

算法偏见与机会公平：电商平台人力资源管理中的算法伦理研究

董祥婷

江苏大学管理学院, 江苏 镇江

收稿日期: 2026年5月16日; 录用日期: 2026年5月29日; 发布日期: 2026年7月17日

摘要

随着人工智能技术在电商平台人力资源管理中的广泛应用, 算法偏见已成为威胁机会公平的核心伦理挑战。本文基于技术伦理学与组织公平理论, 系统剖析电商平台算法人力资源管理中的偏见来源、表现形式及其对机会公平的侵蚀机制。研究发现, 算法偏见主要源于历史数据偏见、特征选择偏见、模型设计偏见与反馈循环偏见四个维度; 在招聘筛选、任务分配、绩效评估、薪酬定价等环节呈现出歧视性分配效应; 并通过“黑箱”不透明性、计算不对称性与权力失衡等机制侵蚀劳动者的机会公平。本文提出“算法向善”的伦理治理框架, 从数据治理、算法审计、透明解释、人机协同、制度保障五个维度构建机会公平的保障机制, 为电商平台实现技术效率与伦理价值的平衡提供理论支撑与实践路径。

关键词

算法偏见, 机会公平, 电商平台, 人力资源管理

Algorithmic Bias and Equality of Opportunity: Algorithmic Ethics in E-Commerce Platform Human Resource Management

Xiangting Dong

School of Management, Jiangsu University, Zhenjiang Jiangsu

Received: May 16, 2026; accepted: May 29, 2026; published: July 17, 2026

Abstract

With the widespread application of artificial intelligence technology in e-commerce platform human

文章引用: 董祥婷. 算法偏见与机会公平: 电商平台人力资源管理中的算法伦理研究[J]. 电子商务评论, 2026, 15(7): 598-606. DOI: 10.12677/eci.2026.157797

resource management, algorithmic bias has become a core ethical challenge threatening equality of opportunity. Based on technology ethics and organizational justice theory, this paper systematically analyzes the sources of bias, manifestations, and erosion mechanisms of equality of opportunity in algorithmic HR management of e-commerce platforms. The study finds that algorithmic bias mainly stems from four dimensions: historical data bias, feature selection bias, model design bias, and feedback loop bias; presents discriminatory distribution effects in recruitment screening, task allocation, performance evaluation, and compensation pricing; and erodes workers' equality of opportunity through mechanisms such as "black box" opacity, calculative asymmetry, and power imbalance. This paper proposes an "Algorithm for Good" ethical governance framework, constructing a safeguard mechanism for equality of opportunity from five dimensions: data governance, algorithmic auditing, transparent explanation, human-machine collaboration, and institutional guarantee, providing theoretical support and practical pathways for e-commerce platforms to achieve a balance between technical efficiency and ethical values.

Keywords

Algorithmic Bias, Equality of Opportunity, E-Commerce Platform, Human Resource Management

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

1.1. 研究背景与核心问题

数字经济时代，以网约车、外卖配送、直播电商为代表的电子商务平台正在经历人力资源管理的“算法化”转型。算法不仅替代了传统的招聘筛选、任务分配、绩效评估、薪酬计算等 HR 职能，更创造了全新的管理逻辑与决策模式。然而，算法在提升效率的同时，也带来了深刻的伦理挑战——算法偏见 (Algorithmic Bias) 正在系统性地侵蚀劳动者的机会公平。

算法偏见是指人工智能系统在数据输入、模型训练、决策输出过程中产生的系统性偏差，导致对某些群体的不公平对待。研究表明，算法偏见在电商平台的 HR 实践中已呈现出多重形态：招聘算法因训练数据的历史偏见而对特定群体产生系统性排斥；派单算法因地理位置偏好而边缘化基础设施薄弱区域的劳动者；评分算法因客户评价的偏见性倾向而压低部分劳动者收入；定价算法因“算法工资歧视”而对相同工作支付差异化报酬。

更为严峻的是，算法偏见在电商平台中具有独特的隐蔽性与扩散性。传统歧视往往表现为显性的规则排斥，而算法偏见嵌入在“黑箱”模型中，难以被识别、举证与纠正。同时，平台经济的跨地域性使算法偏见能够跨越国界快速传播，形成“数字歧视”的全球网络。研究表明，算法控制下的数字劳动过程呈现出“自我剥削”与时空修复的新特征，算法管理中的偏见问题已成为零工经济劳动者权益保护的核心挑战[1]。

核心问题在于：当电商平台宣称“技术中立”与“机会平等”时，算法系统却在实践中复制并放大了社会既有的结构性不平等。劳动者面临“算法歧视”与“算法不透明”的双重困境——既遭受不公平对待，又无从知晓歧视如何发生、如何申诉。这种“机会公平的算法悖论”亟待理论回应：算法偏见如何在电商平台 HR 管理中生成？以何种形式表现？如何侵蚀机会公平？又如何通过伦理治理予以矫正？

1.2. 研究意义与创新点

本研究的理论意义在于：第一，将算法伦理研究引入电商平台人力资源管理领域，构建“算法偏见

- 机会公平”的分析框架,拓展技术伦理学的应用边界;第二,系统揭示算法偏见在平台 HR 全链条中的生成机制与表现形式,丰富算法歧视的理论认知;第三,提出“算法向善”的伦理治理路径,为负责任的人工智能应用提供规范性指引。

实践价值在于:为电商平台识别与纠正算法偏见提供操作指南;为政府制定算法治理法规与公平就业政策提供理论依据;为劳动者维护算法公平权益提供认知框架与行动策略;为构建包容、公平、可持续的数字经济生态贡献伦理智慧。

2. 算法伦理与机会公平的理论基础

2.1. 算法伦理学的核心议题

算法伦理学关注人工智能系统在决策过程中产生的道德问题,核心议题包括公平性、透明性、问责性、隐私保护与自主性。在人力资源管理领域,算法伦理的核心在于确保技术决策不损害人的尊严、权利与发展机会[2]。

公平性要求算法决策不因年龄、地域等受保护特征而产生歧视性结果。研究表明,公平性具有多维内涵:个体公平(相似个体应得相似结果)、群体公平(不同群体应得平等机会)、过程公平(决策过程应无偏见)。然而,这些维度在实践中往往存在张力,完全的公平难以实现,只能在特定情境中寻求“足够公平”。

公平的多维内涵存在内在冲突:个体公平与群体公平的张力、过程公平与结果公平的矛盾。算法审计主要保障群体公平,透明解释主要保障过程公平,人机协同试图在两者之间寻求平衡。这些工具之间存在取舍关系,伦理治理需审慎权衡。

透明性要求算法决策过程可被理解、解释与审查。“黑箱”算法(Black Box Algorithms)因不可解释性而难以被监督,劳动者无法理解派单逻辑、定价规则、评分机制,丧失知情权与申诉权。透明性不仅是技术问题,更是权力问题——算法不透明性强化了平台的信息优势与劳动者的弱势地位。

问责性要求明确算法决策的责任主体。当算法产生歧视性结果时,是算法开发者、平台运营者、数据提供者还是算法本身应承担 responsibility?平台常以“技术中立”“算法自主”为由规避责任,导致“责任真空”。

2.2. 机会公平的理论内涵

机会公平是分配正义的核心原则,指个体应基于自身努力与能力而非先赋因素(如出身)获得发展机会。罗尔斯区分了“形式的机会公平”(法律面前人人平等)与“实质的机会公平”(公平的机会平等),强调后者要求补偿社会偶然因素对个体发展前景的不利影响[3]。

在电商平台人力资源管理中,机会公平包含三个层面:准入公平——劳动者应有平等机会进入平台、获得工作资格,不因算法筛选而被系统性排斥;分配公平——任务、收入、发展机会应在劳动者之间公平分配,不因算法偏见而向特定群体倾斜;评价公平——绩效评估应基于真实工作表现而非偏见性数据,评价结果应可被质疑与纠正。然而,算法偏见正在系统性地侵蚀这三个层面的机会公平,形成“数字不平等”的新形态。

2.3. 算法偏见与机会公平的关联机制

算法偏见侵蚀机会公平的机制可概括为“输入-处理-输出-反馈”的闭环:输入层偏见表现为历史数据携带的社会不平等被编码进算法系统;处理层偏见源于模型设计中的特征选择、权重分配、优化目标隐含价值判断;输出层偏见体现为算法决策产生歧视性分配结果;反馈层偏见则是算法输出重塑社会行为与数据分布,形成“自我实现的预言”。这一闭环使算法偏见具有自我强化、自我扩散的特性,对

传统反歧视机制构成严峻挑战。

3. 电商平台 HR 管理中算法偏见的四维来源

3.1. 历史数据偏见：过去不平等的数字化延续

历史数据偏见是算法偏见的最深层来源。机器学习模型从历史数据中学习模式，而历史数据往往反映了社会既有的不平等结构。当电商平台使用历史数据进行算法训练时，过去的歧视被编码进未来决策[4]。

在招聘筛选环节，若训练数据来自特定群体主导的技术团队招聘历史，算法将习得该群体与“高绩效”的虚假关联，对其他群体候选人系统性打低分。亚马逊 2018 年曝光的 AI 招聘工具歧视事件正是典型案例——该工具对包含特定群体标识的简历进行降级处理，因其训练于公司特定群体主导的历史招聘数据[5]。

在任务分配环节，历史数据中的地理偏见被算法继承。若历史数据显示某低收入社区订单取消率较高(可能源于配送基础设施不足而非劳动者能力)，算法将习得“该社区 = 高风险”的关联，减少对该区域劳动者的派单，形成“数字红线”。历史数据偏见的危险在于其隐蔽性与“客观性”伪装。算法决策看似基于“数据驱动”的客观分析，实则复制了历史的主观歧视。

3.2. 特征选择偏见：相关性与因果性的混淆

特征选择偏见源于模型输入变量的选择不当。算法开发者基于主观判断选择“相关”特征，但这些特征可能与受保护属性高度相关，成为歧视的代理变量。

在招聘算法中，“大学背景”“邮政编码”“社交媒体活动”等特征看似中性，实则与社会经济地位高度相关。使用这些特征相当于间接使用社会阶层进行筛选，构成“代理歧视”。

在网约车派单算法中，“接单响应时间”被作为效率指标。然而，该指标忽视了不同群体面临的基础设施差异——部分劳动者可能因安全顾虑或无障碍设施缺乏而响应较慢。算法将结构性不平等转化为个体“能力缺陷”，加剧机会不公平。

3.3. 模型设计偏见：优化目标的价值负载

模型设计偏见源于算法优化目标的价值负载。算法并非价值中立的技术工具，其设计目标隐含开发者的价值优先序。电商平台 HR 算法普遍以“效率最大化”“利润最优化”为核心目标，将劳动者视为“人力资源”而非“发展主体”。这种工具理性导向使算法倾向于选择“成本最低”“产出最高”的劳动者，忽视公平、包容、可持续等社会价值[6]。

“算法工资歧视”是模型设计偏见的典型表现。平台基于劳动者的地理位置、工作时段、收入目标、历史接受率等数据，预测其“最低可接受工资”，实施个性化定价。相同工作对不同劳动者支付不同报酬，形成“统计性歧视”。这种歧视并非基于偏见，而是基于“理性计算”的利润最大化，但其结果同样侵蚀机会公平。

3.4. 反馈循环偏见：自我强化的歧视闭环

反馈循环偏见源于算法输出对社会现实的重塑。算法决策不仅影响资源分配，更改变行为者的激励与选择，进而改变数据分布，形成“自我实现的预言”。

在评分系统中，算法偏见引发“评分差距 - 收入差距 - 参与差距”的恶性循环。研究显示，部分劳动者群体面临显著的评分差距，导致收入降低；收入降低减少平台参与积极性；参与减少使算法获得更

少正向数据,进一步强化负面刻板印象[7]。算法偏见不仅反映不平等,更生产不平等。

在招聘推荐中,算法基于历史点击率优化推荐结果。若特定群体更常点击某类岗位推荐,算法将更多向该群体推送此类岗位,其他群体更少看到相关机会,形成“信息茧房”与“机会隔离”。这种“过滤气泡”效应使算法成为社会分层的加速器。反馈循环偏见使算法歧视具有“路径依赖”特性——初始的微小偏见经循环放大,最终产生显著的结构性的不平等。

4. 算法偏见在电商平台 HR 管理中的四重表现

4.1. 招聘筛选:从“人岗匹配”到“偏见匹配”

电商平台劳动者准入环节普遍采用算法筛选,从简历解析、背景审查到技能评估,算法偏见渗透全流程。简历筛选偏见表现为自然语言处理(NLP)算法解析简历时,可能因训练数据的偏见而对特定群体候选人系统性降权。面部识别偏见表现为部分平台采用 AI 视频面试与面部识别进行身份验证与情绪评估,研究表明面部识别算法对部分群体的错误率显著高于其他群体[8]。背景审查偏见表现为算法基于社交媒体、信用记录、司法数据进行“全面背景审查”,这些数据本身携带社会经济地位偏见。

4.2. 任务分配:从“智能派单”到“歧视性调度”

任务分配是电商平台 HR 管理的核心环节,算法偏见在此表现为地理歧视、时间歧视与身份歧视。地理歧视表现为派单算法基于“效率最大化”原则,优先向订单密集、基础设施完善区域派单,基础设施薄弱区域因订单密度低、配送成本高而被算法“遗忘”。时间歧视表现为算法根据历史数据预测不同时段的“劳动者价值”,在高峰时段优先派单给“历史高产出”劳动者,形成“马太效应”。身份歧视表现为研究表明部分平台算法存在隐性群体偏好,特定劳动者因时间限制而接单时间碎片化,算法将其标记为“低活跃度”,减少派单优先级。

4.3. 绩效评估:从“数据驱动”到“偏见驱动”

算法绩效评估基于客户评分、接单率、取消率、响应时间等数据,这些数据本身携带社会偏见。客户评分偏见表现为客户评价受社会经济地位刻板印象影响,研究显示相同服务质量下,部分劳动者群体获得低评分的概率显著高于其他群体[9]。评价指标偏见表现为算法选择的评价指标本身隐含价值判断,“接单率”惩罚因正当理由拒单的劳动者,“响应时间”惩罚因安全顾虑谨慎响应的劳动者。评价权力不对称表现为客户拥有评价权,劳动者仅有被评价权,差评可立即影响收入,好评仅缓慢积累信誉。

4.4. 薪酬定价:从“按劳分配”到“算法歧视”

算法薪酬定价通过动态定价、个性化定价、游戏化激励实现差异化报酬,侵蚀分配公平。动态定价歧视表现为 surge pricing (高峰定价)算法根据实时供需调整服务价格,劳动者收入高度不稳定,研究表明算法倾向于在劳动者“最需要收入时”压低单价,利用其经济脆弱性实施“算法工资歧视”[10]。个性化定价歧视表现为平台基于劳动者的收入目标、历史接受率、替代选择等数据,预测其“保留价格”,实施差异化定价,劳动者被剥夺了“同工同酬”的基本权利。游戏化激励偏见表现为平台通过“冲单奖励”“等级勋章”等游戏化设计诱导过度劳动,然而奖励门槛往往基于全职劳动者的能力设定,部分劳动者因时间限制或身体限制难以达成,被排除在奖励体系之外。

5. 算法偏见侵蚀机会公平的三重机制

5.1. “黑箱”不透明性:认知不对称与权力失衡

算法“黑箱”是不透明性的核心来源。深度学习模型的复杂性使决策逻辑难以被人类理解,劳动者

无法知晓算法如何评估自己、为何获得特定结果。这种不透明性构成“认知不对称”——平台掌握算法知识，劳动者仅能看到输出结果。认知不对称强化了权力失衡，劳动者无法预测算法行为，只能被动适应；无法质疑算法决策，只能接受结果；无法改进自身处境，只能自我归因。这种“算法无力感”侵蚀了劳动者的主体性与能动性。不透明性还阻碍了问责与救济，当歧视发生时，劳动者难以举证算法如何运作、偏见如何产生，面临“举证责任倒置”的困境。

5.2. 计算不对称性：数据权力与算法主导

计算不对称性指平台与劳动者在数据收集、处理能力上的巨大差距。平台拥有海量数据、强大算力、专业团队，能够进行复杂的预测分析与行为干预；劳动者仅拥有个人经验与有限信息，难以与算法系统抗衡。数据权力使平台能够实施“全景监控”与“精准控制”，GPS 定位、行程追踪、实时评价构成对劳动者的全面数据化，其工作过程、行为模式、甚至情绪状态都被转化为可计算的数据点。算法主导表现为平台对“合理性”定义权的垄断，算法将特定价值(效率、利润)编码为“最优解”，将其他价值(公平、尊严)边缘化为“非理性”。

5.3. 制度性逃避：劳动关系重构与责任规避

电商平台通过“去劳动关系化”策略规避雇主责任，使算法歧视难以被传统劳动法救济[11]。个体工商户注册表现为平台要求劳动者注册为“个体工商户”或“独立承包商”，将雇佣关系重构为“合作关系”，这种“制度创新”使劳动者丧失最低工资、社会保险、反歧视保护等法定权益。算法中介化表现为平台声称算法是“中立的技术工具”，决策由算法自主做出而非平台意志，以此规避歧视责任。跨国监管套利表现为平台的跨地域性使其能够利用不同国家的监管差异，在劳动保护较弱地区实施更严格的算法控制，形成“逐底竞争”。

6. 机会公平的伦理治理框架

在提出治理对策之前，有必要客观分析平台算法管理的商业逻辑与技术限制。平台面临市场竞争与盈利压力，算法优化以效率最大化为目标是理性选择；算法模型基于历史数据训练，数据偏见难以完全清除。伦理治理应在承认其合理性的基础上，通过制度设计引导其向更加公平的方向演进。

有效的算法伦理治理需要明确不同利益相关方的角色与责任：平台应承担算法公平的首要责任，劳动者应提升算法素养并通过集体行动参与治理，政府应完善算法治理法规，学术界应开展独立算法审计研究。

6.1. 数据治理：从源头消除偏见

数据审计与清洗表现为建立训练数据的偏见审计机制，识别并纠正历史数据中的歧视性模式，采用“公平性约束”的数据采样技术，确保训练数据的群体代表性。合成数据补充表现为在真实数据不足时，采用合成数据生成技术补充代表性不足的群体数据，改善模型训练的平衡性。数据最小化原则表现为遵循“数据最小化”与“目的限制”原则，仅收集与 HR 决策直接相关的数据，避免过度监控与隐私侵犯[12]。

具体实施路径包括：建立训练数据来源追溯机制，采用对抗性去偏技术引入公平性约束；定期审查地理特征与派单结果的关联性，识别并纠正“数字红线”效应。

6.2. 算法审计：过程透明与偏见检测

算法影响评估(AIA)表现为在算法部署前进行系统性影响评估，识别潜在的歧视风险与社会后果，评

估应涵盖不同群体、不同情境、不同时间段的公平性表现。持续性偏见监测表现为建立算法运行中的公平性监测机制,追踪不同群体的决策结果差异,采用“人口统计均等”、“机会均等”等公平性指标量化评估。第三方独立审计表现为引入独立第三方机构进行算法审计,验证公平性声称的真实性,审计结果应向社会公开,接受公众监督[13]。

具体实施路径包括:算法部署前委托第三方进行影响评估,运行中建立实时公平性监测仪表盘;采用差异隐私技术保护个体隐私,每年至少进行一次全面算法审计。

6.3. 透明解释:打破黑箱与赋能劳动者

算法可解释性表现为采用可解释 AI 技术(XAI),如 LIME、SHAP 等,使算法决策过程可被人类理解,对劳动者提供“为什么获得此结果”的解释。透明度报告表现为平台定期发布算法透明度报告,披露算法的基本逻辑、关键参数、公平性表现,报告应涵盖不同群体的机会分配数据。劳动者数据权利表现为保障劳动者的数据访问权、更正权、删除权与可携带权,劳动者有权获取平台持有的个人数据,有权纠正错误数据[14]。

具体实施路径包括:对派单异议提供简明解释,每年发布算法透明度报告;采用 SHAP 值解释模型预测贡献度,将技术解释转化为劳动者可理解的说明。

6.4. 人机协同:人类监督与价值把关

人类在环表现为在关键 HR 决策中保留人类最终决策权,算法仅提供决策支持而非自主决策,人类审查算法推荐的边缘案例、争议案例,纠正明显偏见。多元伦理委员会表现为建立包含技术专家、伦理学者、劳动者代表、社会公众的多元伦理委员会,审查算法设计的价值取向,监督算法运行的公平表现。申诉与救济机制表现为建立算法决策的申诉渠道,劳动者可对歧视性结果提出质疑,要求人工复核,设立独立仲裁机构,处理算法歧视争议[15]。

具体实施路径包括:算法初筛后保留边际案例进行人工复核,违规行为须经人工审核确认;采用“双人复核”机制,建立审核决策的可追溯机制。

6.5. 制度保障:法律规制与政策引导

算法透明度立法表现为制定专门的算法透明度法规,要求平台披露核心算法的逻辑与参数,保障劳动者知情权,参考欧盟《数字服务法》¹《人工智能法》²的透明度要求。反算法歧视法表现为将算法歧视纳入反就业歧视法的规制范围,明确平台对算法歧视的法律责任,建立“算法歧视”的举证责任倒置规则。集体议价权保障表现为保障劳动者组织工会、进行集体议价的基本权利,通过集体行动与平台协商算法规则。数字税与算法红利分享表现为对平台征收数字税,将部分税收用于劳动者技能培训、社会保障与数字素养教育[16]。

具体实施路径包括:制定“算法透明度与公平法”,开展算法治理试点;采取渐进式监管策略,建立监管沙盒机制,设立算法治理专项基金。

7. 研究结论与展望

7.1. 研究结论

本文基于技术伦理学与组织公平理论,系统分析了电商平台人力资源管理中的算法偏见与机会公平

¹<https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>

²<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

问题。研究发现：第一，算法偏见主要源于历史数据偏见、特征选择偏见、模型设计偏见与反馈循环偏见四个维度，形成“输入-处理-输出-反馈”的歧视闭环。第二，算法偏见在招聘筛选、任务分配、绩效评估、薪酬定价等环节呈现出歧视性分配效应，系统性地侵蚀劳动者的准入公平、分配公平与评价公平。第三，算法偏见通过“黑箱”不透明性、计算不对称性与制度性逃避等机制侵蚀机会公平，形成“数字不平等”的新形态。第四，“算法向善”的伦理治理框架从技术、组织、制度三个层面构建机会公平的保障机制。

7.2. 理论贡献与实践启示

本研究的理论贡献在于：构建了“算法偏见-机会公平”的分析框架，揭示算法歧视在平台 HR 管理中的生成机制与表现形式；提出算法侵蚀机会公平的“黑箱-不对称-逃避”三维机制；设计“算法向善”的伦理治理路径，为负责任的人工智能应用提供规范性指引。

实践启示在于：对平台企业，应将算法公平纳入核心治理目标，建立算法伦理审查机制，从“效率优先”转向“公平兼容”；对政府监管部门，应加快算法透明度立法，建立算法审计制度，保障劳动者算法公平权益；对劳动者群体，应提升算法素养与组织能力，通过集体行动争取算法治理参与权；对社会公众，应关注算法歧视问题，推动形成“算法向善”的社会共识。

7.3. 研究局限与未来展望

本研究主要采用理论分析，未来可通过案例研究、深度访谈、算法审计等方法，empirically 验证算法偏见的具体表现与治理效果。此外，不同平台模式(自营型 vs.聚合型)、不同行业(网约车 vs.外卖 vs.直播)、不同文化情境下的算法公平实践可能存在显著差异，值得进一步比较研究。未来还可探索生成式 AI、元宇宙等新技术对 HR 公平性的挑战，构建数字时代机会公平的动态理论。

在数字经济高质量发展背景下，“算法向善”不应停留在口号层面，而应转化为具体的制度安排与实践行动。让算法真正成为促进机会公平的工具，而非复制不平等的技术，是平台经济可持续发展的必然要求，也是数字文明的基本伦理底线。

参考文献

- [1] 黄再胜. 算法控制、“自我剥削”与数字劳动的时空修复——数字资本主义劳动过程的 LPT 研究[J]. 教学与研究, 2022(11): 97-112.
- [2] 孙萍. “算法逻辑”下的数字劳动：一项对平台经济下外卖送餐员的研究[J]. 思想战线, 2019, 45(6): 50-57.
- [3] 陈龙. “数字控制”下的劳动秩序——外卖骑手的劳动控制研究[J]. 社会学研究, 2020(6): 113-135.
- [4] 刘善仕, 裴嘉良, 葛淳棉, 等. 在线劳动平台算法管理：理论探索与研究展望[J]. 管理世界, 2022, 38(2): 225-239.
- [5] 吴清军, 李贞. 分享经济下的劳动控制与工作自主性——关于网约车司机工作的混合研究[J]. 社会学研究, 2018(4): 137-162.
- [6] 李胜蓝, 江立华. 新型劳动时间控制与虚假自由——外卖骑手的劳动过程研究[J]. 社会学研究, 2020(6): 91-112.
- [7] Ribeiro, M.T., Singh, S. and Guestrin, C. (2016) Why Should I Trust You? Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 1135-1144. <https://doi.org/10.1145/2939672.2939778>
- [8] Siau, K. and Wang, W. (2018) Building Trust in Artificial Intelligence, Machine Learning, and Robotics. *Cutter Business Technology Journal*, **31**, 47-53.
- [9] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A. (2021) A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, **54**, 1-35. <https://doi.org/10.1145/3457607>
- [10] Li, P., Ahmed, F. and Ruggieri, S. (2023) Fairness and Bias in Algorithmic Hiring: A Multidisciplinary Survey. *ACM Computing Surveys*, **56**, 1-36.

- [11] 张恩典. 反算法歧视: 理论反思与制度建构[J]. 华中科技大学学报(社会科学版), 2020, 34(5): 60-71.
- [12] 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.
- [13] 丁晓东. 论算法的法律规制[J]. 中国社会科学, 2020(12): 138-159.
- [14] 王莹. 算法侵害类型化研究与法律应对——以《个人信息保护法》为基点的算法规制扩展构想[J]. 法制与社会发展, 2021, 27(6): 133-153.
- [15] 汪庆华. 人工智能的法律规制路径: 一个框架性讨论[J]. 现代法学, 2019, 41(2): 54-63.
- [16] 冯向楠, 詹婧. 人工智能时代互联网平台劳动过程研究——以平台外卖骑手为例[J]. 社会发展研究, 2019, 6(3): 61-83.