E-Commerce Letters 电子商务评论, 2025, 14(12), 5754-5762
Published Online December 2025 in Hans. https://www.hanspub.org/journal/ecl
https://doi.org/10.12677/ecl.2025.14124546

Hans汉斯

电子商务平台AI推荐算法歧视的生成机制与治理路径

张浩然，周丽萍

上海理工大学管理学院，上海

收稿日期：2025年11月28日；录用日期：2025年12月11日；发布日期：2025年12月31日

摘　要

在数字经济快速发展的背景下，电子商务平台AI推荐算法已经成为电商平台提升用户体验和转化率的关键因素。然而，基于大规模数据驱动的AI推荐算法在提升效率的同时，也暴露出一系列算法歧视所导致的不公平问题，如差异化定价、推荐结果对特定群体不利等。本文以电商平台AI推荐算法为分析主体，从数据、模型、平台与用户四个维度分析算法歧视的生成机制，揭示其结构性成因与自我强化逻辑。在此基础上，从数据处理、技术设计、评估体系搭建与用户参与等维度提出一套多层次的治理路径，为电商平台实现“效率－公平”双重目标，构建公平、负责的电商智能推荐体系提供支撑与参考。

关键词

电子商务平台，AI推荐算法，算法歧视，推荐公平性

The Generative Mechanisms and Governance Approaches of Algorithmic Discrimination in AI Recommendations on E-Commerce Platforms

Haoran Zhang, Liping Zhou

Business School, University of Shanghai for Science and Technology, Shanghai

Received: November 28, 2025; accepted: December 11, 2025; published: December 31, 2025

Abstract

In the context of the rapid development of the digital economy, AI recommendation algorithms on

文章引用: 张浩然, 周丽萍. 电子商务平台 AI 推荐算法歧视的生成机制与治理路径[J]. 电子商务评论, 2025, 14(12): 5754-5762. DOI: 10.12677/ecl.2025.14124546

e-commerce platforms have become a key driver for improving user experience and conversion rates. However, while these data-driven algorithms enhance efficiency, they also expose a series of fairness issues arising from algorithmic discrimination—such as differential pricing and systematically unfavorable recommendation outcomes for certain groups. Focusing on AI recommendation algorithms used by e-commerce platforms, this paper analyzes the generative mechanisms of algorithmic discrimination from four dimensions: data, models, platforms, and users, revealing its structural causes and self-reinforcing logic. Building on this analysis, the paper proposes a multi-level governance framework covering data processing, technical design, evaluation system development, and user participation. This framework aims to support e-commerce platforms in achieving the dual goals of efficiency and fairness, and in constructing a fair and responsible intelligent recommendation system.

Keywords

E-Commerce Platforms, AI Recommendation Algorithms, Algorithmic Discrimination, Recommendation Fairness

1. 引言

近年来，随着数字经济的蓬勃发展和人工智能技术的快速进步，电子商务平台已从早期承担用户间交易功能的中间商，逐步演变为集流量分配、价格形成与用户体验优化于一体的综合性数字生态平台。在这一转型过程中，以个性化推荐为代表的人工智能技术[1]，被广泛部署于用户浏览、搜索、下单及售后等全流程环节[2]，AI 推荐算法的运用显著提升了商品与用户需求之间的匹配精度，优化了平台整体资源配置效率[3]。

然而，AI 推荐算法在提升电商平台效益的同时，也催生出了新型且隐蔽的算法歧视现象。基于 AI 的个性化推荐与差异化定价机制，在用户的使用中暴露出诸如“大数据杀熟”[4]等问题。同一商品或服务针对不同用户呈现出不同价格，不同性别、年龄或地域的用户被划分为不同群体施加溢价，以及老用户与高忠诚度客户反而支付更高的商品价格等现象，频繁引发司法争议、舆论质疑与学术反思。在此背景下，“算法歧视”已逐渐成为理解平台算法治理困境与数字市场秩序失衡的关键因素。

现有研究对算法歧视的关注多集中于法规制定与责任界定等规范层面，而对歧视现象在技术系统内部的具体生成逻辑，尤其是从数据获取、模型设计到平台运营及用户交互的整体动态机制，缺乏系统性的整合分析。为弥补这一研究缺口，本文从 AI 推荐算法的技术特性与电商平台的商业运行逻辑出发，构建了一个“数据层－模型层－平台层－用户层”的四维分析框架，旨在揭示算法歧视在电子商务环境中的生成与强化原因。在此基础上，本文进一步从技术优化与管理协同两个维度，提出具有操作性的治理路径，为电商平台 AI 推荐算法治理提供理论参考。

2. AI 推荐算法与算法歧视界定

2.1. AI 推荐算法基本原理与技术特性

AI 推荐算法的技术核心在于利用神经网络模型预测用户偏好，进而从海量商品库中实现个性化的筛选与排序。目前，主流的 AI 推荐算法模型主要基于协同过滤、深度学习与强化学习三大技术路径[5]。

协同过滤算法建立在“相似用户偏好相似物品”的基础假设上，通过分析历史用户–物品交互数据以挖掘用户的潜在兴趣[6]。该方法能够有效利用群体行为模式实现个性化商品推荐，但其推荐效果高度依赖于历史数据的完整性与可靠性。基于深度学习的模型多采用融合了知识图谱、图卷积与注意力等机制的复杂 AI 模型[7]，以捕捉用户与商品之间复杂的非线性关系。深度学习模型显著地提升了预测精度与特征表征能力，并且能够自动从多维数据中深层地学习用户特点、偏好。强化学习则将推荐系统建模为一个与环境(平台用户)持续交互的智能体[8]，其决策通常基于马尔可夫决策过程，以优化长期累积收益为目标。该框架使系统能够统筹即时反馈与长期价值，有助于实现可持续的用户体验优化。

从系统架构上看，大型推荐系统多采用“召回–排序–重排”的多层架构[9]。召回层负责从海量商品中快速初筛出以千或百为量级的商品候选集。排序层运用深度学习模型对候选商品进行精准打分。重排层则进一步结合平台规则与用户体验等因素进行最终调整，最终将商品推荐给用户。这一分层处理机制在保障系统效率的同时，也兼顾了推荐的准确性与多样性。

2.2. AI 推荐算法在电商平台中的角色

AI 推荐算法的核心任务是从海量商品与内容中筛选出最可能被点击或购买的部分，并以一定优先级呈现给用户，其应用可以很好地满足平台用户需求与资源分配。随着深度学习与强化学习等技术的发展，结构更为庞大、复杂的 AI 模型被各大电商平台采用，以捕捉用户与商品之间的关联，提高用户的消费概率。

在此结构下，AI 推荐算法不仅决定了“用户能看到什么”，还会影响“用户看到什么价格”。其目标函数往往被设计为最大化点击率、转化率或总商品交易额(GMV)，这使得推荐算法天然倾向于强化商业收益，而对公平性、可解释性等公共价值关注不足。

2.3. 算法歧视与公平性

在电子商务领域，“算法歧视”通常指 AI 推荐算法在自动化决策过程中，因数据偏差、模型设计或平台价值取向等因素，导致特定用户群体在价格或服务质量上遭受不公正待遇的现象。与传统商业歧视相比，AI 推荐算法歧视具有更强的隐蔽性、动态性和扩展性，具体表现为：(1) 决策过程被封装在“算法黑箱”之中，用户难以理解定价与推荐逻辑[10]；(2) 算法可以基于实时数据不断更新参数，使歧视结果持续调整，甚至产生新的歧视；(3) 相同推荐算法逻辑可以被快速复制到多个场景，扩大影响范围。“大数据杀熟”则是算法歧视在价格维度的典型表现，即平台基于用户画像，将忠诚度高、支付意愿强的老客户识别出来，向其收取更高的价格或减少优惠。在电商环境中，这种差别待遇往往与个性化推荐、优惠券包配置相叠加，使得“杀熟”行为更加隐蔽而难以察觉。

综合以上，本文所讨论电商推荐算法的“公平性”，在技术层面主要指平台在分配曝光机会、价格优惠和服务质量时，应避免因性别、年龄、地域等身份属性及其它强相关特征，导致系统性、持续性的不利结果。在规范层面，则体现为我国现行法律对电商推荐场景的相关要求，包括《电子商务法》禁止平台滥用技术手段损害消费者权益、《个人信息保护法》与《数据安全法》确立数据处理活动的“合法、正当、必要、公平”原则，以及《互联网信息服务算法推荐管理规定》中明确不得利用算法在交易条件上实施不合理差别待遇等。本文后续对算法歧视生成机制与治理路径的分析，均基于上述有关技术与规范公平性的定义展开。

2.4. 算法公平性研究评述

从现有研究上看，“算法公平性”已经成为算法治理领域的重要议题。有学者从评价指标出发，对不同类型的公平性进行了系统划分，代表性指标包括个体公平、群体公平、机会均等和人口统计平等，

文章通过约束分类或排序结果，降低了算法对特定 群体的不利影响[11]。这一研究路径在技术层面为评估和优化算法公平性提供了解决方法，但以上研究多以信贷、医疗、司法等场景为对象，对电商平台推荐与差异化定价的关注相对有限。

而另一些学者则聚焦于算法歧视带来的社会风险，强调平台责任、消费者权益保护以及政府监管制度的完善[3] [4] [10]。近年来，随着《电子商务法》《个人信息保护法》《数据安全法》以及《互联网信息服务算法推荐管理规定》等法律法规的出台，算法公平性逐渐被纳入数字经济治理的制度体系之中。但现有文献在技术机制和平台运营之间仍存在一定断裂，对算法歧视在“数据–模型–平台–用户”多维互动中的生成与强化过程缺乏系统性解释。

在此基础上，本文的主要贡献如下：基于“从数据和模型内部识别偏差与约束决策”的思路，将公平性问题与电商平台 AI 推荐系统的技术架构相结合，并将电商平台的商业逐利逻辑、内部治理结构以及用户行为博弈纳入分析，构建“数据层–模型层–平台层–用户层”的四维生成机制框架。本文在技术性算法研究与规范性规制研究之间搭建起了一个解释模型，并据此提出兼顾效率与公平的电商平台 AI 推荐算法治理路径。

3. 电子商务平台算法歧视的生成机制

在分析算法歧视的生成机制之前，有必要对电商平台类型及其主要参与群体做出基本区分。当前我国电商生态中，既包括以平台自营或品牌商为核心的 B2C 平台，也包括以中小商家为主、商品与服务高度多样化的 C2C 平台，以及以直播带货、内容种草为代表的社交化电商平台。不同平台在流量分配规则、商家准入门槛和用户画像精细化程度上存在显著差异。

从参与主体上看，用户群体可大致划分为高消费、高活跃用户与低消费、低活跃用户，对价格高度敏感的用户与对品牌、服务更为敏感的用户等。商家群体则可分为具有较强品牌影响力和营销能力的头部商家，以及依赖平台自然流量、影响力有限的中小商家。算法歧视在不同平台类型与不同用户/商家群体中呈现出明显差异，这些差异不仅与数据和模型有关，也与平台的商业和治理策略密切相关。下文将从数据层、模型层、平台层和用户交互层四个维度展开分析，并在各层中结合算法歧视在不同情境下的差异化表现予以讨论。

3.1. 数据层：偏差数据的累积与放大

在数据层面，电商平台中的算法歧视往往起源于多重数据偏差的叠加，以深度学习神经网络为主体的 AI 推荐算法模型高度依赖历史数据进行训练和迭代。在电商场景中，用户行为数据的采集本身具有不均衡的特点，高活跃、高消费用户留下的数据远多于低活跃、低消费用户，这使得模型在训练过程中容易将“注意力”聚焦于头部群体。其次，从数据采集环节上看，平台通过 Cookie、实名制个人信息认证与多平台关联登录等方式广泛收集用户特征数据，在用户缺乏充分知情与选择的情况下，形成高度细化的用户画像。因此，即便模型并未显式地设置“敏感参数”，不同参数在神经网络高维空间的组合也可能会导致意想不到的算法歧视的出现，例如通过设备类型、浏览习惯、地域等数据标签可能使模型隐式地学习并固化对特定群体不利的推荐模式。

同时，数据标签的设置可能本身带有人为偏见。例如，平台可能将某些地区、设备或性别的用户视为“高风险”或“高支付意愿”群体，这些标签被加以更大的权重后纳入用户数据持续进入训练集，导致 AI 推荐算法在后续决策中不断重复甚至放大现有的算法歧视。

在此基础上，数据与用户行为之间还存在着持续运转的循环。推荐结果不断影响用户的浏览与购买行为，而这些新的行为又被写回数据库作为后续训练数据，一旦系统在某一阶段对某类用户或商品减少

曝光，后续该用户或商品的数据将进一步减少，模型对其“理解”程度也会随之下降，最终形成“愈加边缘化”的循环效益。因此，在以 C2C 为主的交易平台上，头部商家通过促销、广告投放等方式获得更多交互数据，更容易在训练数据集中占据主导地位，从而挤压中小商家的曝光空间。而在以自营或大品牌为主的 B2C 平台中，高消费、高活跃用户行为数据则更容易主导模型更新，使得低消费用户在商品推荐上处于长期不利地位。

3.2. 模型层：以收益最大化为导向的目标函数设计

在模型层面，算法歧视主要源于目标函数的设定与模型学习的偏向。在电商平台应用中，大量 AI 推荐模型以点击率、转化率、GMV 等为核心优化目标，而缺少对公平性相关指标的考虑，模型因而优先服务于对收益贡献更大的用户与商品，而对低价值用户或小商家的需求关注不足，逐渐形成一种以经济利益为导向的算法歧视。在这一过程中，如果模型没有剔除敏感属性或进行校正，那么黑箱模型中的参数很可能与性别、年龄或地域等因素关联，从而对弱势群体造成不利影响。例如，在以 GMV 为核心考核的 B2C 平台中，模型往往优先推荐客单价高、复购率高的商品，并倾向将更多资源分配给高活跃、高消费的高价值用户。在以流量变现和广告收入为主要盈利模式的平台中，则更容易出现对投放预算高的头部商家给予额外曝光权重的情况。前者可能导致低消费用户的商品推荐范围缩小、优惠力度减弱，后者则会使中小商家即便具备较高商品性价比，也长期得不到与其商品质量相匹配的展示机会。

在此基础上，个性化定价与差别优惠策略进一步加深了这种歧视，这在模型层具体表现为模型使用与支付意愿高度相关的特征(例如用户停留时长、历史购买金额、设备类型等)进行价格预测，并在此基础上对被判定为“愿意支付更高价格”的用户给出更少优惠或更高价格。

与此同时，模型的框架结构设计也会对推荐算法造成影响。即便输入数据相同，不同模型结构及其更新算法的选择都可能导致对不同群体的推荐存在差异，从而在相同特征空间下输出截然不同的推荐结果，进一步加剧模型层面的隐性偏差。

3.3. 平台层：商业逐利逻辑与内部制度缺失

从电商平台角度看，推荐算法的设计往往围绕收益增长、客单价提升和用户留存率等商业指标展开。运营团队倾向采用能够快速提升数据的策略，这类策略最终会被转化为模型的推荐规则，成为算法系统的内部偏好。在不同类型平台中，这种以商业指标为导向的运营逻辑具有不同表现形态。对于依赖品牌溢价和会员体系的 B2C 平台，算法歧视往往体现为老会员在不知情的情况下支付更高的价格或享受更少的隐性优惠。而对于 C2C 或直播电商平台，算法歧视则更多体现为流量高度集中于头部主播或大商家，使得中小商家及新入驻商家难以获得足够的冷启动流量。

另一方面，公平性问题具有跨职能的特性，往往需要运营、算法、法务等多部门协同，但在许多平台内部，沟通与协作机制并不完善。在责任分工上，算法开发者往往强调“技术中立”，而运营方则以“AI 算法自动生成”为由弱化自身责任，从而使平台难以及时感知和量化某些特定用户群体在体验上的负面感受与不公平待遇。再加上实际运营中普遍依赖 A/B 测试来评估新算法模型，如果实验评价以点击率、转化率、GMV 等指标为优化目标，就自然会放大对高价值用户的偏向，而忽视用户长期信任、满意度的变化，从而在看似“数据驱动”的持续迭代中，将平台商品推荐偏好逐渐固化为系统性的算法歧视。

3.4. 用户交互层：认知、适应与逆向迭代

在用户交互层面，算法歧视并非单向地施加给用户，而是在用户认知与行为的适应中不断博弈、不断迭代的。一方面，用户在长期使用过程中会形成对平台算法的主观判断，并据此调整自己的行为，例如刻意多浏览促销信息、频繁使用比价功能等，这些行为被系统记录后又被转化为“价格敏感”、“高支

付意愿”等标签，反向强化了模型对其身份与偏好的既有推断。另一方面，部分用户在察觉可能存在差异化对待后，会通过更换账号、设备或收货信息等方式，规避不利的推荐与定价，但这种行为在增加数据噪声和行为异质性的同时，也迫使算法更加依赖使用更为复杂的数据组合做出推断，这可能导致引入新的偏差，形成新的算法歧视。而且，若用户在使用平台的过程中持续感到自己被区别对待，即便无法理解算法机理，这种不公平感也会逐步沉淀为对平台的不信任与负面情绪，最终通过评价、留存率和口碑等指标体现出来，对平台的长期用户信任度和品牌价值造成实质性损害。

3.5. 四维框架的动态传导和反馈机制

从动态上看，本文所构建的“数据–模型–平台–用户”四个层级之间并非线性单向关系，而是通过一系列反馈回路彼此联动并相互强化。具体而言，平台在收益最大化、流量变现等商业逻辑驱动下形成的运营策略与考核指标，首先通过特征选择、样本构造和目标函数设定等方式固化到模型层之中。模型输出的推荐结果在排序和定价上对高价值用户、头部商家以及高利润商品给予更多曝光和优惠机会，进而影响用户的浏览、点击和购买行为，以及商家的投放和经营策略。这些经过算法“筛选”后的新行为数据又作为训练样本回流至数据层，使原本就不均衡的数据分布进一步向优势群体倾斜，弱势用户和中小商家的交互记录则日益稀疏，在数据与模型层面逐渐“边缘化”。与此同时，用户和商家的适应性行为与策略性反应(如更换个人信息以规避不利推荐与定价、弱势群体的被动迁移等)，一方面通过改变点击、停留、搜索和下单路径，在数据层生成带有噪声与偏差的交互记录，结合部分弱势群体的离去，既进一步集中优势群体的有效样本，又使弱势群体的真实偏好被噪声化和碎片化，形成新的歧视。另一方面用户与商家的适应性行为又会反馈到平台对运营规则、流量分配机制的调整之中，在平台层进一步巩固既有的偏好结构。

在此基础上，如果缺乏外部规范约束与内部公平性治理，这一“数据–模型–平台–用户”的闭环迭代过程，往往会将最初看似微弱的差别逐步放大为结构性、制度化的算法歧视，使低消费用户、中小商家等弱势群体在长期平台使用中处于越来越不利的地位。反之，一旦在数据采集与标注、模型训练与评估、平台制度设计以及用户反馈与申诉机制等关键环节嵌入公平性约束与纠偏机制，则原本放大不平等的传导链条可以在一定程度上被改造为抑制偏差扩散的传导链条。通过多轮迭代，四维框架有望从复制和加剧现有不公平结果，逐步转向识别、修复并缓解已有的不公平结果。

4. 针对 AI 算法歧视在技术与管理视角下的治理思路

4.1. 完善数据处理机制

在数据层面，为了缓解电商平台算法歧视，首先可以从数据处理入手，通过对数据收集与特征构建体系的改进，获得相对公平的模型训练数据。在数据收集阶段，可以对低活跃、低消费用户以及中小商家的样本进行适度多次采样或加权，避免整体数据分布被高活跃、高消费用户和头部商家占据，从而使模型在学习时不仅仅围绕头部群体优化。同时，平台有必要对收集数据进行整体的偏差分析，即便这些变量不直接包含敏感参数，不同参数在高维空间的组合也可能会导致意想不到的算法歧视出现，对此，需要检验现有数据结构是否会引发显著的群体差异，并对识别出的有偏特征进行重构或降低权重处理。例如，在构建训练样本时，可以对不同地区、不同消费层级用户进行分层采样，确保各类群体在训练数据中占有基本比例。在特征标注阶段，对与敏感属性高度相关的特征进行相关性分析，减少模型在高维空间中间接学习到对特定群体不利的决策。在此过程中，还应特别关注人为设置数据标签的环节，例如运营团队在定义“高价值用户”等标签时，容易将主观经验甚至刻板印象直接固化到数据中，平台需要通过制定统一的标注规范等方式，对这类人工标签加强过程监管和效果监督，防止歧视在算法中被放大

和长久固化。

4.2. 以公平性为约束的算法优化

在算法层面，首先需要在优化目标设置与模型训练阶段，将以点击率、转化率、GMV 为优化目标的模型，联立与公平性相关的约束，例如不同群体间曝光差异等公平性指标，使推荐算法从一开始就在不损害公平的前提下进行学习与调参。具体而言，可以将不同用户群体之间的曝光差异、优惠力度差异等指标设置为约束条件，将其嵌入到模型的损失函数或优化目标之中。一方面，通过设置公平性相关的“硬约束”，限定某些不公平结果的最大容忍边界。另一方面，也可以在目标函数中加入“软约束”，以一定权重惩罚导致群体或中小商家不公平待遇的输出结果，使模型在追求收益的同时兼顾公平性目标。

在具体方法实现上，可以在现有强化学习训练框架的基础上，引入安全强化学习(safe reinforcement learning)的思路，将公平性约束设定为训练过程中的 cost 输出，通过 CPO (Constrained Policy Optimization) [12]等安全强化学习算法，在不违背约束的前提下优化策略性能。此时，推荐策略不再只是单纯最大化长期收益，而是在“reward-cost”空间内寻找满足公平性约束的最优解。此外，还可以在优化目标内增加与公平性相关的项，对长期处于不公平状态的用户或商家适度提高公平项的权重，并在推荐多样性与覆盖率上设定必要下限，避免用户群体受到差别对待、少数头部商家长期垄断流量。

在具体电商推荐场景中，可以将上述公平性约束与平台现有的“召回－排序－重排”架构相结合。在召回阶段，对长期曝光不足的用户群体或中小商家的商品设置最低召回比例，以避免其被过早过滤。在排序阶段，将群体曝光差异、优惠力度差异等指标纳入损失函数，对导致不公平结果的排序输出施加惩罚。在重排阶段，则可以结合平台业务规则，对处于长期不利地位的用户或商家设置一定的“公平性加权系数”，在不显著牺牲整体点击率与转化率的前提下，提高其被推荐到前列位置的概率。通过上述方式，公平性约束不再停留在原则层面，而是被嵌入到推荐系统的各个关键决策环节中，成为可实现、可度量、可优化的闭环。

4.3. 构建系统化的公平性评价体系

在算法歧视治理中，仅在数据处理与模型训练阶段考虑公平性还不够，还需要一套适用于电商场景的公平性评价体系以评估推荐算法的公平性，使“公平”能够被显式地评价和管理。具体而言，可以围绕用户与商家两个维度进行设计。从用户角度出发，重点关注不同消费层级、不同地区、不同设备类型用户之间平均优惠力度，以及高质量、高性价比商品的曝光机会是否存在差异。从商家角度出发，则可以比较中小商家与头部商家在商品品类、质量相近的条件下，其曝光量、点击量和成交量等指标是否长期失衡。上述评价体系不应停留在一次性研究或专项排查中，而应被纳入平台的常规数据报表，与点击率、转化率、GMV 等业务指标一并进行定期监测和考核，从制度上避免平台在决策与迭代中陷入“只看增长、不看公平”的单一导向。

4.4. 加强算法可解释性设计与用户参与

在加强算法可解释性设计与用户参与方面，平台可以在不涉及商业机密的前提下，为用户提供简要的商品推荐和定价理由提示，例如通过“根据你的浏览记录”、“根据你的收藏偏好”等说明提示用户，还可以为用户在购买商品时提供同类商品最近成交价格选项，减轻用户对“被推荐算法暗箱操作”的主观感受。而且，应当赋予用户一定的控制权，使其可以主动调整算法推荐偏好(如少看某类广告、多看某类商品)，并快捷地标记“价格不合理”、“推荐不相关”等不良体验，将这类反馈纳入模型迭代和策略优化流程，形成用户与算法之间的良性互动机制。在此基础上，平台还可以通过问卷调查、访谈等方式持续开展用户体验调研，定期了解不同群体对平台公平性和透明度的主观感受，并将这些信息与算法层

面的量化指标结合起来，构建“定性 + 定量”的联合评估框架，为后续的算法调整和产品设计提供更全面的依据。

4.5. 治理路径合规性与可行性分析

电商平台 AI 推荐算法歧视的治理并非孤立进行，而是嵌入在我国现有数字经济治理的法治框架之中。从法律层面看，《电子商务法》明确要求平台经营者保障交易相对人的合法权益，不得利用技术手段对消费者实施不合理的差别待遇。《个人信息保护法》《数据安全法》则分别从个人信息处理合法性、数据分级分类保护等角度，对平台的数据收集与利用行为提出了更为严格的规范要求。《互联网信息服务算法推荐管理规定》进一步针对算法推荐服务，提出不得利用算法实施不正当竞争、不得通过算法对用户进行不合理差别对待等要求。

在这一背景下，本文提出的治理路径在合规性与可行性方面具有较强的法律支撑。一方面，通过完善数据处理机制和开展数据偏差分析，有助于平台落实“最小必要”、“公开透明”等个人信息处理原则，降低在数据收集与用户画像构建环节违反《个人信息保护法》的概率。另一方面，在算法优化中联立公平性约束、构建系统化的公平性评价体系，有利于平台证明其已尽到义务，为后续可能出现的争议提供技术与规章上的证据支撑。同时，加强算法可解释性设计与用户参与，不仅有助于回应《算法推荐管理规定》中关于“保障用户知情权和选择权”的要求，也为监管部门、社会公众开展外部监督提供了基础。

5. 结语

本文系统地探讨了电子商务平台中 AI 推荐算法歧视的生成机制与治理路径。算法歧视并非单一技术故障的产物，而是一个基于“数据–模型–平台–用户”四维架构所导致的现象。在数据层，数据的偏差与循环不断固化既有的歧视行为。在模型层，以收益最大化为单一导向的优化目标设计，牺牲了 AI 推荐算法的公平性与多样性。在平台层，商业的逐利逻辑与跨部门协同的缺失，为 AI 算法歧视的生成提供了条件。在用户交互层，用户应对算法歧视的针对性行为，与推荐算法系统形成了动态博弈，导致了 AI 算法歧视的不断迭代，并在长期上损害了平台的用户信任度和品牌价值。

基于这一生成机制，本文进一步从技术优化与管理协同两个维度，提出了针对 AI 推荐算法的治理路径。在技术层面，通过数据去偏、为算法联立公平性约束以及加强可解释性设计，从系统内部修正歧视的产生。在管理层面，则需要构建涵盖用户与商家双维度的公平性评价体系，并将其纳入平台常规监测与考核流程，同时加强用户反馈，形成推荐算法与用户之间的良性互动。

然而，电子商务平台 AI 推荐算法歧视的治理仍面临诸多挑战，技术上的修复存在内在限度。电商平台推荐算法的问题，最终指向平台商业逻辑与社会价值。未来研究的挑战在于，如何超越平台自身的利益边界，构建一个涵盖政府规制、行业标准、公众监督与学术研究的多元共治生态，从而在电子商务平台的效率与公平之间，寻找到一个可持续的动态平衡点。

参考文献

[1] 李晓明, 王磊. 人工智能在电子商务中的应用与发展趋势[J]. 电子商务, 2022, 18(3): 45-53.

[2] 杨茴杰. 人工智能驱动电子商务创新挑战与完善对策[J]. 电子商务评论, 2025, 14(11): 1812-1816.

[3] 魏天天. 人工智能技术赋能电商的方式与算法歧视问题研究[J]. 电子商务评论, 2025, 14(11): 1839-1845.

[4] 上海市普陀区市场监管局课题组. 网络交易平台“大数据杀熟”的治理问题探究[J]. 中国市场监管研究, 2023(10): 24-29.

[5] Chen, X., Yao, L., McAuley, J., Zhou, G. and Wang, X. (2023) Deep Reinforcement Learning in Recommender Systems:

A Survey and New Perspectives. *Knowledge-Based Systems*, **264**, Article ID: 110335. https://doi.org/10.1016/j.knosys.2023.110335

[6] 钟宴宏. 基于复购行为模式的时序推荐方法研究[D]: [硕士学位论文]. 广州: 广东工业大学, 2025.

[7] Batmaz, Z., Yurekli, A., Bilge, A. and Kaleli, C. (2018) A Review on Deep Learning for Recommender Systems: Challenges and Remedies. *Artificial Intelligence Review*, **52**, 1-37. https://doi.org/10.1007/s10462-018-9654-y

[8] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., *et al.* (2015) Human-Level Control through Deep Reinforcement Learning. *Nature*, **518**, 529-533. https://doi.org/10.1038/nature14236

[9] 文亮. 推荐系统技术原理与实践[M]. 北京: 人民邮电出版社, 2023: 244.

[10] 徐汉明, 孙逸啸. 算法媒体的权力、异化风险与规制框架[J]. 西安交通大学学报(社会科学版), 2020, 40(6): 128-136.

[11] 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

[12] Zhang, Q., Leng, S., Ma, X., Liu, Q., Wang, X., Liang, B., *et al.* (2025) Cvar-Constrained Policy Optimization for Safe Reinforcement Learning. *IEEE Transactions on Neural Networks and Learning Systems*, **36**, 830-841. https://doi.org/10.1109/tnnls.2023.3331304