

CAICT 中国信通院

 新华网大数据中心

算法治理蓝皮书

中国信息通信研究院政策与经济研究所

新华网大数据中心

2022 年 1 月

版权声明

本报告版权属于中国信息通信研究院和新华网大数据中心，并受法律保护。转载、摘编或利用其它方式使用本报告文字或者观点的，应注明“来源：中国信息通信研究院和新华网大数据中心”。违反上述声明者，本院将追究其相关法律责任。

前 言

当前，新一轮科技革命和产业变革蓬勃发展，数字化、网络化、智能化加速推进，为政府和企业数字化转型提供重大发展机遇。2021年，《数据安全法》《个人信息保护法》等互联网顶层设计相继面世，网络领域体系框架基本建成。作为人工智能时代一项关键性网络信息技术，算法正日益广泛地应用在新闻推荐、电子商务、无人驾驶、司法判决、智能诊疗等商业和公共领域，成为推动企业数字化转型和国家治理能力现代化的重要力量。然而，在我们悄然步入“算法社会”的同时，算法歧视、算法霸权、算法黑箱等一系列问题愈演愈烈，日益威胁着个人权利、公共利益乃至国家利益。

习近平总书记多次强调，要依法加强网络社会管理，加强网络新技术新应用的管理，确保互联网可管可控。为及时回应广大人民群众关切，贯彻落实党中央要求，2021年9月，国家网信办等九部门发布《关于加强互联网信息服务算法综合治理的指导意见》，确定了“利用三年左右时间，逐步建立治理机制健全、监管体系完善、算法生态规范的算法安全综合治理格局”的总体目标。2021年12月，国家网信办、工信部、公安部、市场监管总局四部门联合出台《互联网信息服务算法推荐管理规定》，对算法推荐服务做出全面规范，积极回应当下公众反映强烈、较为突出的算法问题，科学构建网络平台问责体系，迈出我国新技术新应用治理的重要一步。

从全球形势来看，算法治理已经成为世界各国共同关注的重要

问题。总体来看，在强化伦理道德规范的同时，算法治理正加速从软法引领迈向硬法规制阶段，算法透明度、算法问责、算法可解释等问题成为算法治理的关键性问题。本蓝皮书遵循算法“场景+赋能”逻辑，从信息内容安全、数据共享利用、市场竞争秩序、社会公共治理四大视角切入，分别梳理算法的典型应用风险、内部机理分析及各国发展态势，对算法治理问题进行逐一剖析和全景化展示。结合我国本土需求，蓝皮书提出以平衡发展与安全为基本遵循、探索算法透明度的落地路径、建立基于风险的算法规制路径、完善算法责任救济机制，以及构建协同互动的算法治理格局五大建议，以期各方提供有益参考。

目 录

一、算法治理成为我国新发展阶段的迫切需求.....	1
(一) 算法治理的趋势背景.....	1
1. 新技术应用推动人类迈入算法社会.....	1
2. 科学把握新发展阶段面临的算法风险.....	2
(二) 算法治理加速从软法迈向硬法.....	3
1. 算法软法治理困境逐步凸显.....	3
2. 各国立法加速推动算法治理落地.....	4
(三) 正确理解算法治理的核心要义.....	5
1. 推动算法创新和风险规制平衡治理.....	5
2. 算法治理的对象和框架体系.....	7
二、算法操纵与信息内容安全.....	8
(一) 信息过载：算法重塑信息分发模式.....	8
(二) 信息内容风险引发全球算法规制热潮.....	9
1. 算法操纵思想成为首要监管对象.....	9
2. 疫情推动国外对虚假信息关注度.....	10
3. 算法推动网络平台责任豁免规则变革.....	11
(三) 我国推进网络平台算法问责体系化.....	11
1. 信息内容安全成为我国算法治理的重点问题.....	11
2. 六大层面细化网络平台信息内容主体责任.....	12
3. 执法、司法等协同推进算法推荐规范化.....	13
三、算法权力与数据共享利用.....	13
(一) 数据喂养：输入环节引发算法问题.....	13
1. 新技术提升个人信息广泛收集的能力.....	14
2. 用户标签成为算法应用的重要数据形式.....	14
3. 数据流通共享不畅影响算法质量.....	15
(二) 数据源头治理成为全球算法监管的共识.....	15
1. 以生物识别信息为代表的敏感信息保护提速.....	15

2.直接明确平台算法的数据处理规则.....	16
3.尝试化解数据流通共享障碍.....	17
（三）我国从数据角度推进算法治理的最新动向.....	18
1.《个人信息保护法》明确自动化决策等规则.....	18
2.《管理规定》首次对用户标签的使用规则作出规范.....	19
3.大数据杀熟成为立法监管中关注的重要问题.....	19
4.我国积极探索数据交易和共享流通制度.....	20
四、算法滥用与市场竞争秩序.....	21
（一）遏制竞争：滥用算法扰乱市场竞争生态.....	21
1.利用算法达成自我优待目的.....	21
2.多国研究报告关注算法共谋问题.....	22
3.算法成为实施限制交易的重要工具.....	23
（二）算法问题进入各国竞争法视野.....	24
1.各国创新规制平台滥用市场支配地位问题.....	24
2.算法共谋的监管难度逐步加大.....	25
3.以透明度作为破解算法竞争损害风险的思路.....	25
（三）我国立法对滥用算法损害竞争秩序的关注.....	26
1.《反垄断法（修正草案）》强化算法规制.....	26
2.国家竞争政策文件强化算法行为指引.....	27
3.《管理规定》为规范市场秩序建章立制.....	27
五、算法驱动与社会公共治理.....	28
（一）治理变革：算法引发公共风险.....	28
1.健康码成为算法在我国社会治理中的创新尝试.....	28
2.算法在劳动领域中的应用成为国内外热点问题.....	29
3.算法在司法、医疗等领域的应用存在待解难题.....	30
（二）欧美高度重视公共领域算法治理.....	31
1.算法引发的劳动权益问题需要理论解答.....	31
2.算法说明解释义务探索落地路径.....	32
3.算法影响评估义务成为算法问责重要手段.....	33

（三）法经济学视角下算法应用的责任机制探索.....	34
六、我国算法治理的未来展望.....	36
（一）以平衡发展与安全作为基本遵循.....	36
（二）积极探索算法透明的落地路径.....	37
（三）建立基于风险的算法规制路径.....	39
（四）建立完善算法责任救济机制.....	40
（五）构建协同互动的算法治理格局.....	41

一、算法治理成为我国新发展阶段的迫切需求

（一）算法治理的趋势背景

1. 新技术应用推动人类迈入算法社会

2020 年，新冠肺炎疫情席卷全球，为经济增长、公众健康、社会秩序等带来诸多负面影响，但同时也成为政府和企业数字化转型的重大发展契机。在 5G、物联网、云计算等新型技术推动下，大量社会经济活动向线上迁移，远程办公、电子商务、在线问诊、网络教学等无接触线上生态呈爆发式增长。在公共治理领域，各地智慧政府建设如火如荼，数据中心、云计算等数字基础设施加速建设完善。在这一进程中，算法逐渐成为经济社会的底层架构，被日益广泛地应用在新闻推荐、电子商务、无人驾驶、司法判决、智能诊疗等商业和公共领域，在不同场景下管理、分类、约束乃至决定整个社会的运行。如耶鲁大学讲席教授杰克·巴尔金（Jack Balkin）所言，我们正逐步迈入一个围绕算法逻辑而组织和运转的“算法社会”。

算法被认为无处不在，但算法究竟是什么？我们却很难明确界定一个具体对象或运作流程来加以解释。算法就像“黑洞”，我们能清晰感受到它的影响，但却并不能对其内部一窥究竟。由于专业差异，不同学科对算法进行了不同视角的理解和观察，例如，计算机科学关注算法模型，哲学强调算法的伦理性质，社会学将算法视为设计者与技术要素互动的产物，法学聚焦算法作为一种规则所体现的作用。大体而言，算法是为了解决某个问题、完成某项任务或达到某种目的而采

取的处理规则、运算指令、策略机制，体现为一种辅助人类决策或自主决策机制。在算法架构中，数据是喂养、训练算法的基础原料，网络平台是提供计算、存储、传输功能的载体，自动化运行是算法应用的重要特点，而目标价值是算法运行的关键。

2.科学把握新发展阶段面临的算法风险

十四五规划《建议》指出，当前和今后一个时期是我国各类矛盾和风险易发期，各种可以预见和难以预见的风险因素明显增多。应当统筹发展和安全两件大事，把安全贯穿国家发展各领域和全过程，防范和化解影响我国现代化进程的各种风险。算法风险作为智能时代一项突出的新型风险，体现出明显的广泛性、复杂性以及不可控性，为传统的监管理念、规制手段、责任机制等带来新的挑战。

算法的多领域应用导致风险的广泛性。当前，算法已经成为平台运行的基础架构，广泛应用于商业和公共领域。九部门《关于加强互联网信息服务算法综合治理的指导意见》指出，算法不合理应用影响了正常的传播秩序、市场秩序和社会秩序，给维护意识形态安全、社会公平公正和网民合法权益带来挑战。与传统的食品安全、交通安全、环境安全等传统安全不同，算法风险已经广泛渗透至各个领域，涉及国家安全、政治生态、市场竞争、人格利益、劳动保障、生命健康等诸多内容。

算法的动态演进导致风险不可控性。相比一般技术形态，算法运行是一种动态过程。在算法运行的数据输入、数据预处理、模型设计及输出等全流程中，算法可能遭遇数据投毒、标签操纵、样本修改、

特洛伊攻击等多种形式的挑战。¹而伴随人工智能发展浪潮兴起，深度学习等算法基于大数据的自我训练、自我学习过程，能够实现自主的参数调整与模型建构，逐步脱离了人类控制，导致风险更加难以预见。

算法运作的不透明加剧风险的复杂性。基于技术鸿沟、商业秘密等原因，算法的运作带有一定的不透明性，监管部门、社会公众和平台用户难以洞悉算法运作的机理，无形中增加了外在监督的难度。算法运行中的风险，既可能是算法本身的设计缺陷导致的，也可能是使用者滥用算法导致的，可能是法律层面需要解决的，也可能是需要伦理层面化解的。这导致算法归责问题打破了以往较为清晰的侵权责任理论框架，加剧了算法问题归责的复杂性。

（二）算法治理加速从软法迈向硬法

1. 算法软法治理困境逐步凸显

一般而言，由于原则、指南、建议、行业标准、最佳实践等软法形式比硬法更加灵活、适应性更强，更容易在算法等新技术领域应用，目前已成为常见的算法治理形式。在世界范围内，IEEE、人工智能合作伙伴组织（Partnership on AI）、生命未来研究所（Future of Life Institute）等均在推动新技术伦理、标准层面做出大量工作。例如，IEEE于2016年发起“关于自治和智能系统伦理的全球倡议”，涉及120多个与人工智能相关的政策、法律和伦理问题。²2019年欧盟《可信

¹ 参见陈宇飞等，《人工智能系统安全与隐私风险》，载《计算机研究与发展》2019年第10期。

² See The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, at <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>.

人工智能伦理指南》《算法责任与透明治理框架》，美国《人工智能倡议》、OECD《关于人工智能的政府间政策指导方针》、我国2021年《新一代人工智能伦理规范》等均提出透明度、问责、隐私保护、反歧视等要求，引导算法安全、透明、负责任发展。总体来看，软法更为灵活高效，相对硬法而言能够快速被采纳修订，而不必经过严格的立法程序，同时软法有利于算法的区分治理，对于低风险算法，软法可有效发挥引导作用，在有效促进创新的同时确保符合伦理框架。

然而，伴随算法技术的快速发展，算法软法治理的困境逐步凸显。一方面，软法缺乏强制约束力，算法透明度、隐私保护、反歧视仅能依赖企业或行业自律，而企业通常以商业秘密、知识产权保护为由，在算法技术鸿沟的遮蔽下拒绝或疏于算法治理，甚至滥用算法技术危害国家利益、公共利益及个人权益。另一方面，算法技术的广泛应用带来责任划分困境，导致责任划分不清、主客观不相一致，或者不符合比例原则，在一定程度抑制了算法创新发展的积极性，亟需在法律层面建章立制、加强规范，促进算法相关行业健康有序发展。

2. 各国立法加速推动算法治理落地

伴随新技术演进迭代，算法在意识形态、经济发展、社会秩序等方面均发挥重要作用。美国大西洋理事会、美国战略与国际研究中心发布多篇报告指出，新兴技术和数据能力将对地缘政治、全球竞争和国际合作产生重要影响，正在加剧地区紧张局势并带来新的安全挑战。在此背景下，从法律层面对算法风险加以规制引导，已成为多国的现实做法。一是规制强度和层级逐步提升。算法治理体现了从国家战略、

道德伦理向行业标准、法律法规逐步推进的态势。美国在联邦和地方层面提出多项算法相关立法，重点关注政府公共部门、雇佣关系、人脸识别等领域产生的种族歧视、政治选举操纵等问题。欧盟相继发布《通用数据保护条例》《数字服务法案》《数字市场法案》《人工智能法案》，对算法和大型平台展开全面监管，体现了欧盟利用规则确保人工智能可信可控、着力维护欧盟技术主权的思路。**二是治理手段更为丰富。**相比通常的事后监管，覆盖全流程的算法评估制度、审查制度、透明度制度等成为各国算法治理的普遍做法。如美国联邦贸易委员会（FTC）要求在训练算法时应对数据集进行提前审计，欧盟强调高风险人工智能系统必须接受事前评估方能投入使用，且应在欧盟数据库中进行事前登记。**三是监管活动日趋频繁。**美国 FTC 在算法监管中发挥关键性作用，以《联邦贸易委员会法》第 5 条、《公平信用报告法》和《平等信用机会法》为依据，提起多项针对算法歧视的执法诉讼。欧盟强化大型平台监管，要求信息内容平台对使用自动工具的行为作出有意义的解释，并强化反垄断领域使用算法进行自我优待、算法共谋行为的处罚力度。

（三）正确理解算法治理的核心要义

1. 推动算法创新和风险规制平衡治理

当前，我国发展的外部环境日趋复杂，提高发展质量、坚持创新驱动发展，实现经济行稳致远、社会和谐安定，成为新发展阶段的重要目标。当前，算法技术为互联网创新发展提供了重要机遇，有助于大幅提升用户信息获取效率，显著改善互联网广告投放效果，进一步

增强我国互联网企业国际竞争力。如字节跳动旗下的 TikTok 和今日头条，凭借算法在国内外竞争中体现出创新优势。网信办等九部门《关于加强互联网信息服务算法综合治理的指导意见》指出，应坚持技术创新，大力推进我国算法创新研究工作，提升我国算法的核心竞争力。在此背景下，**算法创新成为建设网络强国的有力支撑，成为贯彻新发展理念的重要实践。**

与此同时，防范化解各类风险隐患，积极应对外部环境变化带来的冲击挑战，也成为新发展阶段的重要内容。算法的不合理应用引发了一系列危机，例如，**在传播秩序方面**，算法改变了传统的新闻编辑制度，人人变成内容生产者，信息的推送带有“流量为王”、“价值观缺失”的盈利属性，带来信息茧房、虚假内容泛滥、极端事件频发、扰乱政治选举等后果。**在市场秩序方面**，算法的技术壁垒遮蔽了企业的不正当竞争和垄断行为，新型的算法共谋、大数据杀熟、自我优待等问题频出，严重扰乱市场公平竞争，损害中小企业、商家和消费者利益。**在社会秩序方面**，算法霸权下平台过度侵犯劳动者利益，歧视弱势群体，缺乏问责机制，完美避开了公众参与、要求说明解释等正当程序要求，社会公平正义遭受重大冲击。这些风险和危机不断侵蚀国家安全、公共利益和个人权利，也成为制约我国新发展阶段数字经济提速和国家治理能力现代化的瓶颈和障碍。

在此背景下，及时制定具有针对性的算法立法，明确相关主体的法律责任，既是贯彻落实党中央决策部署的重要要求，也是落实相关法律、行政法规、加强网络信息安全管理需要。**统筹发展和安全，**

推进算法治理工作，应注重算法创新和风险规制的平衡治理。

2. 算法治理的框架体系

第一，算法治理的前提是对算法组成要素有清晰认识。在算法架构中，价值取向是关键，如何使用算法、遵循何种价值观决定了算法结果的合法和合理性。网络平台是载体，网络平台的计算、存储、传输能力大幅提升，有助于算法运行结果的实时性和精准度。数据是基础原料，在深度学习算法中，收集数据的体量和种类越多，算法运行的结果越精准。自动决策或辅助人类决策是目的，算法运行体现为自动性、差异性的服务形态。

第二，算法治理的对象是使用算法提供服务的行为。对算法进行规制应当从行为而非技术本身入手。算法由平台开发设计，接入特定数据库，并决定在何时、何种范围，基于何种目的予以使用，实际体现了算法连接各方主体之间的权利关系。因此在算法治理当中，应关注算法背后个人、企业或公共主体大量收集、分析数据，运行算法以达成特定目的的具体行为。技术中立、鼓励创新、行为划线，应当成为算法治理的目标与宗旨。

第三，算法治理的重点是场景化分析。伴随算法下沉到各个领域，算法的场景化分析变得愈加重要。从算法类型来看，《互联网信息服务算法推荐管理规定》（以下简称《管理规定》）区分了应用生成合成类、个性化推送类、排序精选类、检索过滤类、调度决策类等。在具体应用领域中，金融、交通、医疗、新闻等领域之间的风险级别、风险类型、利益平衡需求、主体特点等要素之间存在显著差别。因此，

结合具体场景提出差异化的治理方案和举措，是推进成本收益平衡、分级分类管理的必然要求，具有重要的现实意义。

通过梳理典型应用，本蓝皮书将算法场景划分为信息内容安全、数据共享利用、市场竞争秩序、社会公共治理四个层面，并依次梳理了相关的问题表现、典型案例、机理分析和国内外态势，最终提炼出共性问题，为我国未来算法治理提出建议。具体图示如下：



二、算法操纵与信息内容安全

（一）信息过载：算法重塑信息分发模式

当前，算法已经成为互联网信息分发的重要工具，代表了社会信息传播的新范式。中国社会科学网报告指出，算法正在重新构建起一套全新的传播规则，其将成为整个信息系统运行的“神经”和“命脉”。在早期 Web1.0 时代，门户网站、论坛等运用的是“信息-编辑-受众”的传播结构，信息呈现“千人一面”的特点。在移动互联网、智能互联网时代，算法分发模式则绕过编辑一环，通过协同过滤算法、社交关系、个人喜好等，将不同信息进行“千人千面”的个性化推送。在信息大爆炸的背景下，算法推荐极大地提升了信息分发的效率和精准度，以此颠覆了人与信息的相处方式。

从正面看,算法在海量信息中推送个性化信息,将信息获取从“大海捞针”进入“私人定制”,大幅提高了信息效率,也推动音视频、直播、广告等数字内容产业的快速发展。然而,算法应用中也呈现出一系列突出问题。**一是以获取流量为核心目标,缺乏价值观和客观性。**用户流量是企业盈利的重要来源。在缺乏价值观引领的前提下,算法容易优先推送博人眼球的虚假信息、标题党、极端内容、低俗泛娱乐等信息,成为企业获客、增加用户粘性的重要工具。**二是容易导致信息单一性,形成信息茧房。**一方面,长期困于信息茧房、屏蔽异己信息,不利于个人思想塑造和观点养成。另一方面,信息茧房容易激化矛盾、加剧社会焦虑情绪,如医患关系等问题,甚至导致群体极化。**三是具有操纵用户思想、影响社会舆论的能力。**算法通过大数据深度分析并实施有针对性信息推送,无形中对特定个人或人群进行精准思想操纵。如在剑桥分析事件中,算法通过识别摇摆群体,使用定向信息推送诱导群众支持特定候选人,实现了传统的电视政治广告难以达到的政治营销目的。2021年10月,脸书前雇员 Frances Haugen 揭露脸书为实现盈利目的,故意使用算法放大仇恨言论,在美国国会山事件中起到助推作用。

(二) 信息内容风险引发全球算法规制热潮

1. 算法操纵思想成为首要监管对象

针对近年算法的信息推送乱象,2021年6月,美国参议院商业、科学和交通委员会再度提出《过滤气泡透明度法案》,使用户对信息

获取拥有更多选择权和控制权。一方面，要求大型平台³清楚地告知用户使用算法来确定信息传递顺序的事实。另一方面，应为用户提供非基于算法生成的信息选项。在实践中，Twitter 已提供“闪烁图标”选项允许用户在个性化时间轴和纯时间顺序时间轴之间进行切换。在监管层面，美国联邦贸易委员会（FTC）成为重要监管部门，以指引各方理解算法带来的消费者保护和竞争秩序影响。⁴俄罗斯采用了类似方法，禁止脸书等科技巨头使用算法强加内容，让用户有权关闭内容推荐算法。针对算法操纵，欧盟采取更为严苛的态度，2021 年《人工智能法案》规定，禁止以操纵人类行为或观点为目的的设计或使用高风险人工智能系统。

2. 疫情推动国外对虚假信息的关注度

美国兰德公司（RAND）在《建立基于 AI 的虚假信息规制框架》一文中指出，疫情以来，公共部门和私营部门一直试图解决疫情相关的虚假信息泛滥问题。⁵2021 年 5 月，欧盟委员会发布全球首个针对虚假信息自我审查的指导性文件——《关于加强虚假信息行为守则的指南》，推动政府、平台、用户等多主体共同参与，应对虚假信息带来的系统性风险。⁶针对网络平台的自我审查效果，指南明确要求平台采取措施避免散布虚假信息的操纵行为（如机器人账号、虚假账号、账号收购等），同时要求平台公示自我审查的方法措施，说明自身是

³根据《过滤气泡透明度法案》，大型平台是指每年收集超过 100 万用户数据，总收入超过 5000 万美元的大型互联网平台。

⁴ 参见 https://www.ftc.gov/system/files/documents/reports/ftc-report-congress-privacy-security/report_to_congress_on_privacy_and_data_security_2021.pdf.

⁵ 参见 <https://www.rand.org/blog/2021/03/towards-an-ai-based-counter-disinformation-framework.html>.

⁶ See European commission, *Guidance on Strengthening the Code of Practice on Disinformation*, at <https://digital-strategy.ec.europa.eu/en/library/guidance-strengthening-code-practice-disinformation>.

如何打击虚假信息的。

3. 算法推动网络平台责任豁免规则变革

平台责任豁免是美国《通讯规范法案》第 230 条⁷确立的重要原则，规定平台无须为第三方内容承担侵权责任。2020 年 9 月，美国司法部提出改革 230 条的法律提案，指出平台使用复杂算法推荐内容，具备了审查言论并将特定观点施加于用户的能力。2021 年 10 月，美国众议院能源和商业委员会提出《对抗恶意算法的正义法案》⁸，指出为获取利润，平台主动设计个性化算法宣扬极端主义、虚假信息和有害内容，应当改革美国《通信规范法案》第 230 条对平台责任的豁免规则，当在线平台故意或不顾后果地使用算法技术推荐明显有损于身体或情感的内容时，将解除第 230 条对平台的保护。2020 年底，欧盟也发布《数字服务法案》对平台责任进行改革。法案对超大型平台使用算法推荐系统作出规制，要求使用清晰、易获取且可理解的语言向用户说明推荐系统的运行规则、采用参数，并赋予用户调整参数的权利。同时要求平台应该定期发布并向数据服务协调员和欧盟委员会提交审计报告，其内容包括风险评估、采取的风险控制措施等。⁹

（三）我国推进网络平台算法问责体系化

1. 信息内容安全成为我国算法治理的重点问题

我国算法治理首先聚焦于信息内容安全问题。2017 年，《人民日报》针对算法推荐乱象连发三文，批判算法价值观缺失、信息茧房等

⁷ 其中，(c)(1)项规定：“计算机服务的提供者或用户不得被认定为其他内容提供者的信息发布者或发言人。”

⁸ 参见 <https://www.congress.gov/bill/117th-congress/house-bill/5596/all-info?r=257>.

⁹ 参见 <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>.

问题，提出警惕算法走向创新的反面。为规范新技术新应用，推动信息内容生态有序发展，《网络信息内容生态治理规定》《网络音视频信息服务管理规定》等提出个性化算法的人工干预和自主选择机制，设置安全评估义务等内容，在信息内容领域强化了算法应用的主体责任。2021 年《管理规定》与互联网信息服务相关立法有效衔接、一脉相承，强调促进互联网信息服务健康发展，弘扬社会主义核心价值观，引导算法向上向善。2021 年 8 月，中央宣传部等五部门联合印发《关于加强新时代文艺评论工作的指导意见》，特别强调要加强网络算法研究和引导，开展网络算法推荐综合治理，不给错误内容提供传播渠道。

2.六大层面细化网络平台信息内容主体责任

在内容治理方面，《管理规定》规定了六项“不得”和十六项“应当”要求，从六个层面对网络平台提出具体要求。**一是事前规制。**将把关窗口前移，要求不得将违法和不良信息作为用户标签，并据以推送信息。**二是技术治理。**鼓励运用内容去重、打散干预等技术策略，满足信息提供的多样性。**三是定期审查。**如应当定期审核、评估、验证算法机制机理、模型、输入输出数据等，强化用户标签管理、安全评估义务等。**四是赋权用户。**如告知算法基本原理等，选择、删除用户标签，退出个性化推荐等知情权和选择权。**五是负面清单。**禁止利用算法操纵用户账号，或者虚假点赞、评论、转发等；不得利用算法屏蔽信息、过度推荐、操纵榜单或者检索结果排序、控制热搜或者精选等干预信息呈现，影响网络舆论。**六是强化主流价值导向。**如在首

页首屏、热搜、精选、榜单类、弹窗等重点环节积极呈现符合主流价值导向的信息内容。

3. 执法、司法等协同推进算法推荐规范化

总体来看,《管理规定》明确了算法治理体制机制,确立了网信部门统筹协调,电信、公安、市场监管等有关部门依据各自职责负责的综合治理格局,形成有法可依、多元协同、多方参与的治理机制。近年,我国监管实践也为算法信息推送作出行为指引。在2016年百度魏则西事件中,调查组要求搜索引擎改变竞价排名机制,应以信誉度为主要权重进行排名,对商业推广信息逐条加注醒目标识,并予以风险提示;严格限制商业推广信息比例,每页面不得超过30%,以确保搜索结果的公平性和客观性。2021年11月,因个性化广告推送问题,微信和微博分别被诉至法院。诉状显示,原告认为微信和微博在提供个性化广告功能时,存在未明确区分个性化广告和非个性化广告、设置关闭有效期等违法行为。¹⁰值得注意的是,相关企业已更新隐私政策和相关功能,取消了关闭有效期的设置。这一案件将在司法层面有效推动算法推荐的规范发展,明确企业的合规行为。

三、算法权力与数据共享利用

(一) 数据喂养: 输入环节引发算法问题

海量数据是培育优质深度学习算法的重要基础,也是算法权力形成的重要根源。在云计算、5G、物联网、算法等技术加持下,数据资源得到空前范围的汇聚、挖掘和变现,但同时也带来个人信息滥用、

¹⁰ 参见 https://www.sohu.com/a/498543430_161795。

用户歧视等一系列风险。

1. 新技术提升个人信息广泛收集的能力

随着信息技术快速发展,个人信息、数据的收集途径广、手段多、频次高,可能带来的风险也随之增大。如智能家居等物联网设备终端可实时收集用户活动数据、健康数据,人脸识别可在未经个人同意前提下远距离收集人脸敏感数据,App、SDK 等超范围高频收集个人信息等。类似的绑定收集、隐秘套取用户个人信息成为企业汇聚数据资源、强化算法优势而采取的秘而不宣的手段。与此同时,平台之间、平台内部违规共享数据的现象突出。在工信部 2020 年通报的违规 App 中,便存在知米背单词、应用宝等企业私自将个人信息共享给第三方的情形。¹¹

2. 用户标签成为算法应用的重要数据形式

在算法生产过程中,用户标签发挥关键性作用。用户标签库是一套动态、多样、精准的体系,标签类型可区分为动态标签和静态标签,身份标签、设备标签、消费历史标签和行为标签等。标签生产的方式包括机器自动识别生产、专人生产、用户生产等。例如谷歌 Chrome 上线用户自定义标签功能,可以方便用户自主设置、更改或者消除个人标签。但用户标签的应用也带来一系列风险问题。**一是价格压榨。**例如对价格不敏感用户、苹果手机用户或老用户、会员用户定价更高。2018 年以来,携程、淘宝、滴滴等先后被质疑大数据杀熟,对具有“价格不敏感”等特征标签的用户设定更高价格。**二是用户标签中包**

¹¹ 参见 http://www.gov.cn/xinwen/2020-07/05/content_5524298.htm.

含不良或违法信息,导致反复向该用户推送有害信息,破坏网络生态。

三是歧视问题。例如脸书算法将黑人与灵长类动物混同,亚马逊算法招聘系统给男性简历打分明显高于女性等。

3.数据流通共享不畅影响算法质量

当前,数据流通共享不畅在一定程度影响了算法质量,遏制数字经济产业发展。总体来看,我国数据流通还处在较低水平。在公共数据层面,政府数据开放制度依然处于初步阶段,尚存在思维理念不到位、协调统筹不充分、数据标准不统一、数据权属不明确等“堵点”问题;¹²在商业数据层面,由于数据权属界定模糊、数据交易规则不清、纠纷解决机制不明,商业平台之间亦会出现拒绝分享数据、过度限制数据移植等问题。

(二) 数据源头治理成为全球算法监管的共识

1.以生物识别信息为代表的敏感信息保护提速

生物识别信息、性别、宗教、种族等敏感信息的过度处理,是算法歧视、算法隐私侵犯等问题的重要根源。布鲁金斯学会曾发文敦促监管机构、行业组织尽快解决 Facebook 等平台利用敏感信息的问题,以确保算法对用户的公平性。¹³近年,美欧不断加重对生物识别信息的规制力度,并关注到情绪识别、行走步态等信息采集带来的新型风险问题。美国对数据滥用的监管处罚中,不仅要求 Everalbum 删除其停用账户的照片和视频,而且还需要删除借助非法所得信息训练出的

¹² 《卢向东:加快构建全国一体化政务大数据体系 统筹推进数字政府建设》,澎湃新闻网,
https://m.thepaper.cn/baijiahao_12423309.

¹³ 参见 <https://www.brookings.edu/research/solving-the-problem-of-racially-discriminatory-advertising-on-facebook/>.

人脸识别技术模型，显著加重了违规使用敏感信息的处罚力度。2021年1月，欧洲委员会发布《人脸识别指南》，禁止把情绪识别与员工聘用、保险、教育等挂钩。¹⁴2021年6月，EDPB和EDPS有关《人工智能法案》的联合意见建议，应全面禁止在公共区域使用人工智能自动识别个人特征，包括人脸、指纹、DNA、步态、声音、击键及其他行为信号等。¹⁵

2. 直接明确平台算法的数据处理规则

伴随算法滥用数据风险的加重，更多法案直接明确了算法的数据使用规则。2021年5月，美国参议员提出《数据和算法透明度协议法案》，要求拥有3000万以上用户的公司，在追踪使用用户的喜好、习惯等数据前应获得用户同意，并保障用户撤销同意、删除数据的权利。在向第三方共享或出售数据之前也应当征得用户同意。2021年10月，参议院提出《善意人工智能法案》（Good AI Act），要求建立美国管理和预算办公室与AI工作组协商机制，要求所有联邦承包商采取适当步骤保护通过AI获得的数据，并确保数据仅用于保护国家安全，不得侵害隐私。

欧盟《数字服务法案》《数字市场法案》对平台利用算法规定了严格的数据利用规则。对于“守门人”而言，规定两项“禁为”义务：第一，不得合并使用第三方经营者在其平台运营的数据；第二，不使用非公开可得的数据，既包括由商业用户使用平台服务活动而产生或

¹⁴ See COUNCIL OF EUROPE, *Guidelines on facial recognition (2021)*, at <https://edoc.coe.int/en/artificial-intelligence/9753-guidelines-on-facial-recognition.html>.

¹⁵ See EDPB, EDPS, *Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf.

提供的数据，也包括推断或收集而来的汇总和非汇总数据。同时，还包括三项“应为”内容：第一，确保商业用户或终端用户活动中所生成数据的可携带性；第二，免费提供有效、高质量的汇总数据或非汇总数据，以确保商业用户或终端用户能够持续且实时地访问、使用相关数据；第三，以公平、合理和非歧视的条件，向商业用户提供终端用户在“守门人”在线搜索引擎上产生的数据，包括排名、查询、点击和浏览等数据。此外，在监管权限方面，《数字服务法案》为监管部门赋予了更为灵活的数据访问权限，要求超大型平台有义务建立独立的在线数据库或程序接口，便于监管部门监督和评估其合规水平。

3. 尝试化解数据流通共享障碍

数据开放共享是更好地释放数据价值的前提要件，欧美国家通过不断完善立法，形成了公私合力的多元数据流通框架。**一是确立政府数据开放义务。**美国《开放政府数据法》对数据开放目录、开放方式、审查制度、报告制度与评估制度等都进行了详尽的规定。欧盟《开放数据和公共信息再利用指令》，通过应用编程接口(API) 服务的形式使各成员国可以获取多种高价值数据集。2020年《数据治理条例(草案)》，支持公共部门数据在匿名或安全处理环境下进行内部或远程访问，公共部门和使用之间应签署保密协议，避免披露个人信息，同时应征得个人或法人同意。**二是完善企业数据共享规则。**《非个人数据自由流通条例》一方面授权服务商建立关于数据自由转移的自我规制规则，政府不施加强制性干预，另一方面要求协调成员国废除各地方本地化的数据特殊规定，增强数据流通规则的一致性。《数据治理

条例（草案）》则提出通过数据共享服务提供商进行非营利性数据共享。三是构建政企数据共享体系。2018 年《欧洲数字经济中的私营部门数据共享指南》，对企业到政府（B2G）数据共享规定了数据赠予、设立奖项鼓励开发、行政合作等多元共享方式。《数据治理条例（草案）》则创设性提出“数据利他主义”，允许私营部门为追求公共利益与非盈利主体共享数据，例如电信数据被用于预测西非的埃博拉疫情等。¹⁶

（三）我国从数据角度推进算法治理的最新动向

1. 《个人信息保护法》明确自动化决策等规则

相比一般的个人信息处理行为，自动化决策¹⁷的重点在于由算法直接作出决策，如决定是否解雇员工、是否发放贷款，分配哪个网约车司机，确定什么销售价格等，在脱离人工审查的情况下，对个人利益直接产生影响。因此，《个人信息保护法》对自动化决策作出专门规定，主要从三个层面提出了明确要求。第一，赋予用户在部分场景下自由选择退出自动化决策的权利。对于向个人进行信息推送、商业营销的，应当同时提供不针对其个人特征的选项，或者向个人提供便捷的拒绝方式。第二，赋予用户要求解释说明的权利。对于通过自动化决策方式作出对个人权益有重大影响的决定，用户有权要求个人信息处理者予以说明，并有权拒绝个人信息处理者仅通过自动化决策的方式作出决定。第三，要求个人信息处理者承担个人信息保护影响评

¹⁶ See European commission, *Guidance on private sector data sharing*, at <https://digital-strategy.ec.europa.eu/en/policies/private-sector-data-sharing>.

¹⁷ 自动化决策，是指通过计算机程序自动分析、评估个人的行为习惯、兴趣爱好或者经济、健康、信用状况等，并进行决策的活动。

估等义务。此外,创设可携权,保障用户对个人信息的转移利用行为,促进数据共享利用。¹⁸

2.《管理规定》首次对用户标签的使用规则作出规范

我国《管理规定》从平台义务、用户赋权角度,首次对用户标签进行法律规制。对用户标签的设立、审查、修正等作出规定,实际上是对算法的输入数据进行风险规制,以确保算法输出结果的公平性、非歧视性,并保障用户自主选择权。其中,第10条要求平台加强用户标签管理,不得将违法和不良信息关键词作为用户标签并据以推送信息。第17条则赋予用户对用于算法推荐服务的针对其个人特征的用户标签享有选择或者删除的权利,帮助用户在新闻阅读、电商购物等领域可以实现更为自主的选择权利,满足信息的多样性。然而,值得注意的是,在实践中,用户底层标签可能数以万计,需要进一步对标签的属性、层级和设置要求进行明确,确保条款的可操作性,并有效平衡用户体验、平台成本和社会公共利益。

3.大数据杀熟成为立法监管中关注的重要问题

在立法层面,《个人信息保护法》《管理规定》均对大数据杀熟作出规制。《个人信息保护法》第24条第1款规定,“个人信息处理者利用个人信息进行自动化决策,应当保证决策的透明度和结果公平、公正,不得对个人在交易价格等交易条件上实行不合理的差别待遇。”

《管理规定》第18条规定,“算法推荐服务提供者向消费者销售商品或者提供服务的,应当保护消费者公平交易的权利,不得根据消费者

¹⁸ 第15条规定,个人请求将个人信息转移至其指定的个人信息处理者,符合国家网信部门规定条件的,个人信息处理者应当提供转移的途径。

的偏好、交易习惯等特征，利用算法在交易价格等交易条件上实施不合理的差别待遇等违法行为。”在司法实践中，2021年7月，我国出现首例成功维权的大数据杀熟案，在“胡某诉携程”一案中，用户作为携程的星级会员不仅没有享受到优惠，反而支付了两倍以上的房价。法院审理认为，携程作为中介平台对标的实际价值有如实报告义务，却无价格监管措施，向原告展现失实价格，故认定被告存在虚假宣传、价格欺诈和欺骗行为，支持原告退一赔三。¹⁹

4.我国积极探索数据交易和共享流通制度

当前，数据正在成为关键生产要素，推进数据确权和分类分级管理，畅通数据交易流动，是促进算法创新的重要前提。一是在立法层面规范数据交易。我国《数据安全法》《广东省数字经济促进条例》《上海市数据条例》等立法中都对数据交易进行了规定，行业主管部门、行业组织积极引导规范企业间的数据共享。二是在实践层面开展积极探索。贵阳、武汉、上海、江苏等大数据交易中心围绕数据资源定价、数据权利归属、数据质量标准等规则进行了初步探索，一些企业也自主开展数据流通合作，如京东和腾讯推出“京腾计划”。三是在技术层面推进数据共享利用。通过隐私计算及区块链等技术变革传统数据流通范式、解决数据安全问题，受到广泛关注。隐私计算和区块链等技术在保护安全的同时，不仅可以实现多源数据跨域合作，也可以在一定程度上规避数据产权的争议，成为保证数据在可信、风险可控的环境下流通的有效技术路径。四是在标准层面提升数据质量、

¹⁹ 《浙江一女子以携程采集非必要信息“杀熟”诉请退一赔三获支持》，载人民法院报网，http://rmfbyb.chinacourt.org/paper/html/2021-07/13/content_207319.htm?div=-1。

引导数据有序流通。国家行业主管部门、行业组织等针对数据分类分级、数据质量、个人信息保护等相继出台多项标准指南，推动发挥数据资源价值，同时也为算法技术的迭代创新提供了重要数据资源。

四、算法滥用与市场竞争秩序

在市场竞争领域，算法有助于促进服务创新，降低决策成本，在改善商品定价模型、提供定制化服务、分析预测价格趋势等方面发挥积极作用。与此同时，平台经济领域自我优待、算法共谋、限制交易等问题对市场竞争秩序带来严重负面影响，算法在其中的角色和作用也进入大众视野。

（一）遏制竞争：滥用算法扰乱市场竞争生态

1. 利用算法达成自我优待目的

在发展业务时，平台经营者依托所获取的交易数据等资源，**利用排序算法、搜索算法等对自营业务予以“优待”，以获得竞争优势。**2020年10月，美国众议院司法委员会发布《数字市场竞争状况调查报告》指出，谷歌通过自己的“守门人”角色进行自我优待，将其在搜索市场的支配地位扩展至相邻市场，削弱相邻市场内纵向服务提供商的竞争及创新动力，损害了竞争。²⁰2021年，英国竞争及市场管理局（CMA）发布《算法：它们如何减少竞争并损害消费者》报告指出，平台利用算法限制和排除竞争，表现形式之一是在线平台利用算法操纵商品展示、推荐和排序的优先级进行自我优待。²¹**在监管层面，**

²⁰ See Committee on the Judiciary, *INVESTIGATION OF COMPETITION IN DIGITAL MARKETS*, at https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf?utm_campaign=4493-519.

²¹ See Competition Markets Authority, *Algorithms: How they can reduce competition and harm consumers*, at

欧盟委员会已对大型平台的自我优待行为作出多起处罚。2017 年，欧盟委员会认定谷歌滥用其在搜索引擎市场的支配地位，利用算法系统性地将自己的比较购物服务放在突出位置，并对竞争对手的服务进行降序展示，极大损害了竞争对手和消费者的合法利益，最终对其处以高达 24.2 亿欧元的巨额罚款。2020 年 11 月，欧盟委员会通过对亚马逊订单取样发现，平台从第三方卖家获取交易数据、物流信息、消费者评价等信息，并用于训练亚马逊平台算法系统，进一步优化自己的产品推荐、供应链管理等。据查，该平台自有商品仅 10%，却占有整个平台 50% 的销售额。欧委会认为，平台扮演了“参赛选手”和“裁判员”双重角色，不当增加自身竞争优势，破坏公平竞争环境。

2. 多国研究报告关注算法共谋问题

英国 CMA《算法：它们如何减少竞争并损害消费者》报告指出，算法可以使得明示共谋更容易达成和维持，也可能造成默示共谋的出现。CMA 分析了算法合谋可能带来的危害：第一，自动定价系统的使用以及数据可用性的增加更容易促成明示共谋。在这一过程中，价格偏差更容易检测和响应，并可以减少错误或者意外偏差的可能性。即使是最简单的定价算法，只要能够获得竞争对手的实时价格数据，便能确保经营者间的明示共谋更为稳定。第二，如果经营者使用相同的定价软件或者定价服务，例如由专门的第三方软件或者服务提供商提供，或者将定价决策委托给共同的中介机构，可能导致轴辐协议的出现。第三，可能存在一种“自主默示共谋”的可能性，随着机器学

<https://www.gov.uk/government/publications/algorithms-how-they-can-reduce-competition-and-harm-consumers/algorithms-how-they-can-reduce-competition-and-harm-consumers>.

习算法的应用，定价算法可在不需要人为参与或者授意的情况下完成共谋。

除此之外，德国联邦卡特尔署和法国竞争管理局于 2019 年联合发布《算法与竞争》研究报告，强调算法作为一种更高效的工具在传统的反竞争行为当中可能起到支持、促进实施、维持和隐藏的作用，认为算法可能促进横向垄断协议和纵向垄断协议的达成和维持，例如促成价格联合行为，检测固定或最低转售价格的偏差。²²除此之外，和英国 CMA 一样，这份研究报告同样关注了第三方促成价格协同行为或者自主学习型算法达成默示共谋行为的风险。

3. 算法成为实施限制交易的重要工具

实践中，平台经营者通过算法进行搜索降权、流量限制等惩罚性措施来强迫商家“二选一”等，损害市场竞争秩序。2021 年 4 月，阿里因“二选一”被处罚 182 亿，成为我国平台经济反垄断的标志性事件。《行政处罚决定书》指出，阿里凭借市场力量、平台规则和数据、算法等技术手段，对不执行当事人相关要求的平台内经营者实施处罚，包括减少促销活动资源支持、取消参加促销活动资格、搜索降权、取消在平台上的其他重大权益等，认定阿里巴巴集团实施“二选一”行为排除、限制了中国境内网络零售平台服务市场的竞争，妨碍了商品服务和资源要素自由流通，影响了平台经济创新发展，构成了滥用市场支配地位的行为。可见，算法作为一种新兴技术手段，成为平台保障二选一、限制用户交易的权力工具，相比传统手段，体现出

²² See the Autorité de la concurrence, the Bundeskartellamt, Algorithms and Competition, at <https://www.autoritedelaconcurrence.fr/sites/default/files/algorithms-and-competition.pdf>.

一定的隐蔽性、智能性和低成本性，为平台反垄断带来新的挑战。

（二）算法问题进入各国竞争法视野

1. 各国创新规制平台滥用市场支配地位问题

面对平台经济带来的反垄断难题，欧盟创新性提出“守门人”概念，避开了反垄断工具对相关市场界定、市场支配地位认定的难题和复杂程序，创设了新的平台监管路径。2020年12月，欧盟发布《数字市场法案》指出，一些大型平台创建了围绕其核心平台服务的联合企业生态系统，加剧了现存的进入壁垒，充当着“守门人”角色。²³在守门人义务方面，《数字市场法案》规定守门人不得使用平台内经营者与终端消费者产生的任何非公开数据，与平台内经营者开展业务竞争；也不得在商品和服务的排名中进行自我优待等非公平和歧视性做法。2019年，欧盟《关于提高在线平台交易的公平性和透明度规则》明确了排序透明度、对区别对待服务的说明义务等内容，要求平台披露决定商品或服务展示排名的主要参数，并要求平台对平台内经营者的差别待遇进行说明解释。

监管层面，2021年6月，欧盟委员会启动对谷歌在线广告技术的反垄断调查，以评估谷歌是否在在线广告供应链中对自己的技术服务进行自我优待，损害其他技术服务提供商和广告商的合法利益。2021年9月，谷歌就该项调查向欧盟委员会提出和解，目前该案仍在进行中。与此同时，美国司法部也准备就谷歌在广告技术服务市场的行为提出反托拉斯诉讼，以德克萨斯州检察长为首的州检察长小组

²³ 法案从市场规模（如年营业额或平均市值）、月活终端用户或年活企业用户，以及是否有持久牢固地位三大量化指标对“守门人”加以界定。

也对谷歌提出了单独的反托拉斯诉讼。

2. 算法共谋的监管难度逐步加大

共谋是竞争对手之间实施的损害消费者利益的共同利润最大化策略行为。算法的应用放大了共谋的社会危害和产生场景。2015年，美国司法部首次针对算法垄断协议提起诉讼，认为亚马逊和其他海报销售公司达成了一项联合价格共谋协议，通过使用算法定价软件收集亚马逊平台上销售的涉案海报的竞争者定价信息，并以此来确定卖家的定价。这一共谋协议实现了确定、提高、维持涉案海报价格的效果，违反了《谢尔曼法》第1条，系以共谋的形式限制贸易。同年，在梅耶尔诉卡兰尼克案中，法院认定Uber公司采用算法协调司机的定价行为，构成横向固定价格共谋和纵向价格共谋。相比传统的共谋行为，算法的加入一定程度上加大了监管部门发现线索、搜集证据和确认违法行为的难度。此外，对于自主学习算法导致的共谋行为，算法可让企业在无需明确沟通或互动的前提下相互依赖，加大了共谋风险，并对反垄断“人类中心主义”规制框架实现了潜在突破。²⁴

3. 以透明度作为破解算法竞争损害风险的思路

各国对算法损害竞争问题提出的解决思路通常以算法透明度为主线。如果算法缺乏透明度，就很难论证应用算法主体行为的合理性，也无法评估是否存在主体为自身利益最大化而对竞争秩序和消费者利益造成损害的风险。具体而言，透明度应当通过以下几方面予以实现。**第一，算法风险影响评估。**企业在利用新算法、优化现有算法开

²⁴ 参见周围，《算法共谋的反垄断法规制》，载《法学》2020年第1期。

展业务之前，应当对应用算法可能造成的风险进行影响评估。**第二，建立合适的算法审计机制。**由独立第三方对企业拟使用或者拟改进的算法进行审计，规定应用某些高风险算法的企业应当对相关算法进行定期审计。**第三，面向消费者和商家的透明度。**在向消费者提供服务时，采用简明的语言告知消费者算法的目的、基本运行原理和可能带来的风险。随着新技术和新应用迭代的速度越来越快，对算法的使用进行事前审计显然不符合数字经济的发展要求，取而代之的应当是与风险控制思想一脉相承的问责制度，在事中及事后归责的阶段为算法使用者施以合规证明义务。

（三）我国立法对滥用算法损害竞争秩序的关注

1. 《反垄断法（修正草案）》强化算法规制

2021年10月，全国人大常委会在对《反垄断法（修正草案）》进行初次审议后公开征求意见。此次修订在立足中国国情的基础上，借鉴欧美等地的成熟经验，反映了近期我国反垄断执法实践中的重点问题和发展趋势。本次修订进一步在法律层面对互联网领域可能涉及的数据和算法、技术、平台规则等问题进行规制（第22条），并强调了国务院反垄断执法机构应当依法加强民生、金融、科技、媒体等领域经营者集中的审查（第37条）。其中，第22条明确强调，具有市场支配地位的经营者利用数据和算法、技术以及平台规则等设置障碍，对其他经营者进行不合理限制的，属于滥用市场支配地位的行为。在法律责任方面，修订稿大幅提高罚款数额，并增加信用惩戒的规定，同时明确了对垄断的公益诉讼制度。

2. 国家竞争政策文件强化算法行为指引

2021 年 2 月，国务院反垄断委员会印发《国务院反垄断委员会关于平台经济领域的反垄断指南》（以下简称“《反垄断指南》”），结合平台经济的发展状况、经营特点和运行规律，进一步明确平台经济领域反垄断执法原则。其中，《反垄断指南》在认定垄断协议和滥用市场支配地位行为方面，有意识地突出了技术和算法的工具价值和作用。例如，针对算法共谋问题，《反垄断指南》对利用算法实现协调一致行为、利用算法对价格及其他交易条件进行限定的行为以及利用算法达成和实施轴辐协议的行为进行了明确。对于平台经济领域经营者滥用市场支配地位实施“二选一”这一热点问题，《反垄断指南》指出，除了常见的直接拒绝与交易相对人展开交易，还包括在平台规则、算法、技术、流量分配等方面设置限制和障碍，使交易相对人难以开展交易，强调了算法在促成和维持这一滥用市场支配地位行为中起到的作用。

3. 《管理规定》进一步规范市场竞争秩序

为避免算法推荐服务带来的遏制市场竞争、降低数字经济运行效率等负面影响，《管理规定》制定了具有针对性的算法推荐规定，进一步明确算法推荐服务提供者的主体责任。其中，第六条指出，算法推荐服务提供者不得利用算法推荐服务从事扰乱经济秩序等法律、行政法规禁止的活动。第十五条指出，算法推荐服务提供者不得利用算法对其他互联网信息服务提供者进行不合理限制，或者妨碍、破坏其合法提供的互联网信息服务正常运行，实施垄断和不正当竞争行为。

《管理规定》对算法推荐服务建章立制、加强规范，着力促进算法相关行业健康有序发展。

五、算法驱动与社会公共治理

（一）治理变革：算法引发公共风险

在公共领域，算法带来的优势是显而易见的，例如智慧交通系统的应用大大提升了通勤效率，无人物流系统的应用将人从机械化劳动中解放出来等。然而，在带来便利的同时，算法应用也引发诸多争议。特别是在医疗、司法、劳动力市场以及疫情防控等关乎公民社会福利、社会公平、健康安全等重大利益场景中，算法应用带来的风险需要被仔细反思与认真对待。

1. 健康码成为算法在我国社会治理中的创新尝试

健康码是信息技术在社会治理方面的一次重要创新实践，为推动国家治理能力现代化提供了有益思路。健康码通过收集个人行为轨迹、人际关系等信息，并比对卫健委、公安部、交通部等部门系统中的数据，对感染风险进行红码、黄码或绿码的自动化判定，对个人能否复工复产、自由行动的范围等后果产生直接影响。在疫情防控中，健康码在排查疑似病患、精准抗疫方面成效显著，大幅改善了防控效率。然而，技术平台基于哪些数据比对分析，如何赋值，如何具体确定风险等级的评定标准等问题均由行政机关设定标准，且具体内容并未公开。健康码在应用过程中，也出现了数据过度收集、判定偏差、说明解释程序不完备等问题，对个人财产权、人格权以及行动自由产生了

较大影响。在疫情有效控制后，有地方政府提出将健康码转变为渐变色健康码并进行常态化利用。这一建议引发了多方质疑，例如，有全国人大代表指出，为了防止疫情传播，公众让渡了一部分个人信息权利，但这种让渡是有边界和时限的。可见，在算法创新应用过程中，出现了用户数据收集范围、应用目的不透明，决策结果缺乏救济途径等多方面争议。

2. 算法在劳动领域中的应用成为国内外热点问题

近年，算法被广泛应用在劳动用工领域，体现了数字经济业态下资本对劳动力的牢牢掌控。从其运行机理来看，借助隐秘且严密的算法数据分析系统，平台可以将劳动者精准划分为若干等级，并配以相应的绩效奖惩规则规训劳动行为，实现劳动力压榨。困在算法中的外卖骑手、被技术支配的网约车司机、被智能手环实时监控的环卫工人等都身处算法规训的牢笼之中。这一应用模式将人类工具化，对人的尊严等主体性价值造成了极大挑战。2021 年《新业态从业者劳动权益保护——北京地区网约配送员职业伤害调查报告》指出，外卖员每天都在接受平台的算法管理，平台通过“完单量”“准时率”等指标对站点、代理商进行考核，而相应的管理处罚措施最终由外卖员承担。此外，招聘算法歧视、利用算法自动解雇员工等问题在国外受到广泛关注，例如，亚马逊因使用算法自动解雇 900 多名员工而被诉至法院；LinkedIn 部分招聘算法在读取到带有女性相关词汇时就会降低简历的评分。可见，算法构造出的数字化劳动控制架构正逐渐侵蚀传统的劳动保障机制，这使得劳动风险愈发突出。

3. 算法在司法、医疗等领域的应用存在待解难题

算法在司法领域中的运用由来已久，部分国家已对此进行诸多尝试，如美国的 COMPAS 犯罪预测工具、英国的 Luminance 法律文件分析工具和新加坡的 CJTS 社区司法和裁判系统等。我国也在积极推进司法人工智能系统建设，如北京市高级人民法院“睿法官”智能研判系统、上海市高院刑事案件辅助办案系统、贵州省高院“法镜大数据系统”等。与此同时，算法风险逐步显露。2017 年在美国 *Wisconsin v. Loomis* 案中，算法辅助量刑被认为**存在种族歧视风险**。调查发现，在 COMPAS 作出的高风险错误判定中，黑人数量是白人的两倍以上。同时，算法会对**司法逻辑和司法功能产生冲击**。数据前置性与技术依赖性使得算法将司法系统由以法官为中心改变为以数据为中心，法官的审判职能也相应的被让渡给算法，同时还能将复杂的价值判断问题通过技术手段进行规避。

在医疗领域，算法已经被用于诊疗方案设计、药物和医疗记录管理、健康检测和预警，甚至还被用于进行外科手术。然而，医疗往往关乎人的生命健康等最为基本的权利，而由私主体开发的算法本质上是利益导向的，因此缺乏管理的医疗算法可能带来多重风险，其中最为迫切需要解决的是**责任归属难题**。医疗算法介入诊断与治疗过程改变了传统的医患关系，同时也带来了医疗事故中责任分配的难题。据统计，美国的外科手术机器人在十年间造成了一千余起医疗事故，造成一百多人死亡。此类事故的责任虽然可以通过签署事前协议的方式确定，但仍无法从根本上解决实质性的责任归属问题。

此外，算法在其他公共领域的应用也带来诸多争议。例如，荷兰曾采用 SyRI 算法系统检测公民的福利欺诈行为。但此后的官方调查发现，该算法将申请人国籍作为判断福利欺诈风险指标的设计导致了歧视性后果。最终，海牙地区法院引用《欧洲人权公约》第 8 条规定禁用了 SyRI 算法。再如，2020 年，英国采用算法对学生考试进行打分，结果导致工人阶级和弱势社区学生处于劣势，私立学校学生处于优势，进而引发广泛抗议。

（二）欧美高度重视公共领域算法治理

从近年国外立法来看，公共领域算法规制占据了重要位置。例如，COMPAS 算法歧视问题直接推动 2017 年纽约市《算法问责法案》（《关于政府机构使用自动化决策系统的当地法》）的出台。加拿大《自动化决策指令》、法国《数字共和法案》等均是针对政府、司法机关等公共部门使用算法的相关立法。这些制度实践在一定程度上为我国提供了有益启示。

1. 算法引发的劳动权益问题需要理论解答

针对算法在劳动领域滥用的问题，美国等纷纷采取措施加以应对。2021 年 8 月，美国数据创新中心发布报告《促进负责任地使用人工智能进行劳动力决策的原则》，敦促各方在利用人工智能进行劳动力决策时担负起应负的责任。该报告提出了引入数据保护法规则、确保不会产生就业歧视、增加系统可信度、实现劳动者数据自由流动、放松对引入系统的限制、重点监管雇主而不是系统供应商等八项原则，

认为对劳动歧视等问题应进行重点关注。²⁵8月12日，美国对外关系委员会（CFR）发布《机器人时代仍要保证工人的劳动公平》一文指出，机器人时代，“幽灵工作”（几乎没有劳动保护的工作）蓬勃发展，其中，黑人、拉丁裔人、女性更有可能从事这类报酬较低的工作，并指出公私部门应合作构建新的劳动框架模式，承认幽灵工作者的价值，并保证应得的福利。²⁶在技术加持下，人们逐步陷入了劳动内卷的局面，劳动空间无限延展、劳动自由被极度压缩，对劳动剩余价值的过分压榨带来主体价值贬损、劳动目标异化等新型问题，亟需理论解答。

2. 算法说明解释义务探索落地路径

借鉴欧盟 GDPR 第 22 条中对于具有重大影响或法律影响自动化决策的说明解释义务，《个人信息保护法》第 24 条第 3 款、《管理规定》第 17 条第 3 款均对“对个人权益有重大影响”的自动化决策规定了说明义务。通常而言，对个人有重大影响的算法通常发生在公共治理领域，如信贷、医疗、应聘、司法等场景。因此，如何针对算法进行说明解释，寻求具体的落地方式，是当下各方关注的问题。早在 2013 年的澳大利亚 Cordover 案中，法院就拒绝了要求公共机构公开源代码的要求，理由是这种行为会损害算法设计企业的商业秘密。然而，随着算法角色的转变，信息公开在部分国家得到了支持。2018 年，德国第 36 届信息官员自由会议发布的《算法在公共管理中的透明度》报告认为政府部门在使用算法时需要公布算法的输入数据、基

²⁵ See Hodan Omaar, Principles to Promote Responsible Use of AI for Workforce Decisions, <https://www2.datainnovation.org/2021-ai-workforce-decisions.pdf>.

²⁶ See The Council on Foreign Relations, The Robots are Coming, but We'll Still Have a Global Digital Underclass, at <https://www.cfr.org/blog/robots-are-coming-well-still-have-global-digital-underclass>.

本逻辑、产生后果等重要要素。法国要求社交媒体必须允许监管机构对平台持有信息的广泛访问，包括对算法进行特殊访问。欧盟 2021 年《人工智能法案》则对算法透明度提出更高要求，监管机构有权访问人工智能系统供应商的训练、验证和测试数据集，并在合理的要求下，访问源代码本身。**下一步**，算法公开的方式和内容需要根据算法使用涉及权利的重要程度、公民知情权的保障效果、公共机构的说明义务等做出综合考量与配合使用。

3. 算法影响评估义务成为算法问责重要手段

算法影响评估（AIA）逐渐受到各国立法认可。算法影响评估是由算法使用者、监管部门或第三方对算法进行风险判定的一种机制手段，包括对个人信息保护、反歧视、公共利益、伦理等多维度的影响判定。GDPR 第 35 条要求数据使用者在自动化决策处理时，应履行数据保护影响评估义务。2017 年，**美国纽约**《算法问责法》首创自动化决策系统的影响评估制度。2019 年 2 月，**加拿大**出台《自动化决策指令》，要求政府部门在生成任何自动化决策系统，或其系统功能、范围发生变化时均需进行影响评估，并系统创建了影响评估指标。**欧盟**则在 2019 年 4 月针对算法在公共部门中的运用提出了专门“算法影响评估（AIA）”框架。该框架规定了算法评估的主要步骤，其中包括公共机关公开算法的目的、预期用途等基本内容，设计与进行算法自我评估、公众参与、公布算法评估结果，以及定期对评估框架进行更新。**我国**《关于加强互联网信息服务算法综合治理的指导意见》第三部分“构建算法安全监管体系”中提出“积极开展算法安全评估”

的要求,包括“组织建立专业技术评估队伍,深入分析算法机制机理,评估算法设计、部署和使用等应用环节的缺陷和漏洞,研判算法应用产生的意识形态、社会公平、道德伦理等安全风险,提出针对性应对措施”。**总体来看**,算法影响评估有助于推动算法透明度和算法场景化治理,形成全流程覆盖的风险评估框架,对于推动算法信任、化解算法风险具有重要意义。

(三) 法经济学视角下算法应用的责任机制探索

责任机制具有多重功能,最直观的功能是矫正失衡的权利关系,使得受侵害一方得到合理救济。同时,它还能通过归责设计对潜在受害者和加害者的行为产生影响。目前我国存在一般过错责任、过错推定责任、严格责任等不同归责形式。有学者指出,未来社会以过错责任为基础建立的“风险分配”责任体系,将在某些领域不复存在。例如,对于算法引发的交通事故的认定,归责事由只有结果的“对与错”,而无主观上的“故意”或“过失”。产品质量领域,不同的归责形式会使生产者根据责任所产生的不利后果调整其生产策略。而这种现象及其在人工智能产品责任设计领域的应用可以通过科斯定理来详细阐述。

科斯定理主要有三层内涵,第一层是当不存在交易成本时,制度设计不会影响资源配置的最大化。也就是说当生产者与消费者完全实现信息对等时,即便不存在产品质量管控规则以及后续的责任分配规则,两者的交易也能高质量的完成。但这其实是一种理想状态,因为在产品的实际功能、隐含缺点、销售价格等方面,消费者始终处于弱

势，并需要付出时间和精力去进行甄别，这就不可避免的产生了交易成本。**第二层**是当交易成本大于零时，对两者的责任分配制度的设计会影响资源配置，并产生社会效益。对于具有较高技术门槛、较低透明度的人工智能系统而言，其中存在的交易成本极高，因此政府参与的责任分配的设计也就变得十分必要。**第三层**是在效果层面，政府进行的精确制度设计优于私人间通过商谈纠正的责任设计。这意味着，在对人工智能系统的责任分配领域，政府需要承担起准确设计制度的义务，从而避免陷入私人救济的低效境地。

就实践中的人工智能系统责任分配而言，不同的策略选择间仍然存在讨论空间。例如在**人工智能医疗领域**的责任分配方面就存在两种观点。第一种观点认为作出最终诊断和决定治疗方针的责任由医生承担，其中一个主要目的是降低医疗人工智能企业的责任，以促进该领域的发展，这种观点也被日本政府所采纳。第二种观点是在主治医生与医疗人工智能系统之间分配责任，确定在诊疗过程中的实际责任方。例如，系统是否存在训练错误、医生是否尽到了诊疗义务等。这种观点坚持了一种务实的态度，也符合科斯定理的主旨大意，但是操作难度较大，目前世界卫生组织主张采取这种观点。而在**自动驾驶领域**，关于责任分配的讨论更为充分。宏观的看，在自动驾驶汽车致害的主要责任承担者的讨论中，存在系统设计者责任、汽车生产者责任、销售者责任、使用者责任等多种不同观点。为鼓励自动驾驶汽车的发展，有观点主张通过引入过错责任的方式减轻汽车研发、销售者的责任，但这种过错责任的使用应受到严格的时间和环境限制，不能改变产品

责任的严格责任本质。

由于算法应用还处于探索中，它的可控性、透明度等要素还需要进一步探索，而企业针对算法治理的合规成本也应当控制在合理范围之内。因此，也有观点认为不必急于为算法的责任分配模式下定论，而应当一方面给予算法控制者较为宽松的责任环境，另一方面通过评估算法的潜在危险性要求使算法控制者缴纳一定的强制性保险金。例如在医疗领域，有观点主张设置医疗强制责任险、医疗产品强制责任险及赔偿基金，实现风险的社会化分散。在自动驾驶领域则有观点主张通过强制保险、机动车责任和产品责任的协调，合理分配损害。因此，利用保险制度不仅可以降低责任分配的难度，还能弥补责任分配的盲区，是一种应对发展中风险的折中路径。

六、我国算法治理的未来展望

当前，数字经济发展不断提速，算法作为数字社会新型生产力的地位和作用日趋明显。国家网信办等九部门《关于加强互联网信息服务算法综合治理的指导意见》指出，利用三年左右时间，逐步建立治理机制健全、监管体系完善、算法生态规范的算法安全综合治理格局。在未来算法技术治理中，应进一步完善算法综合治理机制，探索监管创新和多元协同治理，推动数字经济发展行稳致远。

（一）以平衡发展与安全作为基本遵循

一是明确我国算法治理的本土需求。从全球范围来看，算法治理已经成为国际社会广泛关注的重要议题和普遍做法。一方面，我国算

法立法的出台顺应国际算法治理趋势，体现算法公平正义、安全可信等共同价值，有助于我国参与国际算法治理合作，在全球范围内发挥算法治理引领作用。另一方面，相比欧美，我国更为关注信息内容安全、信息茧房、网络沉迷、大数据杀熟等突出问题，重点关注商业领域算法应用场景，高度体现我国本土需求。

二是注重平衡发展和安全。习近平总书记指出，坚持统筹发展和安全，坚持发展和安全并重，实现高质量发展和高水平安全的良性互动。作为大数据时代底层技术，算法对于推动数字经济发展和国家治理能力现代化具有重大意义。一方面，应当抱有客观、包容的心态给予算法技术足够的创新空间，另一方面，应审慎理性地分析算法风险的根源和特征，并有针对性地总结算法治理经验，探索平衡发展和安全的最佳治理方案。

（二）积极探索算法透明的落地路径

一是算法透明是实施算法监督的重要前提。算法透明是一种针对算法的事前规制原则，要求算法使用方披露算法要素，如算法运行机理、输入数据、输出数据等，以实现确保消费者知情权、自主选择权，增进公众信任、维护竞争秩序、认定损害责任等目的。算法透明包括算法本身的透明，应当披露算法基本原理、运行机制等信息；与算法相关的数据透明，包括用户标签等输入数据、输出数据等；以及与算法相关的其他信息透明，包括开发算法的公共目的、影响评估报告、算法审查报告等内容。

二是依据透明度对象进行精细化设计。算法透明应以平衡监管要

求和产业发展为基本原则，并做精细化的制度设计。在具体落实中，应注重依据义务主体及披露对象进行区分。从义务主体来看，平台私主体的透明度目的在于纠正信息不对称，但应同时兼顾商业秘密与透明度的平衡。对政府公共部门而言，透明度的理论基础是行政公开、正当程序，对透明度要求更高。从透明度披露对象来看，可区分为对公众、用户和监管部门的透明。其中，对公众的算法公开核心目的在于实现公众监督，对用户的透明在于确保用户知情和自主权实现，对监管部门的透明要求最高，在于确保算法穿透式监管的实现。

三是通过备案制度实现对监管部门的透明。算法备案是行政机关作出的一种信息收集备查的行为，旨在获取具有潜在风险的算法系统相关信息。《管理规定》首次确立了我国算法备案制度，尝试为算法风险监测和事后问责提供依据。算法备案制度立足我国国情，是适应我国当前算法发展阶段的重要制度。算法备案应坚持形式审查原则、全周期备案原则和监测预警原则，进一步明确备案对象、备案内容以及备案流程，切实保障算法透明和算法问责。

四是实现对用户的算法透明。对用户而言，算法透明并不是公开算法模型或代码，而是告知算法设计的基本逻辑，告知用户使用算法可能带来的风险，以及为什么会产生风险。《管理规定》提出平台应当以显著方式告知用户其提供算法推荐服务的情况，并以适当方式公示算法推荐服务的基本原理、目的意图、主要运行机制等。下一步，应配合《个人信息保护法》等进一步落实对用户的透明度，需在企业合规、公众体验和监管指引的互动中逐步形成最优方案。

（三）建立基于风险的算法规制路径

一是将风险作为算法规制的依据。自 GDPR 推行“基于风险的路径（Risk-based Approach）”以来，风险成为各国规制信息技术立法的关键词。为避免过度的法律责任为公私主体，尤其是中小企业带来不合理、不必要的负担，多国立法强调优先规制高风险的算法。对于内容推送、个人信息保护、公共治理、竞争秩序等不同场景，风险判定有助于识别算法带来的风险。整体来看，高风险判定一般考虑应用场景、内容类别、用户规模、数据敏感程度、对用户行为的干预程度等判定标准。

二是注重分级分类规制。当前，我国在数据、平台分级分类等方面已有初步探索，但整体而言，互联网领域的分类分级尚缺乏细致的划分标准和成熟经验。分级分类意味着尊重不同场景既有法律秩序的功能、利益、风险设定，权衡实施算法规制的限度及方式。例如，信息内容对公共利益、国家利益的影响更为突出，因此相比之下，个人赋权的效果并不突出，政府监管部门的监管引导更加适合。在竞争领域，则更为强调平台对商家、用户的透明度义务，包括数据收集、目的告知、排名权重等。而在劳动场景下，应提供内部申诉流程与算法修改机制，允许说明导致某项决定的事实，帮助劳动者尽可能重新符合相关要求或保障合法利益。

三是构建算法影响评估制度。算法影响评估是根据算法在不同领域应用的成熟程度及其影响的范围大小，进行分领域、分优先级算法规制的前提。专业的算法技术评估是有效实施算法综合治理的前提和

基础。就评估的具体流程来看，一是选取评估主体，包括自我评估、第三方评估和监管机构评估。二是确定评估基准，包括数据质量要求、验证性能指标、技术提升等指标。三是采用事前、事中或事后评估，依据时间或算法效果纳入准入等要求。未来，算法影响评估制度是我国算法治理中的重要内容，需要配合出台行业标准或指南，对评估制度做细化指引。

（四）建立完善算法责任救济机制

一是重点规制算法使用者主体责任。算法服务带来的损害应归因于算法的设计者和使用者，并重点规制决定目的意图的算法使用者主体责任。算法设计者主要承担算法投入应用前的测试责任，必要时协助解释、及时修改或召回。此外，现阶段不建议授予机器人或自治系统法人资格，并警惕平台以算法运行自动化、结果生成自动化为由逃避法律责任。

二是构建算法审计制度。算法审计是由监管部门、第三方机构或企业自身对算法的流程、要素进行合规性审查，旨在揭示算法是否遵循了适当程序，并发现侵害事实与算法损害结果的关联性，以实现有效追责。当前，在《个人信息保护法》中，第 54 条、第 64 条分别对个人信息处理情况提出个人信息处理者自我审计和专业机构审计的要求。下一步，应注重加强算法审计与个人信息保护审计之间的有效衔接，进一步完善算法审计主体、流程、要素标准等制度内容。

三是应符合主客观相一致、责罚相当的基本原则。针对当前算法主观过错认定机制不清、责任设置不符合比例原则等问题，应当明确

和丰富过错的内涵，将操纵用户思想行为等列入禁止行为清单，提升算法使用者注意义务。对于高风险算法，国家应推动制定强制性技术标准，如有违反，则认定存在主观过错。进一步创新产品责任制度，充分利用保险制度弥补责任分配的盲区，作为应对发展中风险的折中路径。

（五）构建协同互动的算法治理格局

一是完善网络平台算法主体责任。应探索建立科学合理的平台算法责任体系，深入了解算法运行的技术机理和平台权力的形成机制，探究平台运营者的主观意图和特定价值倾向。针对算法技术特点，完善平台算法备案、评估等事前责任机制，开展全流程、全链条治理，进一步探索技术治理、动态监测等治理手段，强化专业算法合规人员配置，任命具有专业知识背景的合规官，并采取必要措施确保合规官独立行使职能。

二是发挥行业协会行业自治功能。鼓励建立算法治理行业自律机制，赋予行业自律组织相应的法律权限，建立伦理道德委员会、调节委员会、监察员等组织机制，充分发挥自治力量。推动行业标准、最佳实践、伦理指南等配套规则落地，明确细化具体要求，引导平台及企业担当社会责任，促进算法应用向上向善，强化算法应用示范引领，以实现硬软法协同治理。

三是注重第三方机构外部监督作用。应当进一步明确第三方机构的参与权限，为数据访问、算法模型等进行监督审查等，明确第三方机构的参与程序，如规定听证、检查、验证等参与方式，规定第三方

机构全流程参与的要求等。此外，还应对第三方机构的职业素养和能力进行认定监督，确保中立性等，有效弥补监管部门和公众的专业能力不足。

四是培育广大公众算法技术素养。在算法治理中，应当强化公众在算法社会所需要的思维培养和风险教育，培育公众与算法社会相匹配的算法素养。一方面，帮助公众认识算法可能带来的风险，对算法带来的歧视、操纵、隐私侵犯等风险具有足够的识别能力。另一方面，应建立算法治理的公民参与机制，赋权用户，如通过听证会等形式吸纳公众意见，参与作出重大决策。此外，可考虑为公众开发算法助理工具，利用市场机制进一步平衡各方利益。

中国信息通信研究院政策与经济研究所

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62301327

传真：010-62302476

网址：www.caict.ac.cn

