前端,并且直接引发了社会问题和制度回应。这种趋势已经十分清晰:无论是我国颁布《中国禁止出口限制出口技术目录》对智能推送算法的直接管制,还是发生在美国密歇根州的、由当地居民对州政府反欺诈算法的综合数据化系统"米达思(MIDAS)"过高的出错率(93%)所提出的集团诉讼[12](P51-60),又或者是美国纽约市所颁布的对公用事业领域算法进行监管的"算法问责法"(Algorithmic Accountability Bill)[13],都直接地将智能算法带到了法律的视野中。虽然各国做法均源于特定国情,但人工智能法律治理由宏观、抽象逐步聚焦于核心技术对象的趋势却是一致的。将人工智能算法治理聚焦于算法本身,既是相关治理活动的应有之义,也是对技术现实的必要回应。

(二)人工智能算法治理所确认的利益

人工智能算法相关风险源于"黑盒化"等技术属性[14]对法律治理造成的障碍,以及更为深刻的技术理性与人的利益之间的天然矛盾:极致的技术追求代表着极致的理性,某种意义上,也会造成对人的利益的疏忽甚至损害。这种价值矛盾的产生,既有产业资本逐利性的因素,也有"技术理性"思维的推波助

① 特殊性表现于:其一,人工智能算法与硬件的分离,使其在现行专利审查体系下因为"不是利用自然规律的技术手段" 被排除于专利保护之外;其二,人工智能算法本身始终处于不断进化的过程中,其中间版本的设计与实施,主要依赖于人工智能系统的自主运行,而较少地牵涉人为干预。

澜。技术理性作为现代社会生产力发展的客观驱动，本身并无可责备之处。但如何确保技术发展最终能够造福于社会公众，则依赖于正确的价值引导。算法治理的首要任务，即需要对相关利益进行确认。

首先，要承认和保护人工智能控制者的投资利益。人工智能并不会单独、不受控制地游荡于现实世界或者虚拟的网络空间，而必然要归属于某些机构或者个人。这种归属关系来源于相关主体的人力、物力等多方面的资源投入。对于人工智能的控制者而言，其具有通过相关技术研发、市场应用获取投资收益的合理诉求，也享有在法律框架下利用自己所掌握的人工智能系统实施各项活动的自由。对应地，人工智能算法必然也要反映这些控制者的主观意愿和投资期望。因此，尊重相关的投资利益，并且以制度的方式将这种尊重转化为对权利的确认，是获取技术进步和产业发展的合理要求。对此，需要法律明确地尊重和保障相关投资利益以及人工智能控制者充分的自主活动和经营利益，以合理的制度设计达成鼓励创新、鼓励投资和预防技术风险的平衡。

其次，应保护相关技术研发者的知识利益。人工智能技术的产生和发展无法脱离自然人的聪明才智。即便在一定程度上实现了机器学习或者自主进化，人工智能也离不开持续的人类智力劳动投入。而在人工智能系统开发过程中，会产生大量的技术或者表达成果，与知识产权制度产生关联。因此，人工智能算法治理也应当尊重和保障研发者的知识利益。一方面，虽然与现行专利制度中的授权标准存在差异，但人工智能算法作为技术方案，仍然有获得专利制度保护的可能。无论是作为专利权人还是发明人、设计人，相关的人身利益和经济利益都应当得到保障。另一方面，人工智能算法在设计、书写、使用和传播过程中所涉及的种种文字、图案、计算机程序也有获得著作权制度保护的可能。而从治理的角度看，人工智能算法背后的知识创造和知识利益分配，也隐含着知识保密与公开、信息独占与开放的深层次利益博弈，同样需要通过知识产权制度予以协调和确认。

第三，应保护社会个体的人格与经济利益。人工智能算法作为现代信息产品或者服务的后台，往往并不能被用户所直接感知。这种社会公众主观感知上的缺失或者隐蔽性，并不影响他们或主动或被动地成为人工智能系统用户，也不否认其“免受人工智能技术负面影响”或者“参与人工智能技术发展并受益”等利益诉求的合理性。现实中，由于人工智能技术的快速发展和法律制度的相对滞后，利用人工智能算法侵害用户利益或者掩盖非法行为的活动愈发猖獗，典型地表现为以庞大的信息化系统个性化地侵害社会个体利益，利用定价算法实施大数据杀熟[15](P111-119)，利用算法偏见实施种族、性别、年龄歧视等屡见不鲜。而一些相对传统的违法行为，也在积极地利用智能技术掩盖非法目的。比如：具有市场支配地位的经营者通过“算法共谋”实施的垄断行为[16](P112-121)以及广泛数据搜集、利用活动中的侵犯个人信息行为等。对应地，突破技术遮掩、切实规制相关违法行为的社会需求已经较为迫切，对人工智能算法的法律治理也必须破除技术干扰，从“人”的角度衡量技术的价值。此外，社会公众作为智能技术的消费者和数据供给的原始来源，所应享有的利益也不应仅仅限于个人信息得以保护或者免受打扰，还应有经济方面的回馈。基于行为效率的考虑，这种经济利益或许并不直接体现为相关经营者对社会公众的经济性给付，也应该通过设计其他的给付方式得以呈现。

此外，从防范风险的角度观察，人工智能算法除了直接涉及的控制者、开发者和用户利益外，还有更长远视角下的社会整体利益需要被纳入法律治理的范畴中。在“网络安全和信息化是一体之两翼，驱动之双轮”[17]的科学论断指引下，社会整体利益可以进一步细化为发展利益和安全利益两个方面。在这当中，发展利益是我国发展和应用人工智能技术所追求的根本利益，是我国相关技术进步、市场拓展乃至经济增长、赢得国际竞争的重要驱动力。对此，需要从促进数据流动与利用的角度，平衡相关经营者和社会整体以及个体间的人工智能技术利益分配，继而进一步平衡人工智能相关产业经营者自主经营、智力劳动投入、资本投入、个人利益以及社会整体利益的关系。对于所牵涉的安全利益，则可以从纵向细分为四个层次：国家安全层面，包括数据信息的本地化和外流风险，以及基于算法漏洞的外部攻击、破坏风险；社会秩序层面，包括关键信息基础设施的正常运行、算法影响下的数据安全和舆论导向；市场充分

和有序竞争层面,包括利用算法实施的各种阻碍竞争、无序竞争和侵犯消费者利益的风险;社会公众个人利益保护层面,包括个人信息不当搜集、使用以及对个人表达的不利影响等。相关安全利益牵涉面广泛、敏感,同样需要法律制度的重点关注。对不同层次安全利益的确认与关注,也是基础性的算法治理活动的应然之义。

(三)人工智能算法治理的路径选择

基于对人工智能算法基本范畴和法律表达的判断,以及对人工智能技术应用所引发的利益变动和制度需求的观察,人工智能算法治理的思路可以进一步明确为基于知识产权制度和基于专门法的两条并行的路径。

一方面,通过知识产权制度来规范人工智能算法开发和应用中的新型对象,对所产生的种种利益或者行为予以确认和规制,具有现实的优先性和便利性。无论是从权利客体还是监管对象的角度,人工智能算法的本质都是一种特定的信息,符合知识产权客体的基本要求。知识产权制度作为客体法,也更适合于对具体和有限的信息对象进行概括和定义。这是因为,虽然人工智能算法的本质是一种信息,但却是一种解决明确技术问题的技术方案。这使得人工智能算法具备明确的物理功能和界限,以区别于抽象意义上的信息。具体而言,由于人工智能算法与现有知识产权客体之间存在着一定的相似性,可以通过知识产权制度对特定形态的人工智能算法提供及时、必要的保护。其一,人工智能算法作为一种解决特定技术问题的方案,而非某种著作权意义上的表达或者某种抽象的思维活动,与专利权客体存在着一致性。虽然在具体的授权标准方面,人工智能算法的专利保护制度还有待完善,但是将部分符合要求的人工智能算法纳入专利制度,已经被专利审查部门所采纳①。其二,人工智能算法作为以伪代码书写、记录的信息,也有可能被作为计算机程序(被视作某种文档),获得著作权制度的保护。第三,考虑到人工智能算法往往以底层技术的方式发挥作用,处于表层产品或者服务的掩盖之下,它也往往被其控制者当作技术秘密,可以得到商业秘密相关制度的保护。

另一方面,更具前瞻性的专门法路径,才具备精准地描述人工智能算法,并及时地应对相关的现实与潜在风险的能力。这是因为,以知识产权制度来定义和规范人工智能算法以及相关活动,只是相关制度供给紧缺情况下的应急。单独地依靠现行知识产权制度,既不能对其进行完整的描述与归类,也无法通过简单创设新型知识产权客体的方式应对相关应用中所产生的种种超出知识产品利益分配范畴的问题。也即是说,人工智能算法作为一项特殊的技术方案,无法被整体地划入专利、计算机软件或者商业秘密范畴。而且,其所引发的利益变动以及现实中的种种问题,也非单独的知识产权法律制度所能应对。在我国相关制度供给不足的现实下,应急之余,也应该尝试通过专门法来对人工智能算法实施综合治理。

三、人工智能算法治理的专门制度设计

通过知识产权制度,可以换取一定程度的算法技术公开,为知识传播或者法律监管提供方便。但是,知识产权权利毕竟属于私权,无法成为法律治理或者更具强制色彩的监管活动的可靠途径。而人工智能算法治理也绝非单一的法律监管所能涵盖。当中,技术理性与社会整体利益的协调发展,既是宏观的治理目标,也是贯通于治理框架搭建与规则设计的具体指导。

(一)治理架构:围绕人工智能算法构建综合治理体系

围绕人工智能算法所构建的综合治理体系,需要对相关对象范畴、治理活动整体框架予以全新设计,并体现出对现行主要治理方式的吸收与革新,在专业性和技术性等方面具有鲜明的制度创新色彩。

---

① 我国2019年底修订的《专利审查指南》第九章新增规定:"如果权利要求中除了算法特征或商业规则和方法特征,还包含技术特征,该权利要求就整体而言并不是一种智力活动的规则和方法,则不应当依据专利法第二十五条第一款第(二)项排除其获得专利权的可能性"。

但是，这并不意味着对现有治理工具的全面放弃。在这一理念下，人工智能算法专门治理制度中最具统领性的架构设计，应遵循以下思路展开。

首先，系统化的治理体系应体现对现有制度的充分利用。虽然当下针对人工智能的专门规定较为稀少和分散，但是，也不排除我国在未来条件成熟时设立专门的人工智能“促进法”或者“监管法”，并在当中设计“算法规则”的可能性。就目前而言，仍然应当注重与现有制度的配合，促使制度资源得以最大程度的发挥作用。一方面，可以通过具有网络治理基本法地位的《网络安全法》以及《数据安全法》等法律规范，对人工智能算法的设计与实施进行监督，对相关对象、行为等基本概念进行界定，为算法治理活动明确合理边界。另一方面，可以通过具体应用所涉领域的专门法实现对智能应用中的算法活动予以监督，或者在“分级分类管理”[①]思路的指引下，通过对数据搜集、传输、使用和存储等方面行为进行规定，可以形成对具体应用领域中算法风险的间接识别和应对。与此同时，也可以通过知识产权制度，从智力劳动成果保护的角度，确认和保障相关的知识产权利益，鼓励和促进人工智能算法由技术底层走向开放，使用户可感知、可理解，使社会公众可以尽量自由地获取和利用相关知识。此外，还可以从加强反垄断执法的角度，对利用算法工具实施的新型滥用市场支配地位或者垄断协议等行为予以规制，规制运用算法工具所实施的垄断行为。

其次，应超越传统的“算法可解释”和“平台责任”治理模式。从操作层面实现对人工智能算法的有效治理、理论探讨向实践转化的必然要求。现有研究所给出的应对策略，往往集中于“算法可解释”和“平台责任”两种模式，却难以真正应对人工智能算法的技术复杂性。当下，“算法可解释”往往被认为是突破人工智能算法黑盒障碍的重要手段——通过赋予社会公众个体要求人工智能控制者解释算法决策过程的权利，或者使相关主体承担解释算法运行是否满足合法性、合理性等标准的义务，使原本难以被察觉的非法活动暴露于监管的视野之内。但是，使人工智能支配者或者使用者解释算法，只是实现治理目标的手段之一，并不是最终的目标本身。一般意义上的“算法可解释”既无法应对人工智能算法迭代所造成的版本干扰，也无法使底层的算法原理和具体应用中的实际效果间形成清晰的映射。而且，人工智能技术以及相关应用的基本原理决定了，单纯地要求人工智能控制者或者使用者向用户个人承担解释义务，也无法真正地实现制度目标：通用的底层人工智能算法解决的是技术问题，而不是进行价值判断，无论其保持开放或者是保密的状态，对于用户或者个人乃至立法者和执法者而言，它仍然是专业性过强的技术方案，并不直接地与其个人利益产生关联；而简单化地对相关主体施加解释算法的义务，虽然从理论上或许可以充分地保障每一次自动化决策或者人工智能活动不至于脱离伦理、法律框架或者公平正义的价值观，但会带来致命的效率损耗，并因此极大地抵消掉人工智能技术的技术优势。与此同时，设想中的“算法可解释”还面临着解释方式、程度和检验标准等方面的现实困扰。另外，也不能把人工智能算法治理简单地等同于以平台责任规范人工智能控制者的行为。强调平台责任确实可以在一定程度上保障平台用户乃至社会公众的整体利益，但是如此处理，除了在治理思路上有应急之嫌以外，并没有真正地将人工智能算法纳入治理范畴，也将视野过窄地限制在了市场经营领域，而忽视了人工智能技术的应用前景。与此同时，在智能化的社会中，非平台化经营的人工智能技术研发、应用规模也十分庞大，单凭平台问责也难以充分应对千变万化的市场。

第三，以创新性的“算法可理解+数据可信+参数可解释”为骨干构建新的治理框架。作为对现有应对策略的继承与发展，在以算法为主要对象的治理体系下，着重强调数据与参数相配合。一方面，“算法可理解”可以被认为是一种低水平的算法开放[②]，它以最低限度的、可以被监管部门的专业人员所理解为

① 参见《中华人民共和国网络安全法》第21条“网络安全等级保护制度”的相关内容，它是我国从网络治理基础性法律规范的层面对网络治理活动“分级分类管理”思路的正式确认。

② 算法开放的内涵包括“算法可解释”，但还有其他的内容，如算法共享、算法许可使用、算法二次开发以及对应的数据开放等。

基本要求,而非对全体社会成员的充分公开,并以这样的方式保障人工智能控制者、研发者的自主性和保密利益。这一设计的目的在于规避明显的算法错误和以人工智能为借口的欺诈或者虚假宣传,而不奢求一劳永逸地解决一切算法问题。另一方面,“数据可信”[①]和“参数可解释”[②]可以较好地发挥各自的作用,同时也保有合适的监管尺度。数据之于信息时代的重要性无须赘述,但数据治理也有其自己的功能与逻辑,并非万能。“数据可信”作为更宽泛意义上的数据治理活动的一个组成部分,应用于人工智能算法治理活动,在发挥基本功能的同时,也应体现出必要的克制:确保人工智能及其控制者所使用的数据拥有清晰、可信的来源,并且遵守法律规范中基本的数据行为规则即可,而不对其施加过多的数据安全或者数据合规义务,为人工智能技术相关的数据利用活动保留足够的空间。而“参数可解释”则可以视为是对人工智能算法支配者,或者更贴近应用端的经营者的一种强约束。在工程技术语境下,参数往往带有较强的技术客观性。但是在底层的通用算法实现了基础智能,以及外部数据供给具备基本的合法和可信度的情况下,相关的参数设置往往会更多地体现相关经营者的主观意愿,更具主观性。此时,相关参数的设定或者选取,会更多地与人们所熟知的杀熟、偏见或者合谋等不合理行为产生关联。将“参数可解释”纳入治理框架,是更具有针对性,更远离技术黑盒、贴近具体应用和用户感受,也面临更小阻力的监管手段。

(二)治理策略:遵循技术与市场规律

在“算法可理解+数据可信+参数可解释”的治理框架之下,还应对人工智能算法治理的基本策略有所判断,以回应人工智能算法相关的技术属性和应用规律。其一,应将治理的时机尽量提前,以应对现代信息社会网络外部性对风险的放大效应;其二,应将通用算法治理与具体应用领域相结合,提升算法治理的灵活性。具体而言,体现为两方面内容:

一方面,应在传统的“事后救济”之外,尝试对人工智能算法风险进行事前防范。这是因为,以人工智能技术发展和应用为代表的智能社会,面临着风险活动隐蔽化和损害结果扩大化的新型技术条件和社会背景,事后补救的难度和成本都在急剧攀升。“在传统社会,很多风险是个别性、局部性、偶然性的,而在信息社会和智能社会,大多数风险具有快速蔓延性、急剧增强性”[18](P4-19)。这种风险特性,要求对算法的治理须尽可能地提前,而不能被动地等到损害发生后再补救。现实中,对于人工智能算法所引发问题的事后救济尚不能及时应对已发生风险,如何采取有效的事前防范措施更显模糊,但相关制度建设仍应对此进行积极尝试。

对于人工智能算法风险的事前防范,关键在于将人工智能系统底层的算法运行从技术的黑盒中提取出来,以降低损害结果发生后的修复成本。它是对事后救济的配合与补充,而非替代。因此,相比于应对已发生损害的事后救济,“事前防范”更注重预防而非惩罚,更应体现法律介入的谦抑性,需以对相关投资利益和自主经营的尊重为前提。对于人工智能算法风险的事前防范,至少应当包含三方面的内容:其一,人工智能的控制者、经营者需要将其所控制、使用的人工智能系统的算法与其他的技术对象或者技术行为区分开来,以满足必要的检查或者开放要求。这种技术对象的独立性要求,将确保算法不至于被掩盖于纷繁复杂的技术活动之下,成为法律治理的盲区。其二,人工智能的控制者应当能够解释其所应用于算法的各种参数的合法性。这种解释的标准较低,只要求相关参数的设定以及其所对应的行为,不得违反现行法律法规的禁止性规定。除此以外,人工智能运行中各种参数设定,可以以相关控制者的主观意愿为依据,以保障其行为的自主性。第三,人工智能的控制者应当能清晰解释其人工智能系统所使用数据的来源。此处的解释标准除了要满足合法性的要求外,还应满足合理性要求,并与外部的

① 即要求人工智能控制者能够对其数据供给来源提供合理的解释。

② 即要求人工智能控制者要能够对其算法设计中各种参数的设定提供合理的解释:是基于技术规格的需要,还是基于经济理性,又或者是单纯地体现其主观好恶等。

数据搜集、使用规则相衔接。此外，数据层面的可解释性，也体现出事前防范与事后监督的过渡——作为事前风险防范对象的人工智能算法本身和事后救济所指向的人工智能损害效果，往往也是通过数据来源的合法性以及数据行为的合理性而产生关联。

另一方面，以具体的人工智能应用领域和方式，以分级分类管理的思路，厘定不同的算法理解标准。具体而言，则体现在人工智能算法治理，还需要根据不同的应用场景，设计差异化的“算法可理解”的方式与程度，理由有三。其一，在人工智能算法的控制者、使用者身份脱离的情形下，二者所应承担的法律义务与责任也是不一样的。对于不同的算法，其输出结果和自身的可理解程度都有差异，“只有监督学习状态下的算法才在设计者的掌控之下，对于无监督学习和强化学习而言，结果是无法预测的”[18](P4-19)，并且只有在控制者可以控制算法的明确输入和输出的情况下，才会对于人工智能所造成的损害负直接责任。而对于输出具有不确定性的真正意义上的人工智能算法仅承担类似于主人对其所拥有的动物的看管责任，并且以“对计算机技术能力的一般预期”为前提[19](P497-521)。也就是说，在不同的应用领域中，人工智能算法可理解性或者开放程度的高低，将影响到人工智能控制者的直接责任或者间接责任的认定与分配。其二，人工智能算法治理意味着一定程度上对相关算法黑盒状态的打破，应当有适宜的尺度。虽然在具体的运营过程中，人工智能算法有时以开放或者开源的方式呈现于市场或者其他领域，但这并不导致相关的保密行为丧失合理性。但是，人工智能算法治理要求其必须具备一定的可理解性。对此，我国在设计相关制度时，可以借鉴美国计算机协会（Association for Computing Machinery）“鼓励使用算法决策的系统和机构对算法过程和特定决策提供解释”[20]的做法，将人工智能算法的可理解问题放到具体的人工智能应用领域中进行综合考虑，并以具体的应用特性为基础，为人工智能控制者和人工智能用户之间划定保密与可理解的边界。第三，也需要将“算法可理解”与我国网络治理的“分级分类管理”理念相结合。对于事关国家安全的算法应用，强调其主动分级管理、接受全面监督和报告风险的义务；对于事关社会秩序的算法应用，设计相关经营者就“通用算法”开放的事前开放义务和事后的风险管理报告义务；对于市场范畴下的算法应用，鼓励相关经营者开放通用算法，保障其对个性化的应用算法施以合理程度的保密；对于事关社会公众个体利益的算法应用，则应构建具有可操作性的个人信息保护和利用方面的权利体系，保障其人格尊严、生活安宁和公平参与数据活动的基本权利。

（三）治理规则：体现技术效果与社会效果的平衡

人工智能算法技术的技术属性与应用方式是相关治理体系构建的客观基础。当宏观的治理理念落实于具体的操作层面，也必然要求相关的基础性规则应当满足技术规律与社会需求的双重要求。具体而言，当中的基础性规则至少应包含以下四方面的内容，才能在识别治理对象、规范基本行为和满足社会利益关切方面，使相关治理体系得以真正实施和发挥作用。

其一，应设计合理的算法识别规则，同时避免“唯算法论”的误区。结合人工智能技术复杂性和对数据供给网络的高度依赖性所喻示的高企的市场准入门槛，实际上也暗示了人工智能相关应用背后的资本因素，并且不排除其成为资本所利用的“算法独裁”[21](P52-58)工具的可能性。因此，有关算法与法律的关系也引发了全社会的重视：基于算法之于现代法律制度的重要影响[22](P64-75)，算法甚至带来了新的权力形态，导致了算力即权力的新现象，同时也使公民权利面临更隐蔽、更无所不在的“权力技术”[23](P66-85)的侵蚀。而这种担忧在极端情况下会将二者的关系归纳为“算法即法律”[24](P6)，认为在人工智能的时代，算法就是法律。诚然，算法已经通过愈发智能的系统，成为帮助甚至一定程度上替代人类作出判断的工具，其与基本权利乃至公权力的关系，理应引起法学研究的关注。但是不能因为人工智能应用所展现出来的种种性能优势，就将其中的人工智能算法给神秘化或者万能化，纵然其被隐匿于层层的技术细节之下，也不影响其被法律制度所识别和规制。在“算法可理解”的基本要求下，实现人工智能通用算法、应用算法的精准技术识别和针对性处置，是人工智能算法治理的必然要求。

其二，应设计数据可信规则，摆脱“唯数据论”的干扰。人工智能算法并不能单独地支撑起相关系统

的运行和应用,还依赖于庞大的外部数据供给。而且,从相关系统自身完备的角度衡量,数据供给的质量和充足程度决定了人工智能算法在单位时间内迭代的次数和质量,并直接影响到相关系统的智能水平。在相关技术不断进步以及采用平台化经营的数据主体不断发展壮大的趋势下,以庞大数据体量为支撑的运算能力,所发挥的作用和影响早已经超出了一般的市场经营范畴,而成为一种新兴的社会力量[4](P398-415)。因此,对于人工智能的现实控制或者法律治理也存在着另一个思路,即通过数据控制或者数据治理的方式,控制人工智能系统的基本功能与性能乃至存续。从技术的角度来看,这种认知并无明显差错,但并不能以单一的数据控制替代对人工智能算法本身的监督。一方面,虽然对于人工智能而言数据至关重要,但二者的功能和作用并不一致,数据与智能之间的鸿沟也不会因为数据规模或者质量的提升而自然消除。对于人工智能而言,具有创造性的、足以解决特定技术问题的算法仍然不可或缺。另一方面,数据治理不能替代算法治理。对于数据治理,无论是我国《民法典》《数据安全法》等直接规定数据保护与利用的法律制度,还是其他相关的规范性文件,都以数据内容的敏感性或者应用领域为划分,进行分门别类的管理。这与人工智能算法的通用性之间存在着天然的差异。而且,在人工智能技术环境下,对于不同的数据输入,同一个算法运行的结果会迥然不同,数据的合法性无法成为判断算法合法性或者合理性的依据。因此,在讨论算法治理的基本架构与基础性规则时,对于数据的重要性应有清晰的定位:数据来源的可信与否,是人工智能算法设计与完善、相关产品或者服务发展与应用所不可或缺的外部环境,但数据与算法并非彼此替代关系,数据治理不能代替其他的治理规则。

第三,应设计算法理解规则,明确"可理解"的尺度与方式。当下,人工智能领域的算法开放仍然出于人工智能控制者的自觉或者自愿,当中既有出于开源情怀的考虑,也有对推广自身算法、获取数据反馈的利益驱动。而相关算法的保密与开放之争,也不全是一个私权行使问题,而更是一个基于信息利益公平分配的社会公平问题。具体而言,包括人工智能在内的现代信息技术不能脱离于人类社会单独存在,它必然要从广泛的社会主体处获取数据资源并且实现投资利益。而广泛的人工智能控制、经营活动之外的社会个体,也应当享有从人工智能技术发展中获益,以及不受人工智能技术影响自身合法权益的权利。此时,信息利益公平分配这一价值导向也可以集中体现于算法可理解规则——对于人工智能控制者而言,这一规则可以使其在承担必要开放义务的同时不至于失去对相关投资和核心竞争力的掌控;对于社会公众而言,通过这一规则可以使其免于遭受由于算法错误或者算法歧视所带来的不公,也可以更主动地参与人工智能创新;对于市场监管者而言,这一规则可以有助于其对潜在的算法滥用以及垄断问题保持足够的警惕并及时评估竞争状况,应对由于信息利益不合理分配所引发的市场竞争风险;对于社会管理者而言,算法可理解规则也是其评估人工智能系统应用效果和潜在社会风险的重要工具。换言之,一个安全、可信和不至于引发信息利益失衡的人工智能系统,其算法必须是以某种形式和程度向监管机关和社会公众所开放的。算法可理解规则的设计,以对相关信息系统控制者、开发者、经营者的投资和经营利益的确认与保护为前提,不会造成人工智能系统控制者或者研发者保密技术的外泄,也不会为相关经营者带来过重的解释负担。

最后,应完善个人利益保障规则,同时也为技术发展保留足够的空间。在现实的视角下,无论是作为整体的人工智能还是其中的算法,它们都是服务于"人"的技术工具,而远不至于获得某种主体资格。这也意味着,无论是宏观的治理体制构建还是对基础性规则的设计,都必须把对个人利益的尊重与保障纳入考虑范围。这既是一种伦理层面的价值宣示,也是具体规则层面的利益分配:无论是欧盟《通用数据保护条例》(General Data Protection Regulation)对自动化决策的警惕,还是我国《电子商务法》对算法推送所导致的对个人生活侵扰的规制,都是对这种理念的积极实践。但是也要看到,在当前社会环境中,无论是否有严格的"知情/同意"或者"通知/删除"规则,每一个社会个体都将或主动或被动地参与广泛的数据活动。这种信息化或者数据化浪潮,也必然为人工智能所用。这也意味着,相关的个人利益保障规则须以更加务实的方式看待人工智能的数据需求,具体体现在两方面。

其一，应保障社会个体对智能产品或者服务提供者的要求解释的权利。这种权利与现有理论中的“算法解释权”有一定的相似，但也存在区别。一般而言，算法解释权更多体现的是伦理规范层面的“内在之善”[25](P71-88)，价值宣示的意味更浓。而笔者所要求的解释，更倾向于是一种现实的操作规则，即对相关算法“算法可理解+数据可信+参数可解释”标准的校验。这种校验，可以通过司法途径实现，也可以通过行政管理的途径实现。既包括实体层面的校验，也应当包括程序性的规则，并为相关主体设计适当的法律责任。其二，在充分实施各上位法、专门法的情况下，谨慎地创设个人信息保护或者隐私保护相关的“人工智能专门规定”。我国《民法典》中专设“人格权编”，已经将我国对“人”的利益保障提升到了前所未有的高度，而在我国《电子商务法》《网络安全法》《数据安全法》以及《个人信息保护法》的配合下，个人信息、个人隐私以及其他或具体或抽象的个人利益已经得到了充分的彰显。在这样的情况下，人工智能技术以及相关应用只要遵守现行法律法规或者政策性文件，就足以实现对个人利益的有力保障，并无针对人工智能再专设规则的迫切必要。此外，从可操作性的角度看，社会个体对智能产品或者服务的数据行为所主张的“解释权”或者其他直接的经济利益，除了面临计算依据匮乏的现实困难之外，也会为相关产业带来高昂的经营成本和违法风险，并会消耗大量的法治资源。在未来可能出现的人工智能专门立法中，也应当规避这一误区，为技术和产业发展保留充足的空间。

参考文献

[1] 卢克·多梅尔.人工智能改变世界，重建未来.赛迪研究院专家组译.北京：中信出版社，2016.

[2] 玛格丽特·博登.人工智能哲学.刘西瑞、王汉琦译.上海：上海译文出版社，2001.

[3] John R. Searle .Minds, Brains, and Programs, Behavioral and Brain. *Sciences,* 1980, 3(3).

[4] Nicholas Diakopoulos. Algorithmic Accountability: Journalistic Investigation of Computational Power Structures. *Digital Journalism,* 2015, 3(3).

[5] 史蒂芬·卢奇，丹尼·科佩克.人工智能.林赐译.北京：人民邮电出版社，2018.

[6] 刘晓力.认知科学研究纲领的困境与走向.中国社会科学，2003, (1).

[7] 佩德罗·多明戈斯.终极算法：机器学习和人工智能如何重塑世界.黄芳萍译.北京：中信出版社，2017.

[8] 陈景辉.人工智能的法律挑战:应该从哪里开始？比较法研究，2018，(5).

[9] 王德夫.知识产权视野下的大数据.北京：社会科学文献出版社，2018.

[10] 吴汉东.人工智能时代的制度安排与法律规制.法律科学，2017，(5).

[11] 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，(6).

[12] 唐林垚.算法应用的公共妨害及其治理路径.北方法学，2020，(3).

[13] The New York City Council. A Local Law in Relation to Automated Decision Systems Used by Agencies. The New York City Council Official Website, 2018-01-11.[2020-04-18]*https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0.*

[14] Jenna Burrell. How the Machine “Thinks”: Understanding Opacity in Machine Learning Algorithms. *Big Data & Society,* 2016, 3(1).

[15] 施春风.定价算法在网络交易中的反垄断法律规制.河北法学，2018，(11).

[16] 韩伟.算法合谋反垄断初探——OECD《算法与合谋》报告介评：上.竞争政策研究，2017，(5).

[17] 中华人民共和国国家互联网信息办公室.中央网络安全和信息化领导小组第一次会议召开 习近平发表重要讲话.新华网，2014-02-27.[2020-04-20] *http://www.cac.gov.cn/2014-02/27/c_133148354.htm?from=timeline.*

[18] 张文显.构建智能社会的法律秩序.东方法学，2020，(5).

[19] Gunther Teubner. Rights of Non-Humans? Electronic Agents and Animals as New Actors in Politics and Law. *Journal of Law and Society*, 2006, 33(4).

[20] ACM U.S. Public Policy Council and ACM Europe Policy Committee. Statement on Algorithmic Transparency and Accountability. Association for Computing Machinery Official Website, 2017-01-12. [2020-05-12] *https://www.acm.org/binaries/content/assets/public-policy/2017_joint_statement_algorithms.pdf.*

[21] 陈鹏.算法的权力:应用与规制.浙江社会科学,2019,(4).
[22] 蒋舸.作为算法的法律.清华法学,2019,(1).
[23] 郑戈.算法的法律与法律的算法.中国法律评论,2018,(2).
[24] 劳伦斯·莱斯格.代码2.0:网络空间中的法律.李旭、沈伟伟译.北京:清华大学出版社,2009.
[25] 陈璞.论网络法权构建中的主体性原则.中国法学,2018,(3).

On the Legal Attribute and Governance Approach Of Artificial Intelligence Algorithm

Wang Defu (Wuhan University)

Abstract Artificial intelligence algorithm aims to solve the technical problem of how to achieve intelligence. As a special technical solution, "artificial intelligence algorithm" has not only the relatively traditional intellectual property path, but also the innovative special law governance path. New knowledge and relevant interests are derived in the process of artificial intelligence algorithm design and self-perfection. Specific solutions should be designed to strike a balance between investment interests and social welfare. Based on the framework of "algorithm comprehensible + data credible + parameter interpretable", the comprehensive governance system and basic rules of artificial intelligence algorithm are designed, and the "individual algorithm interpretable" and "platform responsibility" are clarified in order to ensure the coordination and mutual facilitation of technical reasoning and social development.

Key words artificial intelligence technology; algorithm governance; intellectual property; algorithmunderstandability; data element market; data security; data credibility

■ **收稿日期** 2020-07-21
■ **作者简介** 王德夫,法学博士,武汉大学法学院讲师,武汉大学知识产权与竞争法研究所、武汉大学网络治理研究院助理研究员;湖北 武汉 430072。
■ **责任编辑** 李　媛