论算法规制的价值目标与机制设计

On the Intended Values and Mechanism Design of Algorithmic Regulation

苏宇 /SU Yu

（中国人民公安大学法学院，北京，100038）
(School of Law, People's Public Security University of China, Beijing, 100038)

摘　要：算法规制需要确定以安全和自由价值为主的价值目标，形成内在融贯的价值目标体系。基于这些价值目标，世界范围内的算法规制理论与实践已经确认了一系列相关法律关系，应用了多种制度工具。在此基础上，算法规制应借鉴机制设计理论，结合算法制衡的助力，促成理想的博弈均衡结果，形成符合规制价值目标的基本制度安排。

关键词：算法　规制　机制设计　数据权利

Abstract: The regulation of algorithms should follow a coherent system of values led by safety and freedom. Based on these values, the theory and practice of algorithmic regulation have confirmed a series of legal relations and applied multiple institutional tools. On the ground of current legal practice and with the help of algorithmic balance, the regulation of algorithms should draw on the mechanism design theory and pursue widely welcomed equilibriums to establish an institutional framework that is in accordance with the values mentioned above.

Key Words: Algorithm; Regulation; Mechanism design; Data rights

中图分类号：N0　　**文献标识码：**A　　DOI: 10.15994/j.1000–0763.2019.10.002

人类正在逐渐进入“算法统治的时代”。[1]时至今日，大数据、云计算、区块链、人工智能等信息科技的浪潮正在席卷全球，信息权利正在成为日益重要的权利形态；与此同时，智能家居、智慧物流、图像处理、智能网联汽车、医疗影像识别、人脸识别与轨迹追踪等基于算法的产品或服务迅速发展，使得算法也随之进入了法律的视野。算法的运用为社会经济的发展带来了强劲的动力，也为社会生活带来了巨大的便利，但与算法相伴的风险也如影随形，导致算法规制成为学界与实务界日益关注的主题。

所谓“算法规制”（algorithm regulation），本质上是对算法应用的规制，它不同于“通过算法运行的规制”（algorithmic regulation），是指基于国家安全、社会经济秩序维护或个人权益保障的需要，深入到某些软件或程序所应用的算法层面，对算法进行引导、规范或监管。

算法规制已经在隐私权保障、知识产权保护及反歧视领域逐渐开展，对于规制算法的理论基础或如何规制算法的讨论也时常可见，如爱德华兹（Lilian Edwards）与维勒（Michael Veale）有关算法解释权的讨论、[2]班鲍尔（Jane Bambauer）与扎斯基（Tal Zarsky）有关规范算法自动决策的讨论、[3]雅内拉（Philip N. Yannella）关于欧美对用户画像算法的不同规制路径之探讨[4]等研究中都明确论及上述算法规制问题。质言之，算法规制在学界与实务界正日益成为一个重要的新兴主题。

随着算法不断承担更加重要、更加关键的社会经济角色，算法规制的一系列根本问题也浮出水面：算法规制应当指向何种价值目标？规制的整体制度框架呈现何种状态，目前有哪些规制工具可用？算法规制的机制设计较之一般的风险规制有何特点？这些都是算法规制的议题下最根本、

收稿日期：2019年3月6日
作者简介：苏　宇（1985–）男，广东高州人，中国人民公安大学法学院讲师，研究方向为行政法学、数据法学。Email: leafcityeve@126.com

最关键的问题。

一、算法规制的价值目标

市场主体的活动会产生社会和经济上的负外部性，规制的本质就是对负外部性的消除。“负外部性”是一个价值相关概念，易言之，任何规制活动都必然是在一定法律价值目标的指引之下进行的。自法理上观之，法的“目的价值”主要包括公平、正义、安全、自由、效率等，它们具有多元性，并且与人的需求和社会关系的多样性紧密相连。[5] 由于算法应用的情境众多，涉及的法律价值目标也较为广泛，算法规制所面对的负外部性包含了相当丰富的内容。例如，在算法未受任何外部规制的前提下，自动驾驶汽车可能面临重大安全风险；区块链及其衍生应用可能冲击主权国家的金融安全、引发赌博或逃税等其他违反法律秩序的行为；[6] 将匿名化数据重新整合进行分析可能识别出个人信息而侵犯公民隐私，[7] 用户画像和杀熟方案可能影响交易公平、侵犯消费者的知情权，[8] 等等，都引致了规制的需求。但是，对算法的规制又不能过于僵化或严厉，在消除其负外部性的同时摧毁算法本身发挥作用的基础。例如，假设对区块链强制使用同一种共识机制，将很可能使得区块链的众多类型及中小型区块链不复存在；假设强制用户画像和智能推荐的代码完全开源，则有可能对创新形成负面激励、导致企业撤离或者核心技术流失。因此，算法规制的价值目标具有复杂的内容和结构，也具有精微的尺度及边界，需要通过一定的方法进行梳理与整合，形成一个具有一定解释和证立能力的价值目标体系。

1. 价值目标的确定

算法规制价值目标的确定，取决于规制者对算法的负外部性之认识及价值权衡之判断。经济学上所谓“负外部性”，在法学的视野中体现为对各种法益的侵害。算法运用所可能产生负面影响的法益，既包括个人的人身权、财产权、政治自由和人格尊严，也包括难以通过还原主义方式分解的国家安全、社会管理秩序及公序良俗，等等，都属于算法规制需要考虑的价值目标。与此同时，算法的运用也会惠及一定范围内的法益；在特定的技术条件下，每一种负外部性的出现往往伴随着正面的应用价值，而且算法所促进的法益与所损害的法益可能在类型上即互不相同。此时算法规制的具体价值目标可以通过三种方法来确定：一是通过价值位阶或价值秩序的方法，使某一种价值占据优势位置，如未能为此种价值提供充分保障，则不允许通过损害此种价值而实现其他价值；二是通过价值转换与价值平衡的方法，使不同的价值可以相互比较，选择使整体价值最大化的制度安排；三是通过多目标规划等手段，先设置一定的优化标准（例如帕累托最优），再寻求在各方主体的价值判断体系之间能达成的最优机制设计。

这三种方法都有一定的适用条件和范围，也各有其优劣，具体的抉择应由立法机关进行决断，再由行政机关与司法机关在立法决断的范围内、通过法定的程序进行解释。在立法意图不尽明确时，无论依大陆法系还是英美法系的学理，行政机关均有一定的判断、裁量或解释法律的余地。无论行政机关采取何种理解，行政机关根据此种余地作出的价值目标设定或解释均应接受司法审查，在价值目标问题上使行政行为满足目的正当性原则的要求。

2. 价值目标体系的形成

算法规制不仅需要确定价值目标，而且需要形成内在一致的、融贯的价值目标体系。算法的各种应用在社会经济系统中具有一定程度上的关联性，分散而独立地确定某一算法规制的价值目标，有可能形成规制目标之间的混乱与冲突，进而对信息科技与信息经济的整体生态造成不利影响。例如，人工智能中许多有监督学习（supervised learning）的项目与大数据的利用有密切的关系，如果过于高估数据权利的价值，在数据采集与利用方面采取了较为严厉的限制，很有可能波及到人工智能的发展；程序模块的发展与突破高度依赖于Github一类的开源协作社区，如果由于某些程序模块的风险而否定开源协作社区的正面价值，无疑是因噎废食。因此，价值目标及实现目标的规制工具之间的协调和整合殊为必要。

由于不同应用情境下算法运用所涉及的法益丰富多样，而且算法的运用还在飞速变化发展，尝试明确界定并刻画这一价值目标体系尚不可行。不过，算法规制并不需要从零开始建构价值目标体系，它的价值目标体系可以在数十年来不断完善的风险规制基础上形成，需要加以强调和

变革的仅仅是算法规制有别于一般风险规制的特征。易言之，行政法中的风险规制理论与实践已经为确认受保护的法益目标、评估风险（负外部性影响的程度以及实现的概率）、选择规制工具及平衡成本收益等储备了系列方案，算法规制的价值目标设定与制度建构完全可以借鉴这些方案进行。算法应用风险的主要特殊之处有三：一是范围宽广、形态繁多，风险的危害后果、隐蔽性、传播速度和控制难度等依应用情境和技术路线的不同而千差万别。二是评估难，大型软件工程或程序项目的测试与评估较之一般的风险评估（如毒理学分析、FEMA分析等）更难，无人驾驶等复杂程序的风险评估更需要长期积累经验和实验数据。三是影响深远，算法层面的风险往往伴随着算法运用的巨大收益，也有可能影响到一种算法的未来发展，一些过去未充分显现价值的算法在可能日后有巨大发展潜力（例如人工智能发展史上的类神经网络方法）；不仅如此，某些算法的风险一旦出现，将可能在全世界范围内造成巨大的负面影响，例如SHA–2乃至SHA–3加密算法若被破解，全球区块链生态将受到根本性的冲击。这使得算法规制的价值目标体系与一般风险规制应当有所不同。

首先，算法风险范围宽广、形态繁多，这要求算法规制的价值目标体系能够包容众多的法益。其次，算法风险评估难，而且对算法的评估和规制影响深远，就不应轻易对某一种算法的价值或作用下结论，而是在能够清晰评价部分具体程序设计的基础上，采取更加谨慎也更富弹性的价值目标结构。再者，只有极少数算法可能引起无可挽回的巨大风险或者产生严重的伦理问题，信息社会中大多数算法所引起的风险并不会产生不可逆转的重大危害，相反，算法自身的发展却非常迅速，算法应用给社会带来的增益亦不可估量，这要求算法规制在保证控制重大风险底线的基础上，更多地给算法及其应用以自由发展之空间。

由此，自整体上观之，算法规制的首要价值目标应当是安全，此处的“安全”本身是一种包含多种重要实质性价值目标的价值形式，算法安全就是要避免算法应用在国家安全、司法公正、个人的生命权、身体权、人格尊严等重要法律价值上出现无可挽回的巨大法益损失，或者冲击人类文明的伦理底线。在保证安全价值的基础上，不同的价值目标之间应当更加侧重于自由，此种自由既包括公民的言论和表达自由，[9]也包括企业和个人依法进行科学研究和经济活动的自由。由于信息科技（尤其是人工智能）在国家战略、社会经济发展和人类自身发展问题上的特殊重要性，算法应用的相关经济活动也应获得自由价值之支持，如涉及权利或法益之冲突，可以采取诸如阿列克西之重力公式（die Gewichtsformel）[10]或规制的成本收益分析（cost–benefit analysis）[11]一类的权衡机制加以解决。其余价值目标（如效率、公平、秩序）及法益（如原创性、商业秘密、劳动权）等，应当在法律明文规定的范围内严格依法进行保护，避免法律价值在单纯的法律原则层次过度延伸和解释而影响到具有战略意义的信息科学技术及其应用的研究与发展。算法的发展日新月异，算法之规制影响深远，必须通观全局，以安全和自由的价值统摄算法规制的价值目标，以发展的眼光推进规制的进程，避免算法规制的制度框架建构与机制设计吹毛数睫、刻舟求剑，甚至产生寒蝉效应（chilling effects），阻碍算法创新和科技进步。[12]

二、算法规制的制度框架

世界范围内的算法规制实践已经形成了一系列制度工具，它们与新兴人格权益、数据权利的确立与保障共同构成了算法规制的制度框架。这些新兴权益与制度工具在近年来国内外已经广泛进入法律实践。美国、日本、韩国、德国、新加坡、爱沙尼亚等国家先后就算法规制与治理形成了相关立法实践，[13]欧盟《通用数据保护条例》（*General Data Protection Regulation*，即“GDPR”）则更包含了大量典型的相关规定。不过，算法规制的制度框架远未称得上成熟，其法律关系架构与各种制度工具仍然处于探索发展的状态。

1. 算法规制的法律关系基础

法律关系的内涵非常丰富，以拉伦茨（Karl Larenz）的广义权利义务框架为例，法律关系包括狭义上的权利（Recht）、权能（Befugnisse）、权限（Zuständigkeit）、取得期待（Erwerbsaussichten）、狭义上的法律义务（Rechtspflicht）、法律上的拘束（RechtlicheGebundenheit）、职责（Obliegenheiten）、负担（Lasten）等，展开为一个具有复杂层次的体系。

[14]法律关系的核心是权利义务关系，而权利更是法律关系中最引人瞩目的部分，是构筑法律关系的出发点。除权利以外，各国公法中一般都通过某种形式认可一定范围内的法律上之利益，它们在公法请求权与救济可能性上不如权利，但对于个体而言亦具备类似于权利的意义，故常合称“权益”。算法规制的部分制度工具，可以看作是为保障各种合法权益而规定的若干义务或责任，例如算法解释即主要是基于用户知情权而对算法设计者施加的一项法律义务。因此，算法规制的制度框架需要先确认各种各样的算法相关权益，尤其是确立与算法有关的新兴人格权益与数据权利。

（1）新兴人格权益。新兴人格权益主要包括当代的信息隐私权以及人工智能应用的法律人格。传统上肇始于沃伦和布兰代斯的隐私权概念，最初只包含防止侵扰的安宁隐私权和防止个人信息泄露的信息隐私权，后者经过岁月变迁已日显复杂，而信息时代这一权利的范围更是不断扩展。在信息整合技术和身份识别技术不断发展的条件下，如前所述，传统上不涉及隐私的信息有可能被分析出符合身份识别标准的个人信息，因此隐私权的保护范围也随之扩展。[15]与此同时，人工智能应用的法律地位问题也日益引人关注。最初，人工智能应用的法律地位依附于开发者的法律地位，随着人工智能的发展，不仅有学者明确主张人工智能应用（包括基于人工智能的机器人或其他终端）的法律人格，也有部分国家开始尝试加以承认。（[1]，p.69）由此，算法规制正在超越传统上由人类权利义务关系形成的法律关系框架，而将法律关系延伸到算法本身，这将是一场根本性的变革。

（2）数据权利。自既有国内外法律实践观之，目前已经获得一定程度承认的数据权利，包括个人的数据携带权、数据访问权、数据更正权、数据擦除权（被遗忘权）、数据采集与处理的同意权、拒绝权、用户知情权、限制处理权，等等。数据权利中既包括基于个人信息和隐私的人格权利，也逐渐形成了“个人数据即财产”的财产权利观念及其法理内涵。（[7],p.111）基于数据权利观念，欧美国家在个人、数据处理企业、数据应用方与政府之间之间逐渐形成一个日益精巧的控制权配置结构，致力于平衡数据安全、科技应用、信息经济发展和个人信息保护的需要。易言之，这些权益包含了立法者的价值决断和价值目标设定，构成了算法规制的法律关系边界，也为算法规制提供了部分法律手段。

这些新兴权益在我国法律中亦有若干零散的体现，在立法实践中，个人信息的法律地位已为2017年制定的《民法总则》明确承认；在司法实践中，不少新兴权益亦获得法院的认可，例如王艳春与王茹香、李春香等隐私权纠纷一案（北京市门头沟区[2017]京0109民初4611号人民法院民事判决书）中承认的个人信息权、孙旭东与平安银行股份有限公司深圳市鑫富源投资咨询有限公司隐私权纠纷一案（深圳市福田区人民法院[2016]粤0304民初24741号民事判决书）中承认的隐私维护权、北京百度网讯科技有限公司与上海汉涛信息咨询有限公司其他不正当竞争纠纷一案（上海知识产权法院[2016]沪73民终242号民事判决书）中承认的信息价值等，但这些新兴权益的具体内涵及外延还远未有明确定论，仍亟待深入认识。

2. 算法规制的制度工具

尽管算法规制依然是一个新兴的主题，相关的制度工具发展却相当迅速，在世界范围内已经初步形成了一套全方位的监管框架。自既有法律实践及理论热点观之，算法规制的主要制度工具包括如下数种：

（1）软件登记与材料留存。这是一种基本的规制手段，主要是解决算法责任的可追溯性问题以及提供最基本的风险预判，是实施其他监管措施和追究算法责任的前提条件。例如，我国工信部制定的《移动智能终端应用软件预置和分发管理暂行规定》要求互联网信息服务提供者或其他平台经营者登记应用软件的提供者、运营者及其他信息，留存应用软件及其版本、上线时间、用途、MD5校验值等信息以备追溯。在这一方向上，未来有可能发展出诸如算法登记或算法备案一类更精致的制度工具，但仅限于具备特殊重要风险的应用情境。

（2）算法解释。当一种算法可能包含较大风险时，法律可以通过设定算法解释义务或赋予用户以算法解释权的方式展示算法的运行机理与基本逻辑结构。尤其对于算法自动决策的一些重要应用情境，算法缺乏透明度、决策过程不公开以及决策理由不足等问题，已经形成了需要进行解

释的“算法黑箱”。[16] 例如，在数据处理领域，因为机器学习和自动决策而认为自己即将或已经面临侵害的个人，可以要求知晓个人数据自动处理的逻辑，也可以向算法自动作出的决定提出异议，并要求更正错误的决定。（[13]，p.30）这就是算法解释的一种制度安排。自更宽泛的含义上看，算法解释还应包括应用说明或数据使用政策（data use policies）等用户操作指引，违反这方面的要求可能引致沉重的法律责任。例如，2018年，谷歌因未为用户提供清晰易懂的数据使用政策，违反了《通用数据保护条例》的规定，被法国数据保护监管机构处以5000万欧元罚款。[17] 不过，算法解释在实践中有时并不容易，经常面临所谓的“不可解释隐忧”，[18] 尤其是对于参数、变量动辄以十万计的大型程序，算法解释之于开发者的负担过于沉重，当前开发者在规定期限或合理时间内未必能够作出准确的解释；即使作出了解释，由于知识和技术的鸿沟，算法解释也很可能收效甚微。

（3）权益保障设计及安全措施。对于某些算法应用情境，法律规定算法设计者必须提供针对某些合法权益的权益保障设计，或者其他安全措施。这些设计主要是抽象的功能性要求，个别情况下也会有具体的算法限制。例如欧盟《通用数据保护条例》第二十五条就既规定了一般的权益保障原则（数据保护原则和个别数据处理的特定目的原则），又规定了匿名化、数据最小化、默认不可访问个人数据等具体的数据处理限制规则。美国加利福尼亚州2018年为保护商业数据中的个人信息而制定的《消费者隐私法案》（*The California Consumer Privacy Act of 2018*）1798.125节也有较为完整的反歧视义务规定，保障消费者的平等权。又如，我国公安部《互联网安全保护技术措施规定》第七条及第八条要求互联网服务提供者、联网使用单位及提供互联网接入服务的单位落实一系列安全保护技术措施，包括防范入侵措施、冗灾备份措施等。此种规定较为灵活，在未来的算法规制中有可能成为必不可少的关键法律手段；但是，权益保障机制是否具备、是否达到法律的标准，行政与司法上有时存在较大的判断余地，需要相当程度的法律方法支持。

（4）算法标准。随着某一领域软件工程的日益成熟，一定程度上的算法标准有可能成型。目前国内外已经形成了一系列具体的算法标准，或者由政府制定，或者由社会组织制定。例如在区块链领域，中国区块链产业和技术发展论坛就制定了《区块链数据格式规范》等标准。但是，对于算法规制而言，更值得注意的是包含一定法理内涵的抽象性价值标准。例如，夏格尔–费弗科恩（Karni Chagal–Feferkorn）主张模仿法律中常见的理性人标准或专业理性人标准，主要基于侵权法的法理，对算法决策者（算法自动决策的应用）建立“合理化算法”标准（a “reasonable algorithm”standard）。[19] 更引人注目的一个标准是公平性标准，在反对算法歧视方面，随着“公平机器学习”（fair machine learning）的呼声日益高涨，算法上的公平标准已经越来越深入，要求生成公平合成数据（fair synthetic data）、设计公平分类器（fair classifier）直至进行公平数据披露（fair data disclosure）等的标准化数据处理过程的论述也随之涌现；[20] 也可以用程序正则性（procedural regularity）标准进行统合，以实现法律上正当程序之保护，避免各方主体在智能自动化决策面前被区别对待。[21] 易言之，在算法公平这一总体性标准下还可以形成一系列子标准。在可预见的将来，抽象性的算法标准可能会成为算法规制的重要手段。

（5）技术接口与监管便利条件。经营者和网络平台需要为某些行政机关提供技术接口，例如我国《反恐怖主义法》第十八条、《公共互联网网络安全威胁监测与处置办法》（工信部网安[2017]202号）第六条第二款等；此外还有要求相对人提供技术支持或监管便利条件的一些法律规范，例如《电子商务法》第二十条、《区块链信息服务管理规定》（2019年国家互联网信息办公室令第3号）第十八条第一款等。从严格的意义上看，这些技术接口或技术支持并不是针对某一种算法，而是针对某一类经营者或网络平台，但这一制度工具完全可以用于算法规制之中，例如可编辑区块链（editable blockchain）技术就利用变色龙哈希函数（chameleon hash function）创造一个可以编辑特定区块的“陷门”，配合留存修改记录等技术措施，这一“陷门”可以使可编辑区块链技术适应某些有特殊监管需要的应用情境。

（6）算法责任。算法的不正当应用如果引起法律规定的危害后果，则程序的设计者或运营商须负一定法律责任。国务院于2017年发布的《新

一代人工智能发展规划》（国发〔2017〕35号）中要求“建立健全公开透明的人工智能监管体系，实行设计问责和应用监督并重的双层监管结构，实现对人工智能算法设计、产品开发和成果应用等的全流程监管。”此处就明确表示了“设计问责”的监管结构要求。算法责任的本质是算法设计责任。算法应用引起法益侵害后果的原因，既可能来自于算法自身的不稳定、不可靠，也可能来自于算法调用数据与信息的准确性，还可以来自于对算法漏洞的恶意利用，但归根到底在于算法设计本身缺乏足够的安全性、稳定性与可靠性。在未来，启用完全自动驾驶的智能网联汽车有可能成为算法责任的一个实践焦点。

除以上在算法规制中已经得到广泛应用或密切关注的主要制度工具，还有一些初步见于法律实践或理论研究的法律手段，例如算法审计（audits of algorithms）已经在实践基础上得到了部分学者的积极支持；[25] 又如技术认证或认定已经在一定程度上广为应用，我国的密码技术检测与认证即是一例。在未来，算法规制的更多制度工具有望被创造及应用，这是一个已经可以窥见的必然趋势。

但是，随着算法相关法律关系的丰富和制度工具的不断发展，一个新的问题亦随之出现：面对林林总总的权益保障需求及制度工具，何种机制才能有效地实现算法规制目标？这就使得规制者不能不面对机制设计的挑战。

三、算法规制的机制设计

算法规制面临的机制设计挑战是空前复杂的。尽管算法规制方兴未艾，许多算法（甚至包括编写算法的多种计算机语言）也还处在成长期，算法规制已经显示出较高的专业性与精确性要求。与此同时，算法规制所负载的风险亦在增长。自国内外法律实践观之，当前的算法规制尚主要在智能推荐、智能招聘、产品与服务定价以及刚刚起步的无人驾驶等方面着力；而在未来，算法规制可能需要全面介入无人驾驶、智能投顾、医事服务、工业制造、犯罪侦查乃至司法裁判等领域；与此同时，计算机语言和算法自身也在不断变化发展，使 用C、C++、C#、Go、Java、JavaScript、Julia、Lua、PHP、Perl、Python、R、Ruby等语言中的不同组合完成的程序项目日益常见，程序员的专业领域也开始不断细分，算法审计与监管的难度与日俱增。一些领域算法复杂，关涉重要权益，同时潜藏巨大风险，算法决策的得失甚至有可能影响全球而又不可逆转（例如算法决策结果在大型公有链的主链上广播并被确认）。质言之，信息技术和信息产业的飞速发展使得算法规制面临更深刻的机制设计考验。

面对新的规制对象，同样新兴的机制设计理论为算法规制提供了一个富有意义的参考。机制设计理论（mechanism design theory，或译“制度设计理论”），又可称反向博弈论，通过将机制设计目标转化为一定的博弈结果目标，进而逆推博弈过程和约束条件，从而寻找出最有利于实现目标的制度安排。[21] 自赫尔维茨（Leonid Hurwicz）、马斯金（Eric Maskin）和迈尔森（Roger Myerson）等学者开创这一理论起，数十年来，机制设计理论已经在国外的立法和政策制定中得到了广泛的应用。算法规制的各种制度工具若致力于精确调控算法运行、最优化目标价值的实现，也应当从机制设计视角出发，形成有效的制度工具组合，在保障安全与自由等优先级目标价值的同时，避免对算法设计与应用企业产生过度的规制负担，窒息信息经济与信息社会的生命力。

自机制设计理论的一般原理观之，算法规制的机制设计应当注意形成以下若干基本安排：

第一，尽可能清晰界定算法风险及收益。如果不明确某一规制旨在保障的权益目标及相关权益可能具体承受的风险后果，规制就失去了最本质的意义；如果不明确规制所影响的利益，则规制可能变为单方面的压制，随意施加过于沉重的负担。更重要的是惟有明确算法风险与收益，才能明确反向博弈设计的理想均衡状态，以及精准确定这一均衡中各方的策略选择。尽管有时算法风险难以预测和衡量，仍需要尽可能大致判定受影响的法益、出现某种风险的概率及危害范围；如果有重大风险后果连大致判定概率及危害范围也难以做到，则应谨慎限制此种算法的使用，待算法成熟后再投入应用。因此，首先对存在显著风险的算法要求算法解释和测试是必然的前提，正如规制算法合同（algorithmic contract）的一个重要前提就是使算法之目标清晰化，[22] 由此才能对算法的收益与风险深思熟虑，这是不可或缺的步骤。

第二，为各方主体建立合适的行动策略结构。由于算法风险所影响法益及其归属主体的多样性，规制中也应存在多元化的行动策略结构，也就是多种制度工具的组合。每当算法风险影响一种直接归属于私主体的法益，则为私主体配置一种防御机制或请求权；当算法影响一种归属于社会或国家的法益，则考虑为具有代表性的团体或组织配置请求权，或为规制者配置一项规制权限。所有请求权与规制权限的配置应与法益受侵害的可能性及危害程度大致相称。如果侵害风险较高，则应配置事前的防御机制，例如算法注册、算法审计或安全保障设计等；也鼓励市场自行发展针对算法风险的防御性应用，使社会显示更准确的公共物品供应信息与条件，使各方主体更准确地报告其真实类型，并建立个人用户防御算法风险的补贴机制。

第三，为理想的均衡状态创造条件。机制设计本质上是反向博弈推演过程，从一个理想的均衡解倒推各种行动策略的初始收益空间配置。需要注意的是，各方博弈的均衡状态只是一个稳态，未必是最优状态，在最基本的一些非合作博弈类型中，囚徒博弈和蜈蚣博弈是“双输”的均衡，智猪博弈是搭便车者片面获益的均衡，都不是理想的均衡状态。易言之，理想的均衡状态是符合价值目标的，而自然发生的博弈过程经常不导向这些目标；若要产生符合理想目标（例如帕累托最优）的均衡结果，就要改变各方主体报告信息以及采取其他行动的成本与收益结构。机制设计应当致力于减小各方的沟通与互动成本，增加造假或披露不实信息的负担，改变非合作博弈的收益结构，在有条件的情况下甚至可以尝试变非合作博弈为合作博弈（例如通过建立信用系统或联盟协议的方式），为理想的均衡结果创造条件。具体到算法规制，首先就应当使存在风险的算法按照风险水平设定算法解释水平，保障用户、相对方或相关方的知情权，并且设置算法解释的可验证性标准、检测机制以及相应的法律责任；其次应当基于科斯定理合理配置初始权利，并通过发放补贴等方式改变预算平衡，观察真实的风险防范需求与算法应用收益，进而发现供给公共物品的最优水平；此外，对于机制设计所难以解决的一些问题，如算法伦理问题（反歧视、保障人格尊严等），则通过建立算法标准、设置权益保障机制和监管接口的方式解决。

以上基本安排下可以容纳各种各样的具体机制设计。但是，算法规制仍有其更加独特之处。机制设计理论要求对参与博弈的角色、行动策略、收益与损失等有清晰的界定，在部分应用情境下非常适合于算法规制，甚至直接融入算法设计之中。将机制设计融入算法中运行的应用研究，不仅在世界范围内有大量探索，在我国也早已得到开展。[23] 算法本身已经定义了各种各样的变量、参数与函数，一定条件下可以直接转化为机制设计中收益空间或行动策略的参数，甚至可以直接清晰界定一种行动策略。在此种情形下，机制设计可以直接内置于算法架构之中。以基于POW共识机制的大型区块链为例，由于数学上特定区间哈希值的求解比验证困难得多，而碰撞一定长度的特定哈希值则几乎完全不可能，同时攻击又需要消耗巨大的算力支出，导致所有攻击行为都需要比合作行为付出更为高昂的成本，从而实现了激励相容的均衡结果。基于序列到期可撤销合约（Revocable Sequence Maturity Contract）或哈希时间锁定合约（Hashed Timelock Contract）的侧链交易也是参与者以算法对算法、实现相互制衡的一个典型实例。不仅如此，算法自身就在不断发展出适合直接以算法进行规制的社会关系，例如算法合同，尤其是区块链上的智能合约，通过算法标准、形式检测、时间锁、智能风险预警等算法手段谋求算法之内的制衡，可能比传统的规制手段更有效。因此，在未来，算法规制更应当注重算法制衡体系的建设，把机制设计转化为算法中内置的安全保障设计和权益制衡设计；更借助市场开发的算法风险防御应用和竞争性的算法应用，最大限度地实现算法风险在信息经济生态内部消解，避免规制成本以及规制失灵所带来的额外负担。

我国法律体系中算法规制的内容方兴未艾，也为算法规制的机制设计留下了充分的空间。在我国行政法规层次以上的法律规范中，1991年制定的《计算机软件保护条例》第七条第一次提及“算法”，但却明确表示本条例保护的范围不包括算法；1999年制定的《商用密码管理条例》第五条第二款中要求“编制的商用密码算法具有较高的保密强度和抗攻击能力”，这是我国第一次对算法提出明确的要求。“算法”一词第一次进入我国的法律层面是2004年制定的《电子签名法》，但该法仅仅是在附则中提及了这一术语。即使是在部门规章

层面，算法规制的行政立法范例亦为数不多，工信部的《移动智能终端应用软件预置和分发管理暂行规定》（工信部信管［2016］407号）是罕见的实例。此外，存在大量技术性的标准（如住房和城乡建设部《建筑智能化系统运行维护技术规范》、国家密码管理局《密码模块安全检测要求》等），但还没有将算法规制上升到“硬法”的层面。在此种条件下，我国的算法规制完全可以在现有制度框架与工具基础上尝试各种制度工具及其组合，特别是形成基于算法本身的核心规制结构，以最符合信息科技内在规律的方式，达成算法规制的价值目标。

四、结　　语

人类正在进入一个“算法时代”。算法规制绝非一蹴而就的短期任务，它将持续考验规制者的能力与智慧。随着信息科技的日益发达，算法规制的新问题和新挑战也许会远远超出本文的预期，但只要算法规制在价值目标和机制设计的基本问题上坚持正确的认知，算法规制将成为信息科技服务于人类发展的关键助力。

［参 考 文 献］

［1］郑戈. 算法的法律与法律的算法［J］. 中国法律评论，2018, 20（2）: 67–85.

［2］Edwards, L., Veale, M. 'Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking for'［J］. *Duke Law & Technology Review*, 2017–2018, 16(1): 18–84.

［3］Bambauer, J., Zarsky, T. 'The Algorithm Game'［J］. *Notre Dame Law Review*, 2018, 94(1): 2–47.

［4］Yannella, P. 'The Differing US and EU Regulatory Responses to Rise in Algorithmic Profiling'［J］. *Communications Lawyer*, 2018, 33(1): 1–21.

［5］张文显. 法理学（第五版）［M］. 北京：高等教育出版社，2018, 313–314.

［6］苏宇. 区块链治理之现状与思考：探索多维价值的复杂平衡［J］. 中国法律评论，2018, 20（6）: 186–195.

［7］程啸. 论大数据时代的个人数据权利［J］. 中国社会科学，2018,（3）: 102–122.

［8］陶盈. 机器学习的法律审视［J］. 法学杂志，2018, 39（9）: 55–63.

［9］Benjamin, S. 'Algorithms and Speech'［J］. *University of Pennsylvania Law Review*, 2013, 161(4): 1447–1493.

［10］Alexy, R. 'Die Gewichtsformel'［A］, Jickli, J., Kreutz, P., Reuter, D. (Hrsg.) *Gedächtnisschrift für Jürgen Sonnenschein*［C］, Berlin: De Guyter, 2003, 789.

［11］Masurt, J., Posner, E. 'Cost-Benefit Analysis and the Judicial Role'［J］. *The University of Chicago Law Review*, 2018, 85(4): 936–986.

［12］Chagal-Feferkorn, K. 'The Reasonable Algorithm'［J］. *University of Illinois Journal of Law, Technology & Policy*, 2018, (1): 111–146.

［13］汝绪华. 算法政治：风险、发生逻辑与治理［J］. 厦门大学学报（哲学社会科学版），2018,（6）: 27–38.

［14］申卫星. 对民事法律关系内容构成的反思［J］. 比较法研究，2004,（1）: 42–54.

［15］岳林. 个人信息的身份识别标准［J］. 上海大学学报（哲学社会科学版），2017, 34（6）: 28–41.

［16］张凌寒. 风险防范下算法的监管路径研究［J］. 交大法学，2018,（4）: 49–62.

［17］龚鸣. 未履行〈通用数据保护条例〉谷歌遭法国重罚［OL］，人民网，http://sh.people.com.cn/n2/2019/0123/c138654-32565082.html.2019–02–14.

［18］贾开. 人工智能与算法治理研究［J］. 中国行政管理，2019,（1）: 17–22.

［19］Kroll, J., Huey J., et al. 'Accountable Algorithms'［J］. *University of Pennsylvania Law Review*, 2017, 165(3): 633–705.

［20］Raub, M. 'Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias And Disparate Impact Liability in Hiring Practices'［J］. *Arkansas Law Review*, 2018, 71 (2): 529–570.

［21］苏宇. 机制设计理论与中国行政法学的转型［J］. 财经法学. 2018,（2）: 5–19.

［22］Scholz, L. 'Algorithmic Contracts'［J］. *Stanford Technology Law Review*, 2017, 20(2): 128–169.

［23］樊晓香. 基于机制诚实性的显示原理算法比较［J］. 计算机技术与发展，2008, 18（10）: 99–102.

［责任编辑　李斌　赵超］