

全国中文核心期刊

中文社会科学引文索引(CSSCI)来源期刊

中国人文社会科学期刊AMI综合评价(A刊)核心期刊

中文科技期刊数据库(全文版)收录期刊

RCCSE中国核心学术期刊(A)

2025. 4

VOL.18 NO.4

公共行政 评论

JPA
Journal of Public Administration

●专栏：人工智能与政府信任

当 AI 成为“公仆”：政务服务数字人类型对政府信任的影响研究

..... 李姝婷 邱 茜 1

公民更信任公共部门中的“人”还是“人工智能”：一项考虑任务复杂程度
与公共领域的调查实验研究

和思绮 刘炳胜 22

●论文

算法特征披露何以影响公众决策偏好：基于联合实验的实证分析

..... 王 戈 向 丽 张哲君 41

数字技术与多层级治理模式的变迁：以空气污染治理为例

..... 陈 靖 徐 灿 洪 伟 63

农村数字化转型何以悬浮：基于政策执行过程的案例研究	
..... 傅利平 康开荣 冯华超 董永庆	81
基层政府资源分配行为及其多重逻辑研究：基于山东省 A 县的案例分析	
..... 孙宗锋 林婧淇 张晓涵	102
政策组合强度何以影响政策效能：基于大气污染防治政策组合的分析	
..... 张晓杰 刘晓瑜	118
社会组织发展能否有效缓解居民收入差距：基于政府干预门槛效应的实证检验	
..... 刘 蕾 张新亚 何书仪	138
财政纵向失衡会降低地方政府公共服务注意力吗？	
..... 张 原 刘丽君	155
 ●理论综述	
财政压力下的公共组织削减管理：一个文献综述	 陈 馨 173
 ●英文目录与摘要 196	

算法特征披露何以影响公众决策偏好： 基于联合实验的实证分析

王 戈 向 丽 张哲君*

【摘要】算法决策有效提升了政府决策的科学性和精准性，但也引起了公众对决策偏见、隐私侵犯、道德困境及监督失效等多方面的担忧。如何有效实施算法披露以增强公众对公共部门算法决策的选择偏好，成为当前政府面临的重要挑战。论文以信号理论和披露效应为理论支撑，基于1356名代表性样本并围绕公共医疗服务场景展开联合实验设计，分析了不同维度的算法特征披露和公众算法决策偏好的因果关系，检验了感知算法信任和感知算法透明度在这一过程中的中介作用机制。研究发现，在越关键的阶段披露算法应用，其潜在的负面披露效应越明显。而披露不同水平的算法参与程度、算法责任分配、算法绩效信息则产生了复杂的双重效应。感知算法信任在高水平的算法参与、算法绩效披露中发挥中介作用，而感知算法透明度在不同水平的算法绩效信息披露中发挥中介作用。研究补充了算法特征披露影响公众算法决策偏好的研究内容，并对公共部门有效实施算法披露提供了对策建议。

【关键词】算法决策 披露效应 感知算法透明度 感知算法信任 联合实验

【中图分类号】D63

【文献标识码】A

【文章编号】1674-2486 (2025) 04-0041-22

一、引言

算法是人工智能发展的核心要素，对企业实施数字化管理和国家向智能社

* 王戈，华中师范大学公共管理学院副教授。通讯作者：向丽，华中师范大学公共管理学院硕士研究生。张哲君，北京师范大学政府管理学院博士研究生。感谢匿名评审专家、编委会和编辑部对论文的宝贵修改意见。

基金项目：国家自然科学基金青年项目“人工智能算法嵌入街头官僚决策的行为效应及其认知触发机制研究”（72304110），清华大学国家治理研究院课题“人类命运共同体治下WTO应对全球疫情危机的改革”。

会转型发挥了基础性的推动作用（薛澜、赵静，2024）。算法决策是一种集成了机器学习、自然语言处理、计算机视觉等多种技术的复杂系统，通过分析数据和学习模式来做出或协助做出决策（Mergel et al.，2024）。伴随着人工智能正式成为国家战略的重要组成部分，算法决策已成为公共部门提升公共治理效能、优化资源配置的关键工具，在促进智慧医疗、完善智能交通、优化行政执法、提高应急管理等方面展现出显著的优势与潜力（曾大军等，2021）。然而，算法决策潜在的“黑箱”特性、信息不对称、责任性缺失及可解释性不足等问题，引起了社会对其存在决策偏见、隐私侵犯、道德困境及监督失效等多方面的担忧（吴进进、符阳，2021）。

在此背景下，深入了解公众对公共部门算法决策的选择偏好不仅关系到算法技术嵌入公共治理中的有效性和伦理可接受性，而且对推进政府治理智能化、建设人民满意的数字政府也极为关键。过往研究发现，个体对算法决策的态度与行为呈现出算法趋近与算法排斥两种截然相反的倾向（Acikgoz et al.，2020；Logg et al.，2019；王戈、张哲君，2023）。究其原因，包括不同维度的算法特征、特定情境特征以及个体特征等多重因素会发挥显著的影响作用（Grimmelikhuijsen，2023；König et al.，2024；Nagtegaal，2021；Shin，2020；Wang et al.，2023）。随着公共部门日益采用人工智能技术，算法披露逐渐成为提升公共部门治理透明化与公信力的重要途径。传统的算法披露通常限于告知算法的存在、基本原理或数据来源，而广义的算法披露进一步扩展到展示算法应用的关键信息与流程，包括算法生成内容的时机、AI系统注册信息、训练材料的总结和风险评估等（Bonsón et al.，2023）。广义的算法披露不仅会激发个体的兴趣行为，从而产生积极的效应（Li et al.，2024；Yang & Shi，2022），而且在某些情境下也可能引发消极的披露效应，即算法的透明化披露可能会导致个体对算法的厌恶情绪（Dietvorst & Bharti，2020；Keppeler，2024）。然而，关于广义披露中具体算法特征对公众算法决策偏好产生的披露效应，过往研究尚未展开充分的讨论。

本研究从广义算法披露的视角出发，旨在补充现有关于算法特征披露影响公众算法决策偏好的研究内容，以丰富披露效应的理论内涵，同时为公共部门有效运用算法、人工智能等数字化技术进行治理创新提供实践借鉴。本研究以信号理论和披露效应为理论基础，通过系统回顾相关研究文献，构建了包含算法决策阶段、算法参与程度、算法责任分配以及算法绩效信息的披露效应分析框架。通过在公共医疗服务场景中对1356名代表性样本展开联合实验，本研究系统探索了不同维度的算法特征披露如何影响公众对公共部门算法决策的选择偏好，同时检验了感知算法信任和感知算法透明度在这一过程的中介作用。此举不仅有助于设计更加科学和透明的算法决策系统应用，也是实现技术伦理和

公共信任的关键步骤。

二、理论基础与研究假设

（一）公众对算法决策的态度与行为

算法决策不断嵌入私人与公共部门引起了学界的广泛关注，个体对算法决策的态度与行为成为多个学科关注的前沿热点问题。一方面，基于算法欣赏的理论假设，个体认为在某些领域中算法决策有更高的公平感、信任度、满意度，更少的偏见和歧视，因而更加信任和接受算法决策（Araujo et al., 2020; Logg et al., 2019）。另一方面，基于算法厌恶的理论假设，个体认为在某些领域中算法决策不及人类决策公平、可信，因此不愿意接受来自算法决策的建议（Acikgoz et al., 2020; Dietvorst & Bharti, 2020）。此外，个体对不同类型算法决策表现出差异化的态度与行为倾向（Castelo et al., 2019; Wang et al., 2023; 王戈、张哲君, 2023）。

同时，个体对算法决策的复杂态度与行为背后的触发机制同样引起了广泛关注。一方面，公众对算法决策的态度与行为会因任务熟悉度、决策类型、任务客观性、任务难度、道德情况、决策结果等具体任务情境的变化而变化（Araujo et al., 2020; Nagtegaal, 2021; Wang et al., 2023; 魏昕等, 2021）。另一方面，个体的性别、年龄、人格特质、算法使用经历、知识背景、算法素养等个体特征会发挥显著的影响作用（Shin, 2020; Logg et al., 2019）。此外，算法本身的可解释性、可访问性、可问责性、拟人化程度、自动化程度、人机回圈等算法特征亦会产生显著影响（Grimmelikhuijsen, 2023; Horvath et al., 2023; König et al., 2024; 罗映宇等, 2023）。值得注意的是，已有研究开始从不同维度的算法特征出发探索公众对公共部门算法决策的选择偏好（Gao Zhao, 2023; Horvath et al., 2023）。如 König 等（2024）利用联合实验设计探讨了公共部门算法的有效性、透明度和利益相关者参与的影响作用。研究表明，即使在警务和医疗等敏感领域，人们也愿意放弃利益相关者参与算法设计和算法透明度，以换取算法有效性方面的小幅提升。然而，鲜有研究从广义算法披露的视角出发，系统探索不同维度的算法特征披露如何影响公众对公共部门算法决策的选择偏好，本文旨在填补这一研究空白。

（二）算法决策阶段与公众偏好

信号理论（Signaling Theory）用于描述在信息不对称的情况下，个体或组织

如何传递与解读信息 (Connelly et al., 2011)。发送方决定是否以及如何披露信息,而接收方则基于信号解读其价值。该理论为理解算法披露的影响提供了有力框架,其核心在于减少信息不对称。通常,信号传递者掌握着外界无法直接获取的正负面信息,披露这些信息可降低不道德行为的风险,节省成本并维护声誉 (Li et al., 2024)。在人工智能应用中,算法披露被视为一种信号,可以引发个体的特定行为,尤其在负面情境下,有助于缓解焦虑并改善用户体验。例如, Yang 和 Shi (2022) 发现,披露人工智能代理的存在可减轻公众焦虑对客户满意度、公司评价和购买意愿的负面影响。然而,即使人工智能预测能力优于人类,其披露也可能导致个体对人工智能产生厌恶情绪 (Tong et al., 2021)。如 Keppeler (2024) 发现,在雇主主导的招聘中,披露使用人工智能会显著降低求职者对职位的兴趣。

披露算法在何时生成内容对于算法设计者和利益相关者来说至关重要。过往的研究表明,披露算法决策应用的不同阶段会对公众偏好产生显著影响。Luo 等 (2019) 将聊天机器人的身份披露分为不披露、交谈前披露、交谈后披露和决策后披露四种情境,发现了不披露能使公众购买转化率达到经验不足员工的 4 倍,而交谈前披露会导致购买率下降超过 79.7%。Heim 和 Chan-Olmsted (2023) 的实验研究表明,公众在新闻信息收集阶段对人工智能接受度较高,而在写作阶段的接受度较低。这表明,算法在不同应用阶段的披露可能对公众接受度产生不同影响。在非决策阶段,算法主要用于数据处理,公众更倾向于将其视为辅助工具,因而接受度可能较高。在决策阶段,因算法的不确定性和风险较高,易引发公众对算法在信息处理、公正性和决策能力方面的质疑 (Nagtegaal, 2021)。这种质疑不仅源于对算法偏差和错误的担忧,还在于其缺乏伦理判断与情感考量。尤其是在涉及对社会或个人产生重大影响阶段,算法的披露可能加剧对其公正性和可靠性的担忧,从而影响公众的算法决策偏好。因此,本文提出以下研究假设。

H1: 算法披露所处的应用阶段将显著影响公众的算法决策偏好。

(三) 算法参与程度与公众偏好

算法参与程度指人工智能在决策过程中的应用深度,涵盖从辅助人类决策到完全自动化决策 (Araujo et al., 2020)。人工智能可通过提供决策建议和相关信息提升决策效率,如在医疗诊断和病理学领域,算法辅助决策不仅能提高分析速度和准确性,还能增强公众对结果的信任和接受度 (Raymond et al., 2022)。然而,算法的参与程度可能对公众对其决策的信任和接受度产生不同影响 (König et al., 2024)。Horvath (2023) 在移民签证和停车许可证实验中将人

人工智能参与度分为 25%、50%、75% 和 100%，结果表明，适度的人工智能参与会提升公众对决策的接受度，而参与度过高则会降低申请者的认可度。Huo 等（2024）则发现，在涉及种族歧视倾向的医疗决策中，过度依赖人工智能可能被视为忽视人类价值和伦理考量，引发公众抵触。这表明，人工智能在决策中的应用需要平衡不同水平的技术介入，以维持技术优势与公众接受度的最佳结合。本研究认为，较低水平的披露可能强化公众对人类主导决策的认知，而较高水平的披露则可能加剧公众对隐私、伦理和算法控制力等问题的担忧。基于此，本文提出以下研究假设。

H2：算法披露所呈现的参与程度将显著影响公众的算法决策偏好。

（四）算法责任分配与公众偏好

基于算法的决策或行动通常源自多个参与者之间互动的结果，这种分布式的代理性质导致了责任的分散。此外，算法责任的分配并非总是均匀的，由于参与方可能会歪曲他们的贡献或试图逃避责任，这使确定算法的责任归属变得异常复杂（Taddeo & Floridi, 2018）。过往研究表明，责任归属与公众对人工智能应用的态度有密切联系。Esmailzadeh 等（2021）指出，责任归属不清楚可能阻碍公众采用人工智能，尤其是在医疗领域。当人工智能与医生联合诊断时，患者对责任问题的关注显著增加。此外，Flathmann（2023）探讨了不同人工智能责任分配水平（5%、20%、35%、50%、65%、80%、95%）对团队协作的影响。研究发现，随着人工智能承担的责任变化，公众对其能力的信任程度也随之波动。当责任比例较高时，公众对算法的质疑会增加，感知的工作安全感降低。本研究认为，算法披露所呈现的责任承担水平可能影响公众对算法决策的认知和接受度。当算法承担较少责任时，公众可能更倾向于认为人类仍是主要决策者，从而增强信任感；而当算法承担较多责任时，公众可能因责任归属不清晰而感知更高的不确定性，进而降低对算法决策的信任，导致公众对算法决策的偏好降低。基于此，本文提出以下研究假设。

H3：算法披露所呈现的责任承担水平将显著影响公众的算法决策偏好。

（五）算法绩效与公众偏好

绩效信息理论（Performance Information Theory）认为，绩效信息通过影响公众对绝对绩效和相对绩效的认知，从而影响其满意度和行为（James & Moseley, 2014）。James 和 Moseley（2014）将政府垃圾回收率 58% 和 23% 分别视为高绩效和低绩效，研究发现，相比于不提供垃圾回收率信息，低绩效信息显著降低了公众对政府的绩效感知，而高绩效信息则提升了公众的感知。该理论表明，

负面绩效信息更易引发人们对组织的不满，而高绩效信息则较少带来强烈反应。在算法决策领域，算法绩效主要取决于准确性和个性化（Shin, 2020）。披露高算法绩效信息可能引发双重效应：一方面，较高的人工智能计算能力可能加剧公众的隐私担忧，增强对人工智能的威胁感，进而引发更强烈的负面态度；另一方面，展示高准确率或高可靠性能够增强个体对人工智能的信任，提升其使用意愿。Prentice 等（2020）进一步发现，人工智能绩效的多个维度（如准确性、及时性、灵活性、现实性等）均可提升人工智能质量感知，并积极影响用户满意度与参与度。相比之下，披露低算法绩效信息可能加剧公众对人工智能的不信任，因为其对人工智能决策过程缺乏了解，容易将不利结果归因于人工智能的低效或不公（Yue & Li, 2023）。基于此，本研究认为，披露低算法绩效信息会放大公众对算法不公、低效的担忧，增强其对结果的负面预期，从而降低对算法的接受度。而披露高算法绩效信息有助于缓解公众的不确定和消极感知，增强其使用算法的信心和积极性，因此对于算法决策的评估和使用意愿会产生更积极的效果。因此，本文提出以下研究假设。

H4：算法披露所呈现的绩效信息水平将显著影响公众的算法决策偏好。

（六）感知算法信任与感知算法透明度的中介作用

由于大众对人工智能算法应用的需求和依赖日益增长，确保算法决策的可信度已成为一个重要问题（Grimmelikhuijsen, 2023）。在人工智能背景下，感知算法信任被定义为用户对算法是否具备可靠性、效力以及符合人类价值观等方面的整体信念或态度（Choung et al., 2023）。感知算法信任主要涉及用户对算法在伦理和价值维度上的依赖判断。而算法决策偏好则侧重于用户在具体决策情境中选择算法参与或主导的实际意愿。换言之，信任反映用户的心理依赖，而偏好是信任在决策行为中的具体表现。

过往研究表明，感知算法信任与公众对算法决策的态度与行为密切相关。一方面，个体的感知信任能够直接影响其对算法系统的接受度。如 Stevens 和 Stetson（2023）对基于人工智能的电子健康记录应用程序进行调查，结果显示，临床医生对算法的高信任度显著提升了其使用该应用程序的意愿。另一方面，感知算法信任能够间接影响公众对算法系统的感知。如 Beldad 和 Hegner（2018）在研究影响德国用户使用算法意愿的因素时发现，对算法开发人员的信任不会直接影响人们使用健身应用程序的意愿，而增加对该应用程序的有用性感知，会对用户的使用意愿产生积极影响。基于信号理论，算法特征披露可向用户传递可靠性与透明度信息，增强其对算法的可信度判断（Keppeler, 2024）。当用户的感知信任水平较高时，更倾向于认为算法能符合其期望和价值体系，

并具备在高复杂性或高风险环境下履行承诺的能力（Grimmelikhuijsen, 2023）。这种信任有助于降低公众对算法潜在风险的担忧，尤其在关键任务或高风险情境下，是提升算法接受度的关键因素。基于上述分析，本文提出以下研究假设。

H5：感知算法信任在算法特征披露影响公众算法决策偏好过程中发挥显著的中介作用。

目前，算法透明度已经成为公共部门研究公众对算法评估的核心内容。透明并不一定意味着对公众开放，它还包括代码仅向特定参与者公开的情况，例如用于审计或认证。感知算法透明度是指用户对算法公开程度的主观体验，包括对可解释性、决策过程开放性以及数据使用的理解（Yu & Li, 2022）。仅公开决策结果可能被视为不透明，而披露结果、理由和过程可提升透明度。相比之下，算法特征披露侧重客观呈现运行机制，而感知透明度关注用户的主观理解。

随着人工智能在政府中应用的增加以及人工智能应用的复杂化，算法透明度的研究具有很强的现实意义。尽管有研究发现，过多的算法信息披露会引发公众的怀疑和担忧，并且对算法决策偏好产生负面影响（Schmidt et al., 2020）。但大量实证研究表明，适度增加算法透明度感知可以增强公众对人工智能应用的信心，从而进一步提高其对算法服务的评价和使用意愿（Shin, 2020；Yue & Li, 2023）。当实际情况与公众对于算法公平性和透明度预期不一致时，会引发公众强烈的负面情绪感知，导致算法的接受度降低（Schiff et al., 2022）。相反，当算法系统披露更多信息（提高客观透明度）时，公众的感知透明度（主观透明度）会相应提升（Yu & Li, 2022）。信息披露加深了公众对算法决策的理解，增强其监督能力，并降低其不确定性感知，从而提高公众的算法接受度。基于此，本文提出以下研究假设。

H6：感知算法透明度在算法特征披露影响公众算法决策偏好过程中发挥显著的中介作用。

三、研究设计

本研究采用联合实验的方法来检验研究假设。联合实验是基于联合分析技术，在联合列表中构建多个相互竞争的概述情境，让受访者对一组备选方案进行选择 and 评分，收集受访者数据，使用统计分析技术实现多个因素对同一结果因果效应的分析（Hainmueller et al., 2014；陈国良等, 2023）。因此，研究人员可以分析不同属性如何影响参与者的评价和偏好。与传统的实验方法相比，联合实验具有较高的外部效度，能够降低多属性、多水平研究设计的成本，并且缓解社会期望偏差（Zhang & Wang, 2024）。自 20 世纪 70 年代以来，联合分析

被广泛用于估计消费者偏好。近年来,政治学和公共管理学领域广泛采用该方法调查公民和官僚对移民、政策选择、官僚招聘及算法决策的偏好(Gaozhao et al., 2023; Horvath et al., 2023; König et al., 2024)。

本文预先登记了研究设计和调查问卷,并提交给相应的伦理审查委员会审批。本文的研究设计主要分为5个阶段:第一,通过系统回顾公共部门算法决策的现有研究,本文确定了4个影响公民算法决策偏好的关键因素作为研究重点(Flathmann et al., 2023; Horvath et al., 2023; James & Moseley, 2014; Keppeler, 2024; Luo et al., 2019; Prentice et al., 2020)。第二,基于算法决策在公共服务领域的实际应用,本文选择了线上医疗问诊作为主要的研究背景。随着人工智能技术的发展,医疗领域中的人工智能算法决策已经成为被广泛研究的场景(Angerschmid et al., 2022)。第三,基于研究内容,本文设计了一份调查问卷,包括I-Doctor在线问诊平台的背景信息、3个实验选择任务和部分个人信息问题。第四,本文对50名社区居民进行了预调查,根据他们的反馈建议,进一步完善了问卷内容。第五,本文在收集到的具有全国代表性的样本基础上,正式开展了在线联合调查实验。

(一) 样本

本研究的实验设计得到了相关伦理审查委员会的批准,所使用的数据来源于专业调查公司。本文通过配额抽样,将样本按年龄、性别和地区进行分层,以代表2020年中国人口普查中18~59岁的成年人,进一步提高了本研究样本的代表性。这一群体占中国人口的绝大部分,他们的意见能够较好地体现整个社会对于公共服务中算法决策的选择偏好。同时,为了提高数据集的统计稳健性和可靠性,降低Ⅱ型错误的风险,本研究实施了操作检查和填写时间评估,删除填写时间过长或过短的问卷。正式调查总共发放问卷1650份,剔除未能通过测试项目的无效样本,最终有效样本为1356名^①,问卷有效率为82.18%。

(二) 联合实验设计

一个典型的联合实验要求受访者从包含多种属性的假设性概述中进行选择或评分,从而使研究人员能够估算出每个属性及其水平对所产生的选择或评分的相对影响(Hainmueller et al., 2014)。首先,本研究创建了为社会公众提供医疗诊断服务的场景,具体描述如下:“‘I-Doctor’是一款利用AI技术、由全国各大三甲医院支持的线上问诊系统,免费向社会公众开放使用,能够为医疗

^① 由于篇幅原因,正文中未呈现样本信息,若读者感兴趣可以联系作者索取。

决策提供精确、快速的辅助方案，助力医生与患者实现优质医疗体验。晚上你突然感到身体不舒服，打算通过‘I-Doctor’线上问诊平台寻求病因，并期望得到有效治疗。”基于 Gaozhao 等（2023）的建议，本研究让每个参与者进行三轮实验，每一轮实验都会以表格的形式向受访者提供两种不同假设下的线上问诊平台的信息，其差异是由算法属性的不同级别组合产生的，而非情境本身的变化。参与者的任务是在两个问诊方案中选择一个他们更愿意接受的方案，并进行评分。为了构建披露效应分析框架并测量公众对算法决策的选择偏好，本研究基于过往研究文献设计了 4 类属性和相关水平：第一，算法决策阶段参考了西蒙的决策模型，用其“三阶段”结构提供了逻辑框架，并借鉴了 Luo 等（2019）对决策披露阶段的划分，将其分为信息收集、信息分析和决策制定 3 个阶段，以考察公众在不同决策环节中的态度差异。第二，算法参与程度依据 Horvath 等（2023）的划分标准，设置了从完全由人类主导（0%）到完全由算法主导（100%）的 5 个参与水平，旨在分析不同人机协作比例对公众接受度的影响。第三，基于 Flathmann 等（2023）的框架及其对应算法参与程度的划分，本文将算法担责比例细化为 5 个层次，以探究不同责任分配模式对公众信任和接受度的作用。第四，参考 Prentice 等（2020）以及 James 与 Moseley（2014）关于绩效信息的分类方法，本文将满意度百分比划分为低、中、高 3 个等级，以分析算法绩效信息如何影响公众对算法决策的偏好。表 1 详细展示了所有可能的属性、水平。值得注意的是，虽然本文对每组任务的属性水平进行了随机化，但也设置了一些约束条件，以消除不切实际的属性组合（Hainmueller et al., 2014）。

表 1 联合实验属性、水平概览

属性	水平
算法决策 阶段	①信息收集阶段：你被告知 AI 充当的是助手的角色，其功能严格限于收集病人的基本健康信息和症状
	②信息分析阶段：你被告知 AI 会独立处理患者数据，以进行初步临床评估，帮助人类医生进一步评估
	③决策制定阶段：你被告知 AI 可根据复杂的算法自主进行医疗诊断并制定治疗策略，只需极少的人工监督
算法参与 程度	①在 I-Doctor 问诊中，你被告知 AI 参与度为 0%，医生参与度为 100%
	②在 I-Doctor 问诊中，你被告知 AI 参与度为 25%，医生参与度为 75%
	③在 I-Doctor 问诊中，你被告知 AI 参与度为 50%，医生参与度为 50%
	④在 I-Doctor 问诊中，你被告知 AI 参与度为 75%，医生参与度为 25%
	⑤在 I-Doctor 问诊中，你被告知 AI 参与度为 100%，医生参与度为 0%

(续上表)

属性	水平
算法责任分配	①你被告知 AI 不承担责任，医生需要为病人承担 100% 责任
	②你被告知 AI 承担 25% 责任，医生需要为病人承担 75% 责任
	③你被告知 AI 承担 50% 责任，医生需要为病人承担 50% 责任
	④你被告知 AI 承担 75% 责任，医生需要为病人承担 25% 责任
	⑤你被告知 AI 承担 100% 责任，医生不需要承担责任
算法绩效信息	①在使用 I-Doctor 时，你没有得到任何关于该平台的绩效信息
	②在使用 I-Doctor 时，你得知人们对该平台的平均满意率为 20%
	③在使用 I-Doctor 时，你得知人们对该平台的平均满意率为 50%
	④在使用 I-Doctor 时，你得知人们对该平台的平均满意率为 90%

资料来源：作者自制。

在完成对问诊情境的选择和评分后，实验进一步测量受访者对于感知算法信任和感知算法透明度的态度。在上述评估结束时，本研究提出了几个有关受访者特征的问题，特别是关于受访者算法决策的学习和工作经历、线上医疗问诊经验以及对算法决策的了解程度。由于 1356 名受访者进行了三轮不同假设情境的实验，因此研究样本收集的观察总数是 $1356 \times 2 \times 3 = 8136$ 。

(三) 测项

1. 算法决策偏好

为了评估受访者对算法决策的偏好，本研究参照 Hainmueller 和 Hopkins (2015) 的做法，同时使用选择法和评分法来测量参与者的算法决策偏好。在每轮调查中，受访者需在两种线上问诊情境 A/B 中选择其偏好，分别赋值 1 (选择) 和 0 (未选择)，并根据 7 分李克特量表评分。参考胡赛全等 (2022) 的评分标准，评分大于 4 赋值为 1，否则为 0，生成“使用意愿”变量，用于稳健性检验。

2. 感知算法信任

本研究采用 Angerschmid 等 (2022) 的 5 个题项来测量算法决策的信任感知，参与者被要求用 7 分李克特量表回答对每项陈述的同意程度 (1 = “非常不同意”，7 = “非常同意”)，该量表的 Cronbach’s alpha 值为 0.966。

3. 感知算法透明度

本研究采用 Robert 等人 (2020) 的 4 个题项来评估公众对算法决策透明度的感知^①。参与者被要求用 7 分李克特量表回答对每项陈述的同意程度 (1 =

① 由于篇幅原因，正文中未呈现部分变量的测量信息，若读者感兴趣可以联系作者索取。

“非常不同意”，7 = “非常同意”），该量表的 Cronbach's alpha 值为 0.805。

4. 控制变量

过往研究表明，除了年龄、性别、受教育程度等人口统计学变量能够显著影响公众对算法决策的态度与行为（Gaozhao et al., 2023; Wang et al., 2023; 罗映宇等, 2023），算法决策的学习和工作经历（Janssen, 2022）、医疗服务中的算法系统使用经验（Raymond et al., 2022）以及对算法决策的了解程度（Kawaguchi, 2021）也是重要的影响因素，因此本研究将以上 3 个因素作为控制变量。其中，算法决策经验是通过询问参与者是否学习过计算机技术和相关专业课程（1 = “否”，2 = “是”），以及是否在计算机相关行业工作过（1 = “否”，2 = “是”）来进行测量。线上医疗问诊经验则是通过询问参与者是否使用过线上问诊平台（1 = “否”，2 = “是”），以及对平台的熟悉程度（1 = “完全不熟悉”，7 = “非常熟悉”）来测量。对算法决策的了解程度是通过参与者对算法系统决策了解的自评（1 = “根本不了解”，7 = “非常了解”）来测量。

5. 注意力测试

本研究采用两个综合问题来评估参与者的阅读理解能力和对实验程序的理解能力。受访者被要求识别情境中讨论的主题与算法角色，未能通过注意力测试的结果将会被剔除。

四、结果分析

（一）平均边际成分效应分析

本研究通过估算每个因子的平均边际成分效应（Average Marginal Component Effect, AMCE）来证明所提出的研究假设。根据 Hainmueller 等（2014）的研究，由于子分类和普通最小二乘法（Ordinary Least Squares, OLS）之间存在对应关系，因此 AMCE 可以使用 OLS 进行非参数识别。考虑到 OLS 在处理多维度特征组合时存在一定的局限性，而 AMCE 模型专门用于估计各特征水平对决策偏好的边际效应，能有效地揭示不同特征组合对公众偏好的平均影响，在处理联合分析数据时表现更为灵活，结果也更具解释性，因此本文使用 AMCE 模型进行分析。图 1 展示了披露算法决策阶段、算法参与程度、算法责任分配和算法绩效信息对受访者决策选择的影响（双尾 95% 置信区间， $N = 1356$ ）。借鉴 Zhang 和 Wang（2024）的做法，本研究将分析重点放在受访者的二元选择上，因为二元选择更接近真实世界的决策，而且从线性概率模型中得出的系数也更直观。

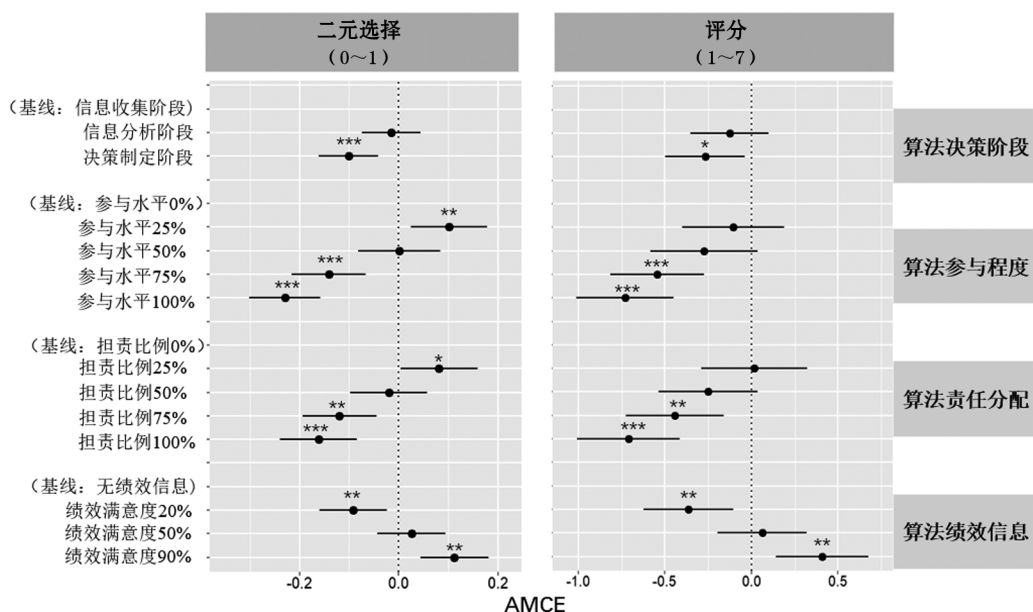


图1 算法披露对二元选择、评分的平均边际成分效应分析

注: * 代表 $p < 0.05$, ** 代表 $p < 0.01$, *** 代表 $p < 0.001$ 。

资料来源: 作者自制。

通过比较各属性水平的 AMCE 值^①, 本文可以得出以下结论: (1) 相较于披露信息收集阶段, 披露信息决策阶段 ($C = -0.101$, $SE = 0.040$, $p < 0.001$) 会显著降低公众的算法选择偏好, 其平均意愿相应降低 0.101 个单位; (2) 相较于披露算法参与水平 0%, 披露算法参与水平为 25% ($C = 0.101$, $SE = 0.039$, $p < 0.01$) 会显著提升公众的算法选择偏好, 其平均意愿相应提高 0.101 个单位。而一旦披露算法参与水平达到 75% ($C = -0.141$, $SE = 0.038$, $p < 0.001$) 乃至 100% ($C = -0.229$, $SE = 0.037$, $p < 0.001$), 则会降低公众的算法选择偏好, 其平均意愿相应降低 0.141 和 0.229 个单位; (3) 相较于披露算法责任分配为 0%, 披露算法责任分配为 25% ($C = 0.081$, $SE = 0.040$, $p < 0.05$) 会显著提升公众的算法选择偏好, 其平均意愿相应提高 0.081 个单位。而一旦披露算法责任分配达到 75% ($C = -0.119$, $SE = 0.038$, $p < 0.01$) 乃至 100% ($C = -0.162$, $SE = 0.040$, $p < 0.001$), 则会显著降低公众的算法选择偏好, 其平均意愿相应降低 0.119 和 0.162 个单位; (4) 相较于无算法绩效信息披露, 披露算法绩效满意度为 20% ($C = -0.092$, $SE = 0.035$, $p < 0.01$) 会显著降低公众的算法选择偏好, 其平均意愿相应降低 0.092 个单位。而披露算法绩效满意度达到 90% ($C =$

① 由于篇幅原因, 正文中未呈现算法披露对二元选择、评分的平均边际成分效应分析表格, 若读者感兴趣可以联系作者索取。

0.112, $SE=0.035$, $p<0.01$), 则会显著提高公众的算法选择偏好, 其平均意愿相应提高 0.112 单位。这一结果部分证实了 H1、H2、H3 和 H4。

(二) 中介效应分析

为了确保感知算法信任和感知算法透明度作为独立中介变量的合理性, 在数据分析之前, 本研究首先进行了交互效应验证。结果显示, 感知算法信任与感知算法透明度的交互项对因变量无统计显著性 ($C=0.009$, $SE=0.005$, $p>0.05$), 支持了将两者视作独立中介路径的假设。为进一步提升中介效应分析的严谨性, 本研究在使用 R 语言进行中介效应检验时, 将感知算法信任和感知算法透明度分别设为控制变量。这种做法确保了在单一中介路径分析中有效排除另一变量的潜在干扰, 使分析结果更加稳健。

1. 感知算法信任的中介效应

感知算法信任在披露算法特征影响算法决策偏好过程中的中介作用如表 2 所示。相对于算法参与水平为 0%, 算法参与水平为 75% ($C=-0.347$, $SE=0.126$, $p<0.01$) 和 100% ($C=-0.314$, $SE=0.121$, $p<0.01$) 时显著降低了公众对算法决策的信任。与此同时, 相对于无算法绩效信息披露, 算法绩效满意度为 90% ($C=0.315$, $SE=0.117$, $p<0.01$) 会显著提高受访者的感知算法信任。除此之外, 感知算法信任与公众算法接受度呈显著正相关 ($C=0.069$, $SE=0.008$, $p<0.001$)。这些结果表明, 在披露算法高水平参与和算法高绩效信息影响公众的算法选择偏好过程中, 感知算法信任发挥显著的中介作用, 而在披露算法决策阶段和算法责任分配影响公众的算法选择偏好中, 感知算法信任没有发挥显著的中介效应, 这一结果部分证实了 H5。

表 2 感知算法信任的中介效应分析

变量	模型 1: 感知算法信任		模型 2: 二元选择	
	Coefficient (SE)	p	Coefficient (SE)	p
信任			0.069 (0.008)	0.000
算法决策阶段 (基线: 信息收集阶段)				
信息分析阶段	-0.047 (0.102)	0.639	-0.011 (0.030)	0.690
决策制定阶段	-0.033 (0.102)	0.743	-0.100 (0.030)	0.000
算法参与程度 (基线: 参与水平 0%)				
参与水平 25%	-0.092 (0.132)	0.484	0.108 (0.038)	0.005
参与水平 50%	-0.123 (0.139)	0.377	0.010 (0.040)	0.812
参与水平 75%	-0.347 (0.126)	0.005	-0.117 (0.037)	0.001
参与水平 100%	-0.314 (0.121)	0.009	-0.208 (0.035)	0.000

(续上表)

变量	模型 1：感知算法信任		模型 2：二元选择	
	Coefficient (SE)	p	Coefficient (SE)	p
算法责任分配 (基线：担责比例 0%)				
担责比例 25%	0.165 (0.136)	0.226	0.070 (0.040)	0.080
担责比例 50%	0.043 (0.129)	0.742	-0.023 (0.038)	0.542
担责比例 75%	0.027 (0.126)	0.830	-0.121 (0.037)	0.000
担责比例 100%	-0.230 (0.129)	0.074	-0.146 (0.038)	0.000
算法绩效信息 (基线：无绩效信息)				
绩效满意度 20%	-0.110 (0.118)	0.351	-0.084 (0.034)	0.015
绩效满意度 50%	0.056 (0.116)	0.625	0.022 (0.034)	0.518
绩效满意度 90%	0.315 (0.117)	0.007	0.091 (0.034)	0.008

注：N=8136。
资料来源：作者自制。

2. 感知算法透明度的中介效应

感知算法透明度的中介效应分析结果如表 3 所示。与本研究的预期不完全一致，没有证据表明算法决策阶段、算法参与水平、算法责任分配会影响公众的算法透明度感知。然而，算法绩效信息显著影响了公众的算法透明度感知。同时，算法透明度与公众的算法选择偏好呈显著正相关。相对于不提供算法绩效信息，绩效满意度分别为 20%（C = 0.270，SE = 0.091， $p < 0.01$ ）、50%（C = 0.289，SE = 0.089， $p < 0.01$ ）、90%（C = 0.405，SE = 0.090， $p < 0.001$ ）时显著提高了公众对算法决策的透明度感知，而算法透明度又会进一步显著提高公众对算法决策的接受度（C = 0.051，SE = 0.010， $p < 0.001$ ）。这些数据表明，感知算法透明度在披露不同水平的算法绩效信息影响公众的算法选择偏好过程中发挥显著的中介作用，而在算法决策阶段、算法参与程度和算法责任分配影响公众的算法选择偏好中，本研究未能发现其显著的中介效应，这一结果部分证实了 H6。

表 3 感知算法透明度的中介效应分析

变量	模型 1：感知算法透明度		模型 2：二元选择	
	Coefficient (SE)	p	Coefficient (SE)	p
透明度			0.051 (0.010)	0.000
算法决策阶段 (基线：信息收集阶段)				
信息分析阶段	-0.010 (0.079)	0.894	-0.015 (0.030)	0.629
决策制定阶段	0.068 (0.079)	0.390	-0.105 (0.030)	0.000

(续上表)

变量	模型 1：感知算法透明度		模型 2：二元选择	
	Coefficient (SE)	p	Coefficient (SE)	p
算法参与程度 (基线：参与水平 0%)				
参与水平 25%	0.044 (0.101)	0.660	0.099 (0.039)	0.011
参与水平 50%	0.018 (0.107)	0.869	0.001 (0.041)	0.994
参与水平 75%	-0.159 (0.097)	0.101	-0.133 (0.037)	0.000
参与水平 100%	-0.032 (0.093)	0.735	-0.228 (0.036)	0.000
算法责任分配 (基线：担责比例 0%)				
担责比例 25%	0.116 (0.105)	0.268	0.075 (0.040)	0.063
担责比例 50%	0.009 (0.100)	0.926	-0.021 (0.038)	0.593
担责比例 75%	-0.068 (0.097)	0.481	-0.116 (0.037)	0.002
担责比例 100%	-0.067 (0.099)	0.502	-0.158 (0.038)	0.000
算法绩效信息 (基线：无绩效信息)				
绩效满意度 20%	0.270 (0.091)	0.003	-0.106 (0.035)	0.003
绩效满意度 50%	0.289 (0.089)	0.001	0.011 (0.034)	0.751
绩效满意度 90%	0.405 (0.090)	0.000	0.092 (0.035)	0.009

注：N=8136。
资料来源：作者自制。

(三) 稳健性检验

本文检验了 AMCE 估计所需的稳定性以及无延滞效应、无概述顺序效应假定，以确保 AMCE 值准确反映了算法披露属性与公众算法决策偏好的因果效应。本文根据任务数量将样本分为三组，分别估计结果并评估一致性。边际均值分析显示，受试者在三轮任务中的偏好基本一致，偏好异质性嵌套模型的 p 值 ($0.699>0.05$) 表明任务轮次间无显著差异，满足稳定性假定，排除携带效应的影响。此外，针对资料顺序效应，本文将样本按资料数量分为两组，分别估计并检验一致性 (Hainmueller et al., 2014)。边际均值分析表明，受试者在不同概述间的偏好基本一致，亚组间无显著差异，从而排除资料顺序效应的影响。这些稳健性检验增强了研究结果的可靠性和实用性。

五、讨论

本研究以信号理论和披露效应为理论基础，从不同维度的算法特征披露出

发，构建了包含算法决策阶段、算法参与程度、算法责任分配以及算法绩效信息的算法披露效应分析框架，围绕公共医疗服务场景展开联合实验设计，并基于 1356 名代表性样本所收集数据进行分析，系统探索了算法披露影响公众对公共部门算法决策偏好的因果关系，检验了感知算法信任和感知算法透明度在这一过程的中介作用机制。

（一）主要结论

第一，本研究分析了披露算法决策阶段对公众算法决策偏好的影响。研究发现，相较于披露信息收集，信息分析阶段的算法应用披露对公众偏好的影响不显著，这可能是由于信息分析阶段的操作相对隐蔽、风险感知低，且不直接影响最终决策，因此公众对此阶段的算法应用容忍度较高，负面偏好较少。而披露算法应用于决策制定则显著降低了公众算法决策偏好，这表明随着算法应用阶段的重要性提升，公众的风险感知和不确定性感知也随之增加，触发了更强的负面算法披露效应，从而导致个体对算法决策产生厌恶情绪（Dietvorst & Bharti, 2020）。该研究发现与过往研究一致，即在消费者购买和新闻生产等领域中，披露人工智能应用于关键阶段会显著影响用户行为和偏好（Heim & Chan-Olmsted, 2023; Luo et al., 2019）。此外，本文在 Keppeler（2024）关于算法披露降低求职者对职位兴趣的研究基础上进行了拓展，特别强调了在公共医疗领域的关键决策阶段披露算法应用会增加公众的不确定性和风险感知，从而引发更消极的算法接受态度。这一结论表明披露算法应用的时机对于公共部门利用算法进行决策尤为关键，恰当的信息披露策略能够有效提高公众的信任和接受度。

第二，本研究分析了披露算法参与程度对公众算法决策偏好的影响。研究发现，相较于披露算法无参与，披露低水平的算法参与会显著提高公众算法决策偏好。研究结果显示，当算法参与度达到 50% 时，其对公众偏好的影响不显著。原因可能在于，该参与水平实现了人类判断与算法自主性之间的适度配合。公众认为在这一模式下，算法既具备一定的独立决策能力，同时仍保留了人类判断的影响，更容易接受这种相互制衡的决策方式，因此不倾向于对该参与水平做出极端正面或负面的评价。而一旦算法参与水平超过人类，则会显著降低公众算法决策偏好。这表明，当决策权主要由算法承担而缺乏其他主体的参与或监督时，公众对算法决策结果的不信任感增强，导致公众较不愿意接受算法决策。一方面，本文的研究结果支持了过高的人工智能参与程度可能导致个体对人工智能决策产生抵触情绪的观点（Horvath, 2023; Huo et al., 2024; König et al., 2024）；另一方面，本文也进一步印证了合理匹配的人机协同决策能够提

供更快速、更准确的分析，进而增强个体对算法决策的接受度（Raymond et al., 2022）。因此，在公共部门应用算法决策时，应优化人机协同的程度，恰当匹配算法与人类决策者的参与，以提高决策效率和质量。

第三，本研究分析了披露算法责任分配对公众算法决策偏好的影响。研究发现，相较于算法不承担任何责任，披露算法只承担少部分责任会显著提高公众算法决策偏好。而一旦披露算法承担责任较重，则会显著降低公众算法决策偏好。这意味着公众对算法承担责任的接受程度存在一个优化点，找到合适的责任分配平衡点是关键。值得注意的是，当算法责任分配为 50% 时，对公众偏好未产生显著影响。这可能是因为算法与人类责任对半划分的平衡状态，使公众在责任归属上难以形成明确的认知和评价。这种模糊的责任界定既不引发极端的正面支持，也不导致强烈的负面反应，从而难以促成明显的偏好变化。本研究佐证了 Flathmann（2023）的发现，即随着人工智能责任的增加，参与者感知的工作安全感也会随之下降。Esmailzadeh 等（2021）的发现也与本文相吻合，即当人工智能与医生共同进行诊断时，患者对责任归属的关注显著增加。该结论也进一步表明，适度分配人工智能的责任可以增强公众对算法决策的偏好。然而，过高的算法责任可能会被认为是推卸责任，从而增加公众对算法决策的厌恶。

第四，本研究分析了披露算法绩效信息对公众算法决策偏好的影响。研究发现，相较于无算法绩效信息披露，披露低算法绩效信息会显著降低公众的算法选择偏好，而披露高算法绩效则会显著提高公众算法决策偏好。而满意度 50% 的中等绩效信息既不足以激发公众的高度信任或质疑，也不能提供足够的正面或负面刺激，因而导致对算法决策偏好的评价温和、不显著。本研究进一步证实了以下两种观点：一是低绩效信息可能会放大公众对算法能力和可靠性的疑虑，从而加剧公众的负面预期和感知（Yue & Li, 2023）。二是披露高性能指标如高准确率、高绩效、优异性等，能传达算法的可靠性和有效性，有助于缓解公众对算法决策过程和结果的消极感知（Prentice et al., 2020）。这表明，在推广算法决策过程中恰当地分享绩效信息对提升公众信任和接受度具有重要作用。

第五，本研究揭示了感知算法信任和感知透明度在算法披露影响公众算法决策偏好过程中发挥的中介作用。研究发现，高水平的算法参与披露和算法绩效披露能够通过感知算法信任的中介作用间接影响算法决策偏好。具体而言，高算法参与度可能引发公众对算法决策的可控性和伦理问题的担忧，导致信任感下降，从而降低了对算法决策的接受度。而高绩效信息的披露缓解了他们对算法失误的担忧，增强了其对算法的信任感，使其更倾向于接受算法参与的决

策。对于算法的不同决策阶段，公众更关注过程的规范性，而非信任问题；同样地，责任分配虽然涉及后果的承担，但并不直接反映算法的能力和可靠性。这种间接性和非直观性使感知算法信任在这两方面的中介效应并不显著。此外，披露不同水平的算法绩效信息能够通过感知算法透明度的中介作用间接影响算法决策偏好。绩效信息作为直观且明确的指标，直接传达了算法的质量和可信度，提升了公众对算法的透明度感知。然而在涉及决策阶段、参与水平和责任分配等复杂、技术性较强的信息时，公众很难在短时间内理解其对透明度的影响，因此这些变量在影响感知算法透明度的中介路径上表现不显著。这与已有研究一致，即个体对算法决策的不同感知，如公平感、有用性、可问责性、可解释性等因素是影响某些算法决策系统接受度的重要影响因素（Grimmelikhuijsen, 2023; König et al, 2024; 王戈、张哲君, 2023）。该结论进一步证实，在算法特征披露过程中强化感知算法信任和透明度，可以有效提升公众算法决策偏好。

（二）理论意义

本研究从广义算法披露的视角切入，系统解析了公共部门算法特征披露塑造公众决策偏好的复杂效应机制。通过识别关键决策阶段的负向披露效应以及披露内容的双重影响，本研究揭示了算法披露并非简单提升透明度即可优化公众认知，而是在特定条件下可能触发公众偏好下降的“信息披露悖论”。具体而言：第一，本研究揭示了在越关键的决策阶段披露算法应用，其对公众算法决策偏好的消极影响越明显，从而触发了显著的算法披露效应。该发现阐明了在不同决策阶段披露算法应用如何影响公众算法决策偏好，进一步深化了本文对算法披露时机以及对其激发公众算法披露效应的理解（Dietvorst & Bharti, 2020; Keppler, 2024; Tong et al., 2021）。第二，本研究探讨了不同维度的算法特征披露对公众算法决策偏好的双重影响，进一步揭示了算法披露的影响并非一律是消极的，而是根据算法特征的具体水平而变化。这些发现不仅加深了本文对算法参与程度、算法责任分配及算法绩效信息如何影响公众算法决策偏好的理解（Flathmann, 2023; Horvath, 2023; König et al., 2024; Prentice et al., 2020; Yue & Li, 2023），还强调了在公共决策中应根据具体情境调整算法披露策略，以优化公众对算法模型的理解和接受度。第三，本研究详细阐述感知算法信任和感知算法透明度的中介作用，展示了它们如何有效地连接不同算法特征披露与公众算法决策偏好之间的关系。这些发现扩展了本文对感知算法信任和感知算法透明度在影响公众接受算法决策行为中所起的作用及其边界条件的理解（Grimmelikhuijsen, 2023; König et al, 2024; Shin, 2020; 王戈、张哲君, 2023），为提升不同特征的算法系统的接受度提供了重要启示。

（三）实践启示

本文对公共部门的算法应用实践有以下启示：第一，公共部门应在算法决策的关键阶段透明地披露算法的功能和限制，以缓解公众对算法的不信任感和厌恶情绪，增强算法决策接受度；第二，应推荐采用合理的人机协同策略，平衡算法与人类决策者的参与度，以增强决策效率和公众接受度；第三，应适当分配算法与人类决策者的责任，避免过度依赖算法自动决策，以提升算法的被接受度和被信任度，减少公众对算法推卸责任的感知；第四，应公开高质量的算法绩效数据以增强公众对算法的信任和接受度，同时避免公开低质量的绩效数据，以减少公众的负面预期和对算法能力的疑虑；第五，可以通过强化高水平算法参与和绩效信息披露中的感知算法信任与感知算法透明度，增强公众对算法决策的偏好。

（四）不足和展望

本研究仍存在以下四点不足：第一，研究利用虚拟实验环境让参与者进行偏好选择，提高了外部有效性（Hainmueller et al., 2014），但这种人为创造的实验场景可能会与现实情况产生偏差，在公民偏好和实际行为之间也可能存在差距。因此，虽然研究测量了参与者的态度和看法，但它可能无法完全预测人们在现实世界中的实际行为。未来的研究可以结合实际行为数据或展开实地实验进行研究，以便更准确地探究公众对算法决策的实际行为态度。第二，本研究部分自变量在中间水平的主效应未能表现出显著影响，且在中介效应分析中，部分中介变量的显著性存在差异。这可能反映了公众在面对复杂信息或多层次特征时，对不同算法特征的认知容易产生分歧。此外，自变量水平的划分和样本特征的多样性也可能影响结果的显著性。未来研究应考虑引入更细化的变量划分，以揭示算法特征的细节层次，提升对公众偏好变化的捕捉能力。进一步的分层样本分析（如算法知识水平、技术接受度和教育背景）可以帮助识别不同群体在算法特征披露下的反应差异。同时，将更贴近现实的情境模拟或实际场景引入实验设计，减少实验室环境与现实情况之间的差异，从而提高研究的外部效度。第三，本研究聚焦于线上公共医疗诊断场景，特定场景的研究结论可能对其他公共服务领域不具有普遍适用性。未来的研究可以探索各种公共服务场景，如行政执法、社会福利、教育或城市规划等，以增强结论的外部性。第四，本研究使用特定网络平台招募参与者，可能限制了样本的多样性，并可能影响研究结果的普遍性。未来研究应囊括来自多个国家或地区、不同人工智能采用程度或政策环境各异的多样化群体，以实现更丰富、更具全球相关性的动态观察。

参考文献

- 陈国良、张书维、陈文娜 (2023). 公共管理研究中的联合实验: 文献回顾、基本原理及结果实现. 公共管理与政策评论, 12(3): 152-168.
- Chen, G. L., Zhang, S. W., & Chen, W. N. (2023). Conjoint Experiments in Public Administration Research: Literature Review, Rationale, and Outcome. Implementation. *Public Administration and Policy Review*, 12(3): 152-168. (in Chinese)
- 胡赛全、刘展余、雷玉琼、刘盼 (2022). 公众对助推型减碳政策的偏好研究: 基于联合实验与机器学习方法. 公共行政评论, 15(3): 40-61+197.
- Hu, S. Q., Liu, Z. Y., Lei, Y. Q., & Liu, P. (2022). Public Preference for Nudge-Based Decarbonization Policies: A Conjoint Experiment with Machine Learning Approach. *Journal of Public Administration*, 15(3): 40-61+197. (in Chinese)
- 罗映宇、朱国玮、钱无忌、吴月燕、黄静、杨智 (2023). 人工智能时代的算法厌恶: 研究框架与未来展望. 管理世界, 39(10): 205-233.
- Luo, Y. Y., Zhu, G. W., Qian, W. J., Wu, Y. Y., Huang, J., & Yang, Z. (2023). Algorithm Aversion in the Era of Artificial Intelligence: Research Framework and Future Agenda. *Journal of Management World*, 39(10): 205-233. (in Chinese)
- 王戈、张哲君 (2023). 任务客观性、情感相似度何以影响算法决策感知公平与接受度? ——基于调查实验的实证分析. 公共管理与政策评论, 12(6): 77-95.
- Wang, G., & Zhang, Z. J. (2023). How Task Objectivity and Emotional Similarity Affect Perceived Fairness and Acceptance of Algorithmic Decision-making? An Empirical Analysis Based on Survey Experiment. *Public Administration and Policy Review*, 12(6): 77-95. (in Chinese)
- 魏昕、黄鸣鹏、李欣悦 (2021). 算法决策、员工公平感与偏差行为: 决策有利性的调节作用. 外国经济与管理, 43(11): 56-69.
- Wei, X., Huang, M. P., & Li, X. Y. (2021). Algorithmic Decision-making, Employees' Justice Perception and Deviant Behaviors: The Moderating Role of Decision Favorableness. *Foreign Economics & Management*, 43(11): 56-69. (in Chinese)
- 吴进进、符阳 (2021). 算法决策: 人工智能驱动的公共决策及其风险. 开放时代, (5): 194-206+10.
- Wu, J. J., & Fu, Y. (2021). Algorithmic Decision: AI-driven Public Decision and the Risk It May Bring About. *Open Times*, (5): 194-206+10. (in Chinese)
- 薛澜、赵静 (2024). 人工智能国际治理: 基于技术特性与议题属性的分析. 国际经济评论, 24(3): 52-69+5.
- Xue, L., & Zhao, J. (2024). International Governance of Artificial Intelligence: Analysis Based on Technical Characteristics and Issue Attributes. *International Economic Review*, 24(3): 52-69. (in Chinese)
- 曾大军、张柱、梁嘉琦、徐楠、王刊良、杨彦武、陈伟运、徐杨、刘鹏飞、谢佳亨 (2021). 机器行为与人机协同决策理论和方法. 管理科学, 34(6): 55-59.
- Zeng, D. J., Zhang, Z., Liang, J. Q., Xu, N., Wang, K. L., Yang, Y. W., Chen, W. Y., Xu, Y., Liu, P. F., & Xie, J. H. (2021). Machine Behavior and Human-Machine Collaboration Decision: Theory and Methods. *Journal of Management Science*, 34(6): 55-59. (in Chinese)
- Acikgoz, Y., Davison, K. H., Compagnone, M. & Laske, M. (2020). Justice Perceptions of Artificial Intelligence in Selection. *International Journal of Selection and Assessment*, 28(4): 399-416.
- Angerschmid, A., Zhou, J., Theuermann, K., Chen, F. & Holzinger, A. (2022). Fairness and Explanation in AI-Informed Decision Making. *Machine Learning and Knowledge Extraction*, 4(2): 556-579.
- Araujo, T., Helberger, N., Kruijemeier, S., & De Vreese, C. H. (2020). In AI We Trust? Perceptions about Automated Decision-Making by Artificial Intelligence. *AI & Society*, 35(3): 611-623.
- Beldad, A. D., & Hegner, S. M. (2018). Expanding the Technology Acceptance Model with the Inclusion of Trust, Social Influence, and Health Valuation to Determine the Predictors of German Users' Willingness to Continue Using a Fitness App: A Structural Equation Modeling Approach. *International Journal of Human-Computer Interaction*, 34(9): 882-

- 893.
- Bonsón, E. , Bednárová, M. , & Perea, D. (2023). Disclosures about Algorithmic Decision Making in the Corporate Reports of Western European Companies. *International Journal of Accounting Information Systems*, 48, 100596.
- Castelo, N. , Bos, M. W. , & Lehmann, D. R. (2019). Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, 56(5) : 809–825.
- Choung, H, David, P. , & Ross, A. (2023). Trust In AI and Its Role in the Acceptance of AI Technologies. *International Journal of Human-Computer Interaction*, 39(9) : 1727–1739.
- Connelly, B. L. , Certo, S. T. , Ireland, R. D. , & Reutzel, C. R. (2011). Signaling Theory: A Review and Assessment. *Journal of Management*, 37(1) : 39–67.
- Dietvorst, B. J. , & Bharti, S. (2020). People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error. *Psychological Science*, 31(10) : 1302–1314.
- Esmailzadeh, P. , Mirzaei, T. , & Dharanikota, S. (2021). Patients' Perceptions Toward Human-Artificial Intelligence Interaction in Health Care: Experimental Study. *Journal of Medical Internet Research*, 23(11) : e25856.
- Flathmann, C. , Schelble, B. G. , McNeese, N. J. , Knijnenburg, B. , Gramopadhye, A. K. , & Chalil Madathil, K. (2023). The Purposeful Presentation of AI Teammates: Impacts on Human Acceptance and Perception. *International Journal of Human-Computer Interaction*, 40(20) : 1–18.
- Gaozhao, D, Wright, J. E & Gainey, M. K. (2023). Bureaucrat or Artificial Intelligence: People's Preferences and Perceptions of Government Service. *Public Management Review*, 26(6) : 1–28.
- Grimmelikhuijsen, S. (2023). Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making. *Public Administration Review*, 83(2) : 241–262.
- Hainmueller, J. , Hopkins, D. J. , & Yamamoto, T. (2014). Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices Via Stated Preference Experiments. *Political Analysis*, 22(1) : 1–30.
- Hainmueller, J. , & Hopkins, D. J. (2015). The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes Toward Immigrants. *American Journal of Political Science*, 59(3) : 529–548.
- Heim, S. , & Chan-Olmsted, S. (2023). Consumer Trust In AI-Human News Collaborative Continuum: Preferences and Influencing Factors by News Production Phases. *Journalism and Media*, 4(3) : 946–965.
- Horvath, L. , James, O. , Banducci, S. , & Beduschi, A. (2023). Citizens' Acceptance of Artificial Intelligence in Public Services: Evidence from a Conjoint Experiment about Processing Permit Applications. *Government Information Quarterly*, 40(4) : 101876.
- Huo, W. , Zhang, Z. , Qu, J. , Yan, J. , Yan, S. , Yan, J. , & Shi, B. (2024). Speciesism and Preference of Human-Artificial Intelligence Interaction: A Study on Medical Artificial Intelligence. *International Journal of Human-Computer Interaction*, 40(11) : 2925–2937
- James, O. , & Moseley, A. (2014). Does Performance Information about Public Services Affect Citizens' Perceptions, Satisfaction, and Voice Behaviour? Field Experiments with Absolute and Relative Performance Information. *Public Administration*, 92(2) : 493–511.
- Janssen, M. , Hartog, M. , Matheus, R. , Yi Ding, A. , & Kuk, G. (2022). Will Algorithms Blind People? The Effect of Explainable AI and Decision-Makers' Experience on AI-Supported Decision-Making in Government. *Social Science Computer Review*, 40(2) : 478–493.
- Kawaguchi, K. (2021). When Will Workers Follow an Algorithm? A Field Experiment with a Retail Business. *Management Science*, 67(3) : 1670–1695.
- Keppeler, F. (2024). No Thanks, Dear AI! Understanding the Effects of Disclosure and Deployment of Artificial Intelligence in Public Sector Recruitment. *Journal of Public Administration Research and Theory*, 34(1) : 39–52.
- König, P. D. , Felfeli, J. , Achtziger, A. , & Wenzelburger, G. (2024). The Importance of Effectiveness Versus Transparency and Stakeholder Involvement in Citizens' Perception of Public Sector Algorithms. *Public Management Review*, 26(4) : 1061–1082.
- Li, T. G. , Zhang, C. B. , Chang, Y. , & Zheng, W. (2024). The Impact of AI Identity Disclosure on Consumer

- Unethical Behavior: A Social Judgment Perspective. *Journal of Retailing and Consumer Services*, 76: 103606.
- Logg, J. M. , Minson, J. A. , & Moore, D. A. (2019). Algorithm Appreciation: People Prefer Algorithmic to Human Judgment. *Organizational Behavior and Human Decision Processes*, 151: 90–103.
- Luo, X. , Tong, S. , Fang, Z. , & Qu, Z. (2019). Frontiers: Machines vs. Humans: the Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. *Marketing Science*, 38(6): 937–947.
- Mergel, I. , Dickinson, H. , Stenvall, J. , & Gasco, M. (2024). Implementing AI in the Public Sector. *Public Management Review*, 1–14.
- Nagtegaal, R. (2021). The Impact of Using Algorithms for Managerial Decisions on Public Employees' Procedural Justice. *Government Information Quarterly*, 38(1): 101536.
- Prentice, C. , Weaven, S. , & Wong, I. A. (2020). Linking AI Quality Performance and Customer Engagement: The Moderating Effect of AI Preference. *International Journal of Hospitality Management*, 90: 102629.
- Raymond, L. , Castonguay, A. , Doyon, O. , & Paré, G. (2022). Nurse Practitioners' Involvement and Experience with AI-Based Health Technologies: A Systematic Review. *Applied Nursing Research*, 66: 151604.
- Robert, L. P. , Pierce, C. , Marquis, L. , Kim, S. , & Alahmad, R. (2020). Designing Fair AI for Managing Employees in Organizations: A Review, Critique, and Design Agenda. *Human-Computer Interaction*, 35(5–6): 545–575.
- Schiff, D. S. , Schiff, K. J. , & Pierson, P. (2022). Assessing Public Value Failure in Government Adoption of Artificial Intelligence. *Public Administration*, 100(3): 653–673.
- Schmidt, P. , Biessmann, F. , & Teubner, T. (2020). Transparency and Trust In Artificial Intelligence Systems. *Journal of Decision Systems*, 29(4): 260–278.
- Shin, D. (2020). User Perceptions of Algorithmic Decisions in the Personalized AI System: Perceptual Evaluation of Fairness, Accountability, Transparency, and Explainability. *Journal of Broadcasting & Electronic Media*, 64(4): 541–565.
- Stevens, A. F. , & Stetson, P. (2023). Theory of Trust and Acceptance of Artificial Intelligence Technology (TrAAIT): An Instrument to Assess Clinician Trust and Acceptance of Artificial Intelligence. *Journal of Biomedical Informatics*, 148: 104550.
- Taddeo, M. , & Floridi, L. (2018). How AI Can Be a Force for Good. *Science*, 361(6404): 751–752.
- Tong, S. , Jia, N. , Luo, X. , & Fang, Z. (2021). The Janus Face of Artificial Intelligence Feedback: Deployment Versus Disclosure Effects on Employee Performance. *Strategic Management Journal*, 42(9): 1600–1631.
- Wang, G. , Guo, Y. , Zhang, W. , Xie, S. , & Chen, Q. (2023). What Type of Algorithm Is Perceived as Fairer and More Acceptable? A Comparative Analysis of Rule-Driven Versus Data-Driven Algorithmic Decision-Making in Public Affairs. *Government Information Quarterly*, 40(2): 101803.
- Yang, Y. , & Shi, X. (2022). Disclose the Bot: Chatbot Disclosure and Anxiety in Customer-Chatbot Interactions(*PACIS 2022 Proceedings*). Taipei/Sydney Virtual Conference: Asia Conference on Information System.
- Yu, L. , & Li, Y. (2022). Artificial Intelligence Decision-Making Transparency and Employees' Trust: The Parallel Multiple Mediating Effect of Effectiveness and Discomfort. *Behavioral Sciences*, 12(5): 127.
- Yue, B. , & Li, H. (2023). The Impact of Human-AI Collaboration Types on Consumer Evaluation and Usage Intention: A Perspective of Responsibility Attribution. *Frontiers in Psychology*, 14: 1277861.
- Zhang, Y. , & Wang, H. (2024). The Origins and Consequences of Administrative Burdens in Mass Immunization Programs: Experimental Evidence Based on the Monkeypox Outbreak. *Governance*, 37(3): 947–967.

责任编辑：张书维

英文目录与摘要

JPA Journal of Public Administration, Vol. 18 No. 4, 2025

●SYNOPSIS: Artificial Intelligence and Government Trust

When AI Becomes a “Public Servant”: The Impact of Digital Humans Types in Government Services on Government Trust Shuting Li & Qian Qiu

Abstract In the context of digital empowerment in government services, AI digital humans, as a new type of intelligent agent, are applied in diverse government service scenarios, contributing to the improvement of service quality and efficiency. However, the question of which type of government service digital human can foster greater public trust in the government, and how different types of digital humans influence public trust through various mechanisms, remains largely theoretical. Based on the theoretical models constructed by the theory of mind perception and trust transfer theory, this study explores the mechanism by which the type of government service digital person (emotion-oriented and task-oriented) influences public trust in government through three experimental studies. The study found that: (1) compared to task-oriented government service digital persons, emotion-oriented government service digital persons can lead to higher levels of public trust in government. (2) there is no significant difference in the public's perception of competence between the two types of government service digital persons; however, compared to emotion-oriented government service digital persons, the public has lower warmth perception of task-oriented government service digital persons. Warmth perception plays a mediating role in influencing public government trust through different types of government service digital persons. (3) the public's past experiences of government-citizen interactions (positive experiences and negative experiences) play a moderating role in influencing warmth perceptions and competence perceptions of different types of government service digital persons. This study not only enriches the research related to digital person types and government trust in government service scenarios, but also provides practical guidance for government departments to balance professionalism and emotional interaction in the process of deploying digital persons for government services, emphasizing the need to balance technological efficiency and humanistic care in the process of digital transformation of government services, to ensure the continuous optimization and effectiveness of government services.

Key Words Digital Humans; Government Trust; Warmth Perception; Competence Perception; Experiments

Do Citizens Trust “Humans” or “Artificial Intelligence” More in the Public Sector? A Survey Experiment Study Considering Task Complexity and Public Domains

..... Siqi He & Bingsheng Liu

Abstract The rapid development of artificial intelligence (AI) is reshaping the operational models and functional scope of public administration. However, theoretical frameworks and empirical studies lag behind practical applications. In recent years, the widespread adoption of AI in public administration has sparked considerable controversy, with growing debates on its value and potential risks. At present,

systematic investigations into citizens' attitudes toward AI applications in public administration remain limited. This study employed a 2×2×2 experimental survey to examine the differences in citizens' trust toward human officials and AI in public sector scenarios, while analyzing the moderating effects of task complexity and domain values. Results show that citizens' generally exhibits higher levels of trust in human officials. Further analysis reveals that in contexts involving solemn values such as life and safety, trust in human agents significantly exceeds that in AI; whereas in domains concerning secular values such as money and material interests, the trust gap is not statistically significant. These findings provide a theoretical foundation for the scientific deployment of AI in public administration and offer practical guidance for its effective governance.

Key Words Artificial Intelligence; Citizen Trust; Task Complexity; Public Domains

●ARTICLES

How Does Algorithmic Feature Disclosure Affect Public Decision-Making Preferences? An Empirical Analysis Based on a Conjoint Experiment Ge Wang, Li Xiang & Zhejun Zhang

Abstract Algorithmic decision-making has effectively enhanced the scientific and precise nature of government decision-making, but it has also aroused public concerns about decision-making bias, privacy invasion, ethical dilemmas, and monitoring failures. How to effectively implement algorithmic disclosure to enhance the public's preference for public sector algorithmic decision-making has become an important challenge for the current government. In this paper, we analyse the causal effect of different dimensions of algorithmic disclosure on the public's preference for algorithmic decision-making based on a representative sample of 1, 356 people and a conjoint experimental design around a public healthcare service scenario with the theoretical support of signaling theory and disclosure effect, and examine the mediating mechanism of perceived algorithmic trust and perceived algorithmic transparency in this process. It was found that the more critical the decision-making stage of disclosing algorithmic applications, the more pronounced its potential negative disclosure effect. In contrast, disclosing different levels of algorithmic engagement, algorithmic responsibility assignment, and algorithmic performance information produced a complex dual effect. Perceived algorithmic trust mediated high levels of algorithmic engagement, and algorithmic performance disclosure, while perceived algorithmic transparency mediated different levels of algorithmic performance information disclosure. This study adds to the research on the impact of algorithmic feature disclosure on the public's algorithmic decision-making preferences and provides countermeasure suggestions for the effective implementation of algorithmic disclosure in the public sector.

Key Words Algorithmic Decision-making; Disclosure Effect; Perceived Algorithmic Transparency; Perceived Algorithmic Trust; Conjoint Experiments

Digitalization and Changes in Multi-level Governance Modes: A Case Study of Air Pollution Governance Jing Chen, Can Xu & Wei Hong

Abstract The different distribution and combination of control rights at different administrative levels have shaped multi-level modes of governance in China. Based on field observations and interviews on air pollution control in DM City from 2018 to 2020, this study finds that the deep embedding of digital technology in air pollution governance has reduced the information asymmetry between multi-level governments, alleviated the difficulty of supervising the lower-level departments by the upper-level departments, compressed the negotiation space between the superiors and the subordinates. These changes have moved the control right up along the three dimensions. In the past, this tight-coupling

mode was usually accompanied by a highly mobilized governance mechanism. However, the application of digital technology normalizes the mobilized mode, making it possible to maintain the tightly coupled mode for a long time. As a result, the governance mode of air pollution has evolved from the traditional administrative contracting mode and loose-coupling mode to a routine mode of tight-coupling, which has greatly reduced the flexibility of governance to a certain extent. By observing the process of introducing digital technology into the governance mechanism, this study reveals the adjustments and changes of the multi-level modes of governance.

Key Words Multi-level Modes of Governance; Control Right; Digital Technology; Air Pollution Control

Why Is the Rural Digital Transformation Is Suspending? Case Studies Based on the Policy Implementation Process Liping Fu, Kairong Kang, Huachao Feng & Yongqing Dong

Abstract Rural digital transformation provides a pathway to modernise agriculture and rural areas. Previous mainstream research has attributed the suspension of rural digital transformation to the lack of farmers' digital literacy and weak rural digital infrastructure. The paper shows, through a stressful case study of goals, conditions and change, that the suspension of rural digital transformation from central policy design to local policy implementation to rural policy effectiveness is essentially due to the governance tension of the hourglass structure in the policy implementation process, which creates 'Softening of Grassroots Institutions'. This means that under the logic of balancing efficiency and effectiveness in policy implementation, the national institutional system has been expanded and alienated in the process of sinking down to the grassroots level and landing in rural areas, and the effectiveness of institutional constraints has been reduced. The study finds that the 'Softening of Grassroots Institutions' triggers the suspension of the digital transformation of the countryside through three logics, namely, the logic of goal advancement of the order of power, downward pressure on tasks, and the administrativeisation of governance; the logic of condition creation of the legitimacy of tasks, procedural compliance, and the marginalisation of small farmers, and the logic of reversal of pressure, conflict of pressures, and the tacticalisation of goals, and the logic of change outcome of reverse pressure, pressure conflict and goal strategisation. From the perspective of the whole process of policy implementation, the paper responds to the question of 'why rural digital transformation is suspending', which helps to expand and deepen the theoretical understanding of rural digital transformation in the local context of China.

Key Words Rural Digital Transformation; Digital Formalism; Policy Implementation; Softening of Grassroots Institutions; Basic Practical Conditions

The Multiple Logics Behind Resource Allocation by Grassroots Governments: A Case Study of County A in Shandong Province Zongfeng Sun, Jingqi Lin & Xiaohan Zhang

Abstract How grassroots governments allocate various fiscal resources is of great significance to regional industrial development and rural revitalization. This study builds a framework of multiple logics to explain these allocation behaviors, focusing on three dimensions: institutional pressure, interest structures, and social relationships. Based on field observations and comparative case studies in County A, the research identifies three distinct logics in resource allocation: a mandatory logic driven by rigid institutional rules, a performance-oriented logic influenced by administrative goals, and a relational logic shaped by local social hierarchies. These logics often overlap and interact in practice. Specifically, the mandatory logic plays a dominant role in allocating both universal and targeted

resources, especially the former. The performance-oriented and relational logics mainly affect how targeted resources are distributed.

Key Words Grassroots Resource Allocation; Multiple Logics; Project-Based System

How Does the Policy Mix Intensity Affect Policy Effectiveness? A Study Based on Air Pollution Control Policy Mix in China Xiaojie Zhang & Xiaoyu Liu

Abstract An appropriate arrangement of policy mix intensity is both the theoretical focus of policy design and a practical challenge in achieving high policy effectiveness. This paper takes China's air pollution control policy mix as a research sample, measuring the intensity of policy mix by using the adapted Index of Policy Activity. Based on the panel data of 31 provincial-level administrative regions from 2010 to 2022, it empirically analyzes the impact and mechanism of the policy mix intensity on policy effectiveness. The findings indicate that the intensity of the air pollution control policy mix exhibits a non-linear inverted U-shaped relationship with policy effectiveness. Both the resource effect and information effect serve as significant mechanisms through which the intensity of the policy mix influences policy effectiveness. The heterogeneity analysis reveals that the inverted U-shape relationship is more significant in the eastern and central regions, while a positive linear correlation is observed in the western region. In areas characterized by a lower degree of marketization, the effect of the intensity of policy mix on policy effectiveness is more remarkable. Moreover, the centrality of regional cooperation networks can significantly influence this relationship. Enhancing the centrality of these networks has the potential to reverse the inverted U-shaped relationship. The analytical results not only contribute to a better understanding of the effects and mechanisms of policy mix intensity on policy effectiveness across different contexts, but also offer valuable insights for improving policy mix design and ensuring the sustainable enhancement of policy effectiveness.

Key Words Policy Mix Intensity; Policy Effectiveness; Air Pollution Control Policy; Resource Effect; Information Effect

Can Social Organizations Development Effectively Alleviate the Income Gap among Residents? An Empirical Test Based on the Threshold Effect of Government Intervention

..... Lei Liu, Xinya Zhang & Shuyi He

Abstract Various types of social organizations play an active role in income distribution through diverse means, such as assisting disadvantaged groups, promoting employment and entrepreneurship, and protecting labor rights. However, most existing studies focus on the role of social organizations in the third distribution, with limited empirical examination of their actual effects on the income gap among residents. Using panel data from 31 provinces (excluding Taiwan Province, HongKong Special Administrative Region, and Macao Special Administrative Region) in China between 2005 and 2021, this paper empirically investigates the impact of social organizations development on the income gap among residents and explores the threshold effect of government intervention through panel benchmark regression and threshold regression models. It is found that the current development of social organizations has failed to curb the persistent widening of the income gap. However, increased government intervention can reverse this effect. When the intervention level is below the threshold of 0.2662, social organizations have failed to contain the widening income gap. Once this threshold is exceeded, social organizations contribute significantly in alleviating the income gap. Additionally, through heterogeneity analysis of the impact of social organizations on income gap by type and region, it is found that foundations, social service organizations, and social organizations in Northeast China have

not effectively curbed the widening income gap. The underlying reason lies in their early-stage development, functionally concentrated primarily on the third distribution while not yet engaging in primary distribution or in secondary distribution. The paper reveals the important role of government intervention in the process of income gap regulation by social organizations, further enriches polycentric governance theory, and provides new perspectives on the exploration of the co-governance model of “an active government and a dynamic society”.

Key Words Social Organizations Development; Income Gap; Government Intervention; Threshold Effect

Do Fiscal Vertical Imbalances Reduce Local Government Attention to Public Services?

..... Yuan Zhang & Lijun Liu

Abstract Achieving equalization of basic public services is a necessary step towards common prosperity, and the supply of basic public services is closely related to the fiscal situation. Based on the panel data of 31 provinces (excluding Taiwan Province, HongKong Special Administrative Region, and Macao Special Administrative Region) in China from 2007 to 2020, and based on the current situation of China’s fiscal decentralization system, this paper analyzes the text of the government’s work report and constructs a system of attention measurement indicators. It systematically examines the mechanism through which fiscal vertical imbalance affects local governments’ attention to public services, both theoretically and empirically. The study finds that the fiscal vertical imbalance has a significant inhibitory effect on local governments’ attention to public services. Furthermore, the higher the fiscal autonomy enjoyed by local governments, the weaker this inhibitory effect becomes. The negative impact of fiscal vertical imbalance on local governments’ attention to public services exhibits regional and environmental heterogeneity, as well as an increasingly prominent single-threshold effect. This paper provides new research ideas for clarifying local governments’ attention to public services through the new perspective of quantifying government attention, which is of practical significance for weakening the negative impact of fiscal vertical imbalance and promoting the realization of the goal of equalization of basic public services.

Key Words Fiscal Vertical Imbalances; Public Services; Attention; Fiscal Autonomy

●THEORETICAL REVIEWS

Cutback Management under Fiscal Pressure: A Literature Review

..... Xin Chen

Abstract When public organizations face reduced fiscal and budgetary resources, besides seeking traditional ways to increase financial revenues, how to achieve effective fiscal savings and control budget expenditures through cutback management has become an increasingly important research topic in the field of public administration. This article reviews the academic literature on cutback management both domestically and internationally, discussing the inherent characteristics, strategies, decision-making mechanisms, and impacts of cutback management, aiming to reveal the theoretical and practical progress in this field. The research finds that although the theory of cutback management has constructed a clear and content-rich system, there are still several areas that require further study.

Key Words Cutback Management; Budget Cutback; Public Financial Management; Fiscal Sustainability

公共行政评论

双月刊, 2008年创刊
第18卷, 第4期 (总第106期)
2025年8月15日出版

Journal of Public Administration
Bimonthly, Since 2008
Vol.18 No.4
Published in August 2025

主管单位 中华人民共和国教育部
主办单位 中山大学
协办单位 教育部人文社会科学重点研究基地-
中山大学中国公共管理研究中心
广东省行政管理学会

社 长 肖 滨
主 编 朱亚鹏

联系电话 020-84113029 020-84038746

电子邮件 jpachina@163.com

编辑出版 《公共行政评论》编辑部
(广州新港西路135号; 邮编510275)

印 刷 广州一龙印刷有限公司

国内发行 广东省报刊发行局

国外发行 中国国际图书贸易总公司
(北京399信箱)

传 真 020-84111478

网 址 <http://jpa.sysu.edu.cn>

Administrator Ministry of Education of the People's Republic of China

Sponsors Sun Yat-sen University

Supporters Centre for Chinese Public Administration Research, Sun Yat-sen University
Guangdong Public Administration Society

President: Bin Xiao

Editor in Chief: Yapeng Zhu

Tel: 86 20 84113029 86 20 84038746

Fax: 86 20 84111478

Email: jpachina@163.com

Website: <http://jpa.sysu.edu.cn>

Edited by Editorial Office of *Journal of Public Administration*

(NO.135 Xin Gang Xi Road, Guangzhou, China. 510275)

Distributed by China International Book Trading Corporation

(P.O.Box 399, Beijing, China)



刊号 ISSN1674-2486
CN44-1648/D

邮发 国内46-364
代号 国外BM8839

国内外公开发行
国内定价: 20.00元

ISSN 1674-2486



9 771674 248258