

· 社会救助 ·

社会救助自动化决策的运行机理、 算法偏差及治理路径

匡亚林

【摘要】 随着人工智能的发展，自动化决策系统将在对象识别、数据对比和算法推荐方面革新传统救助机制，预示着未来社会救助的重要发展方向。本研究尝试厘清社会救助自动化决策系统的技术架构、决策过程与人机交互模式运行机理，并揭示其在提升公共服务效率和公正性方面的巨大潜力。研究深入探讨了社会救助自动化决策中算法偏差的识别与分析，并指出主要存在的三种偏差：数据集构建偏差、算法设计和运行偏差、算法透明度和可解释性偏差。这些偏差对行政决策公正性有显著的潜在影响。为应对这些挑战，本研究提出促进社会救助自动化决策高质量发展的综合治理措施，包括数据集的多样化、算法的公平性评估和模型的优化开发、决策机制的透明性和可解释性、人类干预的强化和动态监测机制的建立、法律法规的完善。

【关键词】 社会救助；自动化决策；算法偏差；治理路径；透明度

一、问题的提出

公共行政机构的行政决策自动化是数字政府改革的一个重要要素，在全球公共管理领域使用全自动或半自动的行政决策正在不断地增加和扩张中；^① 中国也同样积极地引领和探索数字化转型道路，以期实现中国式现代化与特色社会保障新制度文明。^② 当前中央政府的“简政放权”改革正如火如荼地开展中，除了在府际层面的向下授权外，部分“简政放权”工作也在“数字中国”的发展战略背景下借助技术治理手段得以实现。人工智能（Artificial Intelligence，简称 AI）作为一种模拟人类智能的技术，涵盖了包括机器学习在内的多个子领域，而自动化决策（Automated Decision Making，简称 ADM）作为 AI 技术应用的一个分支，主要致力于利用算法实现决策过程的自动化。由自动化决策衍生出的自动化偏差（Automation Bias）问题，即对自

【作者简介】 匡亚林，电子科技大学公共管理学院副教授。主要研究方向：数字治理与社会救助。

【基金项目】 贵州省哲学社会科学规划项目“贵州推进分层分类社会救助体系建设路径研究”（24GZYB55）。

① Ulrik B. U. Roehl, "Automated Decision-Making and Good Administration: Views from Inside the Government Machinery," *Government Information Quarterly*, 2023(4).

② 郑功成：《中国式现代化与社会保障新制度文明》，《社会保障评论》2023 年第 1 期。

自动化系统的过度依赖而对社会或政府造成的负面影响,^①其中算法偏差(Algorithm Bias)就是自动化偏差中最常见的一种。^②全球范围内的社会服务自动化改革揭示了技术治理的双刃剑特性,即它既有可能带来革命性的进步,也可能引发一系列问题和风险,美国印第安纳州的社会救助自动化改革中,实施的自动化系统(Automatic Control System,简称ACS)便充分印证了高科技工具如何在社会福利、法律执行等领域中加剧对贫困群体的歧视与合理性不平等(Rational Discrimination),自动化决策系统在设计上未能有效拆解结构性不平等的情况下,通过其快速和大规模的应用而强化了这些不平等。从既往国外的经验来看,人工智能算法的确存在较大的争议,一方面,支持者认为这些算法在许多领域带来了革命性的进步,瑞典和芬兰等国家纷纷在《行政法典》(Administrativlagboken)和《行政程序法》(Hallintomenettelylaki)中支持在公共管理过程中使用ADM,包括欧盟的《通用数据保护条例》和《人工智能法案》均提到自动化决策的法律适用约束条件;人工智能通过机器学习和深度学习技术能够处理大量数据并广泛将其运用在税收、社会福利、医疗诊断、自动驾驶和金融分析等领域,以敏捷算法模式做出更准确的预测和决策,提高生产效率,^③减少人为错误,推动各行各业创新性发展;^④然而,另一方面,批评者指出人工智能算法也存在透明性和可解释性不足等潜在风险,导致人们难以理解其决策过程而加剧社会不平等,例如,奥地利公共就业服务(Austrian Public Employment Service,简称AMS)算法的案例揭示了自动化偏差在公共管理中的影响,该算法通过基于求职者特征的分组优化资源分配,然而,其评分机制对女性、年长者、有健康问题者以及非欧盟求职者存在不公正现象,从而暴露了自动化决策过程中潜在的歧视问题。由于算法运行依赖海量数据,倘若数据本身包含偏差和歧视将导致算法决策对某些群体不公平。^⑤社会各界对人工智能广泛应用可能引发的失业问题广泛关注,普遍担忧众多传统工作岗位将面临被自动化技术取代的风险,尽管人工智能算法在技术层面取得了令人瞩目的进步,社会仍需以谨慎的态度对待其应用,并确保其发展遵循伦理和公平原则,如何在确保技术革新的同时,充分考量弱势群体的特殊需求,^⑥避免加剧社会分层,成为当务之急。在此背景下,强化对ADM应用系统的监管力度,推动建立行业标准,保障民众的知情权和参与权。未来救助体系的构建,不仅需要技术层面的创新,更应在政策制定、伦理审视及法律规制等多维度进行综合考量和平衡,这种平衡的实现有赖于政府、企业、学术界及社会各界的共同努力。^⑦我国正探索多元化的解决方案,如建立第三方评估机制,对ADM系统的公平性进行定期审核,并公开审核结果,以提高社会

① Hannah Ruschmeier, Lukas J. Hondrich, "Automation Bias in Public Administration—An Interdisciplinary Perspective from Law and Psychology," *Government Information Quarterly*, 2024, (3).

② Anja Lambrecht, Tucker Catherine, "Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads," *Management Science*, 2019, (7).

③ 邢泽宇:《数字赋能人社发展提升便民服务质效——人社信息化建设迈向新台阶》,《中国劳动保障报》,2023年8月16日。

④ Markku Suksi, "Administrative Due Process When Using Automated Decision-making in Public Administration: Some Notes from a Finnish Perspective," *Artificial Intelligence and Law*, 2020, (4).

⑤ Hannah Ruschmeier, Lukas J. Hondrich, "Automation Bias in Public Administration—An Interdisciplinary Perspective from Law and Psychology," *Government Information Quarterly*, 2024, (3).

⑥ 匡亚林等:《数字赋能社会救助高质量发展的路径选择》,《中国行政管理》2024年第7期。

⑦ Geert Bouckaert, et al., "Potential for Comparative Public Opinion Research in Public Administration," *International Review of Administrative Sciences*, 2005, (2).

信任度。通过搭建合作平台,提升数据质量,^①逐步消除算法中的偏差,提升社会救助的精准度。确保算法决策过程中的每一步都符合伦理道德和法律规定,让技术在服务社会、造福人民的同时,^②也能兼顾公平与正义。

随着信息技术的持续发展,我国社会救助体系正逐步向信息化、智能化和自动化方向演进,一些地区已经开始尝试和实施社会救助信息化的深入应用,实施以远程申请、^③远程授权、^④需求评估、^⑤主动识别等主要辅助救助决策手段的实践探索,^⑥以促进救助服务的智能化和自动化。持续推动社会救助信息化建设,实现信息的高效流通,减少困难群众的奔波,不仅有助于提升政府的治理能力和效率,也是提高公共服务质量的关键步骤。^⑦当前学术界与民政部门在相关领域的研究与实践主要集中在以下方面:第一,数字技术在社会救助领域的应用价值研究。数字技术,尤其是人工智能、互联网和大数据,在社会救助领域展现出巨大的应用潜力。技术治理通过远程申请和授权简化了救助申请流程,它还能模拟评估救助需求,通过数据分析精准识别需要救助的群体。算法系统(Welfare Eligibility Systems)的主动发现功能有助于识别那些可能未被纳入救助范围的低收入人群,确保资源的公平分配。数字技术的应用极大地推动了社会救助的精细化管理进程,对困难家庭信息的实时、动态管理,能够精准捕捉家庭致困的多元因素及个性化需求。智慧算法服务救助逐渐成为社会救助制度的核心诉求之一,衡量标准也由单一的收入指标转向多元化的需求评估,多维化需求评估从经济状况、教育水平、健康状况、社会关系等多个维度开展政策干预,确保帮扶措施更具针对性和差异性。第二,自动化算法在救助信息核对中的作用。自动化算法不仅限于简化和加速数据处理流程,分析大量历史数据识别出潜在的欺诈行为和异常模式。部分地区以电力大数据为基础,通过算法赋能的探索,识别潜在救助对象实现救助的主动发现机制,且基于水电气具体使用数据的混合算法也在中国的部分地区进行试点,辅助识别那些可能因为技术或知识限制而未能正确申报信息的个体,确保救助资源能够更加公平地分配给真正需要帮助的人。第三,国内外自动化社会救助的实施效果。以美国印第安纳州的福利自动化系统为例,该系统的设计初衷是为了提高福利申请和审批的效率,但由于系统设计缺陷、项目管理不善以及技术问题,数据处理错误、系统崩溃等,导致食品券资格的综合错误率从5.9%增加到19.4%,增长了超过三倍,漏保率从1.5%增加到12.2%。在自动化社会救助实践中,技术实施和项目管理中对于数据的规制、算法的设计和模型的训练值

① 林国钢:《低收入人口常态化帮扶的整体性治理——基于“两项政策”衔接并轨突破口选取的考察》,《行政论坛》2024年第4期。

② 丹尼尔·布尔:《数字化时代的德国福利国家:主要挑战与发展之道》,《社会保障评论》2019年第2期;克劳斯·彼得森、杰尼克·休乌:《数字化时代的丹麦福利国家:政策变革与战略困境》,《社会保障评论》2019年第2期。

③ 孙明扬:《数字技术与公共服务组织的双向互构——基于L县远程医疗数字技术应用的时序性过程分析》,《求索》2024年第5期。

④ 马亮:《中国数字政府建设的理论框架、研究议题与未来展望》,《中共天津市委党校学报》2021年第2期。

⑤ 杨立雄、刘曦言:《从“情景”到“脱域”:数字时代的福利身份与社会权利探讨——基于C市“救助通”的案例研究》,《学术研究》2024年第9期;匡亚林:《需求侧管理视角下社会救助体系分层分类改革研究》,《河海大学学报(哲学社会科学版)》2021年第2期。

⑥ Nuth Maryke Silalahi, "Taking Advantage of New Technologies: For and Against Crime," *Computer Law & Security Review*, 2008, 24(5).

⑦ 高志鹏等:《数字技术对农民工劳动收入和福利的影响——基于产业结构转型视角的分析》,《中国人口科学》2023年第2期。

得关注；同时，数字济贫院（Digital Poorhouse）研究中，学者关注到随着数字化管理和自动化决策系统的广泛应用，^①美国政府和相关机构越来越多地采用算法来分配社会福利、管理社会服务项目、评估个人资格等，带来的新的数字不平等问题，主要表现在数字资源的分配不均、数字技能的差异以及数字设备的普及程度等方面。^②国内实践方面，部分地区民政部门开发了“救助通”小程序或低收入监测系统平台，广东省民政部门开展了AI赋能社会救助的实践，开发了广东省低收入人口动态监测管理系统，该系统汇集了12个部门机构的18类特殊困难群众基本信息，把405万低收入人口纳入监测。系统根据各项监测指标综合赋分计算结果，采用红、橙、黄、蓝、绿5档颜色信号划分预警等级，实现了由“人找政策”到“政策找人”的转变。^③广东省还建立了底线民生信息化核对管理系统，实现全省异地通办，与银行对接，电子授权模式等创新，提升了社会救助的效率和便捷性。其他地区如内蒙古自治区兴安盟民政局社会救助信息化管理系统建设项目，实现了自助申请、在线核算、在线公示等功能。江苏省常州市智慧大救助平台构建了“五化”智慧救助体系，提高了救助的精准度。山东淄博博山区“博山e救助”微信小程序实现了一网通办，主动发现，打造了“物质+服务+e”新模式。

现有文献为社会救助自动化决策中的算法偏差与治理路径提供了丰富的理论基础和实践案例。前人的研究明确了数字治理是一把双刃剑，自动化算法简化和加速了数据处理流程，在识别潜在的欺诈行为和异常模式具有较强的预警功能，但在数字技术赋能与嵌入政府行政决策实务过程中，也植入了值得防范的新风险。尽管现有文献为社会救助自动化决策提供了丰富的理论基础和实践案例，但在算法偏差与治理路径方面亟待深入探讨，例如，社会救助自动化决策的算法偏差及影响，国内外自动化社会救助实践的具体差异是什么，算法驱动的社会救助自动化运行的内在机理是什么，在此基础上，本文旨在通过深入研究社会救助自动化决策中的算法偏差与治理路径，为推动我国社会救助体系的数智化发展提供理论支持和实践指导。

二、社会救助自动化决策的运行机理

社会救助自动化决策系统作为一种现代信息技术的应用，其核心在于通过算法驱动的自动化流程对社会救助申请进行快速、高效的审查和评估，以决定申请人是否符合救助条件。该系统在数字化救助领域的迅速发展，为社会救助工作的现代化转型提供了强有力的技术支持，其运行机理主要体现在技术架构、决策过程和人机交互模式三个核心环节。技术架构作为系统的基石，通过先进的技术手段确保数据处理的安全性、可靠性和高效性，为系统的稳定运行和高效处理提供有力保障。决策过程则通过自动化的方式，实现了对申请人资格的快速审查和精准评估，提高了决策的准确性和公正性。人机交互模式则注重申请人的参与度和互动性，通过用户界面设计和交互流程优化，提供便捷、友好的在线服务，增强了用户体验，并确保信息的透

① 参见 Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, New York, St. Martin's Press, 2018.

② 匡亚林等：《数字技术赋能社会救助：缘起、风险及治理》，《上海行政学院学报》2023年第2期。

③ 民政部：《广东依托大数据构建精准社会救助体系——发现及时 救助主动 精准高效》，民政部官网：<https://www.mca.gov.cn/n1288/n1290/n1316/c1662004999979994752/content.html>，2023年9月5日。

明度和可访问性。通过该运作机理,社会救助工作将实现由传统的人工审核向自动化决策的转变,大大提高处理效率和准确性,减少人为干预的可能性,确保救助资源的公平分配。该系统还将为政府和社会提供全面、精准的数据支持,为制定科学合理的救助政策提供有力依据,从而在公共服务领域实现数智技术赋能的重要发展趋势。

(一) 技术架构: 支撑高效数据处理与系统稳定

在全球范围内,不同地区在自动化社会救助决策的技术架构选择上呈现出多样性。北美一些地区救助系统则更多采用分布式数据汇总的方式和强大的数据处理引擎,^①以确保数据的实时性和准确性;^②欧洲福利系统则并没有像美国等一些国家采纳联邦政府和州政府分化的救助服务,更多地区倾向于采用数据集成系统平台和云计算技术,以强调数据安全性和处理效率;亚洲政府则偏好混合架构模式,^③结合云计算和本地数据中心的优势,若救助的自动化资格审查未通过,人工干预则会出现在平台拒绝救助后,进一步审查并明确原因,以确保数据的可靠性和安全性。亚洲地区在社会救助自动化决策领域的技术发展趋势尤为值得关注,尤其是我国在积极推动云计算技术在社会救助领域的应用的积极探索,利用云计算、大数据、人工智能和物联网等先进数字技术,推动了社会救助管理的现代化。例如,利用云计算技术建立了社会救助信息化平台,为政府部门和社会救助组织提供了强大的数据支持,最大程度上破解了精准识别障碍问题。通过集成多源数据,实现了对救助对象财产状况、经济条件、家庭条件等的精确核查,解决了边缘贫困群体难识别、求助对象需求难掌握等问题,促进了社会救助供需的精准对接,有助于推动救助对象精准化和救助项目的精细化实施。^④

社会救助自动化决策系统在技术架构方面,系统通常基于数字技术,涵盖机器人流程自动化(Robotic Process Automation, RPA)、确定性规则模型(Deterministic Rule-Based Models)、回归分析(Regression Analysis)、大数据分析(Big Data Analytics)、机器学习(Machine Learning, ML)和神经网络(Neural Networks)等多种技术形式。其中,RPA广泛应用于瑞典社会服务领域,特别是社会救助评估方面,能够自动执行重复性任务,如数据输入和查询。确定性规则模型则依据预设的逻辑规则处理数据并作出决策,例如根据申请人的收入水平自动决定是否批准社会救助申请。回归分析用于预测未来趋势或结果,如预测特定申请人未来可能需要的社会救助金额。大数据分析利用大量数据进行模式识别和预测以优化决策过程,通过分析历史数据识别高风险家庭,提前提供干预措施。机器学习工具利用“大数据”和统计模型补充或替代人类决策,减少决策过程中的人为偏见。神经网络作为一种复杂的机器学习方法,模仿人脑工作方式,能够处理更复杂的数据结构和关系。值得注意的是,自动化决策技术并非中立的工具,需要在嵌入特定社会和文化背景中去理解和运用,政策话语中常常忽视了技术的社会建

① 考虑到具体的政治分布特征,美国联邦部门与州部门之间的政府职能有一定的差异性,所以美国更多地采取的是分布式的数据库集成技术方式,福利系统需在多个州或地区之间共享数据,分布式数据库可以更方便地支持跨地域的数据访问和同步;同时,也便于IBM、Oracle等公司和绩效部门在第三方服务技术的支持下开展质量评估与监督。

② 庄琦、黄春柳:《价值链再造:社会救助数字化转型的国际实践》,《中国社会保障》2023年第7期。

③ 新加坡Support Go Where平台的技术采用集成式+微服务+AWS云的三层架构设计模式,既具有集成数据库的优势,又可以扩展数据功能与服务。

④ 参见罗午阳:《探索“互联网+”背景下社会救助精准化发展》,《中国社会科学报》,2020年第2050期。

构性质,将其视为解决已知问题的中立手段。实际上,技术的应用和发展受到救助需求、权力关系和利益驱动等因素的直接影响。

(二) 决策过程:从数据收集到结果反馈的自动化流程

从全球范围内看,社会救助自动化决策过程在整体上遵循了数据收集、分析处理、决策制定和结果反馈的基本流程。在欧洲地区,社会救助领域的自动化决策体系一般涵盖以下四个核心环节:数据集成、预处理、模型构建与决策生成。在数据集成阶段,该系统通过多元化渠道系统性地采集申请人的个人基本资料、经济状况等关键信息;随后,在预处理阶段,对收集到的数据进行深度清洗与标准化处理,旨在保障数据的全面性与精确性;模型构建阶段则依托先进的机器学习算法,对历史数据集进行训练与分析,以构建出具有预测能力的决策模型;在决策生成阶段系统自动形成并输出精准的社会救助决策。^①在北美地区,社会救助领域的自动化决策机制通常涵盖以下关键步骤:数据综合集成、模型效能评估、决策策略制定及结果反馈优化。数据集成阶段通过多个数据源整合信息,形成全面的数据集;模型评估阶段对不同的模型进行对比和验证,选择最优模型;决策制定阶段基于选定的模型生成救助决策;结果反馈阶段将决策结果及时告知申请人,并收集反馈信息用于后续优化。在亚洲,社会救助自动化决策系统通常包括数据收集、分析处理、决策制定和结果通知等步骤。^②数据收集阶段通过在专有系统平台、线下核对等多种渠道明确救助信息;分析处理阶段对数据进行深入分析,识别申请人的具体需求;^③决策制定阶段根据分析结果生成救助决策;结果通知阶段通过短信、邮件等方式将决策结果通知申请人,并提供后续支持。上述流程反映了各地区在技术应用、政策导向和社会需求方面的多样性,同时也展示了全球范围内技术发展和应用的广泛性。

社会救助自动化决策系统的决策过程是一个高度系统化和自动化的流程,其设计、开发和部署均旨在通过精确的数据收集、分析处理、决策制定和结果反馈来提升社会救助的效率和公正性。其一,系统通过多渠道收集申请人的基本信息,包括个人身份信息、家庭状况、教育水平、经济状况以及以往的社会救助申请记录等。这些数据来源广泛,涉及申请人自行提交的材料、政府部门数据库、金融机构、医疗机构、房产管理机构、车辆管理部门以及社会信用体系等。系统对这些信息进行初步筛选和验证,确保其真实性和准确性,同时保证数据的全面性和实时性。其二,系统运用先进的数据分析技术对收集到的数据进行深入分析,以确保决策的准确性和公正性。依据预先设定的救助标准和规则,系统对申请人的信息进行自动化评估和计算。通过机器学习算法、大数据分析、区块链和人工智能技术对申请者的各种数据进行综合评估,识别潜在的风险和需求。^④其三,系统根据多个方面的标准和规则,如收入水平、财产状况、家庭成员数量等,算法和测量模型需要定期动态更迭以适应社会救助政策的变化和新的数据分析技术的发展,根据区域范围内的收入平均水平以及其他救助评估要素条件,计算出申请人是否符合救助条件,系统生成初步评

① Nora Germundsson, Stranz Hugo, "Automating Social Assistance: Exploring the Use of Robotic Process Automation in the Swedish Personal Social Services," *International Journal of Social Welfare*, 2024, (3).

② 数字化救助平台运用自动化算法和规则对申请者资格进行预审,若预审未通过,则转入人工审查环节。审查结束后,系统将通过在线通知向申请者反馈资格结果及相应的救助项目与金额。

③ 霍萱:《以“并库”为突破口:两项政策衔接并轨的路径选择与政策集成》,《社会保障评论》2024年第5期。

④ 祝建华:《智慧救助的要素驱动、运行逻辑与实践进路》,《社会保障评论》2022年第2期;匡亚林:《论区块链技术在社会救助制度中的适用》,《行政管理改革》2021年第2期。

估报告,详细列出申请人的各项信息和评估结果,提交给相关部门或工作人员进行复核。其四,复核确认无误后,通过数字或通讯平台短信通知申请人自动化系统决策的救助申请结果。申请人通过系统提供的详细报告,了解自身是否符合救助资格及具体的救助金额,并按规定的程序领取现金或服务等形式援助。针对特殊申请个案的排异性机器学习,系统可根据实际运行数据和用户反馈进行迭代改进,以不断提升决策质量和系统性能。

(三) 人机交互模式: 增强申请者参与度和用户体验

全球范围内,自动化决策系统的人机交互模式呈现出地域性的差异,各地在界面设计、功能配置及用户体验方面各有侧重,但普遍目标在于增强申请者的自主性和互动性,提供高效、友好的在线服务体验。在北美地区,人机交互设计尤为强调申请者的主动参与和即时互动。通过集成在线平台和自助服务终端,申请者能够便捷地提交资料并实时监控处理状态。系统配套提供详尽的操作指导和视频教程,确保用户充分理解相关政策与流程。例如,美国的 Benefits.gov 平台不仅提供了全面的服务指南,还实现了应用程序访问功能,支持用户追踪申请进展,并设有即时反馈机制,保证用户提交资料后即可迅速获得确认通知。^① 欧洲地区的人机交互模式更侧重于服务的便捷性和用户体验的友好度。申请者借助在线门户或移动应用完成资料提交,同时享有即时查询进度和结果的功能。系统配备了在线客服和电话咨询服务,确保申请者在遇到难题时能快速获得援助。瑞典的社会福利系统以其多语种支持和全天候在线客服著称,用户可通过即时聊天或电话联系客服代表,获取即时帮助;而丹麦的 Borger.dk 平台同样提供了高度个性化的服务选项。^② 亚洲区域人机交互模式的设计着重于提高申请者的参与度和满意度。除了常规的在线申请和进度查询外,系统还提供了多语言版本和文化适应性界面,以满足多元背景申请者的需求,增强了服务的可及性和用户对系统的信任感、归属感。^③

在人机交互模式下,人类工作者与自动化程序系统之间存在嵌入性分工问题,更多依托技术进行智慧治理还是更多依靠人类智力治理的比重,决定了人机交互的几种协同模式,通常需要人或机按比例完成管理系统的维护和故障排除等任务。社会救助自动化决策中的人机交互模式大致可以分为三种类型。其一,完全自动化系统。自动化社会救助系统通过算法处理大量标准化申请,决策完全依靠算法和技术手段自动处理和审批标准化申请,使决策基于预设规则运行而非个人偏好的指令。系统性能高度依赖于输入数据的质量,错误或不完整的数据可能会直接导致决策失误。以美国印第安纳州为例,该州曾尝试通过自动化系统决定福利申请资格时,暴露出对特殊情况处理不足的问题,部分符合条件的申请者未能获得应有的救助,成为了自动化福利决策改革失败的典型案例。其二,混合系统。混合系统结合了自动化工具与人工审查,旨在平衡技术和人之间的关系。印第安纳州后来转向的混合系统保留了电子数据处理和私有化管理的核心功能,同时允许申请人在必要时与公共雇员面对面交流。这种方法试图在技术效率和人性化服务之间找到平衡,

① 具体可访问美国的救助 Benefits.gov 平台验证,平台内设相关服务指南并可以通过 apps 访问并跟踪救助申请进度。

② 瑞典的社会救助系统不仅支持多种语言,还提供了 24 小时在线客服,申请人可以通过聊天窗口或电话与客服人员进行实时沟通,解决申请过程中遇到的各种问题。此外,系统还设计了用户友好的界面,确保申请人能够轻松导航和操作。包括丹麦 (Borger.dk) 的救助平台也提供相应的救助服务。

③ 新加坡的社会救助系统 (Supportgowhere) 不仅支持英语和中文等多种语言,还提供了专门的文化适应性指南,帮助不同背景的申请人更好地理解救助政策和申请流程。此外,系统还设有电话热线和社区服务中心,申请人可以通过多种渠道获得帮助和支持。

但在实践中仍面临诸多挑战。瑞典的特雷勒堡市（Trelleborg）自 2017 年起将机器人流程自动化（Robotic Process Automation, RPA）引入社会救助申请管理中，RPA 在引入公共部门当年就在个案工作者的协助下处理了大约 70% 的救助申请，当年即作出了 41% 的决定并处理了实际的社会援助待遇给付。随着经验的积累，截至 2020 年，特雷勒堡市超过 85% 的社会援助申请人使用了电子申请，其中 30% 完全由 RPA 处理。^①其三，辅助决策模式。辅助决策模式侧重于利用人工智能进行数据分析，辅助工作人员作出更为精准的决策。当社会救助申请者对系统裁决持有异议时，社会救助专业人员需介入实施复审与调停工作。此过程包括详尽考察申请者家计情况并评估其异议的正当性，视情况作出人工修正。辅助决策方式是行政自动化决策探索的初始模式，也积累形成了较多有益的经验探索。我国陕西省西安市探索建立“线上预警+线下响应+资源链接+结果跟踪”的数链合一机制、江苏省江阴市探索的社会救助项目、数据、资源、监管集成式智慧救助信息平台建设，将大数据及人工智能技术运用在救助申请过程中，除了数智赋能外，他们均采纳了人工辅助决策的运行模式，进一步丰富了“大数据+铁脚板”的审查机制。然而，未来的进步仍依赖于数据驱动机制的完善和自动化救助、人工智能领域的进一步发展与公共服务领域的深入结合运用落地。三种 RPA 模式展示了自动化决策的人机交互不同应用层面，从完全依赖技术的自动化系统，到技术与人的混合决策模式，再到以人为主的决策系统，每一种模式都有其特定的应用场景和优缺点。选择合适的人机交互模式，需要综合考虑系统的效率、决策的公正性和透明度，以及对复杂决策和人际沟通的需求。

三、社会救助自动化决策的算法偏差

尽管社会救助自动化决策展现出广阔的前景，人工智能和大数据技术在社会救助申请过程中显著提升了效率和准确性，但算法偏差问题依然构成了全球数字治理的重大挑战并可能形成“创造性破坏”。^②例如，《自动化不平等》一书中提到的美国印第安纳州福利自动化系统失败案例，清晰地揭示了算法偏差可能导致对特定群体的不公平对待，这不仅暴露了系统设计缺陷、项目管理不善和技术问题对系统效能的影响，还深刻反映了算法偏差对社会公正的损害，以及由此引发的公众对人工智能技术信任度的下降。因此，深入探讨社会救助自动化决策中算法偏差的识别与分析，不仅是技术层面的必要之举，更是确保社会公平与正义的迫切需求。

（一）数据质量偏差：数据完整性和准确性对算法决策的影响

算法决策依赖于数据质量，但数据的完整性和准确性常受影响，导致数据质量偏差。社会救助自动化决策系统面临数据集构建偏差、数据标注主观性、训练数据质量问题 and 目标设定与特征选择偏差等多方面挑战。其一，数据集构建的偏差问题。在社会救助自动化决策领域，用

① Anna Ranerup, Lars Svensson, "Actors and Intentions in the Dissemination of Robotic Process Automation in Social Work," in Gertrud Juell-Skielse, et al. (eds), *Service Automation in the Public Sector. Progress in IS*, Springer, Cham, 2022.

② 杨立雄：《数字化转型与“创造性破坏”：社会保障数字治理研究》，《社会保障评论》2023 年第 5 期；任蓉：《算法嵌入政府治理的风险及其防控》，《电子政务》2021 年第 7 期；邱泽奇：《算法治理的技术迷思与行动选择》，《人民论坛·学术前沿》2022 年第 10 期。

于训练模型的数据集的代表性将直接关系到模型决策的公平性。如果数据集中样本选择不当或带有偏差,训练出的模型可能会放大这些偏差,导致对特定群体的歧视性决策。例如,数据集可能偏向于某一收入水平、地区或职业群体,这不仅违背了社会救助的基本原则——公平与正义,还可能加剧社会不平等。其二,数据标注的主观性差异。从传统救助管理出发,不同工作人员基于各自的判断标准,可能会对申请人的资格和需求做出不同的评估。这种主观性不仅会影响模型训练的效果,还会降低决策的透明度和可解释性,损害公众对社会救助系统的信任。不同的工作人员可能对同一申请人的健康、经济状况有不同的理解,导致标注结果的不一致,数据标注过程中的主观性差异是缺乏统一的评估标准引起的。其三,训练数据的质量问题。社会救助数据往往涉及个人隐私和敏感信息,例如,美国的缅因州(Maine)使用 EBT 卡^①数据来追踪临时救助贫困家庭计划受益人在烟酒商店和州外地点的 ATM 提现情况,从 2011 年 1 月到 2019 年 1 月保罗·勒佩奇任职该州州长期间,受助人的交易数据被严格审查,查验其是否存在福利滥用,这表明在一些地区低收入群体的数据隐私往往被视为监控技术的目标,引发了有关福利救助与污名化的广泛讨论。而在全球的大部分地区,社会救助的数据由于涉及隐私问题而难以保证训练数据的完整性和时效性,数据更新机制的缺失和数据清洗技术的不足都可能影响数据的整体质量。数据质量问题可能导致模型学习到错误的信息,进而影响决策结果的可靠性,类似家庭经济状况数据的获取、比对环节可能存在标准不一致的情况,^②影响了受助者救助需求的正当满足。其四,目标设定与特征选择中的潜在偏倚。分类标签与社会刻板印象的使用可能引发对特定群体的误解,反向红区划分策略可能将弱势群体固定于不利位置,加剧社会不平等现象。目标与特征要素直接关系到预测模型(Predictive Risk Models)能否有效精准识别标的救助对象,并展现出优秀的泛化能力。机器学习工具可能会根据与以往成功的申请者相似的特征识别候选人,从而复制了救助过程中的早期偏见。模型目标的设定必须清晰界定其所针对的核心议题,确保与社会救助实践需求紧密对接;且特征选择应聚焦于那些能最佳体现受助者需求与状态的关键要素,以此增强预测风险模型的精确度与决策的科学依据。若模型目标设定过窄,过分集中于经济指标而忽视健康状况、教育水平等非经济维度,可能会导致模型在特定场景下效能低下,产生偏倚。^③

(二) 算法设计与运行偏差:静态算法与动态迭代不足

算法设计中的偏差问题是自动化决策系统面临的另一大挑战。静态算法缺乏动态迭代能力,无法适应复杂多变的实际应用场景,这使得系统难以处理特殊家庭的具体情况,导致部分合理申请需求被拒绝。此外,算法设计者可能在设计阶段无意识地将个人主观倾向嵌入算法,进而影响决策结果的客观性。自动化救助算法的设计开发与执行面临多维度的挑战,包括但不限于静态算法偏差(Static Algorithmic Bias)、技术限制(Technical Limitations)以及设计缺陷(Design

① EBT 卡,全称为 Electronic Benefits Transfer 卡,是一种类似于借记卡的电子支付工具,用于向符合条件的个人或家庭发放政府提供的福利资金。这些福利通常包括食品援助(如美国的补充营养援助计划,SNAP,即原来的食品券计划)和现金援助(如临时救助贫困家庭计划,TANF)。EBT 卡使得受益人可以更方便地在授权零售处购买食品或其他必需品,而无需使用纸质券。

② 在福利自动化决策领域,有较多的案例可印证由于数据比对出错而造成错保率和漏保率上升的情况。

③ 岳平、苗越:《社会治理:人工智能时代算法偏见的问题与规制》,《上海大学学报(社会科学版)》2021 年第 6 期。

Flaws)。①首先,静态算法问题显著。在低收入对象精准识别、注意力分配及自主决策等方面,特定群体可能遭受不公正待遇(Algorithmic Cruelty)。算法训练过程中,若采用的历史救助决策数据未能全面反映救助对象的真实状况,而是侧重于历史救助经验的总结,那么基于此类静态历史结果训练的人工智能救助算法可能忽略反向集合因果机制的动态迭代过程,从而加剧已有偏差,导致部分合理申请需求遭到不公拒绝。此外,算法设计者可能在设计阶段无意识地将个人主观倾向嵌入算法,进而影响决策结果的客观性。数据预处理与特征选择环节的操作不当亦可能导致关键特征被遗漏或赋予过高权重,进一步加深偏差。算法设计与实施阶段若未能充分考虑复杂多变的实际应用场景,系统输出可能偏离预期目标,产生不公正或不准确的结果。其二,技术限制。②自动化技术系统通常依赖于预设的社会救助规则和标准,当面对特定情境时,系统的灵活性受限,难以准确识别并处理特殊家庭的具体情况,导致真正需要援助的对象被排除在外。居住在偏远地区或低收入家庭的个体可能缺乏必要的设备和网络连接,无法利用在线平台提交救助申请,从而加剧其生活困境。算法在处理数据时主要依赖历史数据和现有统计模型,面对新颖或罕见情形时,系统可能无法作出准确判断,进一步限制了技术的有效性和适用范围,例如,2006年至2008年间印第安纳州进行的一项现代化福利实验项目,部分群体由于地址更换等原因没有接收到重新认证的通知文件导致受助人的Medicaid^③福利资格失效,即使她试图纠正这一错误,最终不得不面临药物短缺的问题,该案例展示了技术工具如何在不经意间影响到了依赖这些服务的家庭和个人。其三,算法设计缺陷。系统设计阶段若未能充分考量用户群体的多样性与复杂性,部分用户的需求可能得不到有效满足。系统用户界面若不够直观易用,将影响用户体验,阻碍其有效使用系统。系统的可扩展性和可维护性不足,难以适应未来需求的变化。系统安全性设计若不完善,易遭受外部攻击,威胁用户数据安全与隐私保护。

(三) 算法透明度与可解释性偏差: 算法黑箱对公正性和可接受性的影响

算法透明度与可解释性是确保自动化决策系统公正性和可接受性的关键因素。然而,由于算法模型的高度复杂性,透明度的不足成为了算法偏差的一个重要来源。算法黑箱^④问题核心在于缺乏公开性和透明度,导致责任主体不明确、纠正机制缺失,决策逻辑模糊,影响决策的公正性和可接受性。其一,透明度缺失。由于算法模型的高度复杂性,透明度的不足阻碍了相关利益主体对系统运行机制的有效理解和监督,从而削弱了系统的可信度与稳健性,使受益对象难以对其遭遇的算法结果发起质疑或挑战。例如,当贫困家庭因误判而被不当排除在救助范

① Robyn Repko Waller, R. L. Waller, "Assembled Bias: Beyond Transparent Algorithmic Bias," *Minds and Machines*, 2022, (3); 陈禹衡、陈洪兵:《反思与完善:算法行政背景下健康码的适用风险探析》,《电子政务》2020年第8期。

② 颜克高、高淼:《利益冲突与技术限制:地方社会组织跨部门合作监管失灵的解释》,《中国行政管理》2019年第7期。

③ Medicaid福利是美国的一个联邦和州共同资助的公共援助计划,旨在为低收入和个人提供医疗保健服务。它为符合条件的个人和家庭提供包括医生访问、住院治疗、处方药、实验室服务、X光检查以及其它医疗服务在内的健康保险覆盖。Medicaid的资格标准由联邦政府设定,并允许各州根据自身情况制定更具体或更宽松的标准。该实验以高度依赖于技术工具对Medicaid等社会福利项目进行自动化申请而闻名,然而,这个自动化系统导致了许多问题,论文中案例的受助对象由于癫痫而严重依赖Medicaid福利。

④ 牛津英语大辞典给算法的定义为“一套用于计算和解决问题的程序或规则”,而所谓的“黑箱”作为一种隐喻,它指的是那些不为人知的不能打开、不能从外部直接观察其内部状态的系统。而算法黑箱就是人工智能的数据输入和数据输出之间存在着隐藏层,使他人无法获悉其真正的内涵、规则和运作机理。

围之外时,其往往既无法理解被排除的具体原因,也缺乏有效的渠道来申诉该决策。透明度的缺乏还可能导致数据隐私的泄露和滥用,侵犯受益人的个人隐私权,进而加剧公众对社会救助系统的信任危机。其二,问责机制的不完善。在自动化社会救助系统中,缺乏有效的问责机制以保障系统的公正性和有效性。当系统发生错误或作出不公正的决策时,若无明确的责任主体和问责程序,问题将难以得到及时地纠正,可能损害受益人的权益,还可能削弱社会救助系统的公信力和可靠性。一些专门机构亦建立了独立的监督委员会,负责审查和评估社会救助系统的决策过程和结果,确保其公正性和有效性。其三,算法的可解释性问题。算法决策过程的不透明性使得对其决策逻辑进行有效解释和理解变得极为困难,这种缺乏可解释性的状况不仅加深了算法黑箱的问题,也使得算法决策过程的监管与验证变得极为困难。在使用算法进行社会救助自动化决策时,会按照系统的决策自动分配救助物资与服务,如食品券、住房补贴等,倘若这些算法不可解释,那么申请者就无法得知自己的申请为何被拒,也无法了解如何改进自己的情况以满足资格要求。此外,政策制定者和审计人员也可能难以评估这些系统的公正性和有效性。例如,在某些地区,自动化的福利系统可能会错误地拒绝合格申请者的请求,或者未能及时更新规则变化,导致部分人群失去应有的支持。社会救助自动化决策算法的可解释性直接关系到资源再分配的公平性。缺乏可解释性意味着即使算法存在偏差或错误,^①也难以被检测和纠正,这要求算法设计者在开发过程中采用透明和可解释的算法设计原则,确保算法决策过程的每一步都可以被追溯和理解,从而增强系统的可信度和公正性。

四、社会救助自动化决策发展的治理路径

通过法律与伦理考量、强化人类判断与干预、提高算法透明度与可解释性及动态监督与流程改造等多维度措施,试图构建多层次、多主体参与的治理体系,引领社会救助自动化决策向更加公平、高效的方向发展,以彰显数字政府、数字社会的效度、高度与温度。^②

(一) 法律与伦理考量:负责任创新的算法纠偏

政府在推进负责任创新的过程中,特别注重将负责任创新的原则与理念融入自动化社会救助决策体系,法律与伦理考量是自动化社会救助决策体系构建和应用的核心,须在决策中融入伦理、公平和社会责任。通过这种方式,政府确保自动化社会救助系统在提升效率和准确性的同时,也能充分尊重和保护受助者的权益,防止潜在的歧视和不公平现象。其一,针对算法偏差问题,须深入研究现行法律框架和伦理标准对算法偏差的影响。根据新一代人工智能治理专业委员会发布的《新一代人工智能伦理规范》,在新一代人工智能的开发过程中,必须遵守相关法律法规和标准,提倡加强国家间的合作,共同推动形成国际共识,制定全球性的算法伦理标准和监管框架,并定期进行算法影响评估。同时,坚持伦理先行原则,将负责任创新的治理理念与伦理规范融入人工智能的开发和运行各个阶段。其二,在算法的开发和应用过程中,应

① 丁晓东:《基于信任的自动化决策:算法解释权的原理反思与制度重构》,《中国法学》2022年第1期。

② 郑磊:《数字治理的效度、温度和尺度》,《治理研究》2021年第2期。

秉持负责任的创新理念,包括关注算法的社会影响、尊重用户隐私、保护数据安全等方面。积极寻求技术创新与伦理道德的平衡点,确保算法的应用符合救助申请者的切身利益。针对隐私保护法律可能对数据访问的限制,需要在立法和实践中寻求平衡之道,既尊重个人隐私,又确保数据的合理利用。此外,自动化决策算法开发应用的规制应结合我国社会保障制度新文明的规律、特点,^①在数字化转型中通过技术创新推动社会救助体系的现代化,特别在技术嵌入治理过程中充分考量创造性破坏的风险点,并强化技术创新的预期性、反身性、包容性和响应性,以实现政府负责任地创新发展。其三,对算法偏差行为进行必要的惩处。算法问责法案需提上人大立法议程,针对根据情节的轻重与主观故意与否设定惩处方式,从算法伦理失范到刑事问责,形成政府负责任的自动化决策规制。倘若检测到算法在运行过程中表现出偏差,从而导致决策结果的不公平或不合理,必须采取相应的纠正和惩罚措施。必要时对算法设计者和使用者问责,情节严重的还要对相关机构或个人进行法律或行政上的处罚,促使算法开发者与运用者在实践过程中严格遵守相关法律法规及伦理准则。例如,欧盟的通用数据保护条例(GDPR)强调数据处理的透明度、用户权利和数据保护,GDPR对违规行为设定了严厉的罚款,要求企业在设计和使用算法时遵守严格的伦理准则。

(二) 强化人类判断与干预: 优化技术与算法

在自动化环境中进一步持续优化技术和算法,人类的判断和干预在处理复杂情境时仍然具有重要意义,在复杂和模糊性治理过程中救助的行政自由裁量权依然可以与技术治理形成有益的互补。其一,通过开展专门的培训活动,对社会救助决策者进行系统的教育和指导,以加深他们对算法偏差的认识和理解。培训内容应包括算法偏差的基本概念、表现形式及其潜在影响,并涵盖实际案例分析,帮助决策者识别和理解社会救助领域中可能出现的算法偏差问题。此外,培训还应提供有效的应对策略和工具,使决策者能够采取具体措施,防范和纠正算法偏差,确保社会救助政策和措施的公平性和有效性。其二,确保利益相关方的积极参与以实现自动化决策的有效治理。在自动化决策的社会救助应用中,政府机构、技术供应商、社会团体及受益人等多方利益相关者均扮演关键角色。政府部门作为政策的核心制定者在自动化决策系统设计和实施中起关键作用,包括确保算法符合国家政策导向,平衡技术创新与社会稳定之间的关系,与技术供应商紧密协作,共同研发并实施符合社会救助目标和需求的技术解决方案。社会团体与受益人作为重要的参与群体,也应参与到算法的监督和反馈过程中,让更多的人能够对算法的公正性进行审视,从而增强决策的透明度和公信力。其三,算法优化是一个至关重要的研究领域,其核心目标在于不断提升和改进社会救助识别与决策算法的性能和效率。^②算法优化主要通过改进算法结构、优化算法参数、引入新的算法技术等路径,优化算法可以增强系统面对复杂数据和任务时的适应性和稳健性。^③国际范围内,也有基于相关算法研发与优化的成功先例。例如,IBM公司开发的公平性约束算法能够在训练过程中自动检测和纠正偏差,IBM通过算法优化使得算法在处理各种数据

① 郑功成:《中国特色社会保障制度论纲》,《社会保障评论》2024年第1期。

② 匡亚林、蒙春雨:《数字赋能社会救助分层分类改革》,《学习与实践》2024年第1期。

③ 丹麦的社会救助管理部门工作人员受访的访谈记录中谈到对于算法优化的态度,指出:“我们计划至少每两年不断改进一次回归模型,这是自动决策的基础。”参见Ulrik B. U. Roehl, "Automated Decision-Making and Good Administration: Views from Inside the Government Machinery," *Government Information Quarterly*, 2023, (4).

时能够更加公正和无偏差,通过技术创新和算法优化提升决策系统的有效性。

(三) 透明度与可解释性:提升算法信任度

提高算法透明度和可解释性是解决社会救助算法偏差问题的关键措施。透明度指的是算法的设计、数据来源以及决策过程应当对所有救助利益相关方保持开放和易于理解的状态。无论是政策制定者、执行者还是受助者,都应该能够清楚地了解算法是如何运作的,数据是如何收集和处理的,以及最终的决策是如何做出的。其一,建立透明的人机混合决策系统登记册。算法的救助决策过程能够被清晰地追溯和解释具有重要意义,以深度学习模型为主要的自动化算法为例,其决策逻辑和机制往往不透明且难以被外界理解,其影响就是降低了民众对于自动化系统的信任度,为了提高透明度和可解释性,可建立透明的人机混合决策系统登记册,该登记册应详细记录自动化决策系统的设计和操作参数,包括但不限于算法模型、数据来源、训练过程、评估指标、决策逻辑等。救助决策的算法逻辑和依据能够被相关利益方理解和审查,供政策制定者、研究人员和社会各界参考。其二,公开算法模型的源代码和决策逻辑。为了实现这一目标,公开算法模型的源代码和决策逻辑是非常必要的。这不仅可以使相关利益方(如政策制定者、研究人员和公众)更好地理解 and 监督这些系统,还可以促进算法的持续改进和优化。例如,Google 在其搜索算法方面提供了大量的透明度报告,解释了其搜索排名的基本原理和更新情况。虽然具体的算法细节仍然保密,但这些报告帮助用户了解算法的主要工作方式。此外,Google 还在其广告平台上提供了详细的解释性报告,说明为什么某些广告会出现在特定用户的搜索结果中,这有助于用户理解广告决策的过程。其三,强化算法信任的验证机制。“双因素认证”(Two-factor Authentication),也称为“四眼原则”(Four-eyes-principle),强调重要的操作或决策都需要两个人独立审查和同意,以减少救助错误和欺诈的可能性。引入双人审核机制来增强救助决策的准确性和公正性。在社会救助算法模型的设计与实施过程中,通过两个或更多专家的独立审查,可以有效避免单一决策者的偏见或失误,确保算法模型的合理性和公平性。同时,为了减少用户对自动化决策系统的过度依赖,应当透明地展示系统的置信区间和误差范围。让用户更清楚地认识到系统决策可以提升救助效率,但并非错误率为0。结合可视化技术将复杂的模型转换为直观易懂的形式,可以帮助利益相关方更好地理解和评估算法模型的工作原理及其潜在影响。

(四) 动态监督与流程改造:动态监管下的持续进步

通过持续监督与改进,确保科技的发展与社会的价值相协调,共同推动形成一个更加公正、透明、可信赖的数字未来。其一,引入数据质量提升机制。数据是社会救助决策的基础,任何数据的偏差或错误都可能导致不公平或不准确的决策结果。应借助先进的数据分析工具,对数据进行清洗和验证,确保用于算法模型的信息准确无误;数据的质量监测须对收集到的基准数据进行实时监控和周期性评估,以识别和纠正数据中的潜在问题,并制定明确的数据更新流程和时间表以确保在规定的时间内完成数据的更新;针对异常数据,特别是救助数据中出现的数字格式不一致、缺失值和异常值,须通过复测加以确认。其二,引入 workflow 管理系统(Workflow Management System,简称 WFMS),并实现流程再造。通过 WFMS 系统进行低收入群体信息的比对与采集,实现跨部门和跨层级的数据流动、共享,用于管理和优化工作流程,以改进政府福利项目的管理效率。在此基础上,政府部门可更加精准地识别低收入群体,通过比对和分析

不同部门的数据，可以发现潜在的社会救助需求，实现资源的精准投放。公开透明的数据监控报告将有效促进社会监督和公众参与，定期向社会公布审计结果和数据监控报告，接受公众的监督和评议。通过定期发布，让民众了解到社会救助工作的实际成效和存在的问题，从而增进公众的理解与支持。其三，建立多方参与的持续监督机制，确保自动化决策系统在实施过程中持续接受评估和改进。鼓励民众参与到社会救助监督中来，建立多元化的监督体系，发挥社会的力量，共同揭露和抵制“人情保”“关系保”等救助顽疾。进行公平性评估和优化也是必要的步骤，通过定量分析和模型调整，确保算法在不同群体中的表现尽可能公平。在持续监督方面，建立独立的评估委员会，负责对自动化社会救助政策的执行情况进行全面审查；制定详细的评估标准和程序，内容包括救助对象的满意度、救助资金的使用情况以及救助效果等，将评估结果及时反馈给政策的利益相关者，以便其能够根据评估结果进行调整和改进。

Algorithmic Bias and Governance Pathways in Automated Decision-Making for Social Assistance

Kuang Yalin

(School of Public Administration, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract: With the advancement of artificial intelligence, automated decision-making systems are set to revolutionize traditional social assistance mechanisms by enhancing beneficiary identification, data comparison, and algorithmic recommendations. This marks a significant direction for the future of social assistance. This study examines the technical architecture, decision-making process, and human-machine interaction mechanisms in automated decision-making, emphasizing its potential to improve efficiency and fairness in public services. However, algorithmic bias presents significant challenges to fair administrative decision-making. This study identifies three primary forms of bias: dataset construction bias, algorithm design and operational bias, and bias in algorithm transparency and interpretability. To address these issues, the study proposes a comprehensive governance framework for the high-quality development of automated decision-making in social assistance. Key measures include enhancing dataset diversity, conducting fairness assessments, optimizing model development, improving transparency and interpretability, strengthening human oversight and dynamic monitoring, and refining legal and regulatory frameworks.

Keywords: social assistance; automated decision-making; algorithmic bias; governance pathways; transparency

(责任编辑：李 莹)