算法偏见对金融普惠性的抑制效应研究

刘明辉

(武汉纺织大学 管理学院 湖北 武汉 430077)

摘要 随着人工智能和大数据技术在金融领域的广泛应用，算法已成为推动金融普惠的重要驱动力。本文以算法偏见对金融普惠性的抑制效应为主题，探讨了在金融服务中算法偏见产生的原因及其对公平性和社会包容性的影响。研究通过构建概念框架，分析算法在信贷评分、风险评估及客户分类中的应用场景，结合实例揭示了数据偏倚、设计偏见和反馈放大等因素如何导致算法偏见，进而加剧对特定群体的金融排斥。本文采用实证分析与定性研究相结合的方法，从社会经济数据集入手，验证不同算法模型在群体间的表现差异，量化了算法偏见对低收入群体及边缘化人群金融服务获取的限制程度。研究表明，算法偏见在金融领域的存在不仅恶化了资源分配的不平等，而且削弱了金融普惠性目标的实现效果。最后，本文提出通过优化算法设计、改进数据质量、引入公平性评价指标以及建立监管体系等途径来缓解算法偏见的负面影响，推动金融普惠的可持续发展。研究对于算法偏见问题的揭示以及解决路径的探讨，为促进金融服务的公平与普惠提供了理论支持与实践参考。

关键词 算法偏见；金融普惠性；低收入群体；资源分配不平等；公平性评价指标

文章编号 057-2025-0501

Research on the Inhibitory Effect of Algorithmic Bias on Financial Inclusion

Liu Minghui

(School of Management, Wuhan Textile University, Wuhan 430077, China)

Abstract With the widespread application of artificial intelligence and big data technologies in the financial sector, algorithms have emerged as a significant driving force for promoting financial inclusion. This paper focuses on the inhibitory effect of algorithmic bias on financial inclusion, exploring the causes of algorithmic bias in financial services and its impact on fairness and social inclusivity. By constructing a conceptual framework, the study analyzes the application scenarios of algorithms in credit scoring, risk assessment, and customer segmentation. Through practical examples, it reveals how factors such as data bias, design bias, and feedback amplification lead to algorithmic bias, which in turn exacerbates financial exclusion of specific groups. Combining empirical analysis and qualitative research, the paper starts with socio-economic datasets to verify the performance differences of different algorithm models across groups. It quantifies the extent to which algorithmic bias restricts access to financial services for low-income groups and marginalized populations. The research findings indicate that the presence of algorithmic bias in the financial field not only worsens inequality

收稿日期：2025-01-09 录用日期：2025-02-13

通讯作者：刘明辉；单位：武汉纺织大学 管理学院 湖北 武汉

in resource allocation but also undermines the achievement of financial inclusion goals. Finally, the paper proposes several approaches to mitigate the negative impacts of algorithmic bias, including optimizing algorithm design, improving data quality, introducing fairness evaluation metrics, and establishing a regulatory framework, thereby promoting the sustainable development of financial inclusion. The study's revelation of the algorithmic bias problem and exploration of solutions provide theoretical support and practical references for enhancing the fairness and inclusivity of financial services.

Keywords Algorithmic bias; Financial inclusion; Low-income groups; Inequality in resource allocation; Fairness evaluation metrics

人工智能和大数据技术的快速发展和广泛应用，正在对各种行业产生深远影响，金融领域尤为明显。应用这些技术的算法模型已成为金融服务，尤其是普惠金融的重要工具。然而，算法模型在客观客观地提供金融服务方面的便利性和高效性并不总是顺风顺水的，问题和挑战常常并存。算法偏见就是其中之一，这不仅可能导致服务质量下降，而且可能在更大范围内影响到社会经济的公平性和包容性，特别是对于弱势群体、低收入群体和边缘化群体。因此，深入理解算法偏见对金融普惠性的影响是至关重要的。在这个背景下，本文的研究主题是“算法偏见对金融普惠性的抑制效应”。本文将通过实证和定性研究，从多种角度深入探索算法偏见在金融服务领域的现象、影响和解决策略，期待能对调节金融普惠性的制约因素提供有益的理论参照和实践启示，从而对金融服务的公平与普惠发展产生积极的推动作用。

1 算法与金融普惠性的理论关联

1.1 人工智能与大数据在金融领域的推动作用

人工智能与大数据技术在金融领域已经展现出了极大的推动作用^[1]。从自动化交易到个性化金融服务，这些技术显著提升了金融机构的效率和精准度。通过机器学习和深度学习算法，大量的金融数据得以快速处理和分析，从中识别出隐藏的模式和趋势。这不仅提高了金融产品的风险管理能力，还优化了资源配置，

降低了金融服务的成本^[2]。金融科技公司利用大数据和人工智能开展新型服务，如在线信贷、智能投顾等，极大地提升了金融服务的普惠性，使得更多的个人和小微企业能够获得及时和有效的金融支持。

1.2 金融普惠性的内涵及其重要意义

金融普惠性指的是通过提供易获取、成本低廉且安全的金融产品和服务，使全社会各阶层，尤其是传统金融服务难以覆盖的低收入群体及边缘化人群，能够平等享受金融资源和便利^[3]。金融普惠性对于促进社会经济发展、提升居民生活质量具有重要意义，不仅能够减少贫困、缓解经济不平等，还能增强经济体系的稳定性和韧性。金融普惠性能够激发创业活力、提高劳动生产率，进而推动整体经济的可持续增长，有助于实现更为广泛的社会包容和公平^[4]。

1.3 算法技术如何影响金融普惠性发展

算法技术通过数据处理、模式识别和预测分析，为金融普惠的发展提供了重要支撑。人工智能和机器学习算法在信贷评分、风险评估和客户分类中显著提升了效率与覆盖范围，使得金融机构能够触及传统上处于服务边缘的群体。算法的运作以数据为基础，其对信息的依赖可能加剧已有的不平等。不合理的特征选择或模型设计可能导致某些群体被系统性排除，从而影响金融普惠性的公平性目标。算法技术对金融普惠的影响包含机遇与挑战，其关键在于设计与应用的合理性。

2 算法偏见的来源及表现

2.1 数据偏倚的根源与传递机制

数据偏倚是算法偏见的重要来源，其根源主要体现在数据的收集、选择及处理过程中存在的不平衡与片面性。在金融领域，数据偏倚通常表现为样本的代表性不足、特定群体数据的缺失以及数据标注中的人为主观性。数据偏倚往往在算法训练阶段被传递并放大，导致算法在应用中对特定群体的预测结果产生显著误差。信贷评分系统中，若训练数据偏向于高收入群体或特定区域，算法会对低收入及边缘化群体的信用评估产生偏颇^[5]。数据历史中的结构性不平等因素也可能被算法继承并进一步强化，形成负向反馈循环。数据偏倚的传递机制不仅影响算法决策的准确性，也对金融服务的公平性和普惠性构成深远挑战。研究数据偏倚的根源与传递机制对于缓解算法偏见具有重要理论与实践价值。

2.2 算法设计中隐含的公平性问题

算法设计中隐藏的公平性问题主要源于在模型开发过程中对特定群体特征的忽视和刻板印象的固化^[6]。设计者的主观偏见、历史数据的不平衡以及缺少对多样性和包容性的考量，往往导致生成的算法在决策过程中倾向于强化已有的社会不平等现象。这些隐含问题不仅难以被察觉，而且难以通过常规的算法优化方法加以纠正，从而在无形中加剧了对边缘化群体的排斥。

2.3 反馈循环对边缘化群体的加剧排斥

算法偏见通过反馈循环机制进一步加剧了对边缘化群体的金融排斥。当算法基于历史数据作出决策时，数据中的偏倚会导致边缘化群体被持续性地低估或忽视，形成循环性歧视。此过程使得这些群体在信贷评分和风险评估中处于劣势，难以获得公平的金融服务^[7]。这种

反馈机制强化了算法的既有偏见，造成资源配置更加不平等，并减弱了金融普惠性所追求的社会包容目标。破解这一循环对于提升金融服务的公平性和普惠性至关重要。

3 算法偏见对金融普惠的负面效应

3.1 信贷评分与风险评估中的群体不平等

信贷评分与风险评估是金融服务中应用人工智能与算法的核心环节，但算法偏见在此过程中可能引发群体不平等。潜在的算法偏见来源于历史数据的偏倚、模型设计中的固有主观性及持续反馈的影响，导致信用评分和风险评估结果对特定群体不公平^[8]。例如，训练数据若包含对某些性别、种族或收入水平群体的歧视性信息，将使这些群体在得到信贷机会时遭遇不公平待遇。进一步，算法偏见会放大现有社会不平等，抑制金融普惠性的发展，影响社会资源的公正分配。

3.2 客户分类中的社会包容性挑战

在客户分类领域，算法偏见对社会包容性的挑战主要体现在两个方面^[9]。基于历史数据的机器学习模型可能会继承并放大已有的社会偏见，导致特定群体在金融服务获取中的劣势地位被进一步固化^[10]。例如，低收入群体和少数族裔容易在信用评估和贷款发放中面临系统性歧视。算法在分类过程中可能会忽视边缘化群体的特殊需求，导致其难以获得个性化和公平的金融服务。这种偏见的累积效应加剧了金融资源分配的不平等，对社会的整体包容性构成了严重威胁。

3.3 对低收入及边缘化群体的金融服务限制

低收入及边缘化群体在金融服务获取中面临显著限制，主要表现为信贷审批和融资机会的显著不足。算法偏见通过数据偏倚将这些群体进行负面分类，并在信用评分与风险评估中

夸大其信用风险，从而使其更难获得贷款及其他金融服务。这种算法偏见进一步削弱了金融包容性，导致资源分配的不平等。这些限制不仅限制了这些群体的金融发展机会，还加剧了社会经济的不平等，严重阻碍了全面实现金融普惠目标的进程。

4 缓解算法偏见的路径与策略

4.1 数据处理与质量优化的关键举措

数据处理与质量优化是缓解算法偏见的重要前提^[11]。优化数据质量需从源头着手，确保数据的代表性与完整性，避免特定群体在样本中的低比例或缺失^[12]。应注重多样性数据的采集与整合，减少由于数据单一性导致的偏倚^[13]。数据预处理阶段应剔除隐含歧视特性的变量，并通过标准化、平衡化措施降低可能存在的偏差。在数据标注过程中，需强化准确性与客观性管理，杜绝因人为偏见引发的潜在问题。引入去偏算法，通过重采样、权重调整或生成对抗网络等技术手段，改善数据分布的公平性。动态监测数据质量，及时识别和解决可能的新型偏倚风险，可持续地保障数据在算法中的可靠性与中立性。这些优化举措为减少算法偏见奠定了坚实的基础，并在推动金融普惠性过程中发挥关键作用。

4.2 算法公平性评价指标的引入与应用

算法公平性评价指标的引入与应用是缓解算法偏见的关键步骤。通过构建具体的公平性评价指标，能够对算法进行全面评估，检测其不同群体间的表现差异。这些指标包括对算法输出结果的公平度测量、对不同群体的影响评估以及对潜在偏见的识别和校正。这些公平性评价指标在算法设计和实施阶段的应用，可以有效降低因数据和设计偏见导致的不公平效应，保障金融服务的社会包容和普惠性目标。

4.3 跨部门协作与监管体系的建立

建立跨部门协作与监管体系要求金融机构、技术公司、监管部门等多方共同参与，建立透明的数据和算法审核机制，确保开放、沟通畅通，及时发现与纠正偏差，有效控制算法偏见对金融普惠的负面影响。

5 推动金融普惠性的可持续发展框架

5.1 技术创新对普惠目标的正向推动

技术创新在推动金融普惠性目标实现中起着关键作用。人工智能和大数据技术的发展，使得金融服务在覆盖面和精准度上得到显著提升。智能化工具如机器学习算法的应用，不仅提升了风险评估的准确性，还能通过动态数据更新及时调整信贷决策，降低服务成本，提高服务效率。金融科技的创新还广泛应用于移动支付、数字钱包等领域，解决了传统金融服务难以覆盖的“一公里”问题，使更多人群，尤其是身处偏远地区或缺乏正规金融服务的人群，能够便捷地获得金融服务。

物联网和区块链技术为金融服务带来了新的透明性和安全性，增强了用户的信任感，促进了金融服务的可获得性和可持续性。智能合约通过自动执行预设条件，减少了中间环节，降低了操作风险和成本，有助于普惠金融服务的推广。通过技术创新，金融机构能更好地识别和服务不同群体的需求，提升普惠金融服务的公平性和包容性，有效减少算法偏见带来的负面影响，推动金融普惠性的可持续发展。

5.2 公平性与效率之间的平衡机制探索

在推动金融普惠性的可持续发展过程中，公平性与效率之间的平衡机制是一项关键性议题。金融领域中的算法技术需要在提升服务效率的同时避免对特定群体的歧视和资源的不平等分配。为实现这一目标，需通过优化算法设计，

增强其对公平性目标的敏感性,确保风险评估、信贷审批等环节对不同群体一视同仁。全民金融包容性指标的建立可有效评估普惠目标的实现程度,并为技术改进提供方向。鼓励金融机构引入多元化的数据采集方式,提升传统模式中的不足,将边缘化人口纳入金融体系范围。通过构建协同治理模式,使技术革新与制度监管并行,从而在公平性与效率之间寻求动态平衡,达成普惠金融的长期发展目标。

6 结束语

本文全面探讨了人工智能在金融领域中算法偏见问题及其抑制金融普惠性的影响,特别是对低收入和边缘化群体的不利影响。通过实证分析和定性研究,揭示了数据偏倚、设计偏见和反馈放大等因素如何导致特定群体受到金融服务的排斥。研究表明,算法偏见削弱了金融服务的公平性,加剧了资源分配不平等。尽管提出了优化算法设计、提高数据质量、引入公平性评估指标及建立健全监管体系等多项策略来缓解这些问题,但算法的复杂性和技术的持续更新带来了持续的监管挑战;同时,如何在创新与监管间寻求平衡,确保技术进步时不牺牲社会公平,也是重要的未来研究和政策议题。此外,国际合作和全球性行动也是促进金融普惠中不可或缺的部分。未来研究需进一步探索有效策略,通过持续监测和政策调整,期望实现更公平、更普惠的金融服务体系,为所有社会成员包括边缘群体提供平等的发展机会。

参考文献

- [1] 张呈磊,郭忠金,李文秀.数字普惠金融的创业效应与收入不平等:数字鸿沟还是数字红利?[J].南方经济,2021,(05):110-126.
- [2] 杨城强.数字普惠金融特性对收入不平等状况改善的影响机制——以“数字普惠金融使用意愿”为中介变量[J].武汉金融,2022,(03):78-84.
- [3] 邓金,钱张娜.数字普惠金融缓解城乡收入不平等了吗[J].农业技术经济,2022,(06):77-93.
- [4] 陈文,刘汉成.数字普惠金融对金融行为的影响研究——基于收入不平等视角[J].财会通讯,2022,(06):83-87.
- [5] 王蓉.普惠金融调节收入分配之功能研究[J].南方金融,2023,(03):66-79.
- [6] 庞楷朱韩浩哲.数字普惠金融能缓解消费不平等吗?[J].兰州财经大学学报,2023,39(04):58-69.
- [7] 叶琴,袁歌骋,张呈磊.数字普惠金融与收入机会不平等[J].当代经济科学,2023,45(03):114-126.
- [8] 张莹,安菲,潘欣然.数字普惠金融对低收入群体的影响[J].合作经济与科技,2022,(20):176-178.
- [9] 赵予焐张世林.普惠金融视角下城乡低收入群体金融服务的“双度”测度研究[J].市场周刊,2021,34(06):139-142.
- [10] 王志杰.算法歧视下的普惠金融监管——基于算法歧视与普惠金融风险的耦合性[J].福建金融,2020,(06):21-27.
- [11] 郭雪,雷雨箫,董继华.普惠金融对城乡收入差距抑制效应的检验[J].统计与决策,2020,(05):142-144.
- [12] 郝亮,刘久彪.数字普惠金融对收入不平等的影响——基于数字鸿沟的中介效应检验[J].北京经济管理职业学院学报,2022,37(03):71-80.
- [13] 常向东,尹迎港.数字普惠金融对城乡收入不平等的影响研究[J].郑州航空工业管理学院学报,2022,40(05):23-34.