

工具理性约束与价值理性嵌入:社会救助算法决策的风险审视与防范机制构建

匡亚林,武 丹

(电子科技大学公共管理学院)

摘要:随着人工智能的不断发展,算法赋能逐渐成为社会救助体系高质量发展的重要动力。以工具理性与价值理性为分析视角,把握算法赋能社会救助决策的底层运作逻辑,深入剖析其可能诱发的异化风险,并寻找针对性的优化策略。研究发现,算法通过感知需求、配置资源、协同治理的内在逻辑赋能社会救助决策,实现了隐性需求显性化、资源配置精准化及治理行动协同化。然而,算法技术在提升治理效率、增强救助精准性的同时,也带来了技术逻辑与公共价值之间的张力。当算法模型深入赋能社会救助决策后,产生了错综复杂的风险隐患:其一是技术层面的算法“黑箱”对程序正义的遮蔽,即决策过程透明度低且难以解释;其二是制度层面的权责边界模糊导致的监管真空,即决策失误时难以厘清算法开发者与基层工作人员的责任归属;其三是价值层面过度扩张的工具理性消解了受助者主体地位,甚至引发新型的数字排斥现象。为预防上述风险,构建技术、制度、价值三位一体的治理逻辑,通过构建可解释与可审计的算法机制重塑技术信任,依托全链条问责体系锚定人机协同权责,并推动技术理性向人文价值回归,为数字时代社会救助体系实现效率提升与价值坚守的统一提供研究参考。

关键词:社会救助;算法决策;预见性治理;算法“黑箱”;数字排斥;数字包容

一、问题的提出

推进全体人民共同富裕是中国式现代化的重要特征。实现发展成果更加公平地惠及全体人民,既需要提高初次分配的劳动报酬,也需要发挥社会保障等再分配制度的兜底保障和调节功能。作为社会保障体系中保障困难群众基本生活、维护社会公平正义的最后一道安全网,社会救助的运行效率与制度公平直接关系困难群众的基本权益保障,也关系国家治理现代化和共同富裕目标的实现。相较于传统社会救助主

要依赖基层工作人员入户调查、群众主动申请和部门人工审核的运行模式^[1],算法技术推动社会救助逐步形成了以数据汇聚、智能识别、动态监测和精准匹配为特征的新型治理模式^[2],其广泛应用于救助对象识别、困难风险预警、政策智能匹配和资源优化配置等关键决策环节,显著提升了社会救助决策的效率与精准性^[3]。自1994年我国正式接入全球互联网以来,社会救助领域先后经历了信息化、业务数字化、数字化转型的演进过程^[4]。近年来,各地又纷纷探索建设了以算法决策为核心的智慧救助平台,

引用本文:匡亚林,武丹.工具理性约束与价值理性嵌入:社会救助算法决策的风险审视与防范机制构建[J].河海大学学报(哲学社会科学版),2026,28(4):35-43.

基金项目:四川省哲学社会科学基金重点项目

作者简介:匡亚林(1986—),男,副教授,博士,主要从事数字化社会保障研究。E-mail:kuangyalin@uestc.edu.cn

将算法模型嵌入社会救助业务流程:2017年,浙江省建设的社会救助信息系统,依托一库、一网、一平台和救助大脑,运用数据建模和智能分析算法,实现了困难群众主动发现、精准识别和政策智能匹配^[5];2020年,广州市建设“穗救易”平台,通过智能比对算法整合公安、人社、医保等部门数据,提升了资格审核效率^[6];2021年,上海市徐汇区将空间地理信息(GIS)与民生数据融合,依托全区2300余万条底层数据,运用困难家庭综合评估算法模型,梳理47项个人和家庭属性标签进行多维立体画像,优化救助等级划分与资源配置;2022年,重庆市建设“渝悦·救助通”,利用数据画像、风险预测模型和动态监测算法,对困难群众开展风险预警和分类帮扶^[7];2024年,西安市碑林区创新“救助+社工”模式,利用需求评估算法联合心理疏导辅助AI工具,使算法决策向发展型综合帮扶延伸。然而,多地也陆续出现因算法误判、模型偏差、自动化审核等引发的救助争议。不同于一般行政事务,社会救助直接关系到困难群众基本生活保障,对公平性、透明性和责任归属有更高的要求。目前,社会救助决策越来越依赖数据计算、模型推演和智能预测,当效率提升、精准识别和成本控制成为算法设计的目标时,工具理性^①不断强化,公平、尊严、回应性和包容性等公共价值则受到挤压,算法“黑箱”、数据偏见、责任模糊、技术悬浮及生成式人工智能“AI幻觉”等风险不断显现。因此,如何防范算法效率逻辑不断扩张、如何实现工具理性与价值理性的协调统一成了数字时代社会救助高质量发展亟待回答的重要问题。

围绕社会救助算法决策,学界形成了较为丰富的研究成果。学者们大多围绕算法如何利用数据驱动进行主动识别^[8]、利用数字孪生技术进行实时交互^[9]或是利用认知、行动、反馈的机制主动发现困难群众^[10]等进行了积极探索。部分学者提出,算法嵌入标志着吸收型数字福利国家的兴起,数字技术成为国家调和新自由市场逻辑与传统福利价值、消解民众异议的关键工具^[11],界面优先的官僚制可以重塑公民与国家服务的互动逻辑^[12],实现以数据要素赋能为核心、以系统整合为中介的整体性

治理^[13]。然而,算法在提高行政效率的同时,也带来了深层的治理危机^[14]。一方面,算法的不可解释性催生了算法“黑箱”,公众面临知情权受损与权力失控的焦虑,与行政法中的说明理由义务相冲突^[15-16];另一方面,在基层应用中,数字技术极易产生技术悬浮现象^[17],即技术功能脱离基层社会,技术与治理相互割裂,导致政府公信力受损。此外,“AI幻觉”等问题也逐渐凸显^[18],即算法可能基于虚假或偏差关联产生貌似合理实则错误的判定。在高风险场景中,社会救助算法决策极易引发严重的合法性危机^[19],如澳大利亚无现金借记卡的研究表明,数字化监测可能演变为对底层贫困人口的严密控制,加剧技术排斥^[20]。针对上述风险,部分学者提出优化治理路径需兼顾技术效能与人文关怀,并且可通过法律层面重构程序正当原则,增设算法规则转译程序、自动化决策纠错的更新程序^[21]。在救助力量层面,强调应推动政府救助+慈善帮扶联动转型,实现资源、主体与过程的精准对接。现有研究为理解算法赋能社会救助决策提供了重要参考,但总体来看,已有研究多聚焦算法赋能效果或风险现象本身,而较少深入探讨算法作为一种强工具理性何以嵌入社会救助制度运行,且对于工具理性扩张何以诱发价值理性偏离,其内在生成机制仍缺乏解释框架。基于此,以工具理性与价值理性为研究视角,立足社会救助算法决策实践,在梳理算法赋能社会救助内在逻辑的基础上,深入剖析工具理性扩张所诱发的算法“黑箱”、责任弱化、技术悬浮、价值偏离等治理风险,进一步构建工具理性约束与价值理性嵌入的防范机制,以期为推动社会救助迈向效率与公平兼顾的善治提供研究参考。

二、算法赋能社会救助决策的内在逻辑

在社会救助体系逐渐向发展型、精细化转变的背景下,算法模型在社会救助领域赋能效

① 韦伯在《经济与社会》中将社会行动划分为工具理性行动、价值理性行动、情感行动与传统行动。其中的工具理性强调行动者基于目的实现对手段进行效率计算与最优选择;价值理性则强调行动者基于伦理、信仰、责任等内在价值而采取行动,其行动并不完全以结果效率为导向。

应的内在逻辑可以归纳为数据挖掘、技术驱动与流程优化的动态耦合体系(图1)。

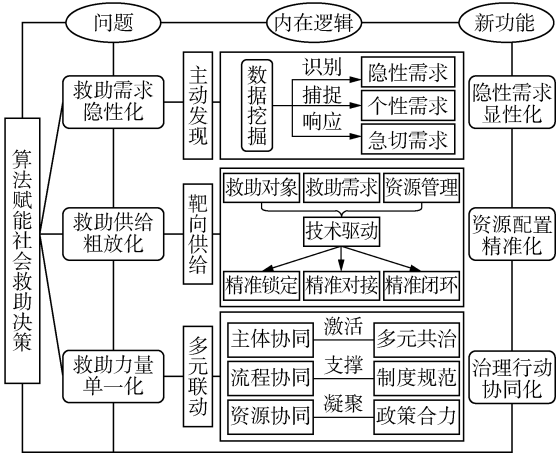


图1 算法赋能社会救助决策的内在逻辑

1. 隐性需求的显性化:基于数据挖掘的主动发现

算法赋能社会救助决策对破解信息不对称导致的治理失灵起着至关重要的作用。其一，主动识别隐性需求。在具体实践中，尽管现阶段跨部门数据无法实现完全互联互通，但部分地方已着力探索集成民政、人社、住建及医保等多部门数据，形成了困难群众多维信息画像，为识别救助对象和资格判定提供决策依据。如浙江省的“浙里救”大救助信息系统已经推动实现了多部门间的数据协同与救助事项联办，通过困难群众经济状况核对、风险预警与主动发现机制，逐步实现低收入群体的动态识别与精准帮扶。此外，也可利用贝叶斯网络等算法动态评估家庭贫困风险，及时识别潜在救助对象，使难以量化的心理与社会特征显性化。如上海市普陀区依靠机器学习等大模型训练算法，将民生数据接入数据中台，发现隐性需求，根据推荐算法绘制待救助人群画像，并进一步链接社会救助子项目^[22]。其二，精准捕捉个性需求。基于无监督学习的聚类分析，算法系统可以综合家庭结构、健康状况、劳动能力和教育负担等信息对救助对象进行分层分类，进而辅助基层部门制定差异化救助方案。算法也能够通过多模态数据分析贫困家庭，提供教育培训、心理辅导和技能提升等非现金救助，提供高质量的发展型救助。如重庆市“渝悦·救助通”平台利用大数据对困难群众的健康状况、就业意愿和

子女教育阶段进行多维画像，精准匹配各部门的个性化救助资源，为患病、失业等不同因由致贫的家庭提供复合型帮扶。其三，敏捷响应急切需求。一方面，面对急切救助需求，可利用监督学习算法构建任务驱动模式，持续监测医疗支出、生活消费和家庭变故等异常信息，及时触发风险预警，并将预警信息推送至基层工作人员，实现急切救助需求的快速发现与及时响应；另一方面，AI大模型可实时采集用水用电异常、医疗支出骤增等线索，自动触发预警信号，并进行智能派单，快速响应并将相关审核后的救助需求推送至属地基层工作人员。如广州市推行“穗就易”，通过接入国家电网的水电数据、卫健部门的医疗自费骤增数据，建立了智能化预警机制，当后台算法监测到空巢老人或困难家庭出现连续多日用水用电量为零或其他异常行为线索时，系统会自动生成预警工单，并智能派单至对应网格的基层民政干部，由其进行及时的线下核实。

2. 资源配置的精准化:基于技术驱动的靶向供给

社会精准救助，是指从对象精准识别、资源精准配置、服务精准供给到过程精准管理的全链条精准^[23]。其一，数据驱动赋能精准识别。算法侧重于建立严密的数字化建档立卡体系，研究发现利用监督学习算法模型并结合少量标注数据与海量卫星图像，能够绘制高精度的低收入群体地图，实时锁定可能处于社会救助的边缘群体^[24]，也可引入空间机器学习算法降低排除误差，有效解决区域发展不平衡带来的识别难题^[25]。其二，需求导向赋能资源配置。精准救助的核心在于解决扶持谁、怎么扶的匹配难题。基层人员采用深度神经网络和集成学习模型相结合的方式，评估劳动能力、教育程度、健康状况等各方面因素后，可自动制定相应的援助措施。如浙江省打造的大救助信息系统通过多维数据建模，在评估申请家庭的财产收入的同时，更能深度分析其家庭成员的劳动技能、身心健康及就学负担。根据算法评估结果，系统在后台将其归类为不同类型，并联动人社、残联等部门精准投放定向岗位推荐或康复增能服务，实现救助项目与个体诉求的靶向衔接。其

三,实时监测赋能闭环管理。资源配置精准化要求构建中央统筹、省负总责、市县落实的责任链条,精准管理救助资源:中央层面依托全国社会救助信息平台,建立统一数据标准与算法模型,实现跨区域跨部门数据共享与风险预警;省级层面负责算法模型本地化适配与动态调优,确保决策符合政策导向与地方实际;市县层面聚焦算法应用的“最后一公里”,通过人机协同将算法输出转化为具体救助行动,并实时反馈执行效果。算法系统通过建立以救助对象满意度、救助效果持续性、社会公平感知度为核心的多维评价指标体系,实现动态追踪与效果量化,最终形成投放、评估、优化的良性循环。

3. 治理行动的协同化:基于流程优化的多元联动

当前,基层社会救助体系普遍存在资源匮乏和力量薄弱等问题^[26],在对象资格审查标准、信用监督机制、违规行为惩处等方面的法律法规尚不明确,同时各部门在经费保障、人员管理、信息共享、政社协同等方面也存在明显障碍,而算法赋能的社会救助决策正成为解决上述问题的关键。其一,主体协同多元共治。纵向上,以市级统筹区级为背景,推进街道、社区网格化管理,并利用非监督学习算法识别社会组织和公益组织的优势领域,根据困难群众需求自动匹配服务主体;横向上,依靠大数据平台整合数据,创建覆盖需求评定、资源调配和服务供给全链条的服务系统。其二,流程协同制度规范。针对基层救助权责边界模糊、缺乏程序规范的痛点,算法赋能通过流程再造为基层工作人员提供支持。一方面,算法模型构建“线上预警+线下响应”的机制,用技术辅助基层人力开展对象核访与资源链接。如西安市通过纵向链接救助资源、横向耦合多元社会力量,构建了“融救联助”服务网络;江苏省江阴市则通过集成式智慧救助平台,实现了“大数据+铁脚板”的联动审查。另一方面,算法介入形成的数字化审计与动态监测体系,为救助过程中的信用监管提供了证据溯源路径。通过标准化决策流程,算法为基层工作人员提供支持,但最终决策仍需要基层工作人员结合具体情境进行人工复核,形成算法辅助+人工裁量的人机协同

决策模式。其三,资源协同政策合力。资源协同的内在逻辑是利用数据比对机制,消除不同救助主体间的职能重叠。算法对民政、卫健、医保、教育及残联等部门数据进行实时关联,精准识别救助对象已享受的保障现状,进而达成资源配置最优化的效果,并通过强化部门间的沟通交流与数据共享,确保各项保障政策同向发力,将分散的救助资源汇聚为共享的资源池。

三、社会救助算法决策的风险审视

数智技术与社会救助体系的深度融合,体现了工具不断发展并提升公共治理效能的趋势。算法依靠数据计算、效率优化与标准化识别提升社会救助的精准性与敏捷性,但当技术逻辑逐渐凌驾于公共伦理与社会公平之上时,也易导致程序正义弱化、责任边界模糊及消解弱势群体主体性等问题。工具理性与价值理性之间的张力,由此成了算法赋能社会救助决策过程中不可回避的议题。技术失灵、制度悬浮与价值异化等风险,正是工具理性过度扩张后在技术、制度与伦理层面的集中体现。

1. 技术失灵:算法“黑箱”对程序正义的逻辑遮蔽

数字化转型过程中,算法被用来促进社会救助领域精准化、高效化管理,其初衷是依靠信息技术解决传统救助模式下信息采集速度慢、响应迟缓等难题。但当社会救助过度依赖算法模型时,技术系统所强调的效率优先与标准化判断就可能逐渐压缩程序正义、公众知情与个体申诉等价值理性空间。一方面,决策逻辑“黑箱”化削减了政策透明度与可信度。由于决策逻辑处于“黑箱”状态,尤其是在救助申请未被通过而又无从查证的情况下,低透明度将大大削弱困难群众的信任基础,从而影响政策公信力。由于算法本身存在指标调和偏差、参数训练约束及技术局限性,当制度主体过度授权算法进行预测或自主决策时,极易产生误导性的救助研判。一旦出现贫困家庭因算法误判而被不当排除的情况,受助者往往无法理解被排除的具体原因,且由于缺乏纠偏补足机制,其申诉权将被搁置。这种缺乏公开性的自动化处理过程在侵犯被救助者知情权的同时,也会因

不规范的数据读取而增加个人隐私泄露的风险,进而引发公众对社会救助体系的信任危机。如美国印第安纳州曾推行福利自动化改革,通过 IBM 构建自动化福利审核系统,对食品券、医疗补助等福利申请进行算法化处理。但由于系统过度依赖自动化决策,大量申请因材料识别错误、信息匹配失败等原因被自动拒绝,甚至出现困难群体在未获得充分解释的情况下被取消福利资格,引发广泛争议。最终,该项目因大量误判、程序不透明与公众投诉而被迫终止。另一方面,数据训练的预设偏见对弱势群体公平权利的侵蚀。算法模型倾向于复刻社会结构的既有偏见,尤其是监督学习算法存在着显著的偏见固化效应,即通过学习历史数据固化偏差。在算法赋能社会救助决策的场景下,那些因文化水平有限而未能留下数字足迹的群体,往往被称为统计上的透明人,其真实需求未被算法模型覆盖,加上缺乏相应的人工复核机制,导致其求助需求难以被及时发现。由于数据质量缺陷及隐性偏见在训练过程中被无限放大,算法带来的歧视在运行中难以消除,且将在长期过程中进一步损害低收入群体的公平权利。技术理性对程序正义的遮蔽,使得救助过程不仅未能消除数字不公,反而催生了新型数字鸿沟,导致救助资源的分配与社会正义的目标产生偏差。美国弗吉尼亚大学的研究也证实了这一观点,在福利算法系统中,缺乏稳定数字记录与数据痕迹的弱势群体往往更容易被排除在算法识别体系之外,形成隐性排斥^[27]。

2. 制度悬浮:人机交互下权责模糊的监管真空

在算法赋能社会救助决策的场景中,人机交互的概念进一步深化为基层工作人员与大模型的权力重组关系,人和技术共同构成救助资格判定与救助资源分配的双重主体。然而,在工具理性主导的算法治理框架下,社会救助决策越来越依赖技术系统的自动化判断,传统行政体系中的责任链条也因此被重新拆解^[28]。一旦出现决策失误,将面临以下两个方面的难题:一方面,执行权异化,行政责任在人机协同过程中被稀释。随着数字技术深入赋能社会救助决策,传统的行政决策权越来越向算法开发

者、平台经营者、一线执行人员等多方面分散。虽然在目前大部分的场景中,算法扮演辅助角色,但其对资源分配及权力分配的影响不容忽视,并且正逐渐变成隐形的决策中心。目前,法律体系对于这类新型责任划分并没有明确的界定,加之信息流传输复杂,原来清晰具体的责任链条也面临互相推诿的风险,追责难度提高。另一方面,归责逻辑失效,缺乏问责机制损害制度公信力与救助稳定性。虽然基于大数据分析可在一定程度上减少人为干扰因素、有效降低人情保等违规行为,但由于缺乏有效的监督核查机制,当算法失误造成纠纷的时候,各方往往会利用技术复杂的理由互相推卸责任,形成集体逃避局面。目前,针对由算法决策失误引发的社会损害行为,我国尚未建立针对性惩罚的法律制度,相关法律制度的缺位既会损害被救助人的权益,也必然会削弱社会救助体系的保障底线。

3. 价值异化:技术理性对弱势群体的数字排斥

算法的应用虽能明显提高行政效率,但是当前的数智技术嵌入过程有明显的工具理性特性,导致技术计算逻辑与人文价值关怀之间出现鸿沟,主要体现在以下方面:一方面,个体主体性的消解体现在量化指标衡量下的救助对象标签化倾向。按照事先设定的规则,无监督学习中的聚类算法将具体的个人转化为抽象的数据点,并以此创建标准化的服务模型,由此虽然可提高数据管理效率,但使得大量的非结构化个人信息流失,使受助对象的特殊需求、个性化特征被标准化的评价所忽略。另一方面,技术准入门槛障碍体现在数字素养差距引发的新型社会不公平。数智技术赋能社会救助决策的目的是提质增效,但技术条件的高智能化要求与弱势群体匮乏的数字素养之间存在严重差距。由于信息获取能力的限制,老年群体、残疾人等特殊人群面临着技术使用困难和身份认同歧视双重障碍^[29-30]。即使在理想环境下,由于数据采集误差等原因,自动化系统也可能会做出错误的判断或遗漏一些被救助者,对于经济条件差、缺乏数字技能的特殊群体来说,其真实需求将更难以被有效捕捉。

四、社会救助算法决策的风险防范
 机制构建

算法赋能是提升基层治理韧性的关键举措,实现工具理性与价值理性的有机统一是其不容回避的议题。在社会救助领域,算法决策既承载着提升治理效率的工具理性诉求,又必须嵌入保障公平正义的价值理性内核。为突破单一技术应用的工具性束缚,亟须构建技术-制度-价值三位一体的系统框架(图2):在算法嵌入社会救助治理过程中,技术失灵、制度悬浮与价值异化成为影响算法有效运行的重要风险来源。对此,需要建立精准应对的防范机制:在技术与制度层面,通过逻辑公开与过程审计实现技术重塑与留痕溯源,界定权责边界、量化贡献比例推动责任复位与尊严重塑;在价值层面,借助多元评价与数字包容开展赋能激活与纠偏补偿,确保算法运作始终坚持以人民为中心的发展方向。

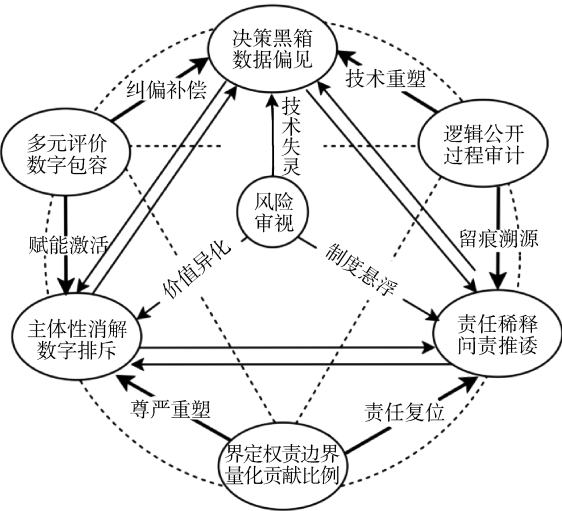


图2 基于算法赋能的社会救助决策优化框架

1. 技术重塑:构建可解释与可审计的算法治理机制

在工具理性主导的算法治理逻辑下,社会救助决策易陷入效率优先与技术中心主义的误区。因此,建立可解释、可审计的算法治理机制是对工具理性扩张的必要约束,也是推动价值理性嵌入社会救助过程的重要前提。一方面,算法逻辑可解释,即构建透明且可追溯的社会救助决策机制。由于算法赋能社会救助决策领域的“黑箱”效应愈发严重,必须通过提高信息

透明度对算法权力进行约束,防止技术系统因不可解释而侵蚀社会救助的程序正义与公众信任。依托金民工程的数据整合能力,将住房、医疗、就业、社会保险等救助相关数据进行关联比对,并通过可视化方式呈现低收入家庭识别依据、风险预警来源与救助匹配过程。在此基础上,政府部门应建立适度透明的算法公开机制,在确保个人隐私与数据安全的前提下,公开救助对象识别标准、指标权重设置原则及算法运行基本规则,并定期发布算法评估与修正报告,以保障社会公众对算法决策的知情权与监督权。另一方面,过程可审计,即以公平导向修正算法偏见的路径依赖。算法偏见往往源于训练样本分布不均衡或权重分配不合理,易忽视边缘群体的特殊需求。因此在数据管理过程中,应全方位引入公平性评价指标,构建包含事前预判、事中监管和事后审计的智能监管机制。为提高模型稳定性和适应性,在算法模型测试阶段应通过历史救助案例模拟、特殊困难群体样本回溯等方式,对低保边缘家庭、残疾人、独居老人等群体进行重点校验,避免因训练样本偏差导致算法偏差。同时,可利用区块链技术,搭建透明的数据追踪平台,用台账记载每一次的决策过程和资金的流动。当检测到异常行为时,系统立即触发警报信号,调动人工队伍执行应急处置任务,形成完善的预警与应急管理机制。

2. 权责锚定:强化人类干预与全链条问责体系

在社会救助算法决策实践中,算法将超越传统技术工具的角色,且应将其作为关键的非人类行动者,深度嵌入困难识别、资格审核、资源分配等流程。随着算法系统与基层工作人员、受助对象、数据平台等多元主体持续互动,人机边界日益模糊,公共责任稀释与价值判断悬浮,一旦出现误判、漏保或歧视性结果,责任归属难以厘清。为防范此类治理风险,亟须在社会救助算法决策中确立清晰的权责锚点,通过制度问责,实现价值理性对技术权力的有效规制。一方面,通过界定权责边界,构建纵向贯通的行政责任矩阵和追溯机制。构建覆盖算法设计、部署、运行与评估的问责体系,明确开发

方、使用单位、监管机构及一线执行人员的职责边界。社会救助领域政策敏感度高、社会关注度高,因此亟须突破技术中立的束缚,在模型运行过程中对算法研发者及政务终端使用人的法律责任作出明确规定。算法研发者应承担算法鲁棒性与逻辑合规性的首要责任,对所用技术框架的安全可靠性、逻辑一致性及面对大量分散数据整合时可能引发的风险进行评估。基层民政部门等执行主体需要提升自身的数据校验与终审控制水平。一旦救助决策出现偏差,通过建立清晰的责任追溯机制,能够依据责任矩阵,迅速定位是技术逻辑缺陷、数据质量问题还是人为操作失误,从而及时纠正,并对违规行为进行精准追责。另一方面,应推进人机协同贡献的量化,破解工具理性下的归责模糊与问责推诿。应当在资格终审、异议复核等关键环节保留必要的人类干预权,确保算法处于辅助角色。在交互过程中,人与机器会相互影响,采用传统的将主体与客体区分的方式处理问题已经难以适应新的需要。因此,司法实践中迫切需要形成人类意志和机器输出之间责任分担的标准。在社会救助的自动化决策过程中,既要承认算法在风险预警与精准识别中的辅助功能,更要警惕技术带来的责任能力虚置。若脱离自然人的监管而单独设立算法责任,将导致问责机制流于形式。为此,应坚持人类干预的原则,对资格审核、停保退出、异常预警等关键决策环节明确人工复核责任,防止基层部门对算法结果直接采纳、一键审批;当算法建议与基层实际调查结果存在明显冲突时,应保留工作人员对救助结果的修正权与最终解释权。

3. 价值回归:推动技术理性的人文价值转向,促进数字包容

在算法赋能社会救助决策的过程中,必须警惕过度追求工具效率而忽视人文价值的错误观念,要着力解决量化评估模型过度简化的问题与社会公平方面数字鸿沟带来的挑战,以确立服务对象主体地位和有效维护个体尊严。一方面,重塑多元评价维度,防范量化指标对救助情境的扁平化刻画。针对算法在处理复杂贫困情境时易产生测量失灵等问题,应推动救助模式多元化,实现物质与情感双重救助的转型,注

重情感关怀。针对独居老人、残疾人等特殊困难群体,可通过定期电话回访、社区走访与基层网格员协同排查等方式,弥补算法识别在情感感知与特殊处境判断中的局限。同时,应继续坚持线上算法预警与线下人工走访的精细化治理模式:线上平台负责常规风险定位,解决覆盖范围的遗漏问题;线下社会工作者则深入具体的生活情境,通过面对面沟通提供精细化的心灵抚慰与帮扶。数智技术与人文关怀的联动,能够有效弥补算法在情感感知与复杂情境响应上的迟滞,从而保障救助对象作为人的主体性与尊严。另一方面,构建数字包容体系,消除弱势群体在智慧救助中的隐性排斥。算法赋能社会救助决策应当拓展人的自由和行动能力,而非构建新的技术门槛。为此,可推进城乡救助服务一体化发展,针对偏远农村地区网络环境脆弱的现状,部署低功耗终端和“云端+终端”模式,避免困难群众因缺乏数字设备或技术使用能力不足而被排斥在智慧救助体系之外。同时,也应从数智赋权的角度出发,降低技术落地的准入门槛。如在救助平台中增加语音引导、大字体模式并简化申请流程,为老年人、残障人士等数字弱势群体提供适老化、适残化服务支持。通过加强对基层社会工作者的数字素养培训,提升其在数据采集、岗位匹配与政策执行中的技术使用能力,消解困难群众与复杂系统之间的数字鸿沟。此外,应秉持以人为中心的设计理念,确保大数据、区块链等前沿技术在算法洪流中始终坚守人文底色。

五、结 语

在国家治理体系和治理能力现代化推进过程中,社会救助对于保障弱势群体基本权益、促进社会公平正义起着不可替代的作用。在数字化转型的浪潮下,以大数据、人工智能为载体的算法技术逐渐深入传统的社会救助服务,通过主动发现、靶向供给与多元联动优化社会救助的流程。然而,当算法深度介入决策时,其技术理性与人文关怀之间也产生了显著张力,为此,本研究主要探讨了在效率优先的技术逻辑下,如何规避算法“黑箱”对程序正义的遮蔽、如何填补人机交互中权责模糊的监管真空及如何防

止弱势群体在智慧救助进程中遭遇新的数字排斥。研究表明,算法赋能社会救助决策通过感知需求、配置资源、协同治理的内在机制,实现了隐性需求显性化、资源配置精准化与治理行动协同化,推动社会救助向主动发现与预见性治理的范式转型。然而,当算法模型深度介入资格判定与资源分配时,极易衍生技术、制度与价值层面的三重异化风险。针对这些问题,本研究构建了技术-制度-价值三位一体的治理路径:构建可解释与可审计的算法机制保障公民知情权,依托清晰的问责体系锚定人机协同的行政权责,并推动技术理性向人文价值的数字包容方向转变。算法赋能社会救助决策的图景在于技术工具与人文价值的有机统一,地方政府在引入算法赋能决策时,应推动数智工具与价值理性的深度契合,从而为推进国家治理体系和治理能力现代化贡献智慧和方案。

参考文献:

[1] 李静,段昆昆. 平台型社会救助服务:实践形态与运作机制[J]. 中州学刊,2025(9):83-93.

[2] 匡亚林,蒙春雨. 从工具理性到价值理性:数字技术驱动下低收入人口动态监测的现实困境与出路[J]. 河海大学学报(哲学社会科学版),2025,27(4):121-131.

[3] 匡亚林,蒙春雨. 迈向智慧救助:AI大模型驱动社会救助数字化转型的实现路径[J]. 内蒙古社会科学,2025,46(5):37-47.

[4] 陈堂,陈光,陈鹏羽. 中国数字化转型:发展历程、运行机制与展望[J]. 中国科技论坛,2022(1):139-149.

[5] 祝建华. 智慧救助的要素驱动、运行逻辑与实践进阶[J]. 社会保障评论,2022,6(2):105-121.

[6] 广州市民政局. 广州市民政局组织开展“穗救易”信息系统轮训[EB/OL]. (2024-01-16)[2026-07-01]. https://mzj.gz.gov.cn/dt/mzdt/content/post_9443919.html.

[7] 杨铃,王燊成. 社会救助平台数字治理的效果评估与生态优化——基于31个省域的实证研究[J]. 社会政策研究,2026(1):121-133.

[8] 王增文,罗佩玲. 分层分类社会救助体系的数智化转型:逻辑转换、衍生风险与化解路径[J]. 行政论坛,2025,32(6):16-25.

[9] 匡亚林,洪思佳. 数字孪生驱动下的社会救助流

程再造——以低收入人口常态化帮扶场景为例[J]. 重庆社会科学,2026(1):102-114.

[10] 李雨荃. 人机协同视角下社会精准救助的实践逻辑及优化路径[J]. 福建师范大学学报(哲学社会科学版),2026(1):64-73.

[11] Busacca M. Bytes, barriers, and logics: the vicious circle of digital welfare in fragmented institutional contexts[J]. International Journal of Sociology and Social Policy,2025,45(13-14):1-18.

[12] Meers J, Halliday S, Tomlinson J. An ‘interface first’ bureaucracy: interface design, universal credit and the digital welfare state[J]. Social Policy & Administration,2025,59(1):119-134.

[13] 许小玲,张高荣. 数智赋能社会救助的逻辑理路与异化风险治理——基于TOE框架的分析[J]. 人文杂志,2025(8):123-131.

[14] 张龙辉. 数字社会建设中的算法安全风险及其化解策略[J]. 东北师大学报(哲学社会科学版),2024(2):134-144.

[15] Busuioc M. Accountable artificial intelligence: holding algorithms to account [J]. Public Administration Review,2021,81(5):825-836.

[16] Grimmeliikhuijsen S, Meijer A. Legitimacy of algorithmic decision-making: six threats and the need for a calibrated institutional response [J]. Perspectives on Public Management and Governance,2022,5(3):232-242.

[17] 徐明强,许汉泽. 乡村数字治理中的技术悬浮及其生态化应对——基于H县“321”基层社会治理大数据平台的案例研究[J]. 电子政务,2025(8):113-124.

[18] 张铤. 人工智能嵌入社会治理的风险及其规避[J]. 浙江工商大学学报,2022(3):120-126.

[19] 胡泳,王昱昊. 技术过程论视角下AI幻觉生成的价值负荷与伦理问题探析[J]. 南京社会科学,2025(3):84-94.

[20] Bielefeld S. Digitalisation and the welfare state-how First Nations people experienced digitalised social security under the Cashless Debit Card[J]. Journal of Sociology,2024,60(3):599-617.

[21] 唐安然. 自动化行政中正当程序原则的重构[J]. 比较法研究,2025(5):177-191.

[22] 匡亚林,赵蜀蓉,孟雅琪. 数字赋能社会救助高质量发展的路径选择[J]. 中国行政管理,2024,40(7):155-158.

[23] 杨玉文. 数智化转型驱动边疆地区治理现代化:价

值意涵、实然样态与行动路径[J]. 云南民族大学学报(哲学社会科学版),2025,42(6):121-130.

- [24] Yeh C, Perez A, Driscoll A, et al. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa [J]. Nature Communications,2020,11(1): 2583.
- [25] 匡亚林. 社会救助自动化决策的运行机理、算法偏差及治理路径[J]. 社会保障评论,2025,9(2): 49-62.
- [26] 李志明. 中国社会救助管理体制与经办机制优化研究[J]. 行政管理改革,2026(1):99-107.
- [27] Wu D, Meyer B D. Certification and recertification in welfare programs: what happens when automation

goes wrong? [R]. National Bureau of Economic Research,2023.

- [28] 樊浩. 算法公共决策“责任鸿沟”风险的生成机理与运行表征[J]. 河南师范大学学报(哲学社会科学版),2026,53(1):45-52.
- [29] Considien M, Mcgann M, Ball S, et al. Can robots understand welfare? Exploring machine bureaucracies in welfare-to-work [J]. Journal of Social Policy,2022,51(3): 519-534.
- [30] 姜燕,钟大年. 智媒时代老龄群体数字素养的生成困境与提升机制研究[J]. 河海大学学报(哲学社会科学版),2026,28(2):94-106.

Constraining Instrumental Rationality and Embedding Value Rationality: Risk Analysis and Prevention Mechanisms for Algorithmic Decision-Making in Social Assistance / Kuang Yalin, Wu Dan (School of Public Administration, University of Electronic Science and Technology of China)

Abstract: In recent years, with the development of artificial intelligence, algorithmic empowerment has gradually become an important driving force for the high-quality development of the social assistance system. This study is devoted to grasping the underlying operational logic of algorithm-enabled social assistance decision-making from the perspective of instrumental rationality and value rationality, deeply analyzing the risk of alienation it may induce, and finding targeted optimization strategies. It finds that algorithms empower social assistance decision-making through the internal logic of perceived demand, resource allocation, and collaborative governance, and realize the explication of implicit demand, the precision of resource allocation, and the coordination of governance actions. However, while improving governance efficiency and enhancing the precision of social assistance, algorithms also give rise to tension between technical logic and public value: first, at the technical level, the algorithmic black box obscures procedural justice, that is, the decision-making process is low in transparency and difficult to explain; second, it is the vacuum in the regulatory field caused by the ambiguity of the boundary of power and responsibility at the institutional level, that is, it is difficult to clarify the responsibility of algorithm developers and grass-roots staff when decision-making is wrong; third, the excessive expansion of instrumental rationality at the value level dispels the subject status of recipients and even triggers a new type of digital exclusion. To solve the above dilemma, this study constructs the governance logic of the trinity of technology, system, and value, reshapes the technical trust by constructing an interpretable and auditable algorithm mechanism, anchors the power and responsibility of human-machine coordination based on the whole chain accountability system, and promotes the return of technical rationality to humanistic value, aiming to provide references for the unity of efficiency improvement and value persistence of the social assistance system in the digital age.

Key words: social assistance; algorithmic decision-making; predictive governance; algorithmic “black box”; digital exclusion; digital inclusion