

# 算法歧视对平等权保护的挑战及对策

矫露露

青岛大学法学院, 山东 青岛

收稿日期: 2024年10月12日; 录用日期: 2024年11月6日; 发布日期: 2024年11月15日

## 摘要

互联网的迅速发展使得人工智能技术突飞猛进, 作为其核心的算法技术的运用范围也越来越广泛。但算法并非中立的, 在面对公平与效率时, 往往会舍弃公平而选择效率。随着算法的广泛应用, 加之算法运行的自身特点, 算法歧视愈演愈烈, 由于算法的不可解释性及自身存在的一些缺陷, 对于平等权的传统保护模式也因此受到危机, 传统法律难以应对算法歧视、识别算法歧视存在困难、司法救济途径不畅等问题使得公民的平等权无法得到有效保护。因此, 应从健全法律机制、加强算法监管以及保障权益救济等方面进行规制, 使之沿着法治的轨道健康发展。

## 关键词

算法歧视, 人工智能, 平等权

# Challenges and Countermeasures of Algorithmic Discrimination to the Protection of Equality Rights

Lulu Jiao

School of Law, Qingdao University, Qingdao Shandong

Received: Oct. 12<sup>th</sup>, 2024; accepted: Nov. 6<sup>th</sup>, 2024; published: Nov. 15<sup>th</sup>, 2024

## Abstract

The rapid development of the Internet makes artificial intelligence technology advance by leaps and bounds, and the application scope of algorithm technology as its core is becoming more and more extensive. As a technical means of computer programming, it can greatly improve social efficiency and bring many opportunities for the development of human society. But the algorithm is not neutral, in the face of fairness and efficiency, often will abandon fairness and choose efficiency.

文章引用: 矫露露. 算法歧视对平等权保护的挑战及对策[J]. 争议解决, 2024, 10(11): 177-182.

DOI: 10.12677/ds.2024.1011451

**With the wide application of algorithm and its own characteristics, algorithm discrimination becomes more and more fierce, and the traditional protection mode of equal rights is also in crisis.**

## Keywords

**Algorithmic Discrimination, Artificial Intelligence, Right of Equality**

Copyright © 2024 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

近几年来，人工智能技术突飞猛进，作为一种崭新的生产力在各行各业得到广泛应用，给人类社会带来许多机遇。算法作为人工智能技术的核心要素更是大大提高了社会效率。作为一种为人类服务已达到解决问题的手段或工具，“技术中性论”的思想被人们所信服。但随着算法技术的广泛应用，实践中逐渐出现许多“技术中性论”的反面案例，算法歧视也渐渐暴露在公众视野当中。例如谷歌公司的照片软件将黑人识别为大猩猩，又如在谷歌搜索引擎中搜索黑人的名字更容易出现犯罪的广告，暗示其可能存在犯罪记录。当算法技术应用在招聘时，也可能引发就业歧视，亚马逊公司使用人工智能进行招聘，但该人工智能在为应聘者进行打分时，给女性的分数明显低于男性，经过调查发现该系统在对简历进行分析时，对有“女性”“女子学校”等关键词的应聘者降低分数，该系统的工程师后来修改了代码和算法也没能实际解决该问题，人工智能已经通过深度学习形成了“男性应聘者比女性更加优秀”的认识。以上种种案例的出现引发了人们对算法技术的思考，“技术中性论”也越来越站不住脚，算法也可能存在歧视的行为，并且因其所具备的难以被识别的隐蔽性，算法歧视对平等权的传统保护路径产生冲击，对平等权的保障提出来了新的挑战。对于算法技术带来的法律风险问题，学界已有诸多讨论，其中往往也包括对公民平等权问题的讨论。然而，对于算法歧视对公民平等权所造成的风险挑战及其对策却少有专门研究。平等是法的核心价值，是人权的核心。在此背景下，有必要对其进行系统的梳理，并提出相应的对策，以保证在人工智能时代，在科技发展的同时，公民的平等权保护也能得到长足的发展。

## 2. 算法产生歧视的原因

### (一) “偏见进，偏见出”

算法歧视具有一套严密的逻辑程序，它的基本逻辑就是通过现有的数据去推算预测。通过对群体的数据特征的分析来预测个体，基于以往既有的数据来预测未来。从而套用特定的行为模式通过算法系统性的决策过程来影响个体。[1]大多数的算法歧视归根结底来自于人的歧视，人们对于算法含有偏见的结果不经审查就作出决策的行为构成了所谓的算法歧视。[2]从算法在数据输入时便可能产生歧视因素，因为算法程序启动的前提便是要输入数据，而该数据大多是基于人类决策者或者算法程序员的主观制定产生的，因此，在算法运作的开始，其所依赖的数据便很有可能存在瑕疵，从而导致算法结果的瑕疵。一旦输入的数据存在瑕疵，那么经过算法固定代码逻辑和程序分析的结果也必然存在瑕疵，最终形成歧视结果。也就是“偏见进，偏见出”。再者，算法的系统性使其运算逻辑只能适用于表达普遍的认知经验，对于特殊个例，也只会因其死板的逻辑分析最终产生歧视后果。对于同类问题，系统基于其固有的逻辑程序对数据进行模式化的分析处理，将这些不同但是相类似的数据使用同样的处理方法，最后形成算法。

但对于这些统一的数据可能会存在多种算法，而具体哪种算法是最好的，它的判定过程并非中立。并且统一数据的多种算法的处理结果往往也是不同的。算法的决策程序都是事先确定好的，他并不能进行区分，因此这种处理过程是不准确的。“当今世界大多数的不公不义，并不是来自个人的偏见，而是来自大规模的结构性偏见。”正如尤瓦尔·赫拉利所言，就当前技术的发展，即使像人工智能这种高进技术能够不断地学习和纠偏，仍不能克服具有特殊性、复杂性特征的事务处理的困难，也很难避免歧视后果的产生。

### (二) 算法黑箱——算法的不可解释性

“黑箱”指一些蕴藏着不为人知，不被公开的内部运行结构的系统，即对数据的分析决策的过程是不透明的。算法技术是在大数据和机器深度学习的基础上进行的，其运行的前提是数据，在输入数据信息后对其进行分析处理，最后得出新的数据信息，而输入数据对其进行处理最后输出新数据的过程，这是一个难以被外界所观察和理解的隐层，这就被称为“黑箱”，这个过程我们无法获知，就连算法设计者都无法解释。以色列学者尼沃等人提出了算法黑箱理论：算法本质上就像一个“黑箱”，信息的输入、决策以及结果的输出都是不公开透明的，公众无法得知其决算过程，更不知道决策的具体标准和依据。微软公司曾在推特上发布了一个人工智能聊天机器人，但在其与人类聊天后，不到一天的时间，这个机器人就变成了一个集性别歧视和种族歧视于一身的种族主义者。算法处理数据的能力之所以提升最重要的原因就在于其具有自主学习能力，也就是“机器学习”。随着科技的发展，海量数据被输入其中，算法能够对现有的数据信息进行有效的学习分析，从而得出相应的结论。但算法本质上是一个“黑箱”，它的学习分析的过程是复杂并且无法预测的，它所依靠的数据中包含了各种歧视性信息，即使是“自动算法”，也经常会得出一些算法设计人员都难以预测的结论。

### (三) 算法自身存在缺陷

算法技术运行的本质是模拟人类决策的一种过程，是在大量数据的基础上总结规律然后进行决策。而这种仅靠数据之间规律得出的结果显然是不够全面的。由于机器并不具有人类所独有的意识形态，仅能客观的分析数据，并不会产生思考，在处理数据时只是遵循固化的逻辑。算法是根据人类设置的一套基本固有逻辑来解决问题的，而这种逻辑一般只适用于一些普遍使用的经验事实，但一些个别的小概率事件则被忽略。这很容易导致经验上的相关性被算法夸大为必然存在的因果关系，造成算法逻辑过于绝对，最终导致对少数群体的算法歧视。<sup>[3]</sup>这样也就不能充分的运用数据之间的相关性，可能会对某些数据产生错误认知，将毫不相关的数据相联系，最终引发歧视。机器算法是人类所发明的，其本身就受研发人员自身技术的局限，算法所计算出的结果也远远无法达到人类认知的高度。

## 3. 算法歧视对平等权现有保护路径的挑战

### (一) 传统法律难以应对算法歧视

由于法律的稳定性，使得法律在制定完成时就落后于社会发展，也就具有了滞后性，而算法歧视的出现使得法律滞后性的这一特点更加突出。在社会发展平稳缓慢的情况下，立法者可以通过解释和修订法律来回应社会的需要。但在现代社会，数字技术成为我国社会进入 21 世纪的新兴产物，正处于强势发展的关键时期，伴随其发展而产生的新兴侵权行为也逐渐增多，其形式、广度以及深度也在持续加强，法律所要规制的范围也随之扩大，但是由于稳定性和滞后性的特点，立法者很难跟上技术的发展速度，让立法程序跟上时代。算法技术日新月异，对于立法者来说则应抓住机遇，把握好时机，随着算法技术的不断更新发展，立法者应把握时机。就我国目前的立法规定来看，制定一部法律大概需要二到五年的时间，而根据“摩尔规律”归纳的信息技术进步的速度一般为两年，这也就意味着算法等人工智能技术一般每两年就会更新升级，这也就说明我国目前三次审议立法程序可能无法及时应对对算法歧视的规制

需要,这就极有可能陷入“新的法律制度生效后其内容在实质意义上过时”的两难境地。

### (二) 算法歧视识别困难

随着平等公正的观念不断深入人心,许多国家纷纷采取措施以根除歧视现象。近年来从表面上来看,显性歧视现象正在慢慢消失,但隐含的歧视并未消失,反而愈加严重。而算法歧视作为一种典型的隐性歧视,正逐步渗透到人们的日常生活中。由于算法技术具有高度的专业性,使得算法技术持有者与用户之间形成了巨大的“数字鸿沟”。普通用户基本只能得到并被动地接受最终的决策结果,根本无法追溯并理解以技术代码的形式呈现的算法决策的过程、决策的依据、决策的结构,并发现其中所存在的算法歧视。<sup>[4]</sup>在人工智能时代,我们越来越生活在我们大都没有意识到的矩阵之中,而围绕在我们身旁的隐性霸主就是算法。<sup>[5]</sup>伴随算法技术的更新升级,我们的生活逐渐被算法包围,生活在一种由算法形成的“拟态环境”中。在人工智能时代,我们越来越生活在我们大都没有意识到的矩阵之中,而围绕于我们身旁的隐性霸主就是算法。也就是说,算法得出的结论在很大程度上已经不是对真实的客观环境的反映,而算法所形成的这种“拟态环境”也极易误导大众,造成歧视现象的产生。同时,算法技术具有高度专业性,从而使其具有复杂性的特征,而算法歧视的复杂性使得算法犹如一个“黑箱”,使得互联网用户与算法拥有者之间存在巨大的数据隔阂,借助技术壁垒将互联网用户甚至法律制度与其隔离开来,为构建符合算法自身特点的权利义务等法律制度设置了更高的门槛。由于人工智能的飞速发展,网络的去中心化致使人们很难在算法空间中识别歧视因素的关键节点,再加之算法能力的不断增强,其歧视性更加隐蔽,同时,对歧视监管的反制能力也更加强化。<sup>[6]</sup>尽管目前我国公力与私力相结合的救济模式已经日渐成熟,但由于算法的不可知性,公民对算法歧视的识别能力远远跟不上算法升级的速度。

### (三) 司法救济存在困难

我国对于平等权的保护在司法救济层面也存在一定的障碍。针对算法的规定,我国仅在2021年出台了《互联网信息服务算法推荐管理规定》(以下简称《规定》)。作为一部规章,其效力位阶较低。其次,目前我国法律并未明确规定算法歧视认定标准应当采取何种过错标准,其主观过错难以认定,相关主体完全可以主张算法歧视是由于算法技术的局限性而产生的,从而主张免除责任。作为以计算机为载体的技术,算法对公民平等权的侵害打破了时间和空间的限制,使其侵害范围不断扩大,在这种情况下,管辖权的确定也成为个极易产生争议的问题。再次,对于算法歧视引发的性别歧视、大数据杀熟等现象,其侵权主体的不特定性就引发诸多问题,例如如何判定侵权主体、如何划分责任标准以及如何启动司法救济等。在算法技术的设计到应用的过程中,存在许多主体,这一过程涉及到算法的设计者、研发者、使用者等等,而算法的运行过程可能会受到各个主体的影响,参与该过程的主体都有可能产生歧视因素,成为算法歧视的实际侵权人,但该侵权实际发生在哪一阶段,如何避免各个主体之间相互推卸责任,我国目前尚未对此做出规定。同时,算法歧视多出现于日常生活之中,一般达不到犯罪的程度,而我国法律对于算法歧视的规制整体偏向于“重刑轻民”,对于一般受到侵害的案件,司法机关在处理时则无法可依。最后,由于对平等的认知差异,使得个人对于算法歧视是否会影响到平等权利的认识上也有所不同,司法机关在处理这类案件时应将这些因素全部纳入考虑的范围,这也无疑会大大增加审判的难度。

## 4. 应对算法歧视的对策思考

### (一) 健全保护平等权法律机制

目前,我国并没有针对算法歧视的制度规制,但算法歧视一直在以一种更为隐蔽的方式侵害着公民的合法权益。平台与企业一方掌握算法技术,同时又拥有大量的数据信息和客户,若想对算法歧视进行有效规制从而完善算法技术治理体系,就应构建一套包括透明管理、责任披露等在内的法律制度。为了确保平台企业在社会中扮演的角色与其影响力相匹配,有必要从法律角度明确其应承担的责任,从而增

强这些企业的主体意识和社会责任感。根据我国于 2021 年发布的《关于加强互联网信息服务算法综合治理的指导意见》，文件强调了平台企业需加强对自身算法管理的责任感。这反映出我国政府对于提升平台企业在算法管理方面责任意识关注度日益增加。因此，平台企业应当主动提升自身的管理水平，确保能够有效地履行社会责任，实现社会责任的具体化和实践化。

### (二) 加强对算法的监管

在数字经济时代，人工智能的发展日新月异，要想有效化解算法歧视给公民平等权带来的侵害风险，一方面要在法律层面对算法歧视和算法权力进行有效规制，另一方面，还应建立相应的监督管理制度和专门机构，对算法运行的全过程进行监管，以此来防止算法权力给平等权带来的威胁。2020 年被认为是开启人工智能监管的第一年，欧盟在《人工智能白皮书》中提到要加强对人工智能算法技术的监管，其目的是想通过事前监管来对算法技术的运行进行规制，以此来保障公民的平等权。而与之相反，美国政府在其发布的《人工智能应用监管指南》中则主张人工智能要进行“软治理”，要充分评估监管对人工智能发展的影响，避免采取阻碍人工智能发展的相关监管措施。而在 2017 年我国国务院发布的《新一代人工智能发展规划》中也要求政府要履行对人工智能的监管职责，建立相应的监管体系，对算法运算过程进行严格监管，避免算法歧视的出现。因此，建立专门的监管机构并赋予其处罚权确有必要。在人工智能监管机构对算法的权利主体进行监管，并将审查结果向公众公开的基础上，若能赋予其相应的行政处罚权力，则将对算法技术进行有效且有力的监管。与此同时，算法监管者在监管的同时，也应当确保其自身的监管效率和监管工作的透明度。当监管者对算法的数据收集、设计以及运算进行监管时，在保护数据安全的情况下，应当尽可能的将相关信息向社会公众公开，并鼓励公众参与监督，这样既可以加强多算法技术的监管，同时也能防止其自身的监管行为对公众的平等权造成侵害。为此，必须保证监管机构的公开性和透明度，以保证其监管的有效性。因此应当坚持公开为基础，不公开为例外的原则，并定期审查算法的运算过程是否符合预期以及是否产生侵害公民平等权的结果。在不能向外公开算法的特殊情况下，监管机构也应为有关的利害关系人提供必要的访问通道，让其能有机会对算法是否侵害其平等权进行审查。

### (三) 保障算法歧视下的权利救济

算法的不平等就在于算法的决策者与相对方之间存在信息不对称，两者之间就产生了不平等。在产生算法歧视时，公民在捍卫其平等权时就陷入无能为力的状态。如果只是直接的单向的对算法进行规制，则很有可能会阻碍人工智能的发展与创新。反之，如果能够通过对算法相对方进行赋权，既能让公民的权利得到救济，同时也能减少对人工智能技术发展的阻碍。对算法相对方赋权，首先要增强算法的可解释性，也即赋予相对方要求对算法过程进行解释的权利。一方面，应要求算法设计者以机器学习和算法的可解释性作为首要原则，使算法或机器学习能够为人们所知晓，或者对公众解释为何要如此操作。另一方面，鼓励使用算法决策的机构对算法决策过程、结果和意义等作出解释。<sup>[7]</sup>算法歧视之所以渗透范围越来越广，难以被发现，其根本原因就在于算法研发者与算法相对人之间存在着数据鸿沟。由于算法具有隐蔽性、不可解释性等特点，算法相对人很难了解其运行过程，即便是知道算法运行的逻辑，由于其复杂性和专业性，也不能完全理解，这也就使得相对方很难意识到自己的权利受到侵害，即使知道权利受到侵害也无法举证证明。针对这一问题，算法解释权也就成为最为关键的一环。另外，算法技术是按照一定的逻辑进行运算，其面向的一定范围内的公众，因此其具有机制性和广泛性。因此，在算法歧视受害者数量较多时引入集体诉讼制度能够有效缓解二者之间由于数据鸿沟而产生的力量失衡。但要求进行算法解释应当成为进行集体诉讼的前置程序，即算法歧视的受害者在权利受到侵害时，应先向算法决策者提出解释算法的要求，若此时算法决策者能够进行详细的解释或者双方能够达成和解，则就没必要进行集体诉讼。若算法决策者无法进行合理的解释，那么算法的受害者就正当性和合理性提起集体诉讼。

## 5. 结论

随着数字经济的发展,算法技术已经逐渐融入到各行各业,并成为推动经济发展的重要力量。但算法歧视正在以一种隐蔽的方式对公民的平等权产生一定的冲击,造成社会公共利益的损失,因此有必要对其进行规制。在社会的智能化转型过程中,确保平等的为社会和公民提供保护至关重要。然而,由于法律的滞后性和当前我国在算法领域的研究尚显不足,对算法歧视问题的规制和对公民平等权的保护工作亦相对滞后。为应对这一挑战,围绕算法技术分层次、分领域的构建一套完整有效的算法治理体系尤为重要,在此基础上对算法歧视进行有效规制,才能彰显公平正义原则,有力维护公民的平等权,同时也能为经济的健康发展提供坚实的法治保障。

## 参考文献

- [1] 张玉宏,秦志光,肖乐. 大数据算法的歧视本质[J]. 自然辩证法研究, 2017, 33(5): 81-86.
- [2] 孟令宇. 从算法偏见到算法歧视: 算法歧视的责任问题探究[J]. 东北大学学报(社会科学版), 2022, 24(1): 1-9.
- [3] 刘朝. 算法歧视的表现、成因与治理策略[J]. 人民论坛, 2022(2): 64-68.
- [4] 石颖. 算法歧视的缘起、挑战与法律应对[J]. 甘肃政法大学学报, 2022(3): 58-68.
- [5] 崔靖梓. 算法歧视挑战下平等权保护的危机与应对[J]. 法律科学(西北政法大学学报), 2019, 37(3): 29-42.
- [6] 吴喜,梁阿敏. 新媒体视域下的平等权保护——对网络仇恨言论的反思[J]. 河北法学, 2020, 38(7): 24-38.
- [7] 刘友华. 算法偏见及其规制路径研究[J]. 法学杂志, 2019, 40(6): 55-66.