

算法推荐下解释制度的优化路径研究

林 曼

浙江理工大学法政学院、史量才新闻与传播学院, 浙江 杭州

收稿日期: 2023年7月9日; 录用日期: 2023年8月1日; 发布日期: 2023年11月13日

摘 要

随着“个性化算法推荐”在互联网时代的广泛运用, 用户在算法社会中面临着严重的法律层面的权益保障问题与道德层面的精神侵害问题。因平台私权利扩张与法律对知识产权的保护, 司法救济等事后补偿途径可操作性较低, 规制推荐乱象的重心应当落于预防性质的立法上。而优化算法可解释性逐渐发展为全球范围内算法立法治理的一致要求。但我国的算法解释制度不完善, 通过深入研究域外的算法解释规定, 分析推荐算法公益性以权衡知识产权下的私权利, 构建算法解释的评估体系并从解释时间、解释内容两个方面对《互联网信息服务算法推荐管理规定》作出合理的补充, 从而推动算法社会健康发展。

关键词

算法解释, 算法推荐, 权利限制

Research on the Optimization Path of the Interpretation System under Algorithmic Recommendation

Man Lin

School of Law and Politics, Shi Liangcai School of Journalism and Communications, Zhejiang Sci-Tech University, Hangzhou Zhejiang

Received: Jul. 9th, 2023; accepted: Aug. 1st, 2023; published: Nov. 13th, 2023

Abstract

With the widespread use of “personalized algorithmic recommendations” in the Internet era, users face serious problems of legal protection of rights and interests and moral infringement in the

文章引用: 林曼. 算法推荐下解释制度的优化路径研究[J]. 法学, 2023, 11(6): 5213-5219.

DOI: 10.12677/ojls.2023.116745

algorithmic society. Due to the expansion of the platform's private rights and the protection of intellectual property rights by law, judicial remedies and other remedies are less operable, and the focus of regulating the recommendation chaos should fall on preventive legislation. The optimization of algorithm interpretability has gradually developed into a consistent requirement for algorithm legislative governance worldwide. However, China's algorithm interpretation system is not perfect, so through an in-depth study of overseas algorithm interpretation regulations, analyzing the public interest of recommendation algorithms to counteract the jurisprudence of trade secrets, constructing an evaluation system of algorithm interpretation and making reasonable supplements to the Administrative Regulations on Algorithm Recommendation of Internet Information Services from two aspects, interpretation time and interpretation content, to promote the healthy development of algorithm society.

Keywords

Algorithm Explanation, Algorithm Recommendation, Rights Restrictions

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

后疫情时代,我国数字化转型进程加快,以算法技术为代表的数字化生存进入深层次、宽领域的发展阶段,算法技术也进入到迭代升级的黄金时期[1]。算法推荐作为算法技术的应用典型,促进了信息的高效率分发,创造了显著的经济效益。然而,侵害用户合法权益的行为却数见不鲜。一方面,学界普遍认定其在获取个人信息以及加工利用以获取利益的过程中可能存在侵犯个体知情权、隐私权等权益的危险[2]。因此我国对该问题出台了《互联网信息服务算法推荐管理规定》、《中华人民共和国电子商务法》(以下简称《电子商务法》)、《中华人民共和国网络安全法》(以下简称《网络安全法》)以及《中华人民共和国个人信息保护法》(以下简称《个人信息保护法》)等法律法规。另一方面,越来越多学者注意到算法制造产生的“信息茧房”与过滤多元化,而封闭式的信息环境对个人价值观和主流意识形态认同产生了不可估量的负面影响。为了扩大用户的知情权,降低算法的不透明性,更好地维护用户的合法权益,利用技术的可解释性完善算法解释的法律规定无疑是大势所趋的有效途径之一[3]。因此,对比域外相关规定,深入探索我国算法解释制度并提出相关完善建议,期望推动后疫情时代的算法法治建设的理论发展。

2. 算法推荐及解释制度的必要性

2.1. 算法推荐

2.1.1. 概念内涵

《规定》明确了算法推荐技术的概念和算法推荐类型,即“利用生成合成类、个性化推送类、排序精选类、检索过滤类、调度决策类等算法技术向用户提供信息”。在法学领域,有学者认为算法推荐是指网站平台通过算法模型将用户信息数据与内容信息数据进行匹配,实现内容高效聚合、精准分发的一种手段[4];另有学者认为算法推荐是通过追踪用户的网络行为,运用一些数学算法计算出个人特征、环境特征等相关信息,并推测出用户可能喜欢的内容[5]。综上,算法推荐为基于对场景、内容、用户偏好

和平台优先级的权重分配和综合分析,将获取的信息通过设定的算法模型对用户进行画像,实现信息间精准匹配,为用户提供推荐服务。

2.1.2. 运行价值

从运行逻辑看,算法推荐流程可以简单地由三个单元表示,即输入单元、运算单元、输出单元。输入单元,一般为数据集和内容池,担任获取记录用户的特征和行为信息,承担过滤数据、分类赋值、粘贴标签的功能[6]。运算单元,一般为深度学习和模型构建,通过程序设计者预置的代码计算出获取数据与潜在需求之间的相似度。输出单元,一般为推荐列表和内容榜单,将运算结果匹配为相对应的内容推送给用户[7]。

推荐技术能够提升平台的运营效率,从而有效节省公司的资源,产生隐形的商业价值。推荐技术主要有三种获利变现方式:广告主通过平台向用户投放广告实现变现并向平台支付一定费用;商家通过入驻平台向用户售卖商品的同时,按照销售额的一定百分比向平台交纳技术服务费,实现电商变现。而推荐技术促进“标的物”的售卖以获取更多经济利益;用户浏览与购买平台产品、会员点击广告,实现增值服务变现。这三种主流的盈利模式的前期成本属于一次性投入,而人工编辑的成本则会线性增长。所以,推荐技术使用带来的规模效应无疑是企业运营可接收范围内最佳的选择。

2.2. 算法解释的必要性分析

2.2.1. 推荐技术弊端

其一,算法推荐的基础是个人信息。在设计与投入使用前,基本的算法架构、后续自我更新和发展以及对用户获取信息的界定都具有固定性。算法的设计者需要在技术伦理规范和设计发展水平的制约下,使算法在最短的时间内运用尽可能少的步骤达到匹配用户偏好的目的。就目前的发展而言,技术自身的准确性和重复率仍难以保证。其二,流量至上的价值取向弱化了信息的公共价值,价值失序使主流意识形态的宣传被低俗娱乐的消遣替代。其三,机器学习的自动化特征使得程序运行的结果难以解释。算法的高专业性使其只能被业内算法技术人员等少数人掌握,易脱离监管的范畴。

2.2.2. 法律规制欠缺

对于个性化算法推荐,我国采取分散式立法,角度集中而规定分散,缺乏高位阶的专门法律,法律规制效力弱[8]。《电子商务法》第17条规定了网络平台的信息披露义务由此保障用户知情权;《网络安全法》设定了七种个人信息权利,均包含知情权。《个人信息保护法》第14条和第25条对义务和权利的规定更加具体,平台处理个人信息应建立在个人知情且同意的基础上,用户有权在特定情况下要求平台作出说明。由此可见,立法通过赋权用户以降低技术的不透明性,而知情权最本质的保障方式就是对技术信息的披露。但是,个体请求解释不仅在《个人信息保护法》中界定模糊,在《规范》中也是由于设计的复杂性而反复修改具体文件,难以清晰界定。

2.2.3. 司法救济困难

个性化推荐算法会侵犯用户的自主权。例如,美国广告公司 MediaBrix 通过技术实时分析互联网玩家的情绪,在最合适的时间点推送个性化广告来诱导用户购买相关产品或服务,增加用户的消费可能性[9]。在表面看来,这属于民法范畴私法自治的内容,用户的意思形成于个人意志并作出向外部的意思表示。实际上,该同意机制针对人性的弱点,已然影响了用户真正的意志,侵蚀了私法自治。而现有的民法却未有相符合要件进行规制。这也是我国司法案件中仅有的十几例算法推荐案件,基本上为企业在技术运用上涉及著作权的纠纷的原因。

综上所述,由私法过渡到公法,利用算法解释,从源头的立法角度降低技术的不透明性,扩大知情

权赋权用户能够更有效地保障个人权益。通过对技术进一步了解,个体也能够更好地意识到推荐技术对自我沉迷、泛娱乐化等现象的不良引导,对此发挥能动性进行调节。

3. 算法解释制度功能发挥的问题检视

3.1. 商业秘密下公开无依据

当前我国算法监管和司法保护发展程度较低,对算法的源头性规定是目前可操作性较强的选择。从技术角度看,算法解释下的公开不完全不利于企业的商业发展。如果系统的开发者通过事前解释将算法运行于公共服务领域的信息公开披露,及时自我评估并纠正算法偏差,从而推出更加完备的算法系统,能够进一步获取消费者的信赖。但知识产权也是法律中重要的权利,算法推荐技术的公开程度成为争议的热点问题,公开难以有一个明确的法理界定。如果将算法技术贸然公开,将对企业的发展造成损害,对市场经济的健康发展不利^[10]。我们需要权衡利弊,并提供权威性的规定。而《规定》中对两项权利的法理分析并不明确,司法上同样没有典型案例可供实践借鉴。

3.2. 算法披露下评估不完备

《规定》第7条集中规定了算法推荐服务提供者的披露义务和责任,包括信息发布等算法审查和审核机制、制定并公开算法推荐相关服务规则、算法安全评估等管理制度等。此外,第8、9、10条分别规定了算法推荐服务提供者定期审核、信息内容安全管理、不得设置歧视或者偏见的义务。这些义务的背后是对算法解释内容披露程度的引导。其中第16条表明算法服务提供者应当披露算法的事项为:算法的基本原理、目的意图和主要运行机制。第23~29条笼统性规定了监管部门应当建立算法分级分类安全管理制度、算法备案制度和算法安全评估制度。但与当前域外立法强度相比力度仍显不足,算法披露相应评估体系的构建落实还有待完善。

3.3. 算法设计下解释不明确

算法解释代码程序的专业性和复杂性使非业内人士难以理解算法推荐的运行机制,即便公开,其作用也相当有限,如果不及时更新,也会出现问题。所以公众需要通俗易懂的算法解释来深入知悉。《规定》第12条对算法推荐服务提供者信息获取的内容与处理手段作出了初步的规范,争取避免不良影响产生,预防和减少争议纠纷。该条文在征求意见稿的基础上限缩了部分内容。此外,《规定》单一地将算法解释作为事后赋予个人请求权的救济方式,具有离散性,在治理效率、行业发展、和公共利益保障等层面存在明显缺陷^[11]。所以解释的内容需要进一步弹性设计,在时间和方式上提出适用场景和程度变化的具体要求。

4. 算法解释制度的域外经验

4.1. 美国算法解释规定

4.1.1. 权衡权利冲突问题

纽约市议会成员詹姆斯·瓦卡曾要求市政府机构运用自动化决策系统维护治安、分配福利和认罪处罚时,必须公布算法的源代码^[12]。英美法律惯例中对迫于公共利益公开的事由作了特别要求,发展并确立公共利益对商业秘密的抗辩原则。美国判例确定的公共利益披露事由包括维护国家安全、便利公众使用公共设施、维护商业道德和自由竞争的秩序以及保护宪法赋予的基本权利等^[13]。事前检测算法系统的适用性、可靠性相较于事后所需社会资源较少,避免了资源浪费。此外,算法解释有助于增加人们的科技信任,加强个体接受算法决策的意愿。

4.1.2. 解释评估体系构建

美国《2019 年算法问责法案》指示联邦贸易委员会覆盖所有互联网企业，对储存、共享和加工个人信息并进行自动化决策的系统开展相关的评估工作。根据规定，算法自动化决策的解释程度应当结合算法的具体运用场景、个人信息处理内容等进行相应的风险评级，并根据风险等级来考虑算法解释的具体形式。2022 年 2 月，美国俄勒冈州、纽约州和新泽西州提出《算法责任法案》，强调企业应当对算法的有效性等其他因素进行深度评估，让算法技术赋权个体，而不为资本冲刷人本价值理念[14]。

4.2. 欧盟算法解释制度

4.2.1. 解释时间方面

从解释时间方面，算法解释可分为事前解释和事后解释。《一般数据保护条例》中的很多条款在规定收集与处理个人信息前知情性权利的同时，也明确赋予了个体以事前解释的权利。如第 12 条第 1 款中规定数据控制者应将第 13、14 条规定的信息和第 15 至 22 条等有关数据处理与算法逻辑提供数据主体[15]。此外，《一般数据保护条例》第 22 条规定的对自动化处理的拒绝是赋予个体以事后解释的权利。德曼和弗莱斯曼也认为其规定了个人对算法决策的事后的、针对具体决定的解释请求权[16]。但在一定程度上表明了欧盟对事后解释的支持态度，也为我国将两者灵活结合提供了理论窗口。

4.2.2. 解释方式方面

从解释方法方面，可分为人工解释和机器解释。欧盟《一般数据保护条例》和相关规定可以看出对人工解释与机器解释运用的不同要求。《一般数据保护条例》在第 21 条中在对全过程自动化决策的反对中实际上就是在侧面上提出了人工介入的必要性要求[17]。此外，倘若相关沟通与解释完全依靠机器，算法解释机制就可能变相为另一种自动化决策。欧洲学者围绕第 22 条进行了激烈的讨论，算法决策本身就具有人工决策和机器决策两种属性。所以对应的解释也应当适配性地采取多样化的方式。这说明不同于我国《规定》的空白，在欧盟政策制定者心中两种解释方式均应被明文规定。

5. 算法解释制度的优化路径

5.1. 限制个体权利扩张

算法推荐通常应用于公共服务领域，应当维护公共利益。若将推荐服务的提供划归为合同性质，从权利义务相对应的角度入手，技术运营者也应在获取商业利益的同时履行提供相应服务与维护公共利益的义务。适度的算法解释是为了规避科技对人的不必要侵害，贯彻以人为本的理念。因此，对公共利益的保护要求算法技术信息的公开与流通。商业秘密作为知识产品也可以受到相对的限制。面对两种权利之间的矛盾与冲突，需要根据一定的利益权衡准则分析算法解释制度与商业秘密制度之间存在的合理性。在美国算法解释的规定的逻辑底层下，国家或公共利益高于个人利益，个人的活动和利益在必要时需要为社会而舍弃。所以，对商业秘密的权利加以限制是合乎法理的[18]。因此，在我国以商业秘密逃避算法解释中技术披露应当以明令禁止的形式规定在具体的法条之中。基于我国的知识产权保护体系尚未达到欧美国家的完备与发展程度，算法解释的内容不宜延伸至技术核心的源代码。

5.2. 强化解释评估系统

我国有必要从算法披露的初级公开过渡到算法具体透明阶段。算法具体透明要求在立法中除了规定算法透明和可释的原则性条款外，对算法披露的具体事项应予以明确规定。算法披露应公开算法自动化决策的存在、公开算法的运行逻辑和运行结果[19]。除了法律明文规定的具体事项以外，可以采用更多元开放的逻辑。其一，学习美国构建算法评估影响机制并充分发挥企业的能动性，引入内部算法评估制度，

从合规角度利用行业标准初步自我评估算法的准确率等数值对算法推荐技术进行量化评估。我国企业数量庞大,行业标准清晰程度难以保证,可以多主体协作配合。引入联合技术专家对算法解释衡量标准予以动态评估,来适应变化发展的内外部环境[20]。其二,根据美国不同场景下建立多元化层次化算法解释衡量标准,能够更有效防止解释流于形式产生消极影响。我们也应当结合算法的具体运用场景、判断是否涉及相关法益目标并进行风险评级,从而明确算法解释的具体程度[21]。其三,特定的运用场景可以利用举例解释的方式,选取少量算法决策实例进行算法解释。举例解释的优势在于充分契合了个体认知新事物的规律,解释效果较好,能够有效提升算法客体对其的信任度[22]。

5.3. 细化算法解释内容

5.3.1. 事前与事后相辅助

基于算法的内涵与特征,事前解释与事后解释理应并存。《规定》要赋予事前解释在法律层面的合理性,全面完善算法解释内容。一方面,事前解释无法适应算法技术的快速发展和更新迭代的特点,这决定了事后解释的必要性。另一方面,如果取缔事前解释,自动化决策者无需提前履行告知义务,这将无法实现个体对自动化决策与个人信息保留的选择权,降低个体对其的信任度,冲击市场自我调节功能。考虑到算法技术涉及的知识产权问题,不可能简单地要求公布源代码等商业细节。所以,事前解释应当站在义务的角度,适当披露算法实质性参与决策的事实,披露应当以清晰易懂、有意义的方式提供算法参与的关键任务的模式。同时可以采取体系化固定的解释内容,包括技术运行、用户权益、平台义务等。而事后解释采取个案解释,处理算法使用后所带来的危害,为特殊主体提供优质服务。

5.3.2. 人工和机器相配合

在算法资本的逻辑思维中,算法本身的维护成本远远低于人工运行的成本的。如果采取纯粹的人工解释,虽能够最大程度地保证个人权益,但高昂的成本将使算法主体不堪重负,其存在也丧失了基本意义。但是纯粹的机器解释未必能消除个人的疑虑,无法迈向真正意义上的沟通信任,人工介入在欧盟也由此成为强制性的要求,使人本价值观注入科技引导科技向善。事物的方式往往由主体与客观条件所决定。据此,《规定》要建立机器与人工的混合机制,再全面地将解释时间与方式结合。例如,系统化的事前解释可以采取机器解释的方式,在固定和稳定的角度两者相契合。而多元化事后解释则可以采取人工解释的方式。不同行业的企业也应具体根据自身能力、行业特征做出相应的调整。比如垄断性强的行业自身的经济能力强,能够在负担人工解释成本的基础上获取用户信任,实现可持续发展。

6. 结语

在互联网高速发展的今日,我们逐步跨入算法社会,算法推荐显著提高了互联网平台信息分发的效率。《算法推荐管理规定》在前有的零散法律规定上进一步整合和完善了算法推荐的规制,有效回应了我国算法推荐治理体系。通过分析算法推荐私法自治的法律逻辑,发现算法解释在算法推荐治理中的基础性地位日益受到学界与实务界的认可,而我国算法治理中算法解释制度的规范缺乏,于是对比域外的法律经验,提出了一些补充优化的意见,希望我国可以尽快建立一套完善的高阶层算法解释制度,来更好的保护互联网用户的合法权益。

参考文献

- [1] 刘璐璐,张峰.后疫情时代数字化生存的技术哲学思考[J].东北大学学报(社会科学版),2021,23(5):1-7.
- [2] 高志宏.隐私、个人信息、数据三元分治的法理逻辑与优化路径[J].法制与社会发展,2022,28(2):207-224.
- [3] 宋青勋,韩奥林.智能场域下算法对人的“圈禁”与“解放”[J].理论导刊,2023,463(6):75-82.

-
- [4] 方师师. 《算法推荐管理规定》中的政策矩阵与交互想象[J]. 青年记者, 2021, 711(19): 81-82.
 - [5] 陈昌凤, 吕婷. “去蔽”的警示: 算法推荐时代的媒介偏向与信息素养[J]. 中国编辑, 2022, 149(5): 35-39.
 - [6] 温凤鸣, 解学芳. 短视频推荐算法的运行逻辑与伦理隐忧——基于行动者网络理论视角[J]. 西南民族大学学报(人文社会科学版), 2022, 43(2): 160-169.
 - [7] 李静辉. 算法推荐意识形态属性的生成逻辑、风险及治理[J]. 理论导刊, 2022, 447(2): 70-76.
 - [8] 卢海君, 徐朗, 由理. 互联网平台算法推荐的法律规制[J]. 中国出版, 2022, 534(13): 22-28.
 - [9] Pritz, S.M. and Tracking, M. (2016) Zur digitalen Selbstvermessung der Gefühle. In: Selke, S., Ed., *Lifelogging: Digitale Selbstvermessung und Lebensprotokollierung zwischen disruptiver Technologie und kulturellem Wandel*, Springer, Berlin, 129-147.
 - [10] 任安麒. 网络服务平台算法推荐的著作权侵权认定规则[J]. 北京航空航天大学学报(社会科学版), 2023, 36(3): 190-198.
 - [11] 周翔. 算法可解释性: 一个技术概念的规范研究价值[J]. 比较法研究, 2023, 187(3): 188-200.
 - [12] Engstrom, D.F. and Ho, D.E. (2020) Algorithmic Accountability in the Administrative State. *Yale Journal on Regulation*, 37, 800-854.
 - [13] Nissenbaum, H. (2004) Privacy as Contextual Integrity. *Washington Law Review*, 79, 119-158.
 - [14] 程莹. 元规制模式下的数据保护与算法规制——以欧盟《通用数据保护条例》为研究样本[J]. 法律科学(西北政法大学学报), 2019, 37(4): 48-55.
 - [15] Rossi, F. (2016) Artificial Intelligence: Potential Benefits and Ethical Considerations. European Parliament: Policy Department C: Citizens' Rights and Constitutional Affairs.
 - [16] 丁晓东. 基于信任的自动化决策: 算法解释权的原理反思与制度重构[J]. 中国法学, 2022, 225(1): 99-118.
 - [17] 张新宝, 赖成宇. 个人信息保护公益诉讼制度的理解与适用[J]. 国家检察官学院学报, 2021, 29(5): 55-74.
 - [18] 汪庆华. 算法透明的多重维度和算法问责[J]. 比较法研究, 2020, 172(6): 163-173.
 - [19] 丛颖男, 王兆毓, 朱金清. 论算法解释权的重构——全算法开发流治理与分级分类解释框架[J]. 计算机科学, 2023, 50(7): 347-354.
 - [20] 苏宇. 优化算法可解释性及透明度义务之诠释与展开[J]. 法律科学(西北政法大学学报), 2022, 40(1): 133-141.
 - [21] 许可, 朱悦. 算法解释权: 科技与法律的双重视角[J]. 苏州大学学报(哲学社会科学版), 2020, 41(2): 61-69.
 - [22] 李亦飞, 梁巧稚, 薛可. 智能算法推荐的治理困境及提升路径研究[J]. 新闻爱好者, 2021, 527(11): 41-44.