财经法学 No. 1，2023
pp. 61 - 75

自动化行政中算法的法律控制

王　宾*

内容提要：自动化行政中的算法可分为转译型算法和自我学习型算法。算法的运用面临合法性危机：处于私主体地位的算法设计师将法律语言转译成机器语言时会嵌入自身的判断，带来改写法律的风险；算法决策有时在事实上超出法律授权范围，且缺乏畅通的救济机制。算法的合法性控制方式应与算法类型适配。针对转译型算法，需结合算法的性质以及技术特点，从转译主体、所译法律的明确性、转译过程的透明度等方面施以控制。针对自我学习型算法，首先应当确立“民主—科学”的合法性框架，其次应当从建立算法信任的角度，围绕保障公众主体地位和算法科学性对算法进行控制。

关键词：自动化行政　形式合法性　行政民主　行政科学　算法信任

一、引　言

随着人工智能技术的迅猛发展，数字技术越来越多地被嵌入公共行政领域，数字法治政府建设已上升为一项国家战略。中共中央、国务院印发的《法治政府建设实施纲要（2021—2025年）》提出健全法治政府建设科技保障，全面建设数字法治政府，坚持运用互联网、大数据、人工智能等技术手段促进依法行政，着力实现政府治理信息化与法治化深度融合，优化革新政府治理流程和方式，大力提升法治政府建设数字化水平。《国务院关于加强数字政府建设的指导意见》提出“将数字技术广泛应用于政府管理服务，推进政府治理流程优化、模式创新和履职能力提升，构建数字化、智能化的政府运行新形态”。数字法治政府建设的潮流不可逆转，其中需要关注的重点问题之一是如何保证“数字”与“法治”共生。由于技术水平的迭代发展，政府开展行政活动所借助的数字技术已经超出了亚里士多德所想象的“按照人的意志或命令自动工作的无生

*　王宾，北京大学法学院博士研究生。

命工具”〔1〕的范畴。机器学习通过算法，让机器可以从外界输入的大量数据中学习到规律，从而进行识别判断。〔2〕机器学习算法都有一定程度的自主性，它们通常无法为证明行政决策合理性的这类因果陈述提供任何依据（例如“因为 X 导致 Y，所以 X 是正当的”）。〔3〕当行政机关使用机器学习系统作出行政决策时，很难保证该决策体现的是特定人的意志。

有观点认为，行政国家的自动化程度越高，就越有可能减损行政国家存在的正当性，理由有二：一方面，自动化系统的复杂性使得利用其执行任务的行政机关工作人员并不了解系统的运行原理，更无法在法庭上给出任何解释说明；另一方面，行政规则以机器语言的形式被嵌入自动化系统之后，行政机关再也无法像过去一样通过行使裁量权来回应变动不居的实际情况。〔4〕在高度自动化行政的背景下，行政机关同时放弃了专业知识和裁量权。行政机关对专业知识和裁量权的放弃，意味着其将行政权力让渡给了自动化系统。而自动化决策的正当性、合理性、合法性是数字法治政府建设中应当认真对待的重要基础和法治前提。〔5〕有学者对无人工介入的自动化行政提出合法性疑问：由第三方主体设计的机器和程序过程取代行政机关一方的意思表示是否具有权力主体上的合法性，用以作出行政决定的算法是否如实复刻了实体法中的内容。〔6〕对此，有观点认为，在自动化行政中，算法设计师在转译法律语言时，可能会扭曲法律法规的内容，导致法律规则的异化执行或适用，不符合形式合法性的要求，而且，由于自动化行政中存在技术壁垒，公众的知情权、异议权和建议权难以实现，行政过程缺乏民主正当性。〔7〕

在传统行政法中，行政机关的权力来源于立法授权，立法机关被设定为代表民意的政治机构，负责民意的汇集和表达，制定法律并承担政治上的责任，行政机关负责法律的执行。行政机关只要严格依法行政，便可以借助政治权威（立法机关）的正当性而获得“合法化”。〔8〕在自动化行政中，自动化系统运行算法的过程相当于执行法律的过程，面对自动化执法合法性不足的问题，该如何对算法施以法律控制？这是本文重点回答的问题。首先，文章结合算法的工作原理对实践中的自动化行政样态和算法类型进行分类；其次，在类型化的基础上具体分析可能面临的合法性问题；最后，针对不同类型的算法，提出合法性控制手段。

二、自动化行政中的算法及其应用分类

（一）自动化行政中的算法工作原理

“自动化行政”是一个描述性用语，指行政程序中特定环节或所有环节由人工智能代为处理，

〔1〕〔古希腊〕亚里士多德：《政治学》，吴寿彭译，商务印书馆 1983 年版，第 11－12 页。

〔2〕参见郭丽丽、丁世飞：《深度学习研究进展》，载《计算机科学》2015 年第 5 期。

〔3〕参见〔美〕卡里·科利亚尼斯：《自动化国家的行政法》，苏苗罕、王梦菲译，载《法治社会》2022 年第 1 期。

〔4〕See Ryan Calo & Danielle K. Citron, The Automated Administrative State: A Crisis of Legitimacy, 70 *Emory Law Journal* 797, 804 (2021).

〔5〕参见马长山：《数字法治政府的机制再造》，载《政治与法律》2022 年第 11 期。

〔6〕参见展鹏贺：《数字化行政方式的权力正当性检视》，载《中国法学》2021 年第 3 期。

〔7〕参见王怀勇、邓若翰：《算法行政：现实挑战与法律应对》，载《行政法学研究》2022 年第 4 期。

〔8〕参见王锡锌：《行政正当性需求的回归——中国新行政法概念的提出、逻辑与制度框架》，载《清华法学》2009 年第 2 期。

而无需人工的个别介入，从而实现部分或全部无人化的行政活动。〔9〕学界也用“算法行政”〔10〕“数字化行政”〔11〕“人工智能算法决策”〔12〕等名称来描述这一现象。自动化行政中的技术载体是计算机，而计算机的工作过程就是执行程序，该程序是由程序开发人员使用某种程序设计语言编写的，以代码形式表示的，能够为计算机识别并予以执行的指令集合，程序的核心是算法。〔13〕程序与算法实际上具有相同含义，计算机执行程序本质上就是执行算法。有学者指出，算法是一套求解逻辑，在计算科学领域，其表现为由代码连结且结构化的一系列问题和求解数学模型的集合，单一代码向计算机传达的是简单的做或不做的指令，若干单一代码有机连结后构成解决具体问题的复杂算法。〔14〕将算法视作求解逻辑是最广义的定义，可以涵盖所有决策程序和步骤，而将其限定于计算科学的定义，是狭义的算法定义。也有学者采用中义的算法定义，将算法界定为人类和机器交互的决策，即人类通过代码设置、数据运算与机器自动化判断进行决策的一套机制。〔15〕本文采取中义的算法定义。

从技术层面来讲，有两种构建算法规则的模式：一是专家系统模式，二是机器学习模式。专家系统是利用人类预先设定的专家知识数据库来解决相应的问题。其发展的近期目标是建造能用于代替人类高级脑力劳动的专家系统。〔16〕以构建认定故意伤害罪的算法为例，专家系统构建路径首先需要法律专家确认构建知识图谱的犯罪构成理论（犯罪三阶层、四阶层抑或其他理论），然后在确定的大框架下，根据故意伤害罪的法律特征，精细化拆分犯罪构成要素，定义基本的法律模式图。定义好数据模式之后，再从大量真实的法律数据中抽取相关知识点以及知识点之间的逻辑关系，将这些实体信息相应挂接在要件要素上，从而形成具有高度逻辑的知识组织形式。〔17〕专家系统具有可理解性，即在执行过程中，系统能解释推理步骤，使之易于理解，其解释的方式应与专家解释他们推理的方式一样。〔18〕

机器学习是通过接收外界信息（包括观察样例、外来监督、交互反馈等）获得一系列知识、规则、方法和技能的过程。这一过程对人类和其他生物而言称为“生物学习”，对计算机而言称为“机器学习”。〔19〕简单来说，机器学习是在样本数据的基础上找出一个公式或者多个公式的组合模型来解决特定的问题。〔20〕中间寻找模型（确定算法）的过程是不可知的、难以解释的。在专家系统模式下，自动化系统的算法是由算法设计师确定的；而在机器学习模式下，算法是由算

〔9〕 参见马颜昕：《自动化行政的分级与法律控制变革》，载《行政法学研究》2019 年第 1 期。

〔10〕 参见虞青松：《算法行政：社会信用体系治理范式及其法治化》，载《法学论坛》2020 年第 2 期。

〔11〕 参见前引〔6〕，展鹏贺文。

〔12〕 参见张恩典：《人工智能算法决策对行政法治的挑战及制度因应》，载《行政法学研究》2020 年第 4 期。

〔13〕 参见刘东亮：《技术性正当程序：人工智能时代程序法和算法的双重变奏》，载《比较法研究》2020 年第 5 期。

〔14〕 参见邱泽奇：《算法治理的技术迷思与行动选择》，载《人民论坛·学术前沿》2022 年第 10 期。

〔15〕 参见丁晓东：《论算法的法律规制》，载《中国社会科学》2020 年第 12 期。

〔16〕 参见张煜东等：《专家系统发展综述》，载《计算机工程与应用》2010 年第 19 期。

〔17〕 参见叶衍艳：《法律知识图谱的概念与建构》，载华宇元典法律人工智能研究院编：《让法律人读懂人工智能》，法律出版社 2019 年版，第 25 页。

〔18〕 参见〔美〕吉奥克等：《专家系统原理与编程》，印鉴等译，机械工业出版社 2006 年版，第 7 页。

〔19〕 参见王东：《机器学习导论》，清华大学出版社 2021 年版，第 2 页。

〔20〕 参见邹劭坤：《机器学习的“黑盒”是什么?》，载华宇元典法律人工智能研究院编：《让法律人读懂人工智能》，法律出版社 2019 年版，第 37 页。

法设计师和机器共同确定的，设计师为机器制定“学习规则”，机器在“学习规则”的指示下，通过对海量数据的学习确定算法。

（二）自动化行政中的算法应用分类

既有研究对自动化行政进行了不同的类型化处理。有学者将特定行政活动区分为识别与输入、分析与决定、输出与实现三个环节，根据自动化系统发挥作用的环节不同，将自动化行政分为 0～4 级，分别为无自动化行政、自动化辅助行政、部分自动化行政、无裁量能力的完全自动化行政、有裁量能力的完全自动化行政。〔21〕自动化辅助行政和部分自动化行政中，分析与决定的权力仍掌握在人类手中；而在完全自动化行政中，行政活动不再需要人类介入。也有学者从是否排除人工介入的角度，将自动化行政分为需要人工介入的半自动行政行为和不需要人工介入的全自动行政行为。〔22〕

还有学者将行政过程中是否有人工干预和自动化对最终决定的实际影响结合起来，将自动化行政分为三类：（1）数字化程序实施，但实体决定仍为人工作出；（2）“程序实施＋实体决定”的完全数字化，但实体决定非以人工智能的方式作出；（3）“程序实施＋实体决定”的完全数字化，且实体决定由人工智能作出。〔23〕“非以人工智能的方式作出”是指自动化系统中的算法是专家系统模式下预先设定好的规则，系统并不进行自主学习。在此语境下，人工智能仅包括能够进行机器学习的自动化系统。

这种兼顾行政活动的实现方式和技术影响力的分类方式，于本文研究而言，更具有相关意义，但其在表述上不当限缩了“人工智能”概念的范围，因此应该稍作修正。按照其分类依据，自动化行政可以分为：（1）自动化程序实施，但实体决定仍为人工作出；（2）“程序实施＋实体决定”的完全自动化，但实体决定是人为设定算法的表达；（3）“程序实施＋实体决定”的完全自动化，但实体决定是机器学习后算法的表达。以下将该三类自动化行政分别简称为自动化行政Ⅰ、自动化行政Ⅱ、自动化行政Ⅲ。

在自动化行政Ⅰ中，自动化系统输出的结果对实体决定发挥作用的方式有两种：一是为实体决定的作出提供参考，例如南京市环保行政处罚自由裁量辅助决策系统。〔24〕二是作为实体决定作出的依据。例如，根据《道路交通安全法》第 114 条的规定，公安机关交通管理部门根据交通技术监控记录资料，可以对有关人员依法予以处罚。依据授权法律的规定，电子警察系统的作用是收集、固定违法事实，为最终处罚决定的作出提供证据。有观点认为，在我国智慧交通体系的建设中，算法可以直接对监控查获的交通违法行为处以罚款，这意味着在此领域，算法已经可以直接作为决策者作出具体行政行为。〔25〕本文在第三部分将对这一观点展开论证。

自动化行政Ⅱ的典型范例是深圳市用于高校应届毕业生引进和落户的“无人干预自动审批”

〔21〕 参见前引〔9〕，马颜昕文。

〔22〕 参见查云飞：《人工智能时代全自动具体行政行为研究》，载《比较法研究》2018 年第 5 期。

〔23〕 参见前引〔6〕，展鹏贺文。

〔24〕 参见《规范执法流程 提升执法精准性 南京辅助决策系统实现全覆盖》，载 https://www.mee.gov.cn/home/ztbd/qt/szhb/201507/t20150713_306216.shtml，最后访问时间：2022 年 11 月 3 日。

〔25〕 参见张凌寒：《算法权力的兴起、异化及法律规制》，载《法商研究》2019 年第 4 期。

系统。审批系统按照既定的规则自动进行数据比对，全程自动办理，无人工干预。〔26〕除此之外，疫情防控中所广泛应用的健康码也属于此类自动化行政的范围，健康码经由机器自动化决策生成，行政机关先将评判标准程式化，然后相对人在线提交信息并申请，最终由系统自动分配不同颜色标识的二维码。〔27〕

自动化行政Ⅲ的实践样本尚未在我国出现。该自动化行政方式意味着系统将在不预设“裁量规则”的前提下代替人类作出裁量性具体行政行为。德国《行政程序法》第35a条将具有不确定法律概念和裁量的行政行为排除于全自动程序的适用范围之外，即只允许羁束具体行政行为适用全自动化程序。〔28〕德国立法例属于自动化行政Ⅱ的范围，即人工为系统设定算法，系统执行。美国劳工统计局使用监督学习系统代替工作人员对收集到的大量关于就业、人力成本等专题信息进行编码。〔29〕尽管在该应用场景中，自动化系统并未直接对公民作出决定，但其的确已经独立完成本应由人类完成的编码工作，该工作将会影响劳工统计局相关的政策制定。

三、算法支配的自动化行政的合法性危机

传统行政法通过依法行政原则建立起用于担保行政机关合法行使行政权的框架性法律制度，依法行政原理的逻辑基点是由人民代表大会及其常务委员会制定的法律为行政机关提供行政权的依据，行政机关必须在法律规定的范围内行使行政权。〔30〕在行政法的传统模式之下，行政机关被设想为一个纯粹的传送带，职责是在特定案件中执行立法指令；行政机关的行为受制于司法审查以符合立法指令。〔31〕当行政机关以自动化的方式执行法律时，其同样需符合依法行政原则的要求，接受合法性检验。本部分将从自动化系统中算法自身的合法性和算法决策的合法性两方面，展开合法性问题的讨论。

（一）算法自身的合法性

自动化行政中算法的生成方式大致可以分为人为设定和机器自我学习生成两种。前者主要依靠算法设计师将法律语言转译成机器语言，可以称为“转译型算法”；后者是以算法设计师设计的学习规则为基础，通过对海量数据的学习生成的新算法，可以称为“自我学习型算法”。自动化行政Ⅱ中涉及的算法是转译型算法，自动化行政Ⅲ中涉及的算法是自我学习型算法；而自动化行政Ⅰ中的算法类型取决于系统的技术应用。

转译型算法的设计者通常是行政机关和私营部门中的算法设计师，转译型算法制定的过程实

〔26〕 参见《推动无人干预自动审批（秒批）改革（深圳做法）》，载 https://www.gd.gov.cn/gdywdt/zwzt/szhzy/jytg/content/post_2906394.html，最后访问时间：2022年11月3日。

〔27〕 参见查云飞：《健康码：个人疫情风险的自动化评级与利用》，载《浙江学刊》2020年第3期。

〔28〕 参见前引〔22〕，查云飞文。

〔29〕 将自然语言转换为统计数据是编码的过程，例如为了回答“门卫人员在工作中最常见的伤害原因是什么”这一问题，工作人员需要阅读每一份描述，以编码的方式将对方的职业与造成伤害的因素关联起来。现在机器学习系统代替劳工局工作人员完成这项任务。参见《采访 Alex Measure：机器学习应用于政府业务场景》，载 https://m.elecfans.com/article/1281070.html，最后访问时间：2022年11月4日。

〔30〕 参见章剑生：《现代行政法总论》（第2版），法律出版社2019年版，第36页。

〔31〕 参见〔美〕理查德·B. 斯图尔特：《美国行政法的重构》，沈岿译，商务印书馆2011年版，第11－12页。

质是把行政规范、行政过程以及自由裁量转化成计算逻辑和代码的自动执行，这一过程无疑会嵌入主观判断、利益选择和价值观设定。〔32〕例如，在设计识别车牌遮挡行为的交通监控系统的过程中，当存在多种识别车牌遮挡行为的技术时，如基于车牌结构特征的检测技术、基于颜色特征的检测技术、基于机器学习的检测技术〔33〕等，算法设计师应该选择何种检测技术实现监控系统的运行目标？不同检测技术的准确率和实现成本不同，受私益驱动的算法设计者可能会和代表公共利益的行政机关作出不同的选择。此时，引发的第一个合法性问题是，不具有行政主体资格的算法设计师转译法律规范、主导自动化行政过程的合法性基础为何。这一问题对自我学习型算法而言更加尖锐。尽管转译型算法的设计者包括除行政机关以外的第三方主体，但仍是特定个人决定了算法的表达，算法仍处于人类的控制之下。自我学习型算法，以“学习规则”为基础，利用海量和非结构化的数据来确定解决既定问题的最优算法。除此之外，系统还可以根据外界环境的反馈持续更新算法，结果输出具有不确定性。自我学习型算法的表达已经超出了行政机关和设计者的严密控制，法律的实施具有更大的不确定性和不可解释性。

算法生成过程存在改写法律的风险。传统法律在制定时存在必要的模糊性，也未考虑到自动化的要求，而自动化系统中运行的算法需要极高的精确度和严格度，这导致人类语言与机器语言的转译过程充满了不确定性。〔34〕丽莎·A. 谢伊和伍德罗·哈特佐格等学者在《机器人欢迎电子法吗？一个法律内部的算法实验》〔35〕一文中构建并实施了一个由 52 位电脑程序员参与的、将特定交通法规以代码方式实现的实验。程序员被分为三组，第一组被要求实现“法律条文”，第二组被要求实现“法律意图”，第三组得到了一份附加的、精心编写的说明书，以此作为其软件实现的基础。无论是参考不同文本的不同组的程序员，还是参考同样文本的同组程序员，其最终设计出的程序都存在较多差异。该实验的结论之一是程序员自身的假设和偏差会体现在代码之中，虽然该问题可以通过构建良好的软件设计说明书来化解，但是对所有可能出现的问题进行预测的完美说明书极难设计。实践中，美国科罗拉多州福利管理系统（The Colorado Benefits Management System，CBMS）是确定申请人是否能够获得公共援助资格的自动化系统，该系统自 2004 年 9 月应用以来，作出了成千上万错误的福利认定，许多错误都可以归因于算法设计者在将法律转译为代码的过程中出现偏差，扭曲了联邦和州政策。〔36〕转译型算法或可通过对转译主体、转译程序等施加严格法律要求的方式来保障其准确性，补强合法性。但自我学习型算法的计算逻辑大多是从训练数据中得来的，很少反映在源代码中，〔37〕因此，难以通过控制源代码的方式证

〔32〕 参见前引〔5〕，马长山文。

〔33〕 参见聂文真：《出租汽车车牌遮挡行为判定与图像取证技术研究》，北京工业大学 2019 年硕士学位论文，第 26－27 页。

〔34〕 参见〔美〕丽莎·A. 谢伊、伍德罗·哈特佐格等：《机器人欢迎电子法吗？一个法律内部的算法实验》，载〔美〕瑞恩·卡洛、迈克尔·弗鲁姆金、〔加〕伊恩·克尔主编：《人工智能与法律的对话》，陈吉栋、董惠敏、杭颖颖译，上海人民出版社 2018 年版，第 278 页。

〔35〕 参见前引〔34〕，丽莎·A. 谢伊、伍德罗·哈特佐格等文。

〔36〕 See Danielle Keats Citron, Technology Due Process, 85 (6) *Washington University Law Review* 1249 (2007).

〔37〕 See Kartik Hosanagar & Vivian Jair, We Need Transparency in Algorithms, but Too Much Can Backfire, Harvard Business Review (July 23, 2018), available at https://hbr.org/2018/07/we-need-transparency-in-algorithms-but-too-much-can-backfire, last visited on Dec. 26, 2022.

成其适用的合法性。

（二）算法决策的合法性

1. 算法决策超出法律的授权范围

自动化行政Ⅰ中的系统可分为两类，一是为人工决定提供参考意见的自动化辅助系统，二是为人工决定提供证据的自动化系统，后者对实体决定的影响甚于前者。在自动化辅助系统应用的场景中，作出实体决定的权力掌握在执法人员手中，即使执法人员事实上高度依赖系统提供的建议，也不能将决策过程称为“算法决策”，因为依赖系统是人的主动选择。在第二类自动化系统的应用场景中，尽管从形式上来看是由执法人员根据系统提供的证据作出决定，但实质上系统在固定证据的同时就完成了对违法行为的认定，剥夺了属于人的裁量空间，也超出了法律的授权范围。

以电子警察系统为例，依据授权法律的规定，系统要实现的目标是收集、固定违法事实，为最终处罚决定的作出提供证据。结合《道路交通安全违法行为处理程序规定》（以下简称《程序规定》）的要求，自动化行政处罚流程可归纳为以下五步：第一，交通技术监控设备收集违法事实；第二，经人工审核无误后录入系统作为证据；第三，通知相对人违法信息；第四，告知相对人处罚事实、理由、依据及权利；第五，实施处罚并送达决定书。〔38〕前两个步骤属于案件事实的认定过程，由系统和人类共同完成，系统用来收集、固定违法行为证据。需要注意的是，系统对行为的记录并上传过程意味着其已经完成了对违法行为的第一次认定，人工审核是一个复核的过程。在认定案件事实的过程中，系统认定的案件事实需要经人工审核无误后方可成为行政处罚决定的证据。结合《程序规定》第 18 条和第 19 条〔39〕的规定，人工审核的内容应当是违法行为记录资料是否清晰、准确地反映机动车类型、号牌、外观等特征以及违法时间、地点、事实；对于系统认定违法行为的标准（即预先设定的算法）是全盘接受的。因此，在案件事实认定阶段，系统与执法人员共同认定违法行为，前者通过算法实质决定了违法行为的认定标准，后者仅能从证据形式是否完备的角度否定不符合形式标准的违法行为。从执法实践来看，交警在大多数情况下仅依靠交通技术监控设备或执法设备所记录的图片或视频就实施处罚。〔40〕虽然形式上执法人员对处罚决定的作出保有审核的权力，但事实上系统已经成为真正的处罚决定实施者。由此观之，电子警察系统在应用过程中，已经超出了法律的授权。

2. 算法决策的救济渠道不畅

自动化行政Ⅱ和Ⅲ是无人工干预下的算法决策，而完全自动化系统在设计时可能缺乏纠错机制。以北京健康宝“弹窗 3”为例，“弹窗 3”产生的原理是系统认定特定个人与京内外风险地区、点位、人员等有时空关联，需要进行风险排查。但是健康宝的决策系统并未给个人提供直接的救济途径，使个人能够通过提供不存在时空关联证据的形式自行解除弹窗。被弹窗的公民只能

〔38〕 参见谢明睿、余凌云：《技术赋能交警非现场执法对行政程序的挑战及完善》，载《法学杂志》2021 年第 3 期。

〔39〕《道路交通安全违法行为处理程序规定》第 18 条规定：“作为处理依据的交通技术监控设备收集的违法行为记录资料，应当清晰、准确地反映机动车类型、号牌、外观等特征以及违法时间、地点、事实。”第 19 条规定：“交通技术监控设备收集违法行为记录资料后五日内，违法行为发生地公安机关交通管理部门应当对记录内容进行审核，经审核无误后录入道路交通违法信息管理系统，作为处罚违法行为的证据。”

〔40〕 参见前引〔38〕，谢明睿、余凌云文。

通过人工申诉的方式解除弹窗，〔41〕而人工申诉解决往往耗时良久，弹窗状态又严重影响公民的正常生活，被弹窗公民的救济渠道并不顺畅。

除此之外，公民事实上难以挑战算法决策的准确性。原因有二：一是公民的专业知识很难与算法所代表的行政机关的专业认定相对抗；二是公民得知被“错误”决策的时间通常晚于决策作出的时间，其难以收集并保留行为发生时的证据以自证清白。例如，在何凯与上海市公安局黄浦分局交通警察支队行政二审案件〔42〕中，何凯鸣喇叭的行为被电子警察记录，交警对其作出行政处罚。何凯具有一定的声学专业背景，在二审时其结合专业认知陈述了异议，即根据照片上有关声波的图案无法对应其车辆喇叭发声的波段。这一异议并未推翻电子警察的认定结果。同样，在高彬与新民市公安局交通行政处罚纠纷案〔43〕中，高彬被电子监控设备认定为超速，并被交警予以顶格处罚。高彬依据监控设备拍摄的照片上显示的时间及其目测的位移，自行计算速度，认为其并未超速，并且提供了相关的学术论文证明雷达测速对其车速的测量是误判。同样这一主张也未得到法院的认可。

自动化行政方式对传统行政法中的法律约束框架提出挑战：第三方设计主体的参与、转译型算法与自我学习型算法改写法律的风险、算法决策超越法律的授权等冲击着立法对行政的约束能力；算法决策的救济途径不畅、算法的难以审查性也使得司法对行政的约束作用减弱。对此，一方面，应当反思传统法律控制框架对自动化行政发挥作用的场域；另一方面，在传统框架规制不足的场域，应当探索新的合法性约束机制。接下来，文章将分别从转译型算法和自我学习型算法的控制角度对前述问题作出回应。

四、转译型算法的控制

转译型算法面临的合法性问题是转译者的主体适当性、转译算法是否能够准确实现法律的要求。在自动化行政中，“代码即法律”〔44〕，转译型算法的制定过程（转译过程）可类比传统行政法中的规则制定过程。转译型算法的裁量存在于转译过程，算法适用过程无裁量空间。对转译型算法系统而言，控制算法制定过程就能够控制算法适用过程。针对转译过程的控制：首先，需要分析转译过程的法律性质为何，应该符合何种主体、程序的要求；其次，应当结合法律语言转译成算法的不确定性特点，探究通过何种方式缩减第三方中算法设计师的判断空间。

（一）转译过程的法律性质

转译型算法作为机器语言，其法律性质与所需要执行的规范条文的性质有关，若其对应的规范条文属于裁量基准，则算法就相当于裁量基准。例如，自动化处罚系统中的转译过程相当于将

〔41〕 参见《收到北京健康宝弹窗 3 怎么办？怎样处理高效便捷，方法来了！》，载 http://beijing.qianlong.com/2022/0919/7635814.shtml，最后访问时间：2022 年 11 月 22 日。

〔42〕 参见上海市高级人民法院（2019）沪行终 204 号行政判决书。

〔43〕 参见辽宁省沈阳市中级人民法院（2016）辽 01 行终 386 号行政判决书。

〔44〕〔美〕劳伦斯·莱斯格：《代码 2.0：网络空间中的法律》（修订版），李旭、沈伟伟译，清华大学出版社 2018 年版，第 1 页。

裁量基准算法化。算法将裁量过程分解为可供机器运行的计算步骤，而代码则以机器语言的形式对计算步骤进行具体化表达。在自动化处罚裁量语境下，算法相当于裁量基准。不同于传统裁量基准，算法化裁量基准将法律适用过程中的事实要素直接纳入，实现事实与法律规范的具体对应。〔45〕

转译过程因存在必须由行政机关和算法设计师填补的判断空间而具有立法的色彩，可以将其类比为行政机关具有较大裁量空间的规则制定过程。例如，在设计CBMS系统时，由于算法设计师对规则进行编码时改变了上百条既定规则，系统相当于在阐明新规则。〔46〕规则制定过程是在阐释语义模糊的立法，在立法规定无法为规则制定提供清晰指引时，该过程会借助公众和专家的参与来增强规则制定的合法性和科学性。转译过程需要减小法律语言与机器语言之间的模糊空间。在将法律语言细化至更容易为算法设计师操作的技术标准和设计说明书过程中，可以借助公众和专家的知识作出价值判断和技术选择。在具体转译算法之时，法律语言到机器语言之间的判断空间，只能由行政机关和算法设计师来填补，此时算法可能偏离其所表达的法律的意图，偏离程度与判断空间的大小有关。下文主要针对法律语言到机器语言的转译过程，从转译过程的主体要求、所译法律的明确性要求和转译过程的透明度要求三方面提出转译型算法的控制方式。

（二）转译过程的主体要求

首先，行政机关采取自动化行政方式应获得立法的授权，即存在授权规范，具体规定何种行政机关在何种行政领域能够以自动化方式开展行政管理活动。当然授权规范的层级、授权的范围和事项，因自动化系统适用的领域、对公民合法权益的影响程度大小而有所不同。例如，电子警察系统的应用就需具备法律、行政法规的明确授权。新修订的《行政处罚法》第41条规定，利用电子技术监控设备收集、固定违法事实的行为，必须有法律、行政法规的授权，且需经过法制审核。〔47〕《道路交通安全法》第114条授予行政机关根据交通技术监控记录资料进行处罚的权力。〔48〕以上可以看作是行政机关使用电子警察系统的授权规范。需要注意的是，前述条款授权的范围限于“利用电子技术监控设备收集、固定违法事实”，不能扩大到利用电子警察系统直接作出行政处罚决定，进行处罚的权力仍然属于行政机关。从当前的立法情况来看，针对电子监控设备的使用问题，只有交通执法和市场监管两个领域有法律和行政法规的授权，环保、海关、农业领域的授权规范位阶是部门规章。〔49〕

其次，转译主体包括行政机关和私营部门的算法设计师，具有立法色彩的转译过程应满足转译主体合法性的要求。以“类裁量基准”的算法为例，裁量基准本身是行政机关根据授权法的旨意，对法定授权范围内的裁量权予以情节的细化和效果的格化而事先以规则的形式设定的一种具体化的判断选择标准，属于行政自制规范。〔50〕行政机关制定裁量基准的权力来自于立法授予的

〔45〕 参见王正鑫：《机器何以裁量：行政处罚裁量自动化及其风险控制》，载《行政法学研究》2022年第2期。

〔46〕 参见前引〔36〕，Danielle Keats Citron文，第1279页。

〔47〕《行政处罚法》第41条规定：“行政机关依照法律、行政法规规定利用电子技术监控设备收集、固定违法事实的，应当经过法制和技术审核，确保电子技术监控设备符合标准、设置合理、标志明显，设置地点应当向社会公布。”

〔48〕《道路交通安全法》第114条规定：“公安机关交通管理部门根据交通技术监控记录资料，可以对违法的机动车所有人或者管理人依法予以处罚。对能够确定驾驶人的，可以依照本法的规定依法予以处罚。”

〔49〕 相关授权规范参见《环境行政处罚办法》第36条、《海关监管区管理暂行办法》第17条、《农业行政处罚程序规定》第37条。

〔50〕 参见周佑勇：《裁量基准的制度定位——以行政自制为视角》，载《法学家》2011年第4期。

行政裁量权，其将裁量基准转译成算法的过程本质上仍是在行使行政裁量权。行政机关选择与私营部门的算法设计师合作共同制定转译型算法的行为也在裁量空间之内，算法设计师的行为也因此具备了合法性基础。此时，算法设计师可以看作是行政机关手脚的延伸，其行为归属于行政机关；行政机关也需要通过细密的规范设计约束算法设计师的行为。

（三）转译法律的明确性要求

为了缩小算法设计师“转译法律”时的判断空间，行政机关应当尽可能地明确法律的含义。具体而言，在设计系统时，算法设计师需要明确系统将要实现的法律目标是什么，即确定“目标规范”，目标规范是系统运行时具体执行的法律。目标规范和算法之间是对应关系，前者是人类世界中由行政机关执行的法律语言，后者是由系统执行的机器语言，二者要实现的是同一行政目标。例如，闯红灯自动记录系统中运行的算法是用来自动认定闯红灯行为的机器语言，相应的目标规范是《道路交通安全法》第 44 条〔51〕和《道路交通安全法实施条例》第 38 条〔52〕中，红灯亮时禁止机动车通行的规定。目标规范是行政机关的执法依据。转译过程实际上是将目标规范这一法律语言转译成机器语言的过程，转译时需要细化、解释具体的法律用语，明确至机器可执行的程度。仍以闯红灯自动记录系统为例，《闯红灯自动记录系统通用技术条件》（GA/T 496—2014）对如何认定闯红灯行为作了更具体的规定：系统需要监测和记录的闯红灯行为是机动车违反交通信号灯红灯亮时禁止通行的规定，越过停止线并继续行驶的行为。〔53〕自动记录系统至少要记录三张反映闯红灯行为过程的图片，图片需符合《闯红灯自动记录系统通用技术条件》的要求。〔54〕为了减小法律语言转译为机器语言时可能出现的偏差，行政机关通常会发布相关技术标准，自动化系统的设计必须符合技术标准的要求。在技术标准的基础上，有必要事先为转译过程设计更为详细的说明书，尽可能地明确可能会引起算法设计师进行独立判断的问题。说明书应当经过法律专家与技术专家的审核，并应当被允许共享以及不断完善，以促使算法设计师的行为合乎规范要求。〔55〕

行政机关通过发布技术标准和设计转译算法说明书的方式减少法律语言的模糊性，为算法设计师提供更为明确的设计方向。但是，即便说明书的表述极尽详细，法律语言转译成算法的过程仍然存在算法设计师的主观判断空间。处于私主体地位的算法设计师受私益驱动，而行政管理活动需将公共利益作为首要考量因素，为了确保公共利益的实现，行政机关应当全程参与系统的设计过程，担任重要问题的最终决策者。

---

〔51〕《道路交通安全法》第 44 条规定：“机动车通过交叉路口，应当按照交通信号灯、交通标志、交通标线或者交通警察的指挥通过；通过没有交通信号灯、交通标志、交通标线或者交通警察指挥的交叉路口时，应当减速慢行，并让行人和优先通行的车辆先行。”

〔52〕《道路交通安全法实施条例》第 38 条第 1 款规定：“机动车信号灯和非机动车信号灯表示：（一）绿灯亮时，准许车辆通行，但转弯的车辆不得妨碍被放行的直行车辆、行人通行；（二）黄灯亮时，已越过停止线的车辆可以继续通行；（三）红灯亮时禁止车辆通行。”

〔53〕参见《闯红灯自动记录系统通用技术条件》（GA/T 496—2014）第 3.1、3.2 条。

〔54〕《闯红灯自动记录系统通用技术条件》（GA/T 496—2014）第 4.3.1.1 条规定：“系统应能至少记录以下 3 张反映闯红灯行为过程的图片：a）能反映机动车未到达停止线的图片，并能清晰辨别车辆类型、交通信号灯红灯、停止线；b）能反映机动车已越过停止线的图片，并能清晰辨别车辆类型、号牌号码、交通信号灯红灯、停止线；c）能反映机动车与 b）图片中机动车向前位移的图片，并能清晰辨别车辆类型、交通信号灯红灯、停止线。”

〔55〕参见前引〔34〕，丽莎·A. 谢伊、伍德罗·哈特佐格等文，第 295 页。

（四）转译过程的透明度要求

首先，转译型算法制定过程应满足公开的要求。转译过程公开的理论基点在于对公民知情权的保障，此处的知情权是指政治上的民主权利，即公民依法享有知道国家的活动、了解国家的事务的权利，国家机关有依法向公民及社会公众公开自己活动的义务，这是人民主权原则的延伸。〔56〕转译过程公开的内容包括公开转译主体、转译目的、转译依据以及源代码等，算法公开体现的是算法透明原则的要求。就具体规制手段而言，算法透明包含告知义务、向主管部门报备参数、向社会公开参数和存档数据、公开源代码等不同的形式。〔57〕算法公开的程序可参照行政规范性文件的公布程序。2008年起实施的《湖南省行政程序规定》率先规定了对规范性文件的统一登记、统一编号、统一公布制度，其后，“三统一”制度被推广至其他省份，目前已被中央层面法律文件纳入。〔58〕

其次，应在算法公开的基础上增强算法的可解释性。反对算法公开的理由之一是“算法透明≠算法可知”，即考虑到披露对象的技术能力、算法的复杂性、机器学习和干扰性披露四重因素，即使向公众公开源代码，公众也未必会理解算法的工作原理。〔59〕对行政机关施加解释算法的义务并非要求其准确地说明算法的工作原理，由于“算法黑箱”的制约，这可能在技术上也是不可行的。行政机关的解释性义务只需要做到提供必要的信息证明系统产生的结果是合理的即可。换言之，行政机关需要提供有关其自动化系统背后的目的及其通常如何运作的基本信息，需要表明在设计系统时已经仔细考虑了关键的设计选项，也可能需要借助公认的审核和验证工作来证明系统确实能够运行并生成预期的结果。〔60〕对行政机关施加公开算法和解释算法的义务，一方面是为了满足自动化行政自身合法性的要求，另一方面也有助于个人对自动化决策提出质疑，引发关于技术的辩论，从长远来看可以促进社会对新技术的接受。

五、自我学习型算法的控制

针对转译型算法，可以通过控制转译过程的合法性来保证算法决策的合法性，确保系统始终处于行政机关的控制之下。此时的规制逻辑是通过形式合法性来解释行政正当性，核心技术是评估行政与法律的一致性。〔61〕但自我学习型算法是根据预先设定的“学习规则”，学习训练数据之后生成的，本身具有不确定性。自我学习型算法无法满足形式合法性的要求，需要探索新的合法性框架，相应控制方式应在新的合法性框架下展开。

（一）“民主—科学”的合法性框架

1. 构建合法性框架的目的

自我学习型算法的适用需要具备合法性基础的本质原因是要保证系统行使行政权时像行政机

〔56〕 参见刘莘：《行政立法研究》，法律出版社2018年版，第167－168页。

〔57〕 参见汪庆华：《算法透明的多重维度和算法问责》，载《比较法研究》2020年第6期。

〔58〕 参见《国土资源部办公厅关于实行规范性文件“三统一”制度的通知》（国土资厅函〔2015〕523号）。

〔59〕 参见沈伟伟：《算法透明原则的迷思——算法规制理论批判》，载《环球法律评论》2019年第6期。

〔60〕 参见前引〔3〕，卡里·科利亚尼斯文。

〔61〕 参见前引〔8〕，王锡锌文。

关一样受到控制。传统法律体系对公权力的控制机制，使得公民可以充分相信行政机关在行使权力时始终以维护公共利益为目的；而逸脱了法律控制机制的系统，难以使公民相信其同样以维护公共利益的方式运转。换言之，控制系统是为了建立起公众对系统的信任。公众对系统的不信任不仅会导致系统本身合法性基础缺失，还会引发公众与系统的提供者——行政机关之间的信任危机。尽管自我学习型算法具有“黑箱”性质，其决策过程难以为人类理解，但这并不意味着人类无法对其建立信任。正如在医疗领域，尽管患者对药物或药物治疗的工作原理不甚了解，但其仍然愿意将生命健康托付给通常难以理解的治疗手段；问题的关键不在于人类是否知道特定药物的作用机理，而是该领域内是否存在充分的规则、制度和专业知识给予我们信心，使我们对治疗手段建立信任。〔62〕

2. 合法性框架分析

行政管理过程偏离形式合法性要求的问题并不限于自动化行政领域，只不过在自我学习型算法上尤为突出。当代行政是目标导向的积极活动，行政机关在目标界定、手段选择等方面，都拥有自主进行权衡和选择的权力；目标导向的行政，意味着法律对行政的控制，通常只能是宽泛的目标指引而非具体的指令控制。立法提出行政活动的宽泛目标，行政对目标进行判断、权衡以及对实现目标的手段进行选择裁量。〔63〕例如，在风险行政领域，由于立法者不具备关于风险的完整知识，需要广泛授予行政机关裁量权，依法行政实际上被依裁量行政替代。〔64〕行政机关规制风险的活动若要符合现代行政法治的基本要求，至少需要满足两个条件：一是价值合理性，即行政机关设定的风险规制目标能够为公众所接受，符合民众的需求，反映民众的偏好，体现卢梭所说的“公意”的要求，从而具有正当性；二是工具合理性，即行政机关规制风险的手段或措施基于精确的计算和预测，追求功效最大化，具有科学性。〔65〕风险行政背景下，行政机关通过增强行政过程中的民主性与科学性，来补强行政活动正当性。“民主—科学”的合法性框架也可以作为自我学习型算法适用的理论基础。

3. 合法性规制目的实现方式

“民主—科学”的合法性规制目的是建立公众对算法的信任。与自我学习型算法相同，诊疗过程对于患者而言同样具有“黑箱”性质，因此，医疗领域信任机制的构建方式可以为算法的规制提供借鉴。医疗领域的信任建立机制有以下三个要点：(1) 医疗服务提供者的能力。以医师为例，医师培训和考核机制、医师资格考试制度、医师执业注册制度、医师的执业规范要求、卫生健康主管部门和医疗卫生机构对医师的监督管理及问责制度等共同建立起一个保证医师专业水准的框架，使得公众即使无法直接评估其实际能力，也能对其建立信任。(2) 保护患者的利益。医学伦理规范和相关制度的存在使公众相信，相较于个人的经济利益，医师会将病人的利益放在首位。美国出台了《联邦反回扣法案》(The Federal Anti-Kickback Statute)、《医师酬劳阳光法案》

〔62〕 See Robin C. Feldman, Ehrik Aldana & Kara Stein, Artificial Intelligence in the Health Care Space: How We Can Trust What We Cannot Know, 30 (2) *Stanford Law & Policy Review* 399 (2019).

〔63〕 参见王锡锌：《行政法治的逻辑及其当代命题》，载《法学论坛》2011 年第 2 期。

〔64〕 参见赵鹏：《知识与合法性：风险社会的行政法治原理》，载《行政法学研究》2011 年第 4 期。

〔65〕 参见戚建刚：《风险规制过程合法性之证成——以公众和专家的风险知识运用为视角》，载《法商研究》2009 年第 5 期。

（Physician Payments Sunshine Act）来监督医师从医药企业获取利益的行为，平衡患者的最大利益与医师个人利益之间的关系。（3）信息的完整性。医师用于诊疗的数据的准确性、诊疗数据使用方式的适当性、诊疗数据的可访问性、可纠错性都有助于增进患者的信任。总体而言，建立信任的路径可以二分：一是建立患者的主体地位保障，对应要点（2）；二是建立诊疗过程的科学性保障，对应要点（1）（3）。两种路径大致可以分别与民主和科学相对应。

（二）自我学习型算法的民主控制要求

与保障患者的主体地位类似，自动化行政中的民主参与是为了使公众获得自尊、自主和自治的心理。〔66〕自我学习型算法的民主控制可以从两方面展开：一是行政机关在制定规制人工智能的法律法规、政策文件时，应当听取公众意见，并提供充分交流意见的平台；二是在算法投入运用阶段，拓宽公众发现、识别算法风险的渠道。

以算法治理为代表的数治主要关注工具有效性和效率，侧重于治理的事实和工具维度，对法治的“价值之治”侧面带来挑战。〔67〕这也导致算法治理中公众意见表达的空间被进一步压缩。反对人工智能立法的理由之一是缺乏精确度的法律难以满足对代码的规制需求。对此的反驳为，法律是在民主程序中妥协的产物，在妥协的过程中，公众不断朝最适当规则的方向达成共识。〔68〕规制算法的规则和政策的形成过程就是一个妥协和不断达成共识的过程，对算法规制的价值选择和目标确定应当以公众的意见为依据。应当规制哪些风险、如何进行价值位阶排序，以及置于何种议程进行规制，体现的是公众希望自己决定生活状态的意愿。〔69〕在参与过程中，公众能够从各种视角了解和理解算法，尽可能地消除对未知风险的疑虑，增进信任。

算法决策过程的瞬时性剥夺了相对人在行政程序中向决策者表达意见的机会，算法决策的黑箱特点使公众难以直接发现算法的技术性错误。对此，有学者提出通过建立“前瞻性基准”（prospective benchmarking）的方式对自我学习型算法的运行情况进行监督，具体而言，在采取算法决策的场景中，行政机关应当随机选取一组同类型的人工执法案例作为基准，公众能够以此作为对比样本，对算法决策结果进行监督，及时发现算法决策中可能存在的错误。〔70〕除此之外，行政机关应向公众提供算法查验途径，即面向用户或公众提供一个公开的查验渠道，使用户、交易者或第三方有机会检验算法能否实现其所宣称的目标，从而对算法的运行机理建立相当程度的了解和预期。〔71〕

（三）自我学习型算法的科学控制要求

自我学习型算法的科学性控制要求体现在两方面：一是对算法的提供者和算法技术的科学性、可靠性的保障；二是对数据可靠性的保障。在对算法提供者的控制方面，行政机关通过算法

〔66〕 参见沈岿：《风险规制决策程序的科学与民主》，载沈岿主编：《风险规制与行政法新发展》，法律出版社 2013 年版，第 308 页。

〔67〕 参见王锡锌：《数治与法治：数字行政的法治约束》，载《中国人民大学学报》2022 年第 6 期。

〔68〕 See Paul Nemitz, Constitutional Democracy and Technology in the Age of Artificial Intelligence Philosophical Transactions: Mathematical, 376 (2133) *Physical and Engineering Sciences* 1 (2018).

〔69〕 参见前引〔65〕，戚建刚文。

〔70〕 See David Freeman Engstrom & Daniel E. Ho, Algorithmic Accountability in the Administrative State, 37 *Yale Journal on Regulation* 800, 849 (2020).

〔71〕 参见苏宇：《算法规制的谱系》，载《中国法学》2020 年第 3 期。

进行治理，是自动化行政行为的直接责任主体，应当承担起对算法科学性的保障责任。第一，行政机关内部应该设立专门的算法审查机构，承担算法审查、算法监测、算法纠错等具体工作。考虑到当前阶段行政机关专业人才不足的问题，有学者建议目前可依托具有相应专业人才、技术支撑和监管能力的行业自律组织，建立起由相关行政机关负责指导、行业自律组织负责实施的算法监管体制。〔72〕第二，行政机关在选择第三方机构共同设计算法时，应当遵循公开透明、公平竞争、公正原则，设计单位的资质、选择单位的程序和标准等信息需向社会公开，并接受监督。

在对算法技术的控制方面，第一，建立算法标准和算法备案制度。统一的技术标准有助于确认某种算法现阶段的科学性和合理性；而算法备案制度便于查明算法风险，明确责任主体。第二，建立算法审查制度。算法设计过程需嵌入算法伦理，因此在设计阶段就应当以立法形式要求算法通过道德审查标准，防止产生不公平后果。〔73〕第三，建立算法影响评估制度，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，在算法系统应用之前就进行独立的社会技术分析。〔74〕第四，开发监督算法运行、监测算法技术可靠性的算法。尽管对算法代码进行实时督导（monitoring）和审计（auditing），需要具备与算法生产和使用相当或超越的技术能力，成本巨大，〔75〕但以技术控制技术既可以推动科技进步，也能有效增进公众对科技的信任。前述控制手段大多是自上而下的机制设计，有可能因为利益关联或认知局限等原因阻碍算法的正常发展，因此，应当鼓励产业界、社会组织及个人创造和发展自下而上的风险识别与防范工具。〔76〕

在对数据可靠性的保障方面，第一，利用数据集缺陷检测技术。目前的人工智能技术已经完全可以为算法开发者提供数据集及训练过程检测工具，主要用于检测训练人工智能的数据集是否存在偏差或缺陷，还可以通过一定的算法检测在数据选取、数据标注、数据清洗以及其他预处理工作过程中是否包含了偏离算法设计目标或足以导致结果发生显著偏差的操作。〔77〕第二，提高数据的互操作性（interoperability）。〔78〕互操作性要求不同行政机关之间共享数据，能够更好地满足自我学习型算法对数据数量的要求，进而提高算法的准确性。

六、结　语

在民主国家，主权统治通过双重形式的透明实现合法性：首先，人民生活在自己制定的规则之下（民主参与）；其次，这些规则的适用能够在打开其解释黑箱的诉讼程序中提出争议（法治）。〔79〕

〔72〕 参见孙清白：《人工智能算法的“公共性”应用风险及其二元规制》，载《行政法学研究》2020 年第 4 期。

〔73〕 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。

〔74〕 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。

〔75〕 参见前引〔14〕，邱泽奇文。

〔76〕 参见前引〔71〕，苏宇文。

〔77〕 参见前引〔71〕，苏宇文。

〔78〕 See Peter K. Yu, Beyond Transparency and Accountability: Three Additional Features Algorithm Designers Should Build into Intelligent Platforms, 13 (1) *Northeastern University Law Review* 263, 290 (2021).

〔79〕 See Mireille Hildebrandt, Law as Information in the Era of Data-Driven Agency, 79 (1) *Modern Law Review* 1, 23 (2016).

这也是传统行政法中行政权获得合法性的途径，行政机关通过严格遵循依法行政原则，获得民主正当性。自动化行政方式面临合法性危机：转译型算法在转译过程中会嵌入算法设计师的判断，而来自私营部门的算法设计师可能尚未获得执行法律的授权，缺乏执法的合法性基础，这一问题在自我学习型算法中更为突出；此外，当前阶段，算法决策有时在事实上超出法律的授权范围，且缺乏畅通的救济机制。

因此，应当结合算法类型对算法进行控制。针对转译型算法，需要保证转译过程的合法性：首先，要有上位法授权行政机关以自动化的方式在某一领域开展行政活动，从而为引入第三方共同设计算法提供法律基础；其次，行政机关有义务细化系统所需执行的目标法律规范，以缩小转译过程的判断空间；最后，转译过程应参考规则制定程序，符合相应程序要求。针对自我学习型算法，传统合法性框架失去作用，应当通过行政过程中的民主性和科学性重构合法性基础，具体控制措施也应从公众参与和算法科学的角度展开。

Abstract: Algorithm in automated administration can be divided into translation algorithm and self-learning algorithm, and the use of algorithm is faced with legitimacy crisis. Algorithm designers in a private position embed their own judgments in translating legalese into machine language, creating the risk of rewriting the law. In addition, algorithmic decision-making sometimes exceeds the scope of legal authority in fact, and lacks the smooth relief mechanism. The legality control method of the algorithm should be adapted to the algorithm type. It is necessary to control the translation subject, the clarity of the translated law and the transparency of the translation process in combination with the nature and technical characteristics of the translation algorithm. For self-learning algorithm, we should first establish a "democracy-science" legitimacy framework, and the algorithm should be controlled from the perspective of building trust in the algorithms by guaranteeing the status of the public and the scientific nature of algorithms.

Key Words: automated administration, formal legitimacy, administrative democracy, administrative science, algorithmic trust

(责任编辑：刘　权　赵建蕊)