

ChatGPT 人工智能技术嵌入数字政府建设的行政伦理风险和优化路径

陈恩

广东工业大学, 广东广州, 515000;

摘要: 数字时代下, 建设数字政府是提升政府治理能力, 助力法治政府建设, 促进经济社会数字化转型, 实现我国治理能力和治理体系现代化的重要举措。ChatGPT 模型作为人工智能领域的一项重大突破, 在数字政府建设中具有巨大的应用潜力, 切实发挥其在我国数字政府建设之中的作用具有重要意义。然而其为数字政府赋能的同时, 新技术的嵌入可能引发行政责任缺失、算法歧视、人性冲突及信息安全等行政伦理困境, 政府应该通过强化责任机制、提升数据多样性、引入伦理审查以及确保人类决策主导权等方式应对这些挑战, 以保障数字政府建设的可持续发展和社会公平正义。

关键词: 数字政府; ChatGPT 模型; 行政伦理

DOI: 10.69979/3029-2700.25.12.059

引言

数字政府是构建新时代服务型政府, 提升国家治理能力, 推进国家治理体系现代化的重大举措。国务院印发的《法治政府建设实施纲要(2021—2025 年)》指出要坚持运用互联网、大数据、人工智能等技术手段促进依法行政, 着力实现政府治理信息化与法治化深度融合, 优化革新政府治理流程和方式, 大力提升法治政府建设数字化水平。党的二十大报告作出加快建设数字中国的重大部署, 次年二月发布的《数字中国建设整体布局规划》指出, 建设数字中国是数字时代推进中国式现代化的重要引擎, 是构筑国家竞争新优势的有力支撑。

1 ChatGPT 嵌入数字政府建设的行政伦理困境

工业化的发展使人的社会生活领域不断扩展, 为了提供更加稳定的社会环境以促进社会化大生产, 政府的职能逐渐扩大, 更加深入地介入到社会微观管理之中, 由此, 传统的家庭伦理关系逐渐向公共领域拓展, 并在政府行政层面上发展出了新的行政伦理关系^[1]。关于行政伦理的内涵界定, 虽然学界目前尚未达成统一的定论, 但在国内研究中形成了三大具有代表性的观点。一是职业道德说, 这种观点将行政伦理等同于行政管理主体的职业道德, 认为行政行为规范是行政伦理学的重心, 使行政人员能忠于职守, 勤政为民, 抑恶扬善^[2]。二是行政责任说, 这种观点认为行政伦理是以‘责、权、利’的统一为基础, 以协调个人、组织与社会的关系为核心的行政行为准则和规范系统^[3]。三是从多维的角度来定义行政伦理, 即从行政伦理的主体性、政治性、层次性、职业性、现实性、体系性等几个不同的角度, 把握中国

行政伦理的基本内涵^[4]。不论是哪一种说法, 都强调了行政伦理体系是以人为主导的, 人在行政过程中发挥了最核心的作用。而 ChatGPT 作为强人工智能时代的代表性产物, 自 2018 年 6 月 OpenAI 公司开始推出 GPT-1 模型以来, 历经多次技术升级和迭代, 其模型在几年的时间里实现了全面升级, 可以在最大程度上实现人机交互的智能化, 达到“类人”的效果, 因此在嵌入数字政府建设的过程中势必会给原本以人为主导伦理体系带来冲击, 进而引发一系列风险。

1.1 政府治理中行政主体责任的缺失

ChatGPT 嵌入数字政府建设过程中会造成行政主体责任的缺失。第一, 对技术的过度依赖会导致行政主体责任意识淡漠。ChatGPT 因其在处理复杂数据和信息方面的强大能力, 可能会导致行政主体在决策过程中对技术产生过度依赖。他们在应用的过程中开始相信, 只要依赖该系统, 就能够保证决策的准确性和效果。这种信念所产生的权力的让渡, 使得决策者在行政过程中沦为了工具的工具, 会使他们逐渐忽视了自身在决策中应有的独立判断和责任意识, 并且会越来越倾向于将责任推卸给技术, 而不是在决策过程中积极地承担个人的道德和行政责任。责任意识一旦出现淡漠, 数字技术的应用就容易陷入权责分离或权责分立的状态, 将对公共理性的秉持造成极大挑战^[5]。第二, 责任界定出现困难。在传统政府治理模式下, 人作为政府治理行为的唯一主体, 政府的任何行政行为都体现着人的主观意志, 并应承担相应的行政责任^[6]。在这种模式下, 行政行为的合法性和责任的明确性可以得到充分保障。然而, 随着

ChatGPT 模型的嵌入,生成式 AI 参与到政府治理中,情况将会发生显著变化。其强大的交互式功能可以在很大程度上取代人类,甚至在某些情况下成为唯一的决策主体,导致机器意志逐渐在行政决策和方案制定中占据主导地位。传统的行政行为建立在人的意志和道德判断基础上,而机器缺乏这种人类特有的主观性和伦理考量,机器意志的介入可能会削弱行政行为的合法性,而一旦发生错误或争议,责任的界定将变得复杂,行政责任主体的不明确和模糊化危机也随之而来。这不仅影响了行政行为的透明度和权威性,还可能引发一系列法律和伦理问题,稍有不慎,就会引发行政权力公共性的结构风险。

1.2 算法歧视带来的行政偏见

ChatGPT 的主体架构遵从“基础语料+预训练+微调”的基本范式^[7],预训练语料包括了海量从不同渠道搜集来的文本数据,通过在大规模训练数据上进行大量通用的训练使得 ChatGPT 具备符合语言逻辑关系的表达方式,而微调语料中所包括的高质量标注文本数据则进一步增强了 ChatGPT 的对话能力,使人机交互顺畅无阻。一方面,这一基本范式决定了 ChatGPT 文本对话和输出观点的基础来源于前期的训练数据,数据来源于生活,归根到底还是对社会思想的反映,因此不可能做到完全中立。那么,如果前期所采集的预训练数据中包含了诸如在年龄、性别、民族、宗教等方面具有的歧视和偏见的观点,那其输出的文本将不可避免地将其中的偏见因素放大,引发算法歧视的风险。应用到政府治理领域,这种歧视将导致行政主体做出一些不公平或歧视性的决策,这样一来部分公众的合法权益将无法得到保障,长此以往必将使得社会矛盾激化。一个很典型的例子是美国芝加哥警察局采用了一种名为“战略性科目名单”(Strategic Subject List, SSL)的算法,用来预测哪些人最有可能成为犯罪的受害者或犯罪者。调查发现,该算法主要基于过去的犯罪数据,而这些数据中已经存在种族偏见。因此,算法过于频繁地将非裔美国人和拉丁裔标记为高风险人群,导致这些群体在警务部署中受到不公平的额外关注,进一步加剧了社区与执法部门之间的紧张关系。相比于人类的偏见,算法的偏见更容易引发大规模的负面效应^[8]。这是因为算法能够在短时间内影响和处理大量的数据,从而放大其偏见的影响范围。ChatGPT 在决策过程中对“数据流量”价值的重视,使得决策者倾向于关注那些能够产生大量数据的领域。这种关注在一定程度上分散了他们对其他不可量化但同样重要的公共需求的注意力。结果是,许多数据流量较低但对民生至关重要的重大社会问题被忽视,例如环境保护、公共卫生和社会公平。这种偏差可能导致资源分

配的不公,进而加剧社会的不平等。

1.3 价值观不合导致的人性的冲突

在人类的治理伦理中,首先应当关注的是正义问题,其次才是效率或其他问题。然而,人工智能不受人类伦理观的约束,它仅基于自身的数理逻辑进行决策和行动^[9]。ChatGPT 的应用依赖其背后的强大算法,这种算法假设代表的是一种机械化的逻辑,与人类的选择存在根本的区别,而人类的选择则常常与社会规范和人际网络密切相关。当 ChatGPT 被嵌入并用于辅助数字政府治理时,决策权将更多地转移到机器手中,依赖的将是算法假设,而不是人类的判断和选择。随着其在数字政府治理领域的不断推广和深入应用,行政行为将越来越依赖这些算法假设,这种依赖可能会导致人类的选择被忽视,甚至在某些情况下与人类的选择发生冲突。在许多决策中,道德和伦理判断非常复杂,而人工智能缺乏理解这些复杂性的能力。道德判断往往涉及多层次的考量,需要综合权衡多种因素。例如,不同个体和群体的利益、短期和长期的后果、以及社会整体的公正和公平等。这些因素之间往往存在复杂的互动和矛盾,无法简单地通过算法或数据模型进行完全准确的衡量和判断。社会中存在多样化的伦理观念和价值体系。不同文化、宗教和社会背景下,人们对同一问题可能会有不同的道德判断和伦理期望。人工智能系统在决策过程中,可能难以兼顾和尊重这些多样化的伦理观念,导致其决策在某些背景下被视为不道德或不可接受。

2 ChatGPT 人工智能技术嵌入数字政府建设的行政伦理困境的优化

2.1 明确责任主体,健全追责机制

在推动 ChatGPT 服务于公共行政的现实价值时,必须同时关注数据安全和行政道德等多重价值需求,最大限度地降低数字技术带来的潜在风险。首先,必须构建责任导向的价值观体系,进一步增强公共行政主体的责任意识。随着 ChatGPT 人工智能技术日益深度融入公共行政,作为公共行政核心要素的行政人员和政府,应更加注重培养自身的责任意识,主动将外在的约束转化为自律,将行政和法律责任提升为更高层次的道德责任,以提升数字素养。责任的内化,作为个体信念和使命感的升华,一方面可以最有效地发挥数字技术的便捷与强大功能,另一方面通过加强道德自律和职业操守,能够有效避免数字技术导致的行政失误与失责。为此,适时进行有针对性的专项培训,帮助行政人员更清晰地认识岗位职责和义务,以更加科学严谨的态度合理使用数字技术,发挥其应有的作用。其次,要健全 ChatGPT 模型嵌入数字政府的追责机制。尽管 ChatGPT 模型在一定程

度上提高了行政决策的自动化效率,但为了防止因其应用于行政决策中而出现“责任无法追溯”的问题,必须健全相关的追责机制,使得行政机关和 ChatGPT 模型的服务提供者都不能以算法为由免除责任。第一,应明确法律框架和监管标准,制定相关政策法规,设立监管机构负责监督该模型的合规性与安全性,此外,公众参与也能够增强透明性和公正性。第二,责任分层与角色明确。各参与方需对系统的开发、使用和维护承担相应责任,同时建立严格的权限控制与操作追溯机制,确保系统操作日志完整,便于问题发生时的追责与调查。为应对可能的故障和风险,需设立故障报告与应急预案,确保快速响应与处理。第三,引入道德与伦理审查机制。为防止系统产生不良社会影响,并明确开发方与政府的社会责任,可通过培训与合规教育,提高使用者对系统的理解 and 责任意识,同时建立健全的问责机制和惩罚措施,确保违规行为受到应有惩罚。此外,定期审查与动态调整制度也有助于持续完善追责体系,确保技术模型在数字政府中的安全、合规与高效应用。

2.2 提升数据多样性,加强政府主导的伦理审查

在 ChatGPT 模型引入数字政府建设的过程中,要避免算法歧视所带来的行政偏见,以最大限度地使社会公义不遭到破坏,提升模型训练数据的多样性以及对其作出的决策进行伦理审查至关重要。首先,在引入模型时提升预训练数据的全面性和多样化。第一,数据应尽可能覆盖不同性别、年龄、种族、经济背景等群体,避免因数据不均衡而导致偏见,还应通过多渠道收集数据、动态更新信息以及清洗去除不公正数据,可以提升数据的质量和代表性,例如,在处理就业、教育等敏感数据时,需警惕其中潜在的种族、性别、年龄等歧视性特征,通过去除这些歧视性数据,避免算法在学习时固化或放大已有的社会不平等。针对小众群体,需加强其数据的代表性,避免算法忽视其需求,为此,政府可以引入加权或重新采样的技术,确保少数群体在算法中的公平性。第二,加强跨部门的数据共享与协作,进一步提高数据的准确性和兼容性,减少信息孤岛的形成。通过打破“数据孤岛”,政府可以更全面地了解公众需求,减少由于信息不对称或数据孤立带来的偏见。同时,通过跨部门的数据协作,可以更有效地识别和消除算法中的潜在偏见问题。其次,要对算法作出的决策进行伦理审查。政府可以成立独立的算法伦理审查委员会,由技术、伦理、法律、社会学等多学科专家组成,负责对政府使用的算法进行伦理审查,确保其公平性、透明性和合法性。委员会需制定明确的伦理准则和审查标准,包括公平性、透明性、隐私保护和责任追究等方面,确保算法不会对弱势群体产生歧视或偏见。同时,政府应推动相关法律

法规的制定,保障算法应用的合法性和合规性。为了增强公众信任,委员会还应公开审查流程和结果,设立公众反馈机制,定期发布算法透明度报告,并通过定期审计和持续监控,确保算法在长期使用中保持公平公正。除此之外,委员会还可以借鉴国际经验,推动全球范围内的伦理标准制定与合作。这一机制将有效减少算法中的歧视问题,确保数字政府 in 提高效率的同时,坚持公平和透明的原则。

3 结语

随着人工智能技术的飞速发展,将其嵌入数字政府建设是未来公共管理领域的重大变革,ChatGPT 模型作为时下最具突破性的人工智能技术,其应用必然是众望所归。现有的研究表明,ChatGPT 模型有望推动数字政府更加智能化,同时也更加亲民,推动行政模式的转型,但与此同时也带来了前所未有的伦理冲击。总之,在应用新技术的同时,也要做好应对新挑战的准备,才能利用好这把双刃剑,使其更好地服务于数字政府地建设之中。

参考文献

- [1]熊远薇,杨钊,张进.数字政府的行政伦理关系解析:发展样态、异化表现与优化路径[J/OL].合肥工业大学学报(社会科学版),2024,(05):1-11[2024-09-26].
- [2]姚恩健.建立行政伦理学的设想[J].理论学习月刊,1994,(05):45-48.
- [3]江秀平.对行政伦理建设的思考[J].中国行政管理,2000,(09):19-21.
- [4]王伟,车美玉.中国现代行政伦理建设与公务员行为规范[J].中国工商管理研究,1999,(01):12-17.)
- [5]李森.数字技术赋能公共行政责任伦理困境的纾解与优化[J].党政干部学刊,2024,(04):73-80.
- [6]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.
- [7]秦涛,杜尚恒,常元元,等.ChatGPT 的工作原理、关键技术及未来发展趋势[J].西安交通大学报,2024,58(01):1-12.
- [8]柳亦博,高洁,李轶倩.人工智能参与决策对行政伦理的影响及其治理之道[J].中共天津市委党校学报,2022,24(06):55-63.
- [9]柳亦博.人工智能阴影下:政府大数据治理中的伦理困境[J].行政论坛,2018,25(03):97-103.

作者简介:陈恩,女(2001—),汉族,籍贯:广东,研究生在读,研究方向:行政管理现代化。