

公共部门人工智能决策的主体构成及责任机制

杨明旭, 曹淑芹

内蒙古大学, 内蒙古 呼和浩特 010021

DOI: 10.61369/VDE.2025280021

摘要 : 随着人工智能深度嵌入公共部门, 传统行政决策的权责体系正面临重塑, 责任模糊、问责失灵等问题逐渐凸显。本文基于数字治理背景, 构建“类型—主体—问责”分析框架, 从替代型、辅助型与人机协同三类应用场景, 梳理技术生产者、决策行为者与监督问责者在责任链条中的角色差异与边界冲突。在此基础上, 从责任界定、责任评估与责任追究三个环节提出系统化治理路径, 强调坚持“最终责任归属于政府”的原则, 通过算法透明、责任清单制度与动态评估机制构建全过程问责体系, 以促进人工智能在公共领域实现规范化、可信化发展。研究旨在为责任型政府建设提供理论支撑与制度参考。

关键词 : 人工智能; 公共决策; 问责机制; 数字治理

Subject Composition and Responsibility Mechanism of Artificial Intelligence Decision-Making in Public Sectors

Yang Mingxu, Cao Shuqin

Inner Mongolia University, Hohhot, Inner Mongolia 010021

Abstract : With the deep integration of artificial intelligence (AI) into public sectors, the power and responsibility system of traditional administrative decision-making is facing reshaping, and problems such as ambiguous responsibility and accountability failure have gradually become prominent. Based on the background of digital governance, this paper constructs a "type-subject-accountability" analytical framework. From three application scenarios—substitutional, auxiliary, and human-machine collaborative—it sorts out the role differences and boundary conflicts of technology producers, decision-makers, and supervision-accountability subjects in the responsibility chain. On this basis, it proposes a systematic governance path from three links: responsibility definition, responsibility assessment, and responsibility pursuit. It emphasizes adhering to the principle that "the ultimate responsibility lies with the government", and constructs a whole-process accountability system through algorithmic transparency, a responsibility list system, and a dynamic assessment mechanism, so as to promote the standardized and credible development of AI in the public domain. The research aims to provide theoretical support and institutional reference for the construction of a responsible government.

Keywords : artificial intelligence (AI); public decision-making; accountability mechanism; digital governance

引言

本文在数字治理背景下, 围绕人工智能在公共部门决策中的实践类型、主体构成及责任机制, 提出了一个三维模型: “类型—主体—问责”分析框架, 该框架不仅关注人工智能本身的技术属性, 更将其置于行政过程与制度嵌套的语境中, 探讨不同责任主体在协同治理中可能面临的责任模糊、角色冲突及治理盲点, 为构建中国特色数字时代的责任型政府提供理论支撑与政策建议。

一、公共部门人工智能决策的类型

与传统决策支持系统不同, 人工智能决策具备自主学习和演进的能力^[5]。随着人工智能技术在公共领域的广泛应用, 其与公共

部门决策的关系日益复杂。目前, 公共部门人工智能决策的应用场景主要包括公共服务咨询与服务、风险预警、行政流程自动化与交通城市管理等。根据其介入决策过程的程度与方式, 公共部门人工智能决策大致可以分为三种类型: 替代型、辅助型与人机

作者简介: 杨明旭, 内蒙古大学公共管理学院硕士, Email: 384856075@qq.com;

通讯作者: 曹淑芹, 内蒙古大学公共管理学院教授。

协同。

(一) 替代决策

替代决策是指人工智能系统完全代替人类作出某些政策或行政决策，由机器自动判断并执行，无需人工干预。替代决策领域的学者认为，随着科技发展，人工智能能够独立完成此前仅由人类承担的决策任务^[9]。其在公共领域的典型应用包括行政审批自动化、交通执法自动判断、福利资格评估与税务风险预警等。影响替代决策的因素主要有三方面：其一，从任务特征看，人工智能更适合结构化程度高、规则明确、变量可量化且高频、重复性强的任务，因此在流程性审批、税务风险初审中得到广泛应用^[7]。其二，技术成熟度决定其可替代程度，模型的准确性、稳定性与可解释性直接影响公共部门的信任水平。若系统为黑箱模型，难以被监管与公众理解，其在高敏感性决策中的应用将受限^[8]。其三，数据质量与可获取性是关键前提，数据若存在偏差、缺失或更新滞后，将影响决策的可靠性与公平性。在涉及法律裁量或道德判断等领域，人工仍需作为最终把关^[9]。

在效果层面，替代决策显著提升了效率与处理能力。人工智能能够高速处理海量信息，实现实时分析与综合判断，缩短决策周期并减少人为延迟^[10]。同时，其依赖既定算法与规则，有助于提升决策一致性与标准化水平，避免因人类主观偏差或情绪波动导致的差异，从而增强政策执行的公平性与可预测性，尤其适用于税收征管、社会保障分配等领域。此外，替代决策还能降低行政成本与人力依赖，减少重复性工作的人力投入，使公共资源转向更具战略性和创造性的治理任务。

(二) 辅助决策

辅助决策是指人工智能系统为人类提供数据分析、结果预测或选项建议，但最终决策由人类做出。辅助决策是公共部门人工智能决策的主要类型。在公共领域，其主要应用包括政策效果仿真、公共卫生趋势预测、应急资源调配、城市规划方案评估与预算分配建议等场景。公共问题往往具有复杂性、模糊性与多目标性，传统方法难以应对多源数据与快速变化的需求，而机器学习与自然语言处理等技术在大规模数据分析、模式识别和情景预测方面具有优势，可为决策者提供更全面、可视化的信息支持。人工智能系统的可解释性与可控性是其进入决策流程的前提。辅助决策能显著提升信息处理的深度与广度。人工智能可整合结构化数据与非结构化文本、图像、传感器数据等，识别模式并预测趋势，使决策者获得更精确的情境认知。同时，算法建模与模拟推演能够为多种政策提供量化评估和外溢效应预测，在公共卫生、城市规划、气候治理等领域增强政策的科学性与稳健性。与全自动决策不同，辅助决策允许结合法律框架、社会规范与政治现实进行裁量，不仅维护合法性与社会接受度，也在技术失灵或数据偏差时保留人类干预的安全阀。

(三) 人机协同

在公共领域，人工智能处理大规模数据、识别趋势并提供方案，人类则进行价值取舍并承担责任。由于公共事务涉及高度不确定性、价值冲突与合法性诉求，人工智能难以独立胜任，因此人机协同体现的是“人类决策增能”而非“去人化”。实现有效

协同需构建系统机制：其一，明确角色分工，人工智能侧重信息处理，人类负责价值与社会判断；其二，嵌入人机共议机制，通过智能生成多方案，人类择优决策；其三，建立双向监督，人类监督算法透明度与偏见，人工智能也可对人类行为提供预警。然而，实现人机协同决策仍存挑战。首先，责任边界模糊，若因算法导致错误，责任主体界定尚无规范。其次，数据偏见与算法歧视可能固化社会不公。再者，部分基层政府缺乏技术理解，易导致依赖或形式化操作。未来发展将呈三趋势：一是从“技术替代”向“能力增强”转变，聚焦提升公共治理效能；二是协同制度化与标准化，包括伦理审查与责任规范；三是以人为本的架构设计，使技术运行始终以公共利益和社会价值为导向，强化参与性、问责性与合法性。

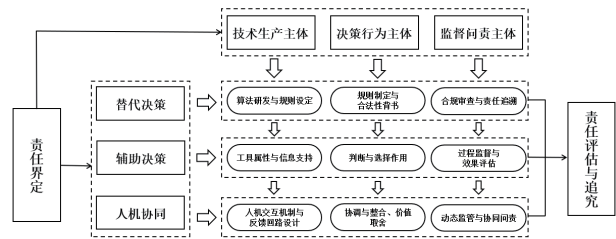


图1: 公共部门人工智能决策的责任机制

二、公共部门人工智能决策的责任机制

(一) 责任界定

责任界定是整个责任机制逻辑链条的起点与基石，不仅是技术治理的前提，也是实现责任评估与责任追究的基础。其目的在于在人工智能参与或自主决策导致不利后果时，清晰回答“谁应为事负责”这一根本问题。如图1，责任界定远非简单的因果归因，而是一个融合了法律、行政与伦理的多维度构建过程。它旨在建立一条从决策产出到责任主体的可追溯链条。在公共部门中，责任界定要求超越简单的技术故障排查，转而构建一个涵盖技术开发者、决策行为者与监督问责者的综合性责任分配框架。清晰的责任划分是进行有效责任评估与追究的前提，也是重建并维持公众对数字化政府信任的基石。责任界定在实践中仍面临诸多挑战。一方面，算法的复杂性与“黑箱”特征使得责任边界模糊；另一方面，多元主体的交织尤其是公共部门依赖外部企业提供技术支持的情况，使责任链条延长，界定难度增加。此外，随着技术迭代与应用扩展，原有的责任划分可能失效，因此有必要建立动态调整与更新机制。公共部门在责任界定中应坚持“决策最终责任归属政府”的原则，确保人工智能不会削弱公共权力的问责基础。

(二) 责任评估

在清晰界定相关责任主体之后，公共部门人工智能决策责任机制的下一关键环节是对其责任进行系统、公正的评估。在公共部门人工智能决策中，责任评估是责任机制的核心环节，其主要任务在于通过科学合理的标准与方法，对责任界定所确定的各主体履行职责的情况进行系统性判断和衡量。责任评估不仅是责任

追究的前置条件，更是公共部门在治理实践中保障人工智能技术合法、合规与高效运行的重要支撑。它要求摆脱单一的结果导向思维，构建一个融合合规性、过程性与因果关联性的复合评估框架。为了能让责任对位落到实处，需要建立贯穿事前、事中、事后的综合评估机制。有效的责任评估不仅为问责提供依据，更是公共部门从人工智能应用实践中学习迭代和优化其治理能力的关键过程。

责任评估的实现需要借助多元化的方法与工具。从形式上看，可通过内部审计、外部监督与独立评估相结合的方式，确保责任评估的客观性与权威性；从技术上看，可以依靠算法审计、可解释性分析与风险评估工具，对人工智能的运行逻辑与潜在风险进行追踪与检验；从程序上看，应建立透明的信息披露制度与公众参与机制，使责任评估过程公开化、民主化，从而增强其合法性与社会认可度。

（三）责任追究

在完成责任界定与评估后，公共部门人工智能决策责任机制的最终环节在于执行责任追究。责任追究是指依据评估结论，对经认定未能履行其应负责任的主体施加正式或非正式的否定性后果，以修复损害、实现矫正正义并震慑未来不当行为的过程。它是责任机制威慑力与有效性的最终体现，但其设计必须审慎，以避免抑制技术创新或导致有失公允的“替罪羊”现象。责任追究是公共部门人工智能决策责任机制能否从纸面走向现实的关键。

一个有效的追责体系必须是多层次、多形式且精准的，它应能同时覆盖公共部门内部与外部，并综合运用行政或法律等多种手段。然而，责任追究在实践中同样存在难点。一方面，人工智能系统的复杂性导致责任因果关系不易认定，使得追责过程可能陷入僵局；另一方面，公共部门与技术企业之间的责任边界若不清楚，容易出现责任转嫁或问责落空的现象。此外，过于严苛或频繁的追责可能使公务人员过度规避人工智能应用，反而抑制技术创新与治理效能提升。

三、结语

当前，人工智能参与公共部门决策领域逐渐拓展。在借助人工智能提高决策效率的同时，应看到数据、算法等可能存在的责任危机隐患。人工智能技术对行政决策的控制越深，就越有可能产生严重的后果，使政府部门频繁陷入责任归属困境之中。在技术生产主体、决策行为主体、监督问责主体三方面共同努力下，推进“以善治促善智”的使命任务，减少人工智能参与决策时的风险。无论如何，人工智能技术都不是万能药，分级使用的弱人工智能和强人工智能都只是一种有效的工具，在人工智能做出的决策中，还是先由人类进行选择，人类始终拥有自主权，公共管理理论发展以及行政决策的制定与执行最关键的推动者仍是人类本身。

参考文献

- [1] 刘育猛, 徐明华. 人工智能与公共决策融合研究: 趋势·风险·实践路径[J]. 常州大学学报(社会科学版), 2019, 20(05): 99-107.
- [2] Kim B, Doshi-Velez F. Machine learning techniques for accountability[J]. AI Magazine, 2021, 42(1): 47-52.
- [3] 汪庆华, 胡临天. 生成式人工智能责任机制的技术与法律建构[J]. 中国法律评论, 2024, (04): 114-130.
- [4] Veale M, Edwards L, Eyers D, et al. Automating Data Rights[J]. Dagstuhl Reports, 2018, 8(4): 126-163.
- [5] 张亚莉, 李辽辽, 丁振斌. 组织管理中的人工智能决策: 述评与展望[J]. 外国经济与管理, 2024, 46(10): 18-38.
- [6] 郭一帆. 人工智能驱动下数字政府自动化行政的挑战与应对[J]. 牡丹江大学学报, 2025, 34(01): 8-18.
- [7] 陈佩, 李晓曼. 人工智能的决策替代与社会缓冲——以X智能电厂为例[J]. 社会学研究, 2024, 39(01): 156-179+229.
- [8] 张康之. 人工智能替代的“能”与“不能”[J]. 中共杭州市委党校学报, 2025, (03): 4-12+2.
- [9] 张宸瑜. 基于公共性算法制度建构的新质生产力治理——以新一代生成式人工智能参与数字政府建设为例[J]. 南京审计大学学报, 2024, 21(04): 97-111.
- [10] 梁健. 人工智能运用于社会公共行政管理创新的优势、领域与路径探析[J]. 兰州学刊, 2024, (03): 85-96.