

人工智能技术在公共管理中的伦理风险及治理对策研究

黄淇梵

广州新华学院(中山大学新华学院), 广东 广州 510000

DOI:10.61369/HASS.2025030002

摘要： 本研究聚焦人工智能技术在公共管理中的伦理风险与治理。研究发现，人工智能在政策制定、公共服务与社会治理等领域应用广泛，显著提升管理效率，但同时引发数据隐私、算法歧视、责任界定模糊等伦理风险。数据隐私风险体现在过度收集、泄露及不当使用；算法歧视源于设计偏见与数据偏差，影响决策公平；责任界定模糊因主体权责不清与追溯机制缺失。针对这些问题，从政策法规细化责任条款与加强数据监管、技术手段提升数据保护与算法透明度、社会监督构建多元主体协同机制等方面提出治理对策，为保障人工智能在公共管理领域的健康发展提供参考。

关键词： 人工智能；公共管理；伦理风险；治理对策

Ethical Risks and Governance Countermeasures of Artificial Intelligence Technology in Public Management

Huang Qifan

Guangzhou Xinhua University (Xinhua College of Sun Yat-sen University), Guangzhou, Guangdong 510000

Abstract： This study focuses on the ethical risks and governance of artificial intelligence technology in public management. The research finds that artificial intelligence is widely used in policy making, public services, and social governance, significantly improving management efficiency. However, it also raises ethical risks such as data privacy, algorithmic discrimination, and ambiguous responsibility definitions. Data privacy risks are reflected in overcollection, leakage, and improper use; algorithmic discrimination originates from design bias and data deviation, affecting decision fairness; ambiguous responsibility definitions are due to unclear subject rights and responsibilities and the absence of traceability mechanisms. To address these issues, governance countermeasures are proposed in terms of refining responsibility clauses in policies and regulations, strengthening data supervision, enhancing data protection and algorithm transparency through technical means, and constructing a collaborative mechanism for multiple entities through social supervision. These provide references for ensuring the healthy development of artificial intelligence in the field of public management.

Keywords： artificial intelligence; public management; ethical risks; governance countermeasures

引言

自20世纪中叶起步，人工智能技术在计算能力提升与数据量激增的推动下，凭借深度学习、自然语言处理等关键技术的突破，迎来发展黄金期，并广泛渗透至公共管理领域。在政策制定环节，其强大的数据分析与预测能力，能挖掘海量政策信息，为决策提供科学依据，优化政策方案；在公共服务层面，智能客服、智能交通系统等应用，显著提升服务效率与质量；于社会治理领域，图像识别与大数据分析技术助力实现风险监控预警，提升治理效能。然而，人工智能技术的广泛应用也带来诸多伦理风险。其在推动公共管理向智能化、高效化、精准化变革的同时，不可避免地公众利益、社会公平正义产生潜在威胁。因此，深入探究人工智能在公共管理中的伦理风险及治理对策，对保障技术可持续发展、维护社会稳定具有重要意义。

一、相关概述

(一) 人工智能在公共管理中的应用研究

如今，得益于神经网络和机器学习领域的应用与探索，以人

工智能为代表的新技术革命到来，再次升级了人类的生产生活方式。人工智能技术带来的进步和相关的价值潜力也逐渐被政府所意识到，各个国家高度关注该领域的技术竞赛，人工智能技术的发展成为首要关切。人工智能已然介入公共行政领域，冲击了传

统的行政理论，引发了公共行政领域的变革与创新^[1]。当前，人工智能已广泛应用于政策制定、公共服务与社会治理等领域。在政策制定方面，通过对大量数据的分析，人工智能能够为决策者提供更精准的预测与建议，提升决策科学性。于公共服务领域，其可实现个性化服务，如智能政务助手为公民提供便捷的咨询与办事服务。在社会治理中，人工智能助力监控与分析社会动态，及时预警潜在风险。人工智能在公共管理中的应用优势显著，它能提升管理效率，降低人力成本，优化管理流程。然而，其应用也面临一些挑战，如技术的不完善、与现有管理体系的融合难题等。

（二）人工智能伦理风险研究现状

国内外学者对人工智能伦理风险的研究涵盖多个方面。在数据隐私方面，研究发现人工智能系统在收集、存储与使用数据过程中，存在过度收集、未经授权收集以及数据泄露等风险，严重侵犯公民隐私。算法歧视也是重点关注领域，学者指出算法设计者主观因素与数据偏差会导致算法存在偏见，影响公共管理决策公平性，在招聘、贷款审批等场景中尤为突出^[2]。责任界定问题同样备受关 注，当人工智能参与公共管理决策出现不良后果时，由于缺乏明确的责任追溯机制，难以界定技术开发者、使用者还是系统本身应承担 责任，这加剧了责任界定模糊的风险。此外，还有研究关注人工智能对人类价值观冲击、就业影响等伦理风险，展现了当前研究的广度。

二、人工智能技术在公共管理中的伦理风险

（一）数据隐私风险

在公共管理过程中，人工智能系统收集公民数据的方式多样，涵盖了线上线下的诸多渠道。例如，通过各类政务 APP、官方网站，公民在办理业务时会被要求填写个人信息，这些信息被人工智能系统收集用于优化服务流程、分析公众需求等。然而，这一过程中可能存在过度收集数据的问题，即收集超出实际业务需求范围的数据。一些政务平台可能会收集与业务无关的公民敏感信息，如个人消费习惯、社交关系等，超出了为提供公共服务所必需的数据范畴^[3]。同时，未经授权收集数据的情况也时有发生，部分人工智能系统可能在公民不知情或未明确同意的情况下，通过技术手段获取其数据，严重侵犯了公民的隐私权。

数据存储环节同样面临诸多风险。由于技术漏洞，如系统存在安全漏洞未及时修补，黑客可能利用这些漏洞入侵数据库，窃取存储的公民数据。此外，管理不善也是导致数据泄露的重要原因，例如内部工作人员操作失误、权限管理混乱等，都可能使数据暴露在不安全的环境中。一旦数据泄露，公民的个人信息将面临被滥用的风险，如被不法分子用于诈骗、骚扰等非法活动，给公民的生活带来极大困扰。在数据使用环节，也存在被不当利用的风险。人工智能系统收集的数据本应用于合法的公共管理目的，但在实践中，可能出现数据被挪作他用，或者被第三方非法获取并利用的情况。

（二）算法歧视风险

算法设计过程中，多种因素可能导致算法存在偏见。一方

面，设计者的主观因素不可忽视。设计者自身的价值观、认知偏差以及所处的社会环境等，都可能影响算法的设计方向。例如，如果设计者在种族、性别等方面存在偏见，这种偏见可能会体现在算法中，使得算法在处理相关数据时产生不公平的结果。另一方面，数据偏差也是导致算法偏见的重要原因。如果用于训练算法的数据本身就存在偏差，如样本不具有代表性、数据存在缺失值或错误值等，那么算法基于这些数据学习到的模式也会存在偏差，从而影响公共管理决策的公平性。

在公共管理实际应用中，算法会将数据中的偏见放大，造成不公平的治理结果。以公共资源的分配为例，如果算法存在对特定群体的偏见，在资源分配决策中，该群体可能会得到较少的资源或不公平的待遇^[4]。例如，在贫困救助资源分配中，如果算法因数据偏差对某些地区或群体存在偏见，可能导致这些真正需要救助的人群无法获得足够的救助资源。此外，在执法、招聘等公共管理领域，算法歧视也可能导致不公平的决策，如对特定种族或性别的群体进行不公正的执法判断或在招聘中筛选掉符合条件的候选人，严重损害了社会公平正义。

（三）责任界定模糊风险

当人工智能参与公共管理决策出现不良后果时，责任界定变得十分困难。从技术开发者角度来看，他们负责设计和开发人工智能系统，但如果系统在投入使用后因设计缺陷导致不良后果，开发者是否应承担全部责任存在争议。因为开发者可能无法预见到所有潜在的使用场景和问题，且系统在实际应用中可能受到多种外部因素影响。对于使用者，即公共管理部门及其工作人员，他们在使用人工智能系统进行决策时，如果因操作不当或误用系统导致不良后果，理应承担一定责任^[5]。然而，如果问题是由系统本身的缺陷引起的，使用者是否应承担 责任以及承担多大责任也难以明确。此外，人工智能系统本身具有一定的自主性，在某些情况下，其决策过程可能超出了开发者和使用者的控制范围，此时将责任归咎于系统本身似乎也缺乏合理性。

当前公共管理领域缺乏有效责任追溯机制，这进一步加剧了责任界定模糊的问题。由于人工智能系统的复杂性和决策过程的不透明性，很难准确追溯不良后果产生的具体原因和责任主体。没有明确的责任追溯机制，当出现问题时，各方可能会相互推诿责任，导致受害者的权益无法得到及时有效地保障，也影响了公共管理决策的公信力和权威性。

三、人工智能技术在公共管理中伦理风险的治理对策

（一）政策法规层面

在全球范围内，诸多国家已针对人工智能伦理问题出台了相关法律法规。例如，欧盟的《通用数据保护条例》（GDPR）对个人数据的保护做出了严格规定，旨在保障公民的数据隐私权益，在公共管理领域，这为政府部门收集、存储和使用公民数据设定了明确的法律边界。美国也发布了一系列与人工智能相关的政策指导文件，试图在促进人工智能技术发展的同时，规范其在公共管理中的应用^[6]。在国内，同样有相关法律法规对人工智能在公

共管理中的应用进行约束，如《中华人民共和国个人信息保护法》等，对个人信息的收集、使用、处理等环节进行了规范，在一定程度上覆盖了公共管理中人工智能技术涉及的数据隐私保护问题^[7]。

然而，当前政策法规仍存在一些不足之处。在责任界定方面，虽然部分法律提及了相关责任主体的概念，但缺乏细化的责任条款，当人工智能参与公共管理决策出现不良后果时，难以清晰判定是技术开发者、使用者还是系统本身应承担 responsibility。在数据保护方面，随着人工智能技术的快速发展，数据的形式和用途不断变化，现有法规对新兴数据类型的保护尚显不足，且在数据跨境流动等方面的监管也有待加强。针对这些问题，应细化责任条款，明确不同主体在不同场景下的具体责任，同时加强数据监管，及时更新数据保护范围，对数据跨境流动制定更为严格的审查机制，以确保公共管理中人工智能应用的合法合规性。

（二）技术手段层面

数据加密技术是保护公共管理数据隐私的重要手段之一。通过对收集到的公民数据进行加密处理，使得数据在传输和存储过程中以密文形式存在，即使数据遭遇泄露，未经授权者也难以获取其中的有效信息，从而降低数据泄露风险。访问控制技术则能够严格限制对数据的访问权限，只有经过授权的人员或程序才能对特定数据进行操作，进一步保障了数据的安全性和隐私性。

算法可解释性技术对于提高人工智能决策透明度至关重要。由于人工智能系统的决策过程往往较为复杂，难以被普通用户理解，算法可解释性技术能够将算法的决策逻辑以可理解的方式呈现给相关人员，使得公共管理决策者能够清楚了解决策的依据，增强决策的公信力^[8]。同时，通过优化算法设计，可以有效减少歧视。在算法设计过程中，充分考虑数据的多样性和公平性，对可能存在偏差的数据进行预处理，避免因数据偏差导致算法产生偏见，进而影响公共管理决策的公平性。

（三）社会监督层面

构建政府、企业、公众、社会组织等多元主体参与的监督机

制是保障公共管理中人工智能应用伦理性的关键。政府作为公共管理的主体，应发挥主导作用，制定监督标准和规范，对人工智能在公共管理中的应用进行宏观监管。企业作为人工智能技术的提供者，应自觉遵守相关法律法规和伦理准则，加强内部自律，确保其提供的人工智能技术在公共管理应用中的安全性^[9]。公众作为公共管理的对象，对人工智能应用有着直接的感受，应积极参与监督，表达自己的诉求和意见。社会组织则可以发挥专业优势，对人工智能在公共管理中的应用进行独立评估和监督。

加强公众对人工智能在公共管理中应用的了解与参与具有重要意义。目前，许多公众对人工智能技术的原理和应用缺乏深入了解，这导致他们在面对人工智能参与的公共管理决策时，难以有效行使监督权力。因此，开展公众教育势在必行。可以通过举办科普讲座、在线课程等方式，向公众普及人工智能技术在公共管理中的应用知识，提高公众对人工智能技术的认知水平^[10]。同时，建立公众参与渠道，如设立意见反馈平台、开展公众听证会等，鼓励公众积极参与公共管理决策过程，提高公众监督能力，从而保障人工智能在公共管理中的应用符合伦理要求。

四、结束语

人工智能技术在公共管理领域的应用推动管理模式创新与效率提升，但数据隐私、算法歧视和责任界定模糊等伦理风险不容忽视。目前，国内外虽已出台相关政策法规，但在责任明确、数据监管方面仍存不足；技术层面，数据安全保障与算法透明性有待加强；社会监督也需构建更完善的多元协同机制。未来，需持续完善政策法规体系，推动技术创新以降低伦理风险，强化政府、企业、公众等多元主体协同监督，形成政策、技术、社会多维度的治理合力，实现人工智能在公共管理领域安全、公平、可持续发展，使其更好地服务于公共利益与社会治理目标。

参考文献

- [1] 王若彤, 柳亦博. 人工智能时代的公共行政: 风险及应对 [J]. 辽宁行政学院学报, 2024, (02): 31-38. DOI: 10.13945/j.cnki.jlac.2024.02.014.
- [2] 彭丽徽, 张琼, 李天一. 人工智能嵌入政府数据治理的算法歧视风险及其防控策略研究 [J]. 农业图书情报学报, 2024, 36(05): 23-31. DOI: 10.13998/j.cnki.issn1002-1248.24-0353.
- [3] 张夏恒. 类 ChatGPT 人工智能技术嵌入数字政府治理: 价值、风险及其防控 [J]. 电子政务, 2023, (04): 45-56. DOI: 10.16582/j.cnki.dzzw.2023.04.004.
- [4] 谭九生, 杨建武. 人工智能嵌入政府治理的伦理风险及其防控 [J]. 探索, 2021, (02): 126-138. DOI: 10.16501/j.cnki.50-1019/d.2021.02.011.
- [5] 何塞·比达·费尔南德斯, 张玉洁, 张玉翠, 等. 数字风险的治理: 人工智能监管的挑战与发展 [J]. 数字法学, 2024, (01): 260-278+342.
- [6] 李文君. 人工智能嵌入政府治理的信任风险内涵、样态及规避 [J]. 党政干部学刊, 2024, (08): 58-64.
- [7] 李文君. 智能技术引发的政府信任风险及其防范研究 [D]. 湘潭大学, 2023. DOI: 10.27426/d.cnki.gxtd.2023.000101.
- [8] 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, (03): 120-126. DOI: 10.14134/j.cnki.cn33-1337/c.2022.03.012.
- [9] 李良成, 李雨青. 人工智能嵌入政府治理的风险及其规避 [J]. 华南理工大学学报 (社会科学版), 2021, 23(05): 1-13. DOI: 10.19366/j.cnki.1009-055X.2021.05.001.
- [10] 陈兵. 人工智能应用的科技伦理与法治化建设 [J]. 人民论坛, 2024, (12): 66-70.