

我国算法歧视的法律规制研究——基于国外规制措施的比较分析

黄麒伦 罗宇 穆紫茜

西南科技大学 法学院 四川绵阳 621000

摘要: 算法歧视伴随着互联网与行业的深度融合,于社会各领域产生了大量新型纠纷问题。本文通过剖析算法歧视运行机理,收集整理相关数据,了解算法歧视现象的成因及困境进行系统分析,提出从设置算法审查制度、新建算法歧视的认定标准、明确算法歧视的归责体系、构建受害者救济体系四个方面对我国的算法歧视规制体系进行完善。

关键词: 算法歧视; 数据法规; 法律规制

Research on Legal Regulation of Chinese algorithm discrimination-- Comparative analysis of foreign regulatory measures

Qilun Huang, Yu Luo, Zixi Mu

Southwest University of Science and Technology, Law School, Sichuan Mianyang, 621000

Abstract: Algorithmic discrimination has led to a variety of new disputes in various fields of society, accompanying the deep integration of the internet and industry. This paper systematically analyzes the operation mechanism of algorithmic discrimination, collects and organizes relevant data, and examines the causes and difficulties of algorithmic discrimination. It proposes four ways to improve China's regulatory system for algorithmic discrimination: establishing an algorithm review system, creating recognition standards for algorithmic discrimination, clarifying the accountability system for algorithmic discrimination, and constructing a relief system for victims.

Keywords: Algorithm discrimination; Data regulation; Legal regulation

一、算法歧视的概念

(一) 算法歧视的定义

基于计算机技术的蓬勃发展,人工智能、数据库、互联网正不断完善自身体系构造,而对于算法歧视的定义,学术界至今难以形成统一认知。

算法歧视的定义需要从算法和歧视两个维度进行把握。算法歧视产生于算法的运行过程当中,传统学者认为具有算法“工具论”的性质,认为算法是在软件中提供控制加逻辑的工具,辅助计算机运行。而先现今一部分学者跳出“工具论”的范畴,将算法当作客观的信息处理方法。在自主学习化的算法时代,应赋予算法学习能力,抽离出必须依附特定客体才能运行的观点。故笔者认为,算法是联通社会数字要素、渗透全民生活的,能够独立运行,具有交互和学习能力的动态机制。

法学角度的歧视是指被法律禁止的、针对特定群体或个人实施的、其效果或目的在于对承认、享有和行使基本权利进行区别、排斥、限制或优待的任何不合理的措施。

综上,笔者认为算法歧视是利用算法进行选择性的数据分析导致对特定群体的不公正待遇。

(二) 算法歧视的成因

算法公平是算法技术设计、实施和有效的核心。因此,有必要从算法歧视形成的原因出发,对其特殊性进行了解。

1. 数据收集的选择偏好

计算机信息处理具有选择、归类、归纳性质,算法运行会对庞大的数据信息进行特定选择,以服务于其它上层的自然语言处理任务。互联网公司高效划分客户群体,会采用算法对数据收集的选择偏好会生成不同用户的使用程度、产品喜好等个人代码,为平台进行区别划价格反馈、差异性服务选择提供参考数据,进而产生“大数据杀熟”等消费者歧视现象。

2. 算法设计者的歧视嵌入

算法本身不具有立场,但通常会受其设计者的情感影响。由于人类思维认知具有天然的局限性,个人偏见往往

会渗透至深层的算法运行。算法设计者将自身的歧视嵌入到算法当中，主观偏见会扭曲正常的算法信息推送处理。

3.模型分类的不合理

算法决策系统的工作原理是使用大量数据训练出准确率很高的机器学习模型来对目标群体进行分类，从而根据预测结果进行决策。在这一过程中，若不施加特殊限制，模型常常会将一些敏感特征纳入分类依据，并赋予较高的权重，从而导致决策结果对某些特定群体产生不公平的结果。

4.算法过程的不公开

算法的搭建和运行始终处于隐秘状态，算法理解具有较高门槛，而在使用过程中，设计者与使用者之间产生坚硬的壁垒，使得用户仅知悉算法的结果，而缺少对信息数据的收集、算法决策的运行的认知，平台往往会利用不对等的信息差异进行“算法歧视”，以满足其盈利需求。

二、算法歧视法律规制的困境

依靠机器计算的自动决策系统在对数据主体做出决策分析时，由于数据和算法本身不具有中立性或者隐含错误、被人为操控等原因，对数据主体进行差别对待，造成歧视性后果。通过分析算法歧视的成因，可见算法歧视具有选择性、过程隐蔽且广泛辐射至不同社会层面。基于其特征，算法歧视在当下必然存在取证困难、认定标准滞后、归责不明确、难以有效救济的规制困境。

（一）算法歧视取证困难

算法最大的特点就是输入数据产生结果。由于算法从输入数据到输出决策结果的逻辑过程并不向外界公开，形成了算法“黑箱”。用户在使用算法时仅能获知算法运行的结果，而算法使用的数据、分析逻辑等关键过程则被算法“黑箱”隐藏。这使得用户在遭受算法歧视后难以获取关键性的过程证据而造成维权困难。

（二）传统认定标准滞后

传统认定歧视的标准以反分类歧视理论为主要内容。反分类歧视理论强调形式上的平等，主张不能基于人的种族、性别、宗教信仰等因素而对人加以区别对待。然而，运用反分类歧视理论难以认定算法歧视的存在。首先，算法歧视的隐蔽性、不透明性使反分类歧视理论难以发挥作

用。其次，反歧视理论的群体性特征不能有效应对算法的个体性歧视，而这一理论在我国并不适用，我国算法歧视的对象是不特定的个体。故需要将个人作为完整的个人加以看待，关注个体公正，而非仅仅聚焦于群体公正。

（三）归责制度有待明确

目前，大多数学者认为算法歧视的担责主体主要是算法设计者和算法使用者。当算法设计者和算法使用者归于同一主体时，毫无疑问全部责任都由该主体承担。然而当算法设计者和算法使用者不归于同一主体时，责任划分却不明确。美国《2022算法问责法案》也提出要规制算法应用者和算法技术提供者，但并未对两者之间的责任划分作出规定。

（四）救济方式亟需完善

目前对于算法歧视受害者的救济方式主要有：引入集体诉讼制度、个人数据赋权等。就集体诉讼而言，算法歧视具有明显的个性化特征，不同的场景下不同的用户被歧视的方面不同，难以适用集体诉讼制度。欧美的很多数据立法都赋予个人数据收集时的知情选择权、数据访问权、数据更正权、数据删除权等一系列权利。个人数据赋权对数据控制者与处理者施加责任，要求数据控制者与处理者满足个人的一系列数据权利，承担维护个人数据安全与数据质量等责任，但通过立法对个人数据赋权需要经过复杂的程序和长久的时间，难以应对日益增加的算法歧视案件的发生。

三、算法歧视法律规制的完善

算法歧视涉及范围广泛及类别多样，因此必须考量各方维度具体的责任分配的规则体系，如此，方能对算法歧视问题进行规制。

（一）构建算法审查制度

算法歧视审查旨在结合既有标准与实际情形，对算法的运行过程、采集模式等方面进行初期审查，能起到一定的预防作用。审查方面应做到程序化，如对算法歧视的主观意图进行认定时，应当注意结合实际情形与间接证据相结合来予以推断证明。同时采取差异化审查，对实施行为进行差异性影响、商业必要性和可替代算法决策进行审查。

（二）构建符合算法歧视特征的歧视认定标准

由于算法歧视的隐蔽性、复杂性，在进行标准认定时，传统的规制方法往往针对性不强，规制精度不够。需要结合具体情形进行有针对性的规制，具体而言可以通过立法和制定行业规则两方面构建认定标准。

从立法层面而言，首先应明确算法价格歧视的违法性认定标准，拓宽价格歧视实施主体的认定范围。从行业规则而言，应组建相关互联网算法发展协会，制定相应行业规则，以行业规则约束经营者行为，对不同情形进行初步认定，形成初步认定标准。同时赋予行业规则一定的认可性，使该认定结果在对经营者行为判定时具有一定效力。同时应向使用者明确平台数据收集和相关算法决策的大致情况，如向用户说明其收集的数据范围、目的、运用的算法等。同时制定算法决策的专门协议，将数据收集及算法决策与其他不得违规的条款内容予以区分。

（三）构建明确的算法歧视归责体系

算法归责在于兼顾从实体和程序两个方面出发。在实体方面，考虑各方主体及其内在关系，根据具体情况及使用程度等确定责任份额，同时适用过错推定原则，保护弱势方的基本权利。算法歧视运行过程涉及算法设计者、互联网平台以及商家。因此，对算法歧视责任主体的考量应当厘清各类主体的法律关系，尤其关注主体参与程度、利益关系、风险控制、主观过错以及算法自主性等因素。在程序层面，基于算法歧视违法认定与举证的难度大、专业化水平较高等情况，应实行举证责任倒置。算法价格歧视的归责仍应当坚守过错原则，建议关注算法本身的特点以及对算法商家合理兼顾，采用过错推定原则。即价格歧视行为且对消费者造成人格权益或财产权益的损害，应当推定算法商家有过错，但允许其进行抗辩。举证责任倒置一方面能保障弱势一方的基本权利，同时也保护经营者一方不受恶意申诉的影响。

（四）构建完善的受害者救济体系

在面对算法歧视所造成的结果，如何完善受害者救济，在于各方主体的行为。法院方面，对算法价格歧视案件应当倒置举证责任，如若消费者提供的证据能够引起法官的合理怀疑，应由经营者证明其不存在价格歧视或价格歧视有正当理由。行业方面，可建立消费者申诉渠道。保障使用者在发现其权益受到平台经营者算法决策的使用和

可能存在损害时，具有留出质疑的权利。同时明确行业内各经营者的义务，如应向使用者履行告知说明义务。

参考文献：

- [1]普通高中课程标准实验教科书（必修）数学（A版）必修3.人民教育出版社
- [2]科尔曼，雷瑟尔森，李维斯特. 算法导论 [M]. 殷建平，徐刚，王刚，等译. 北京：机械工业出版社，2013: 3.
- [3]Bucher T. The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. *Information, communication & society*, 2017, 20(1): 30-44.
- [4]戴维·波谱诺著，李强等译. 社会学. 中国人民大学出版社，1999.
- [5]薛兆丰. 薛兆丰经济学讲义 [J]. 山西财税, 2018(12): 64.
- [6]周伟.《论禁止歧视》[J].《现代法学》2006年第5期.
- [7]谭九生，范晓韵. 算法“黑箱”的成因、风险及其治理 [J]. 湖南科技大学学报(社会科学版)，2020，23(6): 92-99.
- [8]胡元聪，冯一帆. 大数据杀熟中消费者公平交易权保护探究 [J]. 陕西师范大学学报(哲学社会科学版)，2022，51(1): 161-176.
- [9]张欣. 我被美团会员割韭菜: 揭秘大数据杀熟背后隐秘架构 [EB/OL]. (2021-01-08) [2022-04-01]. <https://www.zgghy.org.cn/item/367707728220475392>.
- [10]胡文江，孙龙，高永兵. 一种改进的基于淘宝数据的商品排序算法 [J]. 内蒙古科技大学学报，2013，32(3): 270-272 + 297.
- [11][11]张凌寒.风险防范下算法的监管路径研究[J].交大法学.2018(04)
- [12][12]孙建丽.算法自动化决策风险的法律规制研究[J].法治研究2019(04)
- [13][13]刘友华.算法偏见及其规制路径研究[J].法学杂志.2019(06)
- [14][14]丁晓东.论算法的法律规制[J].中国社会科学.2020(12)

[15][15]Anupam Chander, The Racist Algorithm[J], 115 Mich. L. Rev. (2017): 1023;

[16][16]Jack M.Balkin.2016 Sidley Austin Distinguished Lecture on Big Data Law and Policy: The Three Laws of Robotics in the Age of Big Data, 78 Ohio St. L. J(2017):11239

[17][17]汪庆华:《人工智能的法律规制路径:一个框架性讨论》,《现代法学》2019年第2期。

[18][18]张典恩.反算法歧视:理论反思与制度构建[J].华中科技大学学报(社会科学版).2020(05)

[19][19]Stephanie Bornstein,Antidiscriminatory Algorithms[J].Alabama Law Review,2018,70(2):519-572.

[20][20]Jack M.Balkin,2016 Sidley Austin Distinguished Lecture on Big Data Law and Policy:The Three Laws of Robotics in the Age of Big Data,78 Ohio St.L.J.(2017):1217.

[21][21]刘友华.算法偏见及其规制路径研究[J].法学杂志.2019(06)

[22][22]参见丁晓东:《论个人信息法律保护的思想渊源与基本原理——基于“公平信息实践”的分析》,《现代法学》2019年第3期。