如何实现“黑箱”下的算法治理？

——平台推荐算法监管的测量实验与策略探索

张　楠　闫　涛　张　腾*

【摘要】近年来，平台推荐算法的快速发展和广泛应用深刻改变了互联网信息内容的供给方式，同时也引发了一系列算法风险与现实问题。平台推荐算法治理与监管成为政府规范算法应用的重点内容之一，无论是算法备案制度的探索，还是实现算法透明化的设想，均面临着一定的困难和挑战。在不打开算法“黑箱”的前提下，平台推荐算法规制与监管是否有可行之道？论文基于机器行为学思想，采用实验方法，以用户视角对推荐结果进行跟踪记录，通过实验数据的对比分析，验证了平台推荐算法结果差异的可测性。基于实验结果，论文提出了通过对不同平台推荐结果进行大规模数据测试和检验，测量平台推荐算法运行逻辑与推荐效果，从而实现“黑箱”下有效的逆向监管，以期丰富未来算法治理的选择。

【关键词】推荐算法　算法治理　算法监管　机器行为学

【中图分类号】D63　　【文献标识码】A

【文章编号】1674－2486（2024）01－0025－20

新一轮科技革命驱动人类社会快速进入数字时代，尤其是大数据技术与人工智能技术应用日臻成熟，为智能化、智慧化决策提供了有力支撑。算法作为人工智能的核心要素与影响决策过程的关键因素，不仅具有强大而高效的问题解决能力，并且在商务、政务等领域有着广泛的适用性。然而，算法自身具有“不透明”等复杂特性，这种不可解释的隐忧对现实伦理、个人权益、社会秩序和国家权力造成冲击。如何应对算法风险、规范算法应用，成为算法发展面临的重要问题。其中，平台推荐算法应用所引发的“信息茧房”、算法“歧视”、算法“杀熟”、算法“利维坦”等风险最为突出，这些风险问题对网络内容生态治理提出新挑战，因而加强对平台推荐算法应用的监管、规范算法推荐活动

* 张楠，清华大学公共管理学院教授，清华大学计算社会科学与国家治理实验室副主任；闫涛，山西省委网信办二级主任科员；通讯作者：张腾，清华大学公共管理学院博士后，助理研究员。

基金项目：国家自然科学基金重大项目“大数据驱动的政策多维解析研究”（72293571），国家自然科学基金面上项目“公共部门数字化转型背景下协同治理模式创新与效能评价研究”（71974111）。

变得十分紧迫。

一、引言

（一）内容供给模式的深刻变革引发海量信息处理需求

在传统的信息内容供给模式中，人们主要从报纸、期刊、电视、影院、商业门户网站等渠道获取各类信息内容，而其中的信息内容生产机制并无实质性差异。受限于人员、政策、地理等因素，传统模式信息内容供给能力和范围也始终是极为有限的，这些信息内容供给渠道，无论是在时效性上，还是在丰富性上，都难以跟上“信息爆炸”的时代步伐。早期社交媒体的发展为用户自行撰写、拍摄、制作的图文视频类作品提供了良好机会，也提高了互联网信息内容的多样性。受此类社交模式的启发和影响，各大互联网企业开始搭建账号平台，引导、鼓励网民注册使用公众账号。由分散的用户供给信息内容的方式，能够满足各类人群对信息内容的多样需求，虽然在专业性上可能有所欠缺，但突破了记者、编辑等专业人员信息内容生产的局限性，给“中心化”的传统信息内容供给模式带来了巨大冲击。在传统信息内容供给方式与规则发生革命性改变的背景下（喻国明、韩婷，2018），面对海量的信息内容供给，依靠传统人工编辑和呈现的模式面临两个方面的巨大挑战。一是如何在短期内对信息内容进行挑选、分类，从而发现、捕捉用户感兴趣的信息内容与时事热点；二是当信息内容种类足够丰富时，在有限的手机屏幕当中，该采用何种呈现方式，才可能满足用户多样化的需求，为用户找到“需求内容”，并为不同类型的优质信息内容寻找“需求用户”。

（二）推荐算法成为信息内容分发呈现的主导力量

平台推荐算法恰好解决了“为用户找内容”和“为内容找用户”两个方面的痛点问题，因此成为智能传播时代信息内容供给与呈现的灵魂（张省、蔡永涛，2022）。首先，推荐算法可以对用户上传的信息内容进行标签化分类，并在此基础上建立信息内容库。其次，推荐算法通过收集用户的兴趣喜好、浏览习惯、选择倾向等行为偏好数据，对不同用户进行画像和“打标签”，从而建立用户“需求清单”。最后，对信息内容与用户进行匹配，为需求用户推荐呈现特定类目的信息内容，便可以实现海量信息内容的有效分类和准确分发，进而满足用户的个性化需求，提升用户体验感。

不同类型推荐算法的实现均依赖两项基础条件。一是平台需要具备多样化信息内容供给的能力，二是平台需要具备收集用户行为数据的能力（陈洁敏等，2014）。当商业平台具备了以上两项条件后，才能基于大量真实的行为数据对用

户进行准确画像，为用户精准推送兴趣内容，从而给用户带来更加个性化、智能化的浏览体验，并帮助平台企业在激烈的流量竞争环境中突出重围。可见，推荐算法能够化解信息过载困境，同时又能够满足以用户为中心的信息内容供给需求（陈洁敏等，2014）。事实证明，较早应用推荐算法的APP，正是依靠推荐算法突破了信息呈现方式，对信息内容进行高效处理、为用户提供个性化推荐，才迅速受到广大网民的认可和追捧。

（三）平台推荐算法监管的隐蔽性与“黑箱化”难点

推荐算法技术的广泛应用，使得不同的用户看到的信息内容不尽相同，产生“千人千面”的具体现实表现。诸如分裂国家、暴恐暴力、色情赌博、侮辱英烈等违法违规的信息内容，可能在不同的用户手机端以不同的方式展现。此时，推荐算法成为一个“盲盒”“黑箱”的角色，使得监管部门很难看清、查清推荐结果中的违规内容（张红春、章知连，2022）。如果一个采用了推荐算法的APP，向少数特定的用户群体推送违规信息，向多数普通用户推荐正常合规信息，而接收到违规信息的用户又没有进行投诉、举报，监管部门依靠现有技术手段则很难发现这些违规行为的存在及损害结果的发生。这不仅可能引发网络信息内容与网络空间治理风险，也给信息内容监管带来前所未有的挑战。

在传统市场监管中，对媒体信息内容的监管，监管部门可以进入实体经营场所，查看实际经营流程；而相较之下，在平台推荐算法应用过程中，监管部门无法通过一个公开的入口或渠道，对平台分发的信息内容进行监测、调查和取证。推荐算法具有较强的监管隐蔽性和“黑箱化”特征，对采用推荐算法的APP，监管部门能看到的也仅仅是“冰山一角”，无法判断和认定平台为全部用户推荐了什么内容、有特定兴趣偏好的用户是否浏览了违规内容，以及违规内容被推送至多少用户等情况。

（四）“黑箱”下算法监管与治理的设想

算法“黑箱”给平台推荐算法监管带来诸多难题，例如，违规信息内容推送行为难于取证、精准推荐与“信息茧房”难以区别、“流量至上”扰乱网络生态等。然而，由于当前对算法以及推荐算法的定义较为模糊、算法伦理规范效力不足、相关立法进程漫长且滞后，加之算法可解释性成本和负担过高，这为打开算法“黑箱”、实现算法透明增加了极大的难度。那么，在不打开算法“黑箱”的前提下，算法治理是否具有可行之道？算法“黑箱”下全面有效的平台推荐算法监管应采取何种方式与模式？这些问题亟待解决。

以往的理论研究与治理实践将注意力主要集中在算法透明上，对“黑箱”下算法治理可能性、可行性的关注不足。尤其是对于平台推荐算法而言，短期内无法完全实现算法透明，但监管与治理需求又十分迫切。因此，不能“守株

待兔”，而是要另辟蹊径，在接受算法“黑箱”客观存在的现实前提下，探索当下算法治理的出路。本文面对算法“黑箱”固有特性与算法透明相对性的矛盾，基于机器行为学理论与算法行为生成机制，提出因果“倒推”的逆向算法治理设想。这种“黑箱”下算法治理的理论视角，不仅突破了算法透明治理逻辑的局限，同时也补充、丰富了既有的算法治理体系框架。另外，本文聚焦于平台推荐算法治理，通过实验方法验证“黑箱”下算法监管技术方案的可行性，进而展开监管策略探索，为“黑箱”下平台算法监管实践提供参考。值得注意的是，打开算法“黑箱”与不打开算法“黑箱”两种治理路径互为补充，需要综合考量和利用，两者间关系的调适能够为多层次的算法治理制度构建提供新的思路和选择。

二、文献综述：算法治理的宏观演进与微观进路

（一）算法治理宏观演进中的基本共识

“算法”这一概念，从数学领域到计算机领域再到广泛的社会领域，已发生了多次重要流变，需要从技术、系统、社会等不同层面予以理解（肖红军，2022）。在技术层面，算法是仅限于适合计算机程序来实现的决策技术或解决方案（胡键，2021），其本质是一种计算工具（Ziewitz，2015）。在系统层面，算法是人类通过代码设置、数据运算与机器自动化判断进行决策的一套机制（丁晓东，2020），强调设计者的算法责任（Martin，2019）。在社会层面，算法是建构社会结构与秩序的理性模型（贾开，2019），在此过程中算法被视为一种社会权力（Beer，2017；许晓东、邝岩，2022）。

算法治理概念包含两个维度的内涵。一是应用算法的治理（Algorithmic Governance），二是对算法及其应用的治理（Governance of Algorithms）。正是由于算法被广泛应用于社会和国家治理领域，对个人和组织决策产生了深远影响，并因此导致一系列不良现象与风险，才引发了对算法及其应用的监管、规制、引导与优化的思考（Ebers & Gamito，2021）。本文中算法治理指的是对算法及其应用的治理，它是算法规则的建立、重塑、运行过程，具体而言，这些规则可以分为法律政策规则、社群规则与技术规则等（许可，2022）。同时，对算法及其应用的治理也是算法风险和影响的化解、消除过程，其范式包括个体赋权、外部问责与平台义务（张欣，2019）。面对算法多层次、多维度的定义与内涵，算法治理所关注的不仅包括以代码为载体的技术作用对象及其影响因素，还包括算法运行机制、算法结果、算法塑造的规则与算法对人类社会产生的影响（贾开，2019）。算法治理以提高算法应用准确性、合法性与效率为目标，算法“黑箱”隐忧与算法透明理性是其中最为关键的两大问题（Coglianese & Lehr，

2019）。

算法的应用发展驱动了算法治理研究的演进，算法治理逻辑与治理体系建设逐渐得到重视（孟天广、李珍珍，2022）。不同的算法治理理论主张在制度实施过程中发生着转场与互动，无论是法律规制、行政问责、伦理约束等传统治理方式，抑或是以算法透明为核心的技术治理方式，均强调了算法治理的重要性与必要性。算法治理议题广泛，不同维度的研究涉及的基础理论依据纷繁各异，但学者们对数字正义（马长山，2022）、公共价值创造（昌诚等，2022）、科技向善等基本理念的追求始终是一致的。

当前，多元主体参与下的算法治理体系建设已成为学界共识，算法治理需要政府、社会、企业、公众等多元主体的共同参与（张吉豫，2021），而算法治理体系构建过程中各主体间权责、利益关系的均衡是十分复杂、困难的（杨华锋，2022；金雪涛，2022）。另外，不同国家和地区在治理目标、治理主体、治理对象、治理手段和治理模式等方面也存在一定差异与共性（曾雄等，2022），算法全球治理问题亦开始得到学者关注（贾开等，2022；李龙飞、张国良，2022）。

（二）算法治理实现路径的差异化

算法治理与监管的实现路径被广泛讨论，不同类型的算法治理，以及不同应用领域的算法治理，在监管思路与治理模式上存在较大差异。例如，智能网联车自动驾驶算法治理（Lyakina et al.，2019）与平台新闻分发算法治理（Sehat，2022）所关注的风险问题截然不同；再如，社交网络中算法治理（Lemes de Castro，2018）与金融领域算法治理（Wijermars & Makhortykh，2022）所遵循的原则和理念也迥然相异。对于特定的算法治理问题，学者们根据不同的算法特征与风险机制，结合不同的理论基础分析，提出了具体的治理路径。

推荐算法作为应用最为广泛的算法类型，学者们围绕其风险挑战、特征特性、技术演变、治理对策等展开讨论（Andrews，2019；孟天广、李珍珍，2022）。行政规制被认为是兼顾效力与效率的推荐算法治理方式。例如，政府和社会等利益相关者共同设计实现一种责任机制，用以监督算法设计者行为（Dekker et al.，2022），或者围绕算法推荐的代表性（Representation）、方向性（Direction）和干预性（Intervention），构建一个跨部门、跨地区、跨技术和跨组织的监管框架（Eyert et al.，2022）。其实，问责与处罚并非算法监管的根本目的，纠正算法偏差、防范算法风险、促进算法更好地应用才是算法治理的出发点，这对推荐算法应用的优化改进也是十分必要的。例如，单晓红等（2022）提出通过融合话题特征和目标用户兴趣偏好，改善推荐结果的多样性；王旭娜和谭清美（2020）基于一项对互联网平台用户偏好与平台推荐机理的研究，提出综合集成个体推荐和群体推荐的系统优化建议。

（三）算法“黑箱”与算法透明

算法“黑箱”指的是算法输入、输出及运行过程中不公开、不可知、不可解释、不确定的状态，具体包括两层含义：“一是指源于算法本身的技术复杂性而导致的模型不可解释，这类问题存在于深度学习等算法中；二是指算法设计者不向用户公开其算法原理与机制，导致用户对算法特征与运算过程毫不知情。”（孟天广、李珍珍，2022：16）算法因其技术逻辑及应用方式的特性而带来不可解释的隐忧，导致算法“黑箱”不能为人所知晓或理解，从而产生算法不可监督、难以追责等治理困境（贾开，2019）。西方发达国家在公共服务领域较早地使用了推荐算法，但算法这类深度学习工具固有的“黑箱”困境，给公共管理与决策带来了巨大的不确定性（Busuioc，2021）。还有的大量研究涉及算法在电子商务领域的应用。例如，在利用算法推荐技术的大型电子商务平台中，算法“黑箱”导致的信息不对称和不平等议价能力问题，对中小企业造成严重困扰（Di Porto & Zuppetta，2021）。算法治理是应对算法“黑箱”困境的有效途径，有学者对德国和荷兰警方在预测性警务领域的推荐算法进行了实证研究，发现两种算法系统的不同定位和使用方式，取决于不同的社会主导规范和行政文化，算法“黑箱”不可知的技术特征并不直接影响推荐算法应用，有效的治理能够促进建立信任环境（Meijer et al.，2021）。

算法透明是在一定程度上打开算法“黑箱”，通过提高关于算法目的、算法设计、算法运行、算法结果等方面的可解释性、信息对称性，来保障用户知情权，从而化解算法“黑箱”困境。算法透明能否实现、效果如何，均受技术、机制、信息披露等多方面因素影响。越来越多的学者认识到算法透明是相对的，而非绝对的（贾开，2019；孟天广、李珍珍，2022），透明度不能被视为纯粹的开放性，而应该是一种交流行为与治理方式（肖梦黎，2021）。强化算法解释、算法透明和实现算法祛魅被认为是解决算法监管问题的必要环节（黄静茹等，2022），法学学者们对算法“透明度模型”展开了诸多讨论，算法透明原则成为法律规制的重要内容（Bayamlioglu，2018；徐凤，2019）。有人认为准确的算法认知是有效发挥算法备案和公示制度监管作用的前提，也是相关问题得到良好解释和妥善解决的关键，缺乏准确的算法认知则将导致算法监管过度，并因此制约算法的应用与发展（刘泽刚，2022）。与此同时，算法透明的缺陷被广泛讨论，算法透明的可行性也受到质疑。例如，有学者提出算法透明作为事前规制方式，其规制效力有着天然的不足。通常的算法透明既不可行，也无必要，因此算法透明理念应该处于非普适性、辅助性的地位（沈伟伟，2019）。

（四）文献述评

当前研究对算法风险的探讨热度始终不减，对于算法治理体系构建的思考

也逐渐深入，学者们注意到推荐算法这一类型算法在实践发展中的特点与理论研究中的重要性。算法“黑箱”成为算法治理研究中不可回避的核心问题，多元化治理工具的设计仍然遵循着算法透明的基本逻辑。这种打开算法“黑箱”，以算法解释与算法透明为核心主张的治理模式固然“美好”且重要，但算法透明的合理、合法限度尚不明确，算法透明的实现也并不能等同于算法治理的有效实现。保持对算法透明的理性思考，有必要探索更多的算法治理可能路径和范式。尤其是对于平台推荐算法而言，相关法律规范尚未制定、代码公开尚不可行、算法不可解释、伦理规范效力不足，而平台企业的市场行为与商业活动活跃，导致平台推荐算法的负外部性逐渐在放大，诸多围绕算法透明原则展开的治理设想在平台推荐算法的监管实践中均难以实现。

跳出基于算法透明的打开“黑箱”的治理逻辑，是否可以实现“黑箱”下的算法治理，以及如何实现“黑箱”下的算法治理，此类问题讨论与相关研究仍较为不足。在不打开算法“黑箱”的前提下，伦理约束等柔性规制（Mittelstadt et al.，2016；胡键，2021）、企业与行业自律等社群规则（孟天广等，2022）是防范算法风险的必要治理方式，但这种前置性治理方式和手段对于算法结果有效治理的充分性作用不足，往往需要结合配套体系建设来增强规制能力，延长规制作用周期。另外，以实用主义为导向、以算法问责为代表的事后规制方式，作为一种间接的治理策略（沈伟伟，2019），依赖于法律规则、政策规则、社群规则等治理架构的完善，并且主要应用于产生损害结果的情形，无法直接而全面地解决算法“黑箱”带来的问题与风险。可见，当前缺乏在不打开“黑箱”的前提下，对算法“黑箱”困境的直接回应性治理及监管模式探究。不同于以算法透明原则促进算法行为和推荐结果合法、合理的传统治理思维，本文在机器行为学理论视角下，创新地提出以推荐结果“倒推”算法行为的监管设想，以多层次的行政监管策略直接回应和破解算法“黑箱”不可知性困境，对现有算法治理理论研究形成有益补充。

三、模式思考：现实需求、挑战与机器行为学启发

（一）平台推荐算法监管的现实需求与面临挑战

2021 年 9 月，国家互联网信息办公室联合中央宣传部等九部委出台《关于加强互联网信息服务算法综合治理的指导意见》，从治理机制、监管体系、算法生态三个维度出发，提出了建立算法安全综合治理格局的任务要求。2021 年底，国家互联网信息办公室、工业和信息化部、公安部与国家市场监管总局联合出台《互联网信息服务算法推荐管理规定》，确立了服务提供者算法备案与公示制度。2022 年 3 月 1 日起该规定正式实施，同时互联网信息服务算法备案系统上

线。截至2023年底，国家互联网信息办公室已公开发布303条境内互联网信息服务算法备案信息，其中包括算法名称、算法基本原理、算法运行机制、算法应用场景、算法目的意图等方面的简要信息内容。算法备案与公示制度强调了算法推荐服务提供者的责任与义务，体现了国家和政府对算法治理的重视，同时也是监管部门在算法治理领域迈出的重要一步。

监管部门要求推荐算法服务提供者提交算法信息并描述算法推荐逻辑，可见监管部门有意从代码等技术层面入手，通过推动算法透明来破解算法“黑箱”困境。但在实际操作层面，由平台企业自主提供算法信息的备案方式存在诸多不确定性。这种备案监管模式主要面临以下挑战。一是平台企业提交的推荐算法信息是一整套代码，监管部门对该整套代码如何进行评测，能否通过对代码的审查来印证其推荐逻辑和推荐效果的合理性、合法性，均不确定；二是平台企业在运营中往往会根据用户喜好、时政热点等因素，不断优化推荐算法，这就需要及时动态地调整算法代码，而监管部门难以实现对持续更迭演变的推荐算法进行备案管理和审查；三是监管部门难以搭建真实、庞大的用户环境，不具备对备案推荐算法进行实际测量和验证的客观条件。另外，推荐算法备案制度主要作用于事前监管环节，其事中监管和事后监管有效性较弱。然而，算法推荐结果是对用户权益和公共利益产生直接影响的关键因素，事中监管与事后监管环节不容忽视。

（二）机器行为学视角下平台推荐算法监管逻辑与模式启发

以人工智能为驱动的机器在社会、文化、经济和政治互动中发挥着越来越重要的调节作用，机器行为学（Machine Behavior）成为跨越多个研究领域的新兴学科（Rahwan et al.，2019）。人机协同是机器行为的核心特征，机器决策行为的发生与算法应用之间具有密不可分的关系。同时，学习算法的机器行为也受到个体、集体的社会性因素影响（Borch，2022；Hagendorff，2021）。当前“人—机—物—网”相互融合，在机器行为、机器学习算法、人机协同的共同作用下，复杂的社会联动效应将产生复杂的人机混合决策场景，因此参与决策的个体需要对机器行为有一定的理解，才能形成人机高效协同（曾大军等，2021）。人工智能不断迭变发展，算法与人类社会的深层次融合在一定程度上突破了算法技术的“工具性”，算法也并非简单的人造物或人工现象。在机器行为学理论视角下，人机交互融合，算法虽然属于机器范畴，因其与社会环境互相作用、影响，在一定意义上成为具有“自主性”的行动主体。机器在特定的环境中触发或产生的行为是可以被观察到的，通过算法在特定环境中的行为表现、行为结果，可探究、验证其运行机制和行为动机。算法是机器行为学的主要研究对象之一，机器行为学为算法治理理论发展带来新的契机，对算法治理的理论框架研究与政策实践创新均具有重要意义，尤其在治理原则与方案、算法行

为与环境间关系、算法生命周期整体协同性治理等方面具有启发性（贾开等，2021）。

在人工智能时代，机器行为学颠覆了人类以缔造者的主体身份来研究机器行为，并从中寻求机器改进方法的传统研究范式（孙立会、王晓倩，2022）。机器行为学对机器行为生成机制与逻辑的关注，更多地体现着以结果“倒推”原因的逆向逻辑，而非通过探究机器设计过程，以原因“界定”结果的正向逻辑（贾开等，2021）。就应对算法“黑箱”所引发的一系列风险而言，机器行为学提供的这种“倒推”因果的逻辑，应在算法治理与监管过程中得到重视与应用，从而修正算法透明逻辑的偏差，弥补多元化、系统化治理框架与治理体系构建过程中，机器行为学理论视角的缺失。算法“黑箱”是算法设计与运行因果关系中的关键介入因素，由因推果的过程不可避免地需要打开算法“黑箱”，算法透明原则便遵循这种正向逻辑，通过探究、揭示算法“黑箱”的原因来达到对风险结果治理的目的，算法透明尤其关注算法运行的公正性、公开性（姜野、李拥军，2019）。相反，在以果推因的过程中，面对算法“黑箱”本身的不可解释性特征，将算法“黑箱”与其他复杂的算法作用环境视为整体，基于算法不确定性的多种可能结果之间的差异比较，来分析和探究算法设计目的、功能等方面的差异。这种逆向逻辑能够避免或者不需要打开算法“黑箱”，并且以成熟的算法“黑箱”优化测试理论、技术传统为基础（聂长海、徐宝文，2004；张永鎏、汪镭，2020），能够提高“黑箱”下算法监管的适应性、可行性，有助于实现以原因端为落脚点的算法源头治理，而不是以结果端为落脚点的不确定性风险治理。

基于机器行为学的视角，推荐算法不仅是功能性的机器，同时也是带有动态演化与学习特征的智能行动主体。推荐算法治理与监管的关键在于规避推荐算法风险行为所导致的风险结果。以平台推荐算法作用结果反向推导算法设计及运行机制的合理性、合法性，建立基于算法效果测量的监管模式，或许可以成为当前算法治理与监管的新思路，并将为我们理解和预测平台推荐行为、优化监管策略提供更深刻的洞见。这种全新的、逆向的监管思路仍然需要验证。对不同平台推荐结果进行测量实验，能否监测到推荐算法的差异，能否真实发现推荐结果中的不良问题和实际风险，将为我们提供答案。

四、测量实验：算法差异的比较与发现

（一）从用户视角跟踪记录推荐结果的实验设计

从技术角度来看，实验方法在算法设计、改进、优化、分析等环节有着广泛而普遍的应用。社会科学研究一直以来也不乏实验研究方法的应用（Blom-

Hansen et al.，2015），已有学者在社会学（张钺、李正风，2022）、哲学（黄雪婷，2022）等不同学科领域运用实验方法对算法治理相关问题展开研究。推荐算法涉及相关平台企业的商业秘密，研究者很难对其内部代码进行解剖观察，因此无法准确把握算法的实际推荐机制和过程。为了探究“黑箱”下平台推荐算法逆向监管、治理模式的可行性，本文提出了另外一种推荐算法实验思路：从用户视角切入，向不同平台推荐算法输入用户行为偏好，持续操作、观察、记录平台实际推荐结果，测试算法推荐结果输出，并对推荐结果数据进行分析，从而研究、验证推荐算法的实际运行逻辑及相关特征。同时，本文通过有效的实验控制排除随机变量带来的影响，对实验组与对照组的推荐结果数据进行多维比较，探究平台推荐算法结果差异。

（二）实验过程

1. 实验对象的选取

目前采用推荐算法进行信息内容推荐的互联网平台较多，按照内容形式来看，这些平台主要可以分为图文资讯类和社交短视频类。本文分别选取时下用户最多的两款图文资讯类 APP（分别为 J 客户端和 T 客户端）和两款社交短视频 APP（分别为 D 客户端和 K 客户端）进行实验。以上 4 款 APP 是国内较早应用推荐算法取得竞争优势的平台，在用户市场占据绝对优势，对互联网信息内容推荐、传播与网络舆论有着广泛的影响力；同时，这 4 款 APP 也是平台推荐算法治理与监管的重要对象，具有一定的代表性、典型性，能够在一定程度上反映当前我国平台推荐算法治理与监管过程中主要对象的一般性、普遍性特征。

2. 实验步骤

第一步，注册新用户：申请 7 个新手机号码，在 J、T、D、K 4 款 APP 上分别注册 7 个新账号，模拟 7 个用户。第二步，实验分组：将 7 个用户分为实验组和对照组，实验组设置 5 个不同兴趣偏好的用户，分别是猎奇君、明星娱乐君、美食君、军事君、历史君。对照组设置两个不带有个人兴趣偏好的用户，分别是对照 A 和对照 B。第三步，浏览与互动操作：实验组用户和对照组用户每日登录各 APP 两次，均浏览前 30 款推荐内容。实验组用户在阅读到自身兴趣类别内容时，进行点赞、评论、收藏等互动操作。对照组用户仅浏览，不进行互动操作。第四步，记录数据：记录实验组用户和对照组用户每次登录时所浏览到的内容。第五步，数据分析：对 7 个模拟用户在各 APP 中接收的各类推荐内容结果数据进行整理与分析。

3. 实验控制

内容控制：根据观察统计，将图文资讯类推荐内容分为奇闻轶事、国际政治军事、社会时政新闻、体育运动、明星娱乐、厨艺美食、历史文化、财经股

市和其他等9个类目；将短视频类推荐内容分为搞笑剧情、体育、购物、财经股市、日常生活分享、社会时政新闻、美女、游戏、美食、知识科普、汽车、景色旅游、明星娱乐和其他等14个类目。

时间控制：7个模拟用户于2022年2月21日至26日，每日在13：00－18：00和19：30－24：00两个时间段内进行登录。

操作控制：7个模拟用户分别登录J、T、D、K各10次，每个模拟用户登录后浏览30款推荐内容。5个实验组用户严格按照兴趣偏好设定，仅对兴趣类内容进行点赞、收藏、关注等操作，对非兴趣类内容进行忽视。两个对照组用户无差别地浏览推荐作品，不做任何操作。在实验期间，每个模拟用户在各APP上共计浏览300个信息内容作品；全部模拟用户在4款APP上共计浏览8400个信息内容作品。

（三）实验结果分析

1. 实验组用户结果

实验组用户在J、D、K客户端实验中，经过互动操作后，均被算法准确捕捉到兴趣类内容并进行持续推送。以J客户端内容推荐数据为例，5类兴趣内容均被算法捕捉并进行持续高占比推荐。如表1所示。

表1　J客户端实验组用户兴趣类内容推荐数据

登录次数	登录日期	实验用户1：猎奇君	实验用户2：明星娱乐君	实验用户3：美食君	实验用户4：军事君	实验用户5：历史君
1	2022.2.21	9	3	6	3	2
2	2022.2.22	15	7	11	4	8
3	2022.2.22	14	9	17	19	17
4	2022.2.23	20	18	7	24	28
5	2022.2.23	20	16	6	26	20
6	2022.2.24	27	11	10	27	25
7	2022.2.24	21	17	18	27	23
8	2022.2.25	27	16	25	26	25
9	2022.2.25	24	13	28	29	26
10	2022.2.26	21	22	29	27	23

资料来源：作者自制。

再以J客户端实验中的用户4为例，模拟设置的兴趣类内容为“国际政治军事类”，在经过两次登录后，用户阅读倾向基本被算法掌握，兴趣类内容推荐数量增多且上升趋势明显，随后一直保持较高的推荐量。在第9次登录时，兴趣类内容数量最高达到29个，在全部浏览内容中占比达到96.7%。如图1所示。

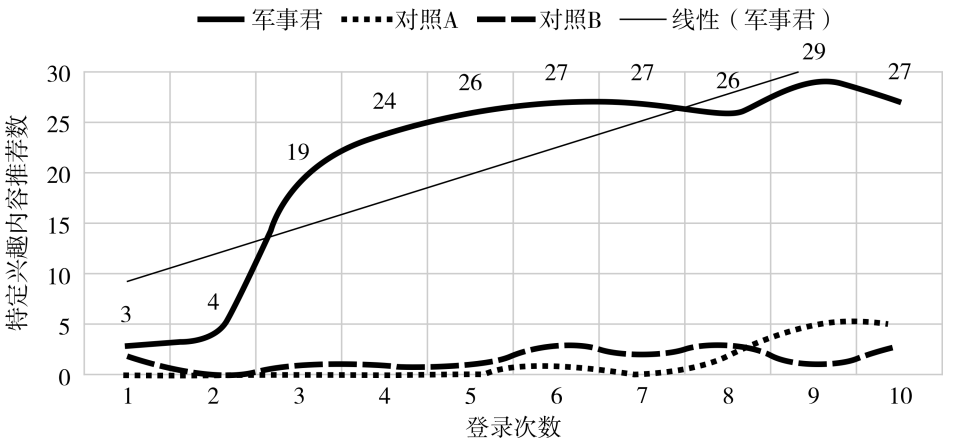

图1　J客户端实验中用户4“军事君”与对照A、对照B的数据对比

资料来源：作者自制。

而在T客户端实验中，本文发现该APP是基于手机硬件设备进行内容推荐的。在同一款手机上登录不同的账号，即使进行不同的用户兴趣偏好操作，但被推荐的内容没有显著变化或差别。

2．对照组用户结果

在J、D、K客户端实验中，对照用户因为没有进行互动操作，最终呈现出较为多元的信息展示，在实验初期、中期、后期所呈现的各类推荐内容比例也大致相同，反映出该推荐算法一直按照既定的推荐策略进行推荐，并未随机变换内容比例，也未增加新类型内容来进一步开发用户兴趣偏好。以对照用户A在D客户端第2、4、6、8、10次登录数据为例，实验过程中推荐内容类型与数量变化较小，总体上相对集中于搞笑剧情、用户日常分享、社会时政新闻三大类。如图2所示。

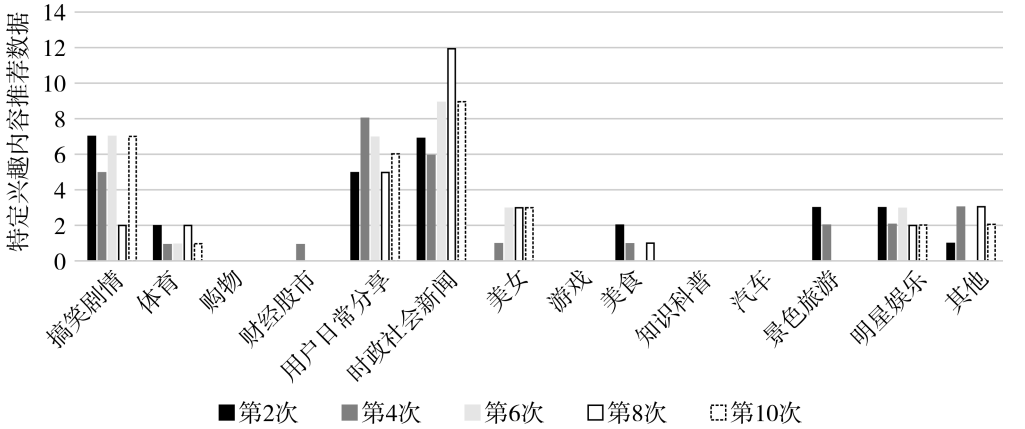

图2　对照A在D客户端实验中第4、6、8、10次实验数据对比

资料来源：作者自制。

（四）平台推荐算法的差异测量与问题发现

通过实验研究，本文对J、T、D、K 4款客户端平台推荐算法结果进行差异比较，发现不同推荐算法在用户识别和推荐机制方面均有所不同。J、T、K客户端均采用针对账号的内容推荐，而T客户端则是依据手机设备进行内容推荐。不同的推荐算法对不同类型内容赋予了差别化的推荐策略，即使用户都具有较强、较明显的兴趣偏好和频繁的互动操作，但不同的平台推荐算法还是体现出了不同的反应敏锐度和内容推荐度。

将J、D、K客户端中的实验用户兴趣类推荐内容进行对比，发现J客户端采用的推荐算法对奇闻轶事类、历史文化类、国际政治军事类内容的反馈更为强烈，而明星娱乐类、厨艺美食类的内容在前中期的推荐反馈略显弱势，特别是明星娱乐类内容推荐占比没有超过70%。D客户端的推荐算法对体育运动类和美女类的兴趣捕捉更为敏锐，对厨艺美食类、明星娱乐类、财经股市类的兴趣捕捉稍显迟钝。K客户端的推荐算法仅对体育类的兴趣捕捉较为敏锐，拥有较高的推荐占比，而对其余4类兴趣内容并没有强烈的推荐反馈。如图3所示。

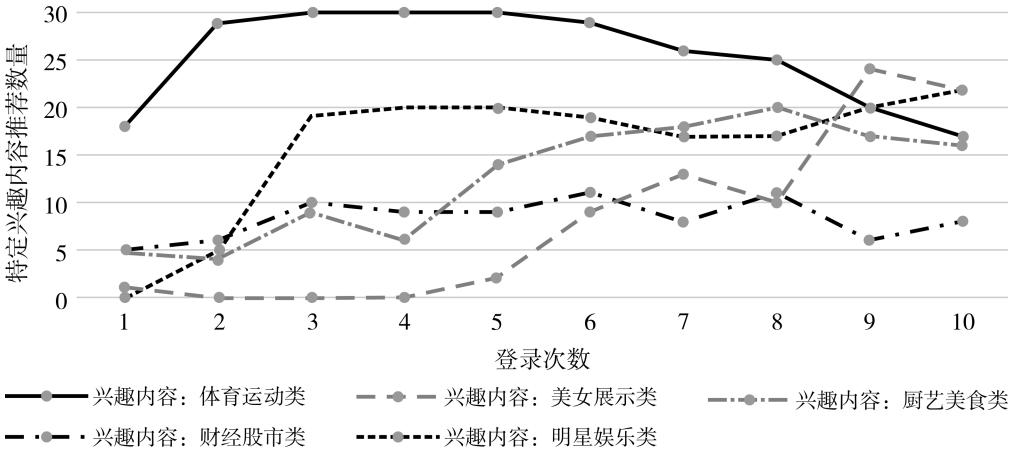

图3　K客户端实验用户兴趣类内容推荐对比

资料来源：作者自制。

在平台推荐算法对用户兴趣偏好内容的推荐方面，推荐算法决定着内容的选取、赋值和分发推送力度，其背后实现逻辑具有显著的内容选取和推荐力度的偏向性、策略性。由于可以直接地决定内容的选取和推送，在面对初始新用户、广泛兴趣用户、单一兴趣用户等不同群体时，推荐算法均展现出了对平台设定的某些类型内容推荐的倾向性。即使是有特定兴趣偏好的用户，平台推荐算法在满足其兴趣类内容之外，也会根据设定的策略倾向推荐某些类型内容，而不是随机推荐其他类别内容。特别是在J客户端实验中，推荐算法对奇闻轶事、标题党、吸引眼球类内容进行了强烈推荐，展现出显著的预设内容选择倾向性，实验用户在10次登录浏览的共计300个作品中，除去自身兴趣类内容，

其余被推荐内容的数量和优先级如表2所示。

表2　J客户端实验用户10次登录中非兴趣类内容数据总和对比

用户	第1推荐优先级	数量	第2推荐优先级	数量	第3推荐优先级	数量
J实验用户1	明星娱乐	30	社会时政新闻	16	厨艺美食	15
J实验用户2	奇闻轶事	96	厨艺美食	19	社会时政新闻	11
J实验用户3	奇闻轶事	57	社会时政新闻	31	明星娱乐	16
J实验用户4	奇闻轶事	28	明星娱乐	13	历史文化	9
J实验用户5	奇闻轶事	31	国际政治军事	29	明星娱乐	16

资料来源：作者自制。

面对“信息茧房”的争议，本文对J、D、K客户端实验组用户最后3次所接收的兴趣类与非兴趣类内容进行对比，发现对于兴趣类别较为单一的实验用户，K客户端兴趣类内容比为60%左右，J、D客户端兴趣类内容占比均为90%以上，非兴趣类内容占比为10%以内，极少接收到其他类别内容。从这一结果来看，平台推荐算法确实加剧了单一兴趣类别用户“信息茧房”的形成，面对兴趣面较窄的用户，算法推荐的逻辑始终与尝试发掘用户多元兴趣内容相矛盾。

本实验虽然样本数据有限，但从以上实验数据的分析结果来看，仍然可以明确地对不同平台推荐算法的运行逻辑和运行结果进行测量。小规模的测量实验证实了不同平台推荐算法在设计、运行过程中，针对不同类型信息内容采取了差异化的推荐策略，并且这种差异是明显的、可测量的。

五、策略探索：基于推荐算法效果测量的监管

本文基于平台推荐算法治理模式思考与测量实验验证，尝试提出应对算法治理与监管难点的新思路。对算法在特定环境中触发的推荐结果实施测量，是未来算力支撑下不打开算法“黑箱”的监管新模式。这一思路从设想到落地需要三个层次上的若干具体策略。首先，在不打开算法“黑箱”的情况下，基于规模测试进行逆向评测和监管的创新理念，能够有效化解当前正向监管、算法透明理念下的监管被动性和“黑箱”难解性困境。其次，监管部门可以通过建构虚拟账号池和规模计算动态监测环境等途径，丰富和完善推荐算法监管手段，并对推荐算法进行全方位刻画，有效地发现推荐算法的症结和风险。最后，在具体应用场景层面，测量不同用户隐私设置和个性选择在算法推荐过程中的履行情况，可作为保障用户知情权和选择权的检查方式；通过模拟一些极端的用户阅读浏览行为，测量网络不良内容传播中的个体偏好，并对平台网络安全责任进行判断；针对可能引发社会广泛关注和激烈争论的舆论热点和关键议题，加强风险识别，及时发现网络敏感议题讨论中的潜在外部干预，维护网络意识

形态安全和社会稳定。三个层次下的五方面策略如图 4 所示。

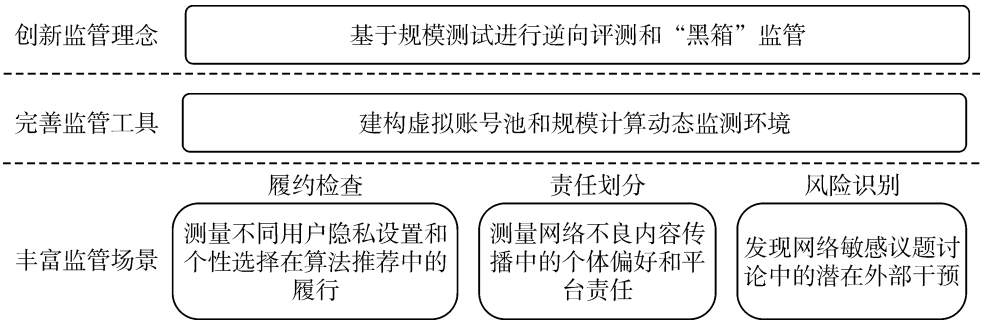

图 4　“黑箱”下算法监管策略的层次结构

资料来源：作者自制。

（一）创新监管理念：基于规模测试进行逆向评测和“黑箱”监管

目前推荐算法的治理思路与监管理念，仍然主要集中于破解“黑箱”、研究算法细节和实现逻辑等方面（张红春、章知连，2022），但此类方式具有一定的被动性和难解性。一方面，政府有关部门作为监管机构，不具备平台企业的实际用户规模和实际运行环境，无法准确验证推荐算法的实际推荐逻辑。另一方面，此类监管模式的实现，需要依靠平台主体的有效配合，方能有效地规制、改进算法。因此，理念创新成为化解当前算法治理与监管困境的重要方向。在不打开算法“黑箱”的情况下，对不同平台推荐算法进行结果的测试和检验，能够实现逆向监管。通过外部评测的手段统计推荐算法的实际推荐结果，以此作为算法监管的参考依据，无须平台企业的配合，也不以风险和损害结果发生为必要条件，因此能够让监管部门在设计和执行相关监管法律法规时掌握主动权。

（二）完善监管手段：建构虚拟账号池和规模计算动态监测环境

受限于实验样本、评测周期等因素，本文中的小规模实验还无法对推荐算法作出准确、完整的评测鉴定。但实践中，监管部门可以通过协调电信运营商，开通大量虚拟手机号，在监管平台上进行账号注册，模拟更加复杂的用户兴趣偏好和浏览互动行为。同时，监管部门可以建构虚拟账号池，制定详细的、有针对性的评测策略，在较长的评测周期内，对算法推荐结果进行统计，进而得到具有统计意义的评测结果。基本实现结果对推荐算法进行全方位刻画，是发现推荐算法的症结和风险的有效途径。监管部门通过构建规模计算动态监测环境，还可以不断丰富、完善平台推荐算法监管手段。平台推荐算法可能存在向少数特定用户、小众圈群推荐违规内容，制造群体对立和传播不良网络亚文化等问题，通过一定规模的虚拟账号进行外部评测，可以更为全面地掌握推荐算

法的特定策略和结果，并以此作为平台推荐算法监管的样本证据。

（三）履约检查：测量不同用户隐私设置和个性选择在算法推荐中的履行

虽然现阶段政府监管部门已经明确要求各推荐算法服务提供者要保障用户的知情权和选择权，为用户设置开启或关闭相关推荐功能的按键。但在实际运行中，平台推荐算法在多大程度上按照用户的选择进行信息内容分发，用户在知情选择下的真实效果又如何，目前还缺乏有效的评估手段和方法。而利用大规模虚拟账号进行推荐结果测量、检验不失为一种有效的履约检查方式。设置若干个实验组和对照组用户，分别选择关闭或开启某种推荐模式，对一定周期内的推荐结果进行统计，便可以有效对真实效果进行验证、评估。特别是可以评估在青少年模式下，平台推荐内容的类型、品质和保护效果等重点因素。监管部门通过主动地评估测量，能够倒逼平台企业切实保护用户的知情权和选择权，从而优化推荐算法整体的行业生态。

（四）责任划分：测量网络不良内容传播中个体偏好和平台责任

平台推荐算法的初衷是“投其所好”，依靠数据标签和用户兴趣标签等数字化指标，向感兴趣用户进行信息内容推荐。一方面，定向的个性化信息推荐将可能产生“信息茧房”（桑斯坦，2018）；另一方面，为了迎合用户，推荐算法的设计更关注用户喜好和时事热点内容，不会将信息内容的价值判断作为重要推荐指标。在这种情况下，我们必须警惕算法将类似的、迭代的不良内容推荐给用户。尤其是在用户并没有显著搜索、浏览等主动操作时，平台仍然可能向其推荐不良内容。平台算法设计者应发挥推荐算法对网络舆论的正向价值（邓杭，2018）。监管部门通过模拟一些极端用户的阅读浏览行为，便可以甄别出推荐算法在迭代收敛的过程中是否“越界”，是否会走向不良、有害甚至违法违规的内容禁地。一旦出现上述情况，监管部门便可严肃追究相关平台网络安全责任。

（五）风险识别：发现网络敏感议题讨论中的潜在外部干预

平台推荐算法在一定意义上具备了互联网信息资源配置的公共权力，并具有一定的意识形态属性（李静辉，2022）。商业资本的控制可能会带来意识形态领域的冲击和风险，特别需要警惕对公共属性较强、争议较广和参与度较高的内容议题的推荐策略。构建大规模的虚拟账号形成一个动态的、主动的推荐结果分析矩阵，便可以进行详细的数据分析，观测出不同类型用户被推荐的热点内容，以及不同议题被推荐的热度。例如，可以加强对青少年群体或老年人群体被推荐内容的动态评估。针对可能引发社会广泛关注和激烈争论的舆论热点和关键议题，监管部门必要时应介入、干预平台推荐算法策略，通过调整推荐

策略来影响信息内容分发，阻断极端思想、错误思潮和虚假信息的传播，阻止可能引发的群体对立，从而维护网络空间的安全与稳定。

六、研究结论与未来展望

本文基于机器行为学思想，采用实验方法，以用户视角对平台推荐算法结果进行跟踪记录，通过实验数据的对比分析，验证了平台推荐算法结果差异的可测性。基于实验结果，本文提出了对不同平台推荐算法进行大规模数据测试和检验，测量平台推荐算法运行逻辑与推荐效果的监管方式。对于平台推荐算法治理而言，不打开算法“黑箱”成为可供选择的新模式，同时也为全面提高监管有效性、建立多层次和多元化的监管体系提供了更为丰富的实施路径。政府监管部门可通过模拟具有统计学意义的、不同行为习惯的规模用户，对算法推荐效果进行跟踪记录，对推荐算法开展评测和监管，并据此对平台企业提出整改意见或作出行政处罚。

本文提出的“黑箱”下算法治理是用技术手段解决技术风险思路的具象化，目前仍处于初步探索层面。后续研究可能将在两个向度进行延伸：在实证层面，探究、验证平台推荐算法运行机制、结果差异与存在的问题，仍需要规模更大、场景更为复杂的数据测试。通过模拟更加复杂的用户行为偏好、增加实验用户数量和延长测试时间等方式，可以提升问题发现的准确度。结合外部评测结果和真实用户数据采样，也可以准确发现当前平台中热度较高、争论激烈、广泛推荐的公共性议题，进而对相关热点内容的推荐情况和效果进行研判，及时防范、化解平台推荐算法风险。在理论层面，鉴于监管行为本身比推荐算法具有更强的可解释性要求，关于“黑箱”下算法治理的理论对话与理论建构同样是亟待讨论与扩展的重要议题。目前，相关监管逻辑、思路、模式、策略的讨论受限于相关理论研究的匮乏，尚无法形成层次性、体系性的理论框架和实践路径。相关理论思考至少将涉及管理学视角中，不同信息对称性条件下的多主体监管博弈理论模型建构，传播学视角互联网情景中群体与个体复杂交互下的行为理论拓展，法学视角算法效果差异背后的主观责任确认机制等议题。这只有多学科背景研究者的广泛关注，才能从现有思路出发系统性设计“黑箱”下算法治理的实践方案。

参考文献

昌诚、张毅、王启飞（2022）. 面向公共价值创造的算法治理与算法规制. 中国行政管理，10：12－20.
Chang, C., Zhang, Y. & Wang, Q. F. (2022). Algorithmic Governance and Algorithmic Regulation: A Perspective of Public Value Creation. *Chinese Public Administration*, 10: 12－20. (in Chinese)

陈洁敏、汤庸、李建国、蔡奕彬（2014）. 个性化推荐算法研究. 华南师范大学学报(自然科学版)，5：8－15.
Chen, J. M., Tang, Y., Li, J. G. & Cai, Y. B. (2014). Survey of Personalized Recommendation Algorithms. *Journal of South China Normal Uniwesity* (*Natural Science Edition*), 5: 8－15. (in Chinese)

单晓红、崔凤艳、刘晓燕（2022）. 融合话题多维特征和用户兴趣偏好的微博话题推荐研究. 现代情报，5：69－76＋97.
Shan, X. H., Cui, F. Y. & Liu, X. Y.（2022）. Research on Microblog Topic Recommendation Integrating Topic Multidimensional Features and User Interest Preferences. *Journal of Modern Information*, 5：69－76＋97.（in Chinese）
邓杭（2008）. 算法推荐的风险防范和导向管理——发挥算法推荐对网络舆论的正向价值. 新闻战线，11：62－64.
Deng, H.（2008）. Risk Prevention and Guided Management of Algorithm Recommendations：Leveraging the Positive Value of Algorithm Recommendations for Online Public Opinion. *New Media*, 11：62－64.（in Chinese）
丁晓东（2020）. 论算法的法律规制. 中国社会科学，12：138－159＋203.
Ding, X. D.（2020）. On the Legal Regulation of Algorithms. *Social Sciences in China*,12：138－159, 203.（in Chinese）
胡键（2021）. 算法治理及其伦理. 行政论坛，4：41－49.
Hu, J.（2021）. Algorithm Governance and Its Ethics. *Administrative Tribune*, 4：41－49.（in Chinese）
黄静茹、莫少群（2022）. 算法监控的逻辑理路、伦理风险及治理路径. 南京社会科学，6：130－136.
Huang, J. R. & Mo, S. Q.（2022）. Logic, Ethical Risk and Governance Path of Algorithmic Monitoring. *Nanjing Journal of Social Sciences*, 6：130－136.（in Chinese）
黄雪婷（2022）. 马克思主义实践观视域下无人驾驶道德算法困境的哲学实验探究. 闽南师范大学.
Huang, X. T.（2022）. A Philosophical Experiment Exploration of the Moral Dilemma of Autonomous Driving Algorithmsfrom the Perspective of Marxist Practice . Minnan Normal University.（in Chinese）
贾开（2019）. 人工智能与算法治理研究. 中国行政管理，1：17－22.
Jia, K.（2019）. Artificial Intelligence and Algorithm Governance Research. *Chinese Public Administration*, 1：17－22.（in Chinese）
贾开、徐杨岚、吴文怡（2021）. 机器行为学视角下算法治理的理论发展与实践启示. 电子政务，7：15－22.
Jia, K., Xu, Y. L. & Wu, W. Y.（2021）. Theoretical Development and Practical Enlightenment of Algorithm Governance from the Perspective of Machine Behavior. *E-Government*, 7：15－22.（in Chinese）
姜野、李拥军（2019）. 破解算法“黑箱”：算法解释权的功能证成与适用路径——以社会信用体系建设为场景. 福建师范大学学报(哲学社会科学版)，4：84－92＋102＋171－172.
Jiang, Y. & Li Y. J.（2019）. Open the Algorithm Black Box：Function Verification and Applicable Path of the Right to Explanation：Based on the Construction of Social Credit System. *Journal of Fujian Normal University* (*Philosophy and Social Sciences Edition*), 4：84－92＋102＋171－172.（in Chinese）
金雪涛（2022）. 算法治理：体系建构与措施进路. 人民论坛·学术前沿，10：44－53.
Jin, X. T.（2022）. Algorithmic Governance：System Building and Measures Algorithmic Governance：System Building and Measures. *Frontiers*, 10：44－53.（in Chinese）
桑斯坦（2008）. 信息乌托邦——众人如何生产知识. 毕竞悦译. 北京：法律出版社.
Sunstein, C. R.（2008）. *Infotopia*：*How Many Minds Produce Knowledge*.（Bi, J. Y. Trans.）. Beijing：Law Press China.（in Chinese）
李静辉（2022）. 推荐算法意识形态属性的生成逻辑、风险及治理. 理论导刊，2：70－76.
Li, J. H.（2022）. The Generation Logic, Risks, and Governance of Ideological Attributes in Recommendation Algorithms. *Journal of Socialist Theory Guide*, 2：70－76.（in Chinese）
李龙飞、张国良（2022）. 算法时代“信息茧房”效应生成机理与治理路径——基于信息生态理论视角. 电子政务，9：51－62.
Li, L. F. & Zhang, G. L.（2022）. The Generation Mechanism and Governance Path of the “Information Cocoon House” Effect in the Algorithm Era：Based on the Perspective of Information Ecology Theory. *E-Government*, 9：51－62.（in Chinese）
刘泽刚（2022）. 论算法认知偏差对人工智能法律规制的负面影响及其矫正. 政治与法律，11：50－65.
Liu, Z. G.（2022）On the Negative Impacts of Cognitive Biases of Algorithm on the Legal Regulation of Artificial Intelligence and Its Correction. *Political Science and Law*, 11：50－65.（in Chinese）
马长山（2022）. 算法治理的正义尺度. 人民论坛·学术前沿，10：68－76.
Ma, C. S.（2022）. The Justice Criterion of Algorithmic Governance. *Frontiers*, 10：68－76.（in Chinese）
孟天广、李珍珍（2022）. 治理算法：算法风险的伦理原则及其治理逻辑. 学术论坛，1：9－20.
Meng, T. G. & Li, Z. Z.（2022）. Governance Algorithm：Ethical Principles and Governance Logic of Algorithm Risk. *Academic Forum*, 1：9－20.（in Chinese）
聂长海、徐宝文（2004）. 基于接口参数的黑箱测试用例自动生成算法. 计算机学报，3：382－388.

Nie, C. H. & Xu, B. W. (2004). An Algorithm for Automatically Generating Black-Box Test Cases Based Interface Parameters. *Chinese Journal of Computers*, 3: 382 – 388. (in Chinese)

沈伟伟 (2019). 算法透明原则的迷思——算法规制理论的批判. 环球法律评论, 6: 20 – 39.
Shen, W. W. (2019). The Myth of the Principle of Algorithm Transparency: Criticism of Algorithm Regulation Theory. *Global Law Review*, 6: 20 – 39. (in Chinese)

孙立会、王晓倩 (2022). 人工智能之于教育的未来图景: 机器行为学视角. 中国电化教育, 4: 48 – 55 + 70.
Sun, L. H. & Wang, X. Q. (2022). The Future Prospect of Artificial Intelligence in Education: From the Perspective of Machine Behavior. *China Educational Technology*, 4: 48 – 55 + 70. (in Chinese)

王旭娜、谭清美 (2020). 互联网平台用户偏好挖掘与推荐机理研究——基于经典扎根理论的探索. 情报科学, 8: 49 – 56 + 87.
Wang, X. N. & Tan, Q. M. (2020). User Preference Mining and Recommendation Mechanism of Internet Platform: An Exploration Based on the Classical Grounded Theory. *Information Science*, 8: 49 – 56 + 87. (in Chinese)

肖红军 (2022). 算法责任: 理论证成、全景画像与治理范式. 管理世界, 4: 200 – 226.
Xiao, H. J. (2022). Algorithmic Responsibility: Theoretical Justification, Panoramic Portrait and Governance Paradigm. *Journal of Management World*, 4: 200 – 226. (in Chinese)

肖梦黎 (2021). 风险视角下对算法偏见的双轨治理——自动决策算法中偏见性因素的规制研究. 理论月刊, 7: 134 – 142.
Xiao, M. L. (2021). Dual Track Governance of Algorithm Bias from a Risk Perspective: A Study on the Regulation of Bias Factors in Automatic Decision Algorithms. *Theory Monthly*, 7: 134 – 142. (in Chinese)

徐凤 (2019). 人工智能算法“黑箱”的法律规制——以智能投顾为例展开. 东方法学, 6: 78 – 86.
Xu, F. (2019). Legal Regulation of Artificial Intelligence Algorithm Black Box: Taking Intelligent Investment Advisors as an Example. *Oriental Law*, 6: 78 – 86. (in Chinese)

许可 (2022). 驯服算法: 算法治理的历史展开与当代体系. 华东政法大学学报, 25(1): 99 – 113.
Xu, K. (2022). Taming Algorithms: The Historical Development and Contemporary System of Algorithm Governance. *Journal of the East China University of Politics & Law*, 25(1): 99 – 113. (in Chinese)

许晓东、邝岩 (2022). 算法权力的形成与风险治理. 郑州大学学报(哲学社会科学版), 3: 18 – 24 + 127.
Xu, X. D. & Kuang, Y. (2022) The Formation of Algorithmic Power and Risk Governance. *Journal of Zhengzhou University* (*Philosophy and Social Sciences Edition*), 3: 18 – 24 + 127. (in Chinese)

杨华锋 (2022). 赋能与均衡: 算法权力秩序的合法性风险及其治理. 新视野, 4: 62 – 68.
Yang, H. F. (2022). Empowerment and Balance: The Legitimacy Risk and Governance of Algorithm Power Order. *Expanding Horizons*, 4: 62 – 68. (in Chinese)

姚凯、涂平、陈宇新、苏萌 (2018). 基于多源大数据的个性化推荐系统效果研究. 管理科学, 5: 3 – 15.
Yao, K., Tu, P., Chen, Y. X. & Su, M. (2018). Research on the Effectiveness of Personalized Recommender System Based on Multi-source Big Data. *Journal of Management Science*, 5: 3 – 15. (in Chinese)

喻国明、韩婷 (2018). 算法型信息分发: 技术原理、机制创新与未来发展. 新闻爱好者, 4: 8 – 13.
Yu, G. M. & Han, T. (2018). Algorithmic Information Distribution: Technical Principles, Mechanism Innovation, and Future Development. *Journalism Lover*, 4: 8 – 13. (in Chinese)

曾大军, 张柱, 梁嘉琦等 (2021). 机器行为与人机协同决策理论和方法. 管理科学, 6: 55 – 59.
Zeng, D. J., Zhang, Z., Liang, J. Q. et al. (2021). Machine Behavior and Human-Machine Collaboration Decision: Theory and Methods. *Journal of Management Science*, 6: 55 – 59. (in Chinese)

曾雄、梁正、张辉 (2022). 欧美算法治理实践的新发展与我国算法综合治理框架的构建. 电子政务, 7: 67 – 75.
Zeng, X., Liang, Z. & Zhang, H. (2022). The New Development of Algorithm Governance Practice in Europe and America and the Construction of China's Algorithm Comprehensive Governance Framework. *E-Government*, 7: 67 – 75. (in Chinese)

张红春、章知连 (2022). 从算法“黑箱”到算法透明: 政府算法治理的转轨逻辑与路径. 贵州大学学报(社会科学版), 4: 65 – 74.
Zhang, H. C. & Zhang, Z. L. (2022). From Algorithm Black Box to Algorithm Transparency: The Transition Logic and Path of Algorithm Governance by the Government. *Journal of Guizhou University* (*Social Sciences Edition*), 4: 65 – 74. (in Chinese)

张吉豫 (2022). 构建多元共治的算法治理体系. 法律科学(西北政法大学学报), 1: 115 – 123.
Zhang, J. Y. (2022). Constructing an Algorithmic Governance System of Multiple Co-governance. *Science of Law* (*Journal of Northwest University of Political Science and Law*), 1: 115 – 123. (in Chinese)

张省、蔡永涛（2023）. 算法时代“信息茧房”生成机制研究. 情报理论与实践，4：67－73.
Zhang, X. & Cai, Y. T.（2023）. Research on the Generation Mechanism of “Information Cocoon” in the Algorithm Era. *Information Studies*: *Theory & Application*, 4: 67－73.（in Chinese）
张欣（2019）. 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，6：17－30.
Zhang, X.（2019）. From Algorithm Crisis to Algorithm Trust: Multiple Solutions and Localization Paths for Algorithm Governance. *Journal of the East China University of Politics & Law*, 6: 17－30.（in Chinese）
张永鑫、汪镭（2020）. 基于协同过滤的连续黑箱优化问题元启发算法选择. 控制与决策，6：1297－1306.
Zhang, Y. W. & Wang, L.（2020）. Metaheuristic Algorithm Selection System for Continuous Black-Box Optimization Problems Based on Collaborative Filtering. *Control and Decision*, 6: 1297－1306.（in Chinese）
张钺、李正风（2022）. 人工智能算法的社会实验及其意义. 清华社会学评论，1：60－77.
Zhang, Y. & Li, Z. F.（2022）. Social Experiments on Artificial Intelligence Algorithms and Their Significance. *Tsinghua Sociological Review*, 1: 60－77.（in Chinese）
Andrews, L.（2019）. Public Administration, Public Leadership and the Construction of Public Value in the Age of the Algorithm and “Big Data”. *Public Administration*, 97（2）: 296－310.
Bayamlioglu, E.（2018）. Contesting Automated Decisions. *European Data Protection Law Review*, 4: 433－446.
Beer, D.（2017）. The Social Power of Algorithms. Information. *Communication & Society*, 20（1）: 1－13.
Blom-Hansen, J., Morton, R., & Serritzlew, S.（2015）. Experiments in Public Management Research. *International Public Management Journal*, 18（2）: 151－170.
Borch, C.（2022）. Machine Learning and Social Theory: Collective Machine Behaviour in Algorithmic Trading. *European Journal of Social Theory*, 25(4): 503－520.
Busuioc, M.（2021）. Accountable Artificial Intelligence: Holding Algorithms to Account. *Public Administration Review*, 81（5）: 825－836.
Coglianese, C. & Lehr, D.(2019). Transparency and Algorithmic Governmance. *Administrative Law Review*, 71(1), 1－56.
Dekker, R., Koot, P., Birbil, S. I. & van Embden Andres, M.（2022）. Co-Designing Algorithms for Governance: Ensuring Responsible and Accountable Algorithmic Management of Refugee Camp Supplies. *Big Data & Society*, 9（1）, 205395172210878.
Di Porto, F. & Zuppetta, M.（2021）. Co-regulating Algorithmic Disclosure for Digital Platforms. *Policy and Society*, 40（2）: 272－293.
Ebers, M. & Gamito, M. C.（2021）. *Algorithmic Governance and Governance of Algorithms*: *Legal and Ethical Challenges*. Cham: Springer.
Eyert, F., Irgmaier, F. & Ulbricht, L.（2022）. Extending the Framework of Algorithmic Regulation: The Uber Case. *Regulation & Governance*, 16（1）: 23－44.
Hagendorff, T.(2021). Linking Human and Machine Behavior: A New Approach to Evaluate Training Data Quality for Beneficial Machine Learning. *Minds and Machines*（*Dordrecht*）, 31(4): 563－593.
Lemes de Castro, J. C.（2018）. Social Networks as a Model of Algorithmic Governance. *Matrizes*, 12（2）: 165－191.
Lyakina, M., Heaphy, W., Konecny, V. & Kliestik, T.（2019）. Algorithmic Governance and Technological Guidance of Connected and Autonomous Vehicle Use: Regulatory Policies, Traffic Liability Rules, and Ethical Dilemmas. *Contemporary Readings in Law and Social Justice*, 11（2）: 15－21.
Martin, K.（2019）. Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*, 160（4）: 835－850.
Meijer, A., Lorenz, L. & Wessels, M.（2021）. Algorithmization of Bureaucratic Organizations: Using a Practice Lens to Study How Context Shapes Predictive Policing Systems. *Public Administration Review*, 81（5）: 837－846.
Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L.（2016）. The Ethics of Algorithms: Mapping the Debate. *Big Data & Society*, 3（2）: 1－21.
Rahwan, I., Cebrian, M., Obradovich, N. et al.（2019）. *Machine Behavior*. *Nature*, 568: 477－486.
Sehat, C. M.（2022）. Journalistic Values and Expertise in Platform News Distribution: The Possibilities and Limitations of Participatory Panels for Algorithmic Governance. *Journalism Studies*, 23: 1225－1246.
Wijermars, M. & Makhortykh, M.（2022）. Sociotechnical Imaginaries of Algorithmic Governance in EU Policy on Online Disinformation and FinTech. *New Media & Society*, 24（4）: 942－963.
Ziewitz, M.（2015）. Governing Algorithms: Myth, Mess, and Methods. *Science*, *Technology*, *& Human Values*, 41（1）: 3－16.

责任编辑：郑跃平