

算法性别歧视的法律规制路径

崔泽姝

青岛大学法学院, 山东 青岛

收稿日期: 2025年12月20日; 录用日期: 2026年1月14日; 发布日期: 2026年1月26日

摘要

本文聚焦人工智能时代的算法性别歧视问题, 探讨其法律规制路径。算法性别歧视是用人单位或平台利用算法技术对女性实施的不合理差别对待, 本质上侵犯女性人身与财产权益。当前问题源于基础数据含偏见、算法设计缺乏监管机制及侵权追责法律规定模糊。本研究指出, 需通过强化算法数据审核、加强个人信息保护来清理算法数据性别歧视因素, 建立健全算法审查与监督制度, 构建算法性别歧视侵权追责路径, 以实现算法公平与性别公正, 推动人工智能技术良性发展, 保障女性平等就业权。

关键词

算法性别歧视, 平等就业权, 妇女保护, 反歧视

Study on the Legal Regulation Paths of Algorithmic Gender Discrimination

Zeshu Cui

Law School of Qingdao University, Qingdao Shandong

Received: December 20, 2025; accepted: January 14, 2026; published: January 26, 2026

Abstract

This paper focuses on the issue of algorithmic gender discrimination in the age of artificial intelligence and explores its legal regulation paths. Algorithmic gender discrimination refers to the unreasonable differential treatment imposed on women by employers or platforms using algorithmic technologies, which essentially infringes upon women's personal and property rights and interests. The current problem stems from biased basic data, lack of supervision mechanisms in algorithm design, and vague legal provisions on liability for infringement. This study points out that it is necessary to eliminate gender discrimination factors in algorithmic data by strengthening algorithmic data review and personal information protection, establish and improve algorithm review and

文章引用: 崔泽姝. 算法性别歧视的法律规制路径[J]. 争议解决, 2026, 12(1): 228-234.

DOI: 10.12677/ds.2026.121029

supervision systems, and construct liability investigation paths for algorithmic gender discrimination infringement. These measures aim to realize algorithmic fairness and gender justice, promote the sound development of artificial intelligence technology, and protect women's equal employment rights.

Keywords

Algorithmic Gender Discrimination, Equal Right to Employment, Women's Protection, Anti-Discrimination

Copyright © 2026 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

在人工智能技术深度应用于社会治理与商业运作的当下，算法决策已成为就业筛选、资源分配等领域的核心机制。然而，算法技术在提高效率的同时，也衍生出新型性别歧视形态——算法性别歧视。这种歧视借助数据训练与算法运算，将传统社会的性别偏见以技术化、隐蔽化的方式渗透到决策流程中，如就业场景中算法基于含偏见的历史数据生成性别偏好结果，或通过不透明的运算逻辑对女性实施差别对待，严重侵害女性平等就业权、人格尊严等合法权益。

我国《宪法》《民法典》等虽确立了性别平等原则，但面对算法歧视这一技术异化现象，现行法律体系存在明显规制短板：基础数据中性别偏见的法律清理机制缺位，导致“偏见数据输入 - 歧视结果输出”的恶性循环；算法设计的黑箱特性使监管难以介入，利益驱动下的歧视性编程缺乏有效约束；侵权责任主体、归责原则的模糊界定，则让受害者救济陷入制度困境。在此背景下，如何构建覆盖数据源头、算法设计到责任追究的全链条法律规制体系，成为破解人工智能时代性别平等难题的关键。本文基于算法性别歧视的侵权本质，剖析现行法律规制漏洞，尝试提出能够切实缓解基于算法就业歧视导致的女性生存困境的可行之策。

2. 算法性别歧视的界定及侵权本质

(一) 算法性别歧视的界定

算法性别歧视是指用人单位或者平台运用算法技术，以算法的语言输入和训练数据为基础对女性做出的不合理差别对待行为，造成了女性合法权益受损的结果^[1]。算法性别歧视打破了传统法律针对“直接歧视”构建的规制范式，逐步转向以“间接歧视”和“结构性偏见”为核心的全新样态。这类歧视实质上是传统性别歧视在数字时代的技术化演化结果，其产生既归因于训练数据里男性主导职业的数据失衡、技术团队性别结构不均带来的设计偏向，也因为算法决策流程的非透明性造成女性举证障碍与责任主体认定困境，最终使得性别不平等问题的突出。

(二) 算法性别歧视侵权本质

《中华人民共和国宪法》明确规定所有公民具有平等的法律地位¹，并对女性权利作出特别规定。²

¹ 《中华人民共和国宪法》第三十三条 凡具有中华人民共和国国籍的人都是中华人民共和国公民。中华人民共和国公民在法律面前一律平等。国家尊重和保障人权。任何公民享有宪法和法律规定的权利，同时必须履行宪法和法律规定的义务。

² 《中华人民共和国宪法》第四十八条 中华人民共和国妇女在政治的、经济的、文化的、社会的和家庭的生活等各方面享有同男子平等的权利。

《中华人民共和国妇女权益保障法》及《中华人民共和国民法典》(以下简称《民法典》)均遵循“法律面前人人平等”的宪法原则。³特别是《民法典》在总则编、物权编、人格权编以及婚姻家庭编等分编中,均体现对女性合法权益的维护。算法歧视现象的本质,主要体现为对女性人身权益及财产权益的侵犯。

1、算法性别歧视侵犯女性人身权益

算法性别歧视对女性的人身权益侵害,核心体现为对人格权的损害,《民法典》已针对人格权制定了相应的法律规范。算法性别歧视引发的侵权情形具体可分为两类:其一,女性的名誉权与人格尊严遭受侵害。比如招聘平台算法因训练数据含性别偏见,自动给女性求职者简历打低分或过滤,标注“适合行政/内勤”等标签,既以算法偏见否定其职业能力与社会评价,又以性别刻板印象贬低其职业价值。其二,女性健康权与生命权的保障存在显著短板。以汽车正面碰撞测试为例,女性假人在测试中出现死亡或重伤的概率明显高于男性假人,这一现象折射出相关科技研发过程未能充分兼顾女性的生理特点与现实需求。

2、算法性别歧视侵害女性财产利益

现实生活中,女性在就业求职时往往遭遇多重阻碍,这一将现实直接冲击其家庭的经济支撑。妇女平等就业权既是一项重要的基本人权,亦是劳动权利范畴内的核心权利。《中华人民共和国劳动法》《中华人民共和国妇女权益保障法》等多部法律规范,围绕着妇女平等就业权的保护与就业性别歧视的禁止,作出了细致且全面的规定。《中华人民共和国劳动法》不仅明确对劳动者在就业时被歧视的相关情形做出了禁止,划定了招录环节的反歧视准则,还进一步确认了女性在劳动权过程中女性对免受歧视合法权益的拥有。

3. 算法性别歧视成因

(一) 算法原始数据存在歧视

1、现实社会存在性别歧视偏见

算法设计是整个社会不断发展向前的产物,其输入的种种算法数据也将反应不同群体、不同价值观、不同意识形态的人类思维与意识。而在当今社会中仍然存在种种歧视现象。日常生活与公共场域中,性别歧视仍以隐蔽且顽固的姿态渗透在社会各个角落。校园里,理工科专业常被默认是男性主导的领域,女生的科研潜力与动手能力易被贴上不如男性的标签,参与竞赛、项目时也容易遭遇隐性的名额倾斜。家庭分工中,育儿、家务的责任仍被惯性推给女性,即便女性承担着与男性同等强度的职场工作,也难逃“理应多顾家”的道德绑架。女性的职业成就常因性别而被无端质疑,而男性若选择幼教、护理等柔性职业,也难免被人以歧视目光审视。这种双向的歧视既固化了非此即彼的性别刻板印象,也严重压缩了不同性别个体的多元发展空间。因此当这些歧视性的内容不可避免地被算法所收集,算法输出的内容也将带上性别歧视的色彩。在算法制造过程收集的数据中,男性就业者的数据比例自然而然地高于女性就业者,对于不足数据,算法设计人往往采用合成数据进行,但是合成数据依然是由原始数据生成的,原始数据当中蕴含的观念趋向依然会存在于合成数据当中^[2]。

2、算法设计存在性别歧视偏见

算法归根结底是人的产物,其运行逻辑与数据基底均刻有人类社会的印记,算法与数据的设计者本身就可能暗藏未被察觉的性别歧视偏见。这类偏见并非源于主观恶意,更多是潜藏在设计者成长经历、社会认知与行业惯例中的隐性倾向,比如在职业推荐算法的训练数据筛选中,不自觉地侧重男性主导领域的样本,或是在简历匹配模型里,默认将某些“女性化”职业标签与低薪资、低晋升空间相关联。即便

³ 《中华人民共和国民法典》第九百九十条 人格权是民事主体享有的生命权、身体权、健康权、姓名权、名称权、肖像权、名誉权、荣誉权、隐私权等权利。除前款规定的人格权外,自然人享有基于人身自由、人格尊严产生的其他人格权益。

设计者主观上秉持性别平等的原则，也很难真正实现算法的公平公正。算法的核心运行逻辑往往围绕利益最优化展开，在效率优先的导向下，极易忽视社会公平等价值维度。更关键的是，算法决策依赖复杂的模型运算与数据处理流程，其“黑箱”特性使得外界难以穿透技术壁垒，对决策依据的合理性、结果的公平性展开有效核验与追溯。这种技术壁垒不仅为算法歧视提供了隐蔽空间，也大幅提升了监管的技术门槛与实施难度。值得注意的是，职业在“算法公平的认知情况”“对技术与策略性举措解决性别平等问题的认识”两方面无显著差异，这或许再次表明，即便作为技术主体，从业者们也较少接触并掌握专业化的技术解决进路[3]。

（二）监管算法设计的法律缺位

尽管我国已经开始对数据算法监管问题采取一系列措施，但是现行的法律体系对于算法结果如何有效规制仍在起步阶段。算法在处理各类现实问题时通常以利益最优化为核心导向，但算法运行逻辑的不透明性使得外界无法对其决策的合理性与公平性展开有效核验。在劳动力市场的实践场景中，女性群体客观上需要兼顾家庭责任与职业发展，但部分企业认定招录女性员工难以实现利益最大化与资源利用高效化，甚至以女性存在产假、哺乳期等阶段性离岗情形、“不擅饮酒”等片面化理由，在招聘环节呈现出明显的男性偏好。此类企业的隐性偏见与逐利诉求极有可能被嵌入算法设计的底层架构，由此生成的算法决策结果本质上是经济利益驱动下的算法性别歧视，这一非公平性后果最终则由对算法运行机制毫不知情的女性劳动者群体承担。

算法歧视侵害女性名誉权与人格尊严的侵权责任，需契合传统四要件并适配技术特性，核心难点在因果关系与过错认定。《个人信息保护法》第24条要求自动化决策透明公平，为要件认定提供关键依据。⁴因果关系上，算法“黑箱”导致数据偏差、模型缺陷等多因素交织，链条断裂且证据不对称，原告难追溯损害关联。过错认定中，算法歧视常以“技术中立”为幌子，开发者无意识偏见、多主体责任边界模糊，主观心态难以界定。而我国针对算法设计环节的监管法律尚处于起步阶段，相关规范仍存在显著缺位，未对算法设计的公平性审查、数据训练的性别偏见规避设置刚性义务，进一步加剧了侵权责任认定与追责的现实困境。

（三）侵权责任法律规定模糊

人工智能算法最大的挑战就在于现行法律对其缺乏有效的追责机制[4]。这将导致算法性别歧视进一步加剧，某种程度说是社会责任混乱的起点。

1、歧视结果归责主体不清

在算法性别歧视产生潜在风险或实际损害后果的前提下，对算法相关责任方启动追责程序具有现实必要性，人工智能算法诱发的性别歧视问题，其形成机理与现实社会中的一般性性别歧视存在本质区别。当人工智能应用场景下的性别歧视事件发生时，需对涉事责任主体开展精准追责，这一过程随即衍生出责任主体认定与责任边界划分的核心难题，但现行立法尚未针对此类问题作出清晰明确的规制，若仅聚焦于权益保护层面，却未能明确算法歧视侵权后的责任主体、责任划分类型、归责原则等关键内容，将直接导致算法治理的权益保护路径陷入实施困境，进而加剧算法性别歧视的恶性循环。

2、责任类型及归责原则不明晰

算法性别歧视的危害主要体现在对女性人身权益以及财产权益的侵害。我国《民法典》侵权责任编中，尚未规定算法侵权的相关条款。现存法律对能否根据一般侵权责任来解决人工智能算法侵权问题，

⁴ 《中华人民共和国个人信息保护法》第二十四条 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。通过自动化决策方式向个人进行信息推送、商业营销，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。通过自动化决策方式作出对个人权益有重大影响的决定，个人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。

不可行时是否能够将其归入特殊侵权责任类型等问题均缺乏明确的法律规定。此外,我国《民法典》侵权责任编的归则责任体系主要以过错责任与无过错责任为核心来进行构建,过错推定责任的适用以及无过错责任的认定,均须以法律的明文规定作为前置依据。这意味着,在缺乏专门法律规范的情形下,相关算法侵权纠纷应当适用一般侵权行为的过错责任原则。而基于算法性别歧视的特殊性与复杂性,其能否归入特殊侵权责任的规制范畴,尚需要进一步的分析论证。

4. 解决算法性别歧视的法律途径

(一) 强化算法数据审核与个人信息保护

1、强化算法数据源头审核

强化算法数据源头审核需构建全流程、多维度的系统性机制,从流程规范、主体协同、标准界定到责任约束形成闭环,切实阻断性别偏见的源头传导。首先,要规范数据全生命周期审核流程,明确数据采集、筛选、清洗、标注各环节的审核节点,对涉及就业、薪酬、晋升等与性别权益密切相关的算法数据实施全链条审核,重点核查数据样本的性别代表性是否均衡,杜绝因样本失衡导致的算法偏见。其次,组建跨领域审核团队,吸纳性别平等领域专家、算法技术人员、法律从业者及社会公益代表共同参与,借助多元视角精准识别显性性别歧视标签与隐性偏见特征,比如对将“女性”与“离岗风险”“效率低下”等不当关联的数据集进行专项清理。同时,建立量化的性别公平审核标准,明确数据中性别相关特征变量的合理使用范围,对可能引发歧视的敏感变量设置使用限制。另外可借助区块链技术对全流程备份,以便日后随时审查,确保数据不含偏见因素[5]。健全审核溯源与责任追究制度,要求算法开发主体完整留存审核记录并定期报备,对因审核缺位、失职导致算法输出性别歧视结果的,依法依规追究相关企业及个人的责任,通过刚性约束倒逼审核机制落地见效。

2、加大个人敏感信息保护力度

我国法律尚未对个人一般信息与敏感信息的收集处理标准作出区分,而两类信息理应采用不同制度规制,敏感信息的应用应该有更严格的规定[6]。加大个人敏感信息保护力度是破解算法性别歧视的关键路径,核心在于通过规范性别相关敏感信息的采集与使用,从数据源头切断偏见传导的可能。算法性别歧视的重要诱因之一,是性别信息被过度采集并作为核心决策变量,进而衍生出差异化评价逻辑。因此,需严格界定算法应用中个人敏感信息的采集边界,明确性别、婚育状况等与能力无直接关联的信息不属于必要采集范畴,禁止企业在招聘、薪酬核算等场景过度索取此类敏感信息。同时,建立敏感信息使用限制机制,对确需采集的性别相关信息,明确其使用范围与目的,禁止将其作为算法决策的核心权重变量,防范算法通过关联分析放大性别偏见。此外,强化敏感信息全流程监管,要求算法开发主体落实数据最小化原则,完善信息加密与访问控制措施,严厉打击滥用性别敏感信息训练歧视性算法的行为。通过筑牢个人敏感信息保护防线,减少算法对性别特征的不当依赖,从数据层面为性别公平提供保障。

(二) 建立算法性别歧视审查监督制度

1、提升算法运行公开透明度

推动算法公开透明,是公众有效对算法进行监督从而建立公众信任的关键之举。在符合法律规定和社会公众利益的前提下,推动算法公开透明是当前应对算法性别歧视问题呼声最高的解决路径。在抵制算法歧视时,无论是通过政策和法律“他律”,还是通过行业进行“自律”,破解算法黑箱都是重中之重。对于一个企业来说首先要做到的就是自律,并致力于针对算法制定透明性标准,保证算法决策是公开透明的[7]。以破解算法性别歧视为目标,提升算法公开透明度需构建靶向性治理体系,核心路径包括三方面。第一,建立场景化分级披露机制,针对招聘、信贷、职业推荐等性别偏见高发场景,强制要求披露算法决策的核心依据、数据筛选规则,明确剔除性别作为非必要差异化变量的条款。第二,规范公开

信息的呈现标准,采用公众可理解的标准化表述,避免专业壁垒导致的知情权虚化,保障不同群体均能有效监督算法是否存在性别偏向。第三,搭建公开化监督与反馈闭环,设立专门的算法性别歧视投诉通道并公示处理流程,定期向社会发布算法公平性核查报告,以全流程公开倒逼算法设计与运行摒弃性别偏见。

2、完善社会共同参与监督模式

现阶段反算法性别歧视执行的面临法律规范并不完善、专业性薄弱以及执行力不足等问题。算法数据输入、数据应用及算法数据收集处理构成了一个错综复杂的产业链,涉及社会各个方面的主体。因此,对算法招致的性别歧视问题不能仅限于政府相关部门的监管,对其他利益方尤其是受歧视群体的监督同样十分重要,需吸纳企业、组织及社会团体等共同参与算法性别歧视治理。可以由政府牵头,明确各方在审查监督中的角色定位以及合理的权限,形成行业自律、政府监管与公众参与相结合的综合治理模式。在针对算法性别歧视的专门制度规范方面,我国应通过制定具体应用领域的算法技术标准将算法和法律相联结,明确算法主体、责任分配和法律后果强化法律问责机制,强化算法设计者、决策者、应用者的解释义务[8]。

(三) 明确算法性别歧视侵权追责路径

1、明确多元责任主体的追责边界

如前所述,算法性别歧视会给女性带来各种人身损害和利益损害,而在侵权行为发生时是否能要求人工智能独立承担侵权责任,前提是要厘清人工智能是否具备民事主体地位,是否能独立承担责任。根据我国现行法律,人工智能在当下阶段并没有类似于人类的社会关系和地位,不具有独立人格,也不具备独立担责的前提,所以目前在我国现有法律体系中,人工智能的民事主体资格不被认可[9]。

2、确定侵权责任类型及归责原则

首先,若行为人通过算法操控、指令输入或利用人工智能应用谋利、实现特定目的并造成损害,该行为人即为侵权主体,可援引我国传统侵权责任规则予以规制。而当人工智能在无外界干预下自主致人损害时,因其致害成因具有多元性,如何合理界定侵权责任便成为研究重点。

若适用一般侵权责任规则,侵权人需存在过错方需承担责任,但人工智能的自主运行特性,使得过错认定与责任归属成为实践难题。现阶段可行的探索方向之一是对产品责任的责任主体范围进行扩充。对于数据收集者来说,当无法举证其进行数据收集存在恶意,就应当认定其数据来源合法合理,该主体无需承担侵权责任。反之,若有明确证据证实原始数据本身存在歧视性倾向,或在侵权后果发生后未积极履行配合调查、补救等义务,则数据收集者需对侵权后果承担无过错责任。

3、构建全流程追责实施机制

解决算法性别歧视,可以在前置审查环节、运行环节、追责实施环节等多个方面入手,确保各个阶段责任落实到位。前置审查环节,应建立强制性性别公平性评估制度,要求开发者提交训练数据性别平衡性说明、算法决策逻辑无偏见论证报告,未通过评估的算法不得上线。运行阶段,强制平台留存完整决策日志,明确数据处理、模型运算的全链路追溯路径,确保性别歧视行为可定位。追责实施环节,实行举证责任倒置规则,由算法相关责任主体证明其决策无性别偏见,同时引入第三方专业机构开展技术核查,出具权威评估意见。

5. 结语

大数据时代对男女性别平等提出了更高要求,这需要对算法导致的性别歧视问题进行重点治理。该现象的产生主要源于基础数据存在偏见、算法设计透明度不足以及侵权追责相关法律法规规范缺失三大层面。因此,从法律视角切入实施针对性干预,是当前兼具可行性与实效性的破解路径。针对数据层面的弊病,

可通过专项立法对基础数据中的性别歧视问题进行清除,同时健全算法设计监督机制,确保监管主体能够及时、有效地应对各类新情况与新问题。此外,还需明确算法歧视侵权后的责任追究规则,为追责措施的切实落地提供制度支撑。唯有通过多维度规制手段的协同发力与综合施策,才能最大程度地规避风险、降低损失,有效消解算法性别歧视问题,实现算法公平与性别公正的既定预期目标,推动人工智能算法技术释放其应有的社会价值与技术价值。

参考文献

- [1] 崔慧慧. 算法性别歧视下女性平等就业权实现的困境与路径[J]. 中国价格监管与反垄断, 2024(12): 62-64.
- [2] 张广智. 算法时代下劳动性别歧视防范的法律规制研究[J]. 文化学刊, 2024(6): 131-134.
- [3] 赵宇超, 周旅军. 人工智能性别歧视的伦理治理——基于 AI 从业者技术性别价值观的实证调查[J]. 西南民族大学学报(人文社会科学版), 2024, 45(10): 155-163.
- [4] 张玉洁. 论我国人工智能的法治化进程: 现状、挑战与革新[J]. 广州大学学报(社会科学版), 2019, 18(2): 101-108.
- [5] 闫文光. 从数据角度治理算法歧视[J]. 五邑大学学报(社会科学版), 2020, 22(1): 70-74, 95.
- [6] 王智慧. 算法社会下个人信息保护的难题[J]. 天水行政学院学报, 2020, 21(2): 93-97.
- [7] 马琬晴. 算法性别歧视的法律规制研究[D]: [硕士学位论文]. 武汉: 中南民族大学, 2021.
- [8] 甘湫, 马亮. 算法性别歧视: 特征、类型与治理[J]. 山东行政学院学报, 2024(2): 97-105.
- [9] 吴汉东. 人工智能时代的制度安排与法律规制[J]. 法律科学(西北政法大学学报), 2017, 35(5): 128-136.