

预训练语言模型引入数字政府建设的风险应对

——基于行政关系的视角

曹冬英

(云南师范大学历史与行政学院, 云南昆明, 650500)

摘要: 从行政主客体关系的角度来看, 将预训练语言模型引入数字政府建设能够推动行政主体与相对人之间的沟通升级、促进行政行为内外流程的多维完善、提高行政决策的理性水平。在正视预训练语言模型带来的正面效能的同时, 需要清醒地认识到它可能引发的行政相对人的数据权利被侵扰、行政伦理基本原则面临挑战、行政责任主体模糊不明等风险。为了应对风险, 行政主体应当以权利保障为核心落实相关法律法规, 以价值理性为取向构建伦理保障机制, 以权责清晰为基点优化责任结构关系。

关键词: 预训练语言模型; 数字政府; 行政关系; ChatGPT; 数字治理

中图分类号: D63

文献标识码: A

文章编号: 1672-3104(2025)04-0143-10

一、问题的提出

受到信息化、数字化与智能化叠加共进的社会发展状态以及政策层面的驱动, 我国政府管理的数字化转型程度不断加深, 数字政府职能应用的覆盖面不断拓宽。《数字政府一体化建设白皮书(2024年)》指出, “数字技术融合催生一体化平台应用”, 为政府履职提供了新的方案。当前, 将大数据、人工智能等技术成果嵌入数字政府建设有利于构建数字化、智能化的数字政府运行新形态。ChatGPT(Chat Generative Pretrained Transformer)是近年来人工智能领域最具争议性的技术突破, 其背后的预训练语言模型(Pre-trained Models, 以下简称“PTMLs”)在自然语言表达和智力工作领域具有强大的应用能力。PTMLs是一种人工智能模型, 它基于大量文本数据进行预训练, 且一般使用无监督学习技术(如语言建模)来构建理解语言的模式和结构。PTMLs表现出的自然语言理解(natural language understanding, 以下简称“NLU”)能力、文本生成能力和知识更新能力构成了AI领域中的关键功能, 三者是现代AI模型在处理和理解人类语言时不可或缺的要素, 它们共同使得AI能够以更加自然和灵活的方式与人类进行交流和互动, 并推动了数字时代下行政关系中主体与行为的样态迭代。2023年7月10日, 国家互联网信息办公室发布的《生成式人工智能服务管理暂行办法》表明了我国政府对生成式人工智能及PTMLs的发展创新与应用所秉持的积极态度。

当前, 对于预训练语言模型与数字政府的相关研究主要从ChatGPT这一具体的产品形式出发, 并沿着三个方向展开。其一, 生成式人工智能在数字政府建设中的应用价值。多数学者已然清晰地认识到生成式人工智能及其技术要素在我国数字政府建设中的应用前景, 认为生成式人工智能技术嵌入数

收稿日期: 2024-04-22; 修回日期: 2024-09-06

基金项目: 国家社会科学基金后期资助项目“数字政府治理研究”(19FGLB032); 福建省习近平新时代中国特色社会主义思想研究中心项目“从‘数字福建’到‘数字中国’的治理变迁研究”(FJ2022XZB026)

作者简介: 曹冬英, 女, 江苏淮阴人, 法学博士, 云南师范大学历史与行政学院副教授, 主要研究方向: 公共治理、比较政治与比较行政, 联系邮箱: caodynsfdx@126.com

字政府建设不仅可以促进政府治理结构、流程与工具的优化^[1],而且能助力数字政府治理决策、提升数字政府治理效率、降低数字政府治理成本^[2]、推动数字政府治理精细化、强化数字政府监管等^[3]。其二,生成式人工智能参与数字政府治理的风险分析。现有研究在充分肯定生成式人工智能技术之于数字政府建设的作用的同时,也从多个层面对相应的风险与挑战进行了阐述,如知识产权风险^[4]、行政公共性解构风险^[5]、算法歧视、算法黑箱、数据泄露^[6],以及其他不确定性风险等^[7]。其三,生成式人工智能参与数字政府建设的完善进路。对于生成式人工智能所可能产生的风险与挑战,既有研究多从规范数据分类分级^[8]、完善责任链条^[9]、规定公民权利、设置国家保护义务、加强相关技术培育等角度出发,试图构建具有可操作性的人工智能应用规制体系^[10]。

然而,我国学界对于生成式人工智能助力数字政府建设的研究仍存在以下缺陷。一方面,现有研究多以 ChatGPT 为核心,简单地认为其能应用于我国数字政府建设。但实际上,ChatGPT 是集成化系统,真正可引入我国政府治理中的的是其所属的生成式人工智能及其核心技术,特别是 PTMLs。PTMLs 是生成式人工智能开发与训练的基础,也是使人工智能系统能够不断自我完善与进化的关键。因此,在探讨生成式人工智能与数字政府的关系时,需对 PTMLs 这一技术基石给予足够的关注。另一方面,我国学界现有研究在层次展开上并不存在清晰、紧密的内核勾连,对价值、风险及完善方案的分析,都没有深度结合数字政府建设中存在的主体、客体、行为等之间的核心关系,论述显得较为零散。事实上,数字赋能改变了做出行政行为的机理与方式,生成式人工智能技术可能产生的风险以及对此所应做出的规制手段也可以在行政关系中得到体现。因此,应以数字时代行政关系的新样态为出发点,深度剖析 PTMLs 的应用对于政府治理过程中的各方主体以及政府行为所产生的影响,并以此为基础有针对性地提出相应的规制路径。

本文试图通过分析在数字政府建设中引入 PTMLs 的风险及其应对,从数字时代的行政关系新样态出发,结合 PTMLs 在当下数字政府建设格局中已经或可能出现的具体应用形态,就以下三个方面展开论述:首先,阐释 PTMLs 在辅助数字政府建设中对行政主体、行政相对人、行政行为的多层影响,从而分析其引发的行政关系样态迭代;其次,以数字时代的行政关系为切入点,剖析 PTMLs 参与数字政府建设所可能引发的各类风险;最后,结合 PTMLs 的技术特点、应用风险和发展前景,依次从行政相对人、行政行为、行政主体的角度出发,提出将 PTMLs 引入建设数字政府的完善进路,即在确保数据安全、减少算法偏见、维持法律伦理合规的基础上,平衡技术辅助与人类判断,构建有效的技术治理和监管框架,以提供公平、高效的公共服务,并维护公民的权益和社会的整体福祉。

二、PTMLs 推动数字时代行政关系的样态迭代

PTMLs 作为技术更新迭代的最新成果,在数字政府建设上的应用价值得到了充分的体现^[11],当前集中表现为实现行政主体与相对人的沟通升级、行政行为做出程序的多维完善、行政决策的理性促成等三个方面。

(一) 行政关系主体间的沟通升级

第一,自然语言理解与生成能力的提高意味着政府与民众沟通对话的智能性的提高。PTMLs 可以通过扫描大量网络内容或知识,生成类似于人类语言模式的回应^[12]。此种回应显然不同于以往数字政府中“电子客服”内预设的僵化问答,它可以根据不同群体的需要,做出有效的、高度符合需求的回答。PTMLs 的底层逻辑能够支持系统与民众进行连续对话,运用“类人化”语言与民众进行人机交互,提高行政服务的精准性和高效性。在此意义上,引入 PTMLs 不仅能够精准满足民众的个性化需求,而且能对公众的疑问做出及时回应,从而提高民众对于公共管理和行政服务的满意程度。如厦门市人

力资源和社会保障部门的“12333 智能政务大模型”通过专业语料的预训练与微调，具备了较强的政务咨询理解能力、政策知识信息实时获取与分析能力，能够直接处理口语化的语音和文字信息，实现便捷化、协同化、亲民化的政务咨询升级。

第二，自然语言理解与生成能力的提高意味着政府与民众沟通对话的亲民性增强。数字政府建设水平提高的重要标志之一是能够为公民提供主动化和定制化的服务，使公民在享受数字化政务服务和公共服务的过程中获得美好的感觉、永久的记忆和值得回味的事物与经历^[13]，即政务服务和公共服务不仅要满足公民的办事需求，还要在这一过程中表达人文关怀，使人民群众体验到作为“新时代国家治理的良政基准和善治标尺”^[14]的“获得感”。相比于冰冷的信息传达，PTMLs 在运行过程中还会特别注重情景式表达，兼顾行政相对人的情感需要和情绪表达，最大限度地增进“以人为本”的民生福祉。而且，当 PTMLs 以自然语言形式与行政相对人进行对话时，将会给行政相对人一种“交往错觉”，即这种与“真人”交流的顺畅性和自然性会使行政相对人忽视其“机械身份”，从而赋予其一种“社会身份”。此交流过程可被视作一种有感情的交流，并能充分照顾到行政相对人的情感与偏好等，实现温情服务与依法行政的统一。

第三，自然语言理解与生成能力的提高意味着在政府和民众的直接沟通之间有了相应的缓冲地带，沟通效果更好，政民互信更加坚实。在以往的行政行为中，行政主体以国家权力为依托，在行政对话过程中可能会呈现出“盛气凌人”“居高临下”之势；行政相对人因为对政务服务不满意、自我诉求无法满足等，可能会表现出情绪及行动上的不理性。但随着 PTMLs 的应用，行政主体与行政相对人之间的交互有了更为融通与舒适的缓冲地带，两者之间的沟通效果能够得到极大的提升。概言之，PTMLs 的技术优势和便利在一定程度上消除了行政相对人的弱势地位，进一步缩小了政民之间的信息差，从而尽可能地使“政民摩擦”降到最低。

（二）行政行为内外流程的多维完善

随着人工智能技术的发展，PTMLs 在自我学习、人机交互、语言能力等方面有了较大提升，这在技术层面上提高了我国数字政府建设的效率。数字政府建设归根到底就是利用数字技术实现政府在履职和运行等层面的转变，具体涵盖了提升政府对内的自身决策管理能力及对外与公众的交流互动能力两个层面。

一方面，就提升数字政府的对内自身决策管理能力来看，PTMLs 在一般文本内容质量上的增强提高了政府文书工作的效率。与以往的人工智能模型相比，PTMLs 能够通过自我学习理解人类的语言内容，并通过大数据分析得出符合人们认知的答案。其回答并不仅仅是对现有材料的总结概括，而是根据现有材料生成的新内容，这种与传统智能模型的巨大代差依赖于自我学习和长篇注意力机制两种技术的运用。从具体流程上看，PTMLs 可以通过以下流程优化政府文书撰写工作。第一，进行数据库喂养。将行政数据和相关信息投喂给 PTMLs 模型，根据其生成的内容不断完善提问内容，生成文书初稿。第二，人工介入初稿完善。通过人工完善 PTMLs 生成的文书初稿，对关键信息进行审查、修改。第三，文稿完善与最终审校。利用 PTMLs 对第二环节中的文稿进行完善和修正，并通过人工进行最终把关及后续流程。这意味着在给 PTMLs 提供充足的政府信息和数据后，它通过对各类政府文书的写作方式和模板进行学习，便能够准确且全面地生成一份合格的文书，这在很大程度上节约了政府文书写作的人力成本与时间成本，其高效的运算能力也提高了文书产出的效率。

另一方面，就提升数字政府对外与公众的交流互动能力来看，第一，针对特定行政相对人的特定需求，PTMLs 的运用提高了文书的个性化特征，满足了公众的个性化需求。随着我国社会经济的不断发展，公众的个性化需求也在不断增加，而 PTMLs 则能以其技术为依托，通过数据分析及时获取相对人的喜爱和偏好，更好地识别和满足公众的需求，保证政务服务和需求之间的良好双向交互^[3]。第

二, PTMLs 的多模态输出能够丰富行政文书的展现方式, 加深行政相对人对文书内容的理解, 降低行政沟通和宣传层面的成本。当公众通过多元化形式准确地理解了文书内容后, 便能够间接地使政府工作人员从枯燥重复的文字解释工作中解脱出来, 降低政府在宣传工作中的人力成本和物力成本。第三, PTMLs 还能够实现内外交互的公民反馈机制。PTMLs 能够主动学习公民对行政管理和公共服务的反馈意见, 并对其反馈意见进行汇总分析, 甚至能够结合分析结果直接给出优化方案, 有效提高公民反馈意见的处理效率。通过利用 PTMLs 处理这些意见数据, 政府可以掌握公民的关注点和趋势, 从而做出更明智的决策并制定出更符合公民需求的政策。更深入地了解公民的情绪, 可以促进政府进行更有效的治理并提高公民的满意度^[15]。

(三) 行政决策的理性促成

行政决策的理性程度取决于决策依据的正确性与充分性, 决策依据需要适应于决策目的且符合比例原则。PTMLs 作为一种先进的自然语言处理工具, 可以为行政主体做出相关决定提供重要支持, 并在以下方面促成行政决策的理性化。首先, 它可以帮助行政主体对大量信息进行快速分析和处理, 为行政决策提供数据来源和实践支撑。例如, 北京计算中心推出的“政务参政议政建言大语言模型”将海量的法律法规和相关制度作为预训练和微调的“数据原料”, 能对不同来源的文件(如政府工作报告、政策法规等)进行分析和总结, 并进行数据溯源, 从而实现提案与社情民意的“同频共振”。其次, 它可以协助行政主体进行文本校验和结果预测, 方便决策者把握决策的准确性和实施效果。如广州市联合百度智能云推出的“海珠政务云脑大模型”以“强算法+大算力+海数据+多场景”为基础, 能够提取企业主体和民众个体需求中的关键信息, 辅助行政主体制定相应的政策。最后, 它还可以运用自动生成算法提供信息概览和分析数据, 缩短决策者的检索时间, 提高决策效率。如华为“盘古”政务大模型能够自动生成摘要, 还可以智能校对、辅助批示、智慧督办等, 以助力政府治理数字化转型。借助 PTMLs, 行政主体能够更加全面、快速、客观地了解事件发生的全过程, 精准把握相对人的个性化需求, 从而做出高效且准确的判断。

但值得注意的是, PTMLs 是数字政府建设的辅助工具, 其提供的数据支撑的有效性和可靠性需要专业的政策分析师和决策者通过互补性的验证审核、人机协作等方式进行保障, 从而避免单纯依赖技术而忽视专业判断的情况发生。ChatGPT、“文心一言”等是强人工智能初级阶段的技术成果, 其回答的生成并非简单地依赖于工作人员的置入和关键词检索, 而是要通过数据喂养、收集人类数据以优化模型、收集对比数据以训练奖励模型, 以及强化算法学习进行自身策略的优化等方式, 借助上下文语境生成类似人类回答的文本^[16]。但从根本上看, 这仍旧不能改变 PTMLs 作为工具的本质属性。因此, 在借助 ChatGPT 促进行政主体做出理性决策的过程中, 要综合运用技术工具与人类智慧, 保证决策过程和决策内容的科学性、合理性。

三、PTMLs 嵌入数字政府建设的风险

PTMLs 在以其技术优势重塑行政关系样态的同时, 也对传统行政法治形成了冲击^[17], 为数字政府建设带来了诸多风险和挑战。从行政关系的不同方面来看, 在数字政府建设中引入预训练语言模型可能引发行政相对人的数据权利被侵扰、行政伦理基本原则面临挑战、行政责任主体模糊不明等问题。

(一) 行政相对人的数据权利受到侵扰

PTMLs 参与数字政府建设, 在一定程度上会提高数字政府建设的水平与效率。然而, 数字政府建设是一个动态、开放的阶段性过程^[18], PTMLs 的嵌入并不能从本质上化解数字政府建设与公民数据权利保障之间的冲突, 反而会进一步加大行政法律关系中行政相对人数据权利被侵犯的风险。

首先，在数据收集阶段，行政相对人面临着权利失衡的困境。PTMLs 的训练和运行依赖于海量的数据，包括但不限于民众的个人信息等。在实践中，ChatGPT 的开发公司 OpenAI 曾在未经授权和同意的情况下，收集了大量的个人敏感信息数据^①。从本质上看，资源集聚在某种意义上意味着权利集聚。PTMLs 在嵌入数字政府建设的过程中，必然会掌握大量的个人信息和敏感数据，极易导致 PTMLs 利用信息数据的集聚效应，对行政相对人形成实际意义上的数字权力压迫。

其次，在数据处理阶段，行政相对人面临着权利内容失控难题。这主要是指 PTMLs 能够通过数据转化、数据推理和数据模拟等技术方式，将海量、乱序的信息数据转化为具有逻辑和规律的处理结果^[5]。PTMLs 的整合系统以其深度性、专业性、高效性和精准性特征，能够有效赋能数字政府建设，但不可忽视的是，PTMLs 在嵌入数字政府建设时，极易偏离公共行政目标，可能在行政主体和行政相对人不知情的情况下，将收集到的、处理后的信息数据用作其他用途。究其根源，主要在于公民虽然能够在表面上感知到算法的运用，但“算法黑箱”的存在使其并不能被追根溯源，公民并不知道涉及自身利益的算法的设计意图及数据来源是否正当，算法如何运行及运行结果是否公平^[19]，由此导致公民数据权利的保障陷入被动局面。

最后，在数据存储阶段，行政相对人面临着数据安全问题。其中最重要的是 PTMLs 未经授权存储个人信息数据和无期限存储问题。在行政关系中，行政机关与作为第三方的人工智能服务提供者都可能因其特定的能力优势而对行政相对人形成不当支配，尤其是在个人信息存储方面。当前，作为 ChatGPT、“文心一言”等 PTMLs 产品的开发者与运营者，OpenAI、百度等公司并没有提供合理的方式或者渠道供个人检查其数据存储库，用户知情同意书中也未明确个人信息的存储限制和权利保护等内容。此外，《中华人民共和国个人信息保护法》第 19 条规定个人信息存储时间应以实现处理目的的最短期间为限。作为数据生产者的个人无法成为数据的实际所有者，而数据平台对数据的不当存储和使用将极有可能产生数据霸权^[17]，并进一步导致个人信息泄露、个人隐私裸露，从而对公民的数据权益造成不可逆转的伤害。

（二）行政伦理基本原则面临挑战

行政伦理是公共行政的灵魂^[20]，它决定着公共行政的目标、性质、内容和责任。政府行政伦理逻辑的原则性要求是政府官员和公务员在执行公务时应当遵循的基本道德原则。这些原则旨在确保政府行为的合法性、公正性、透明性和效率。在大数据背景下，PTMLs 嵌入数字政府建设，能够提高数据的可访问性和利用率，从而为政府做出决策提供支持。但它也可能带来隐私风险、误导行政相对人、导致错误等问题，在一定程度上对数字政府的行政伦理造成冲击，引发数字政府行政伦理风险。

一方面，PTMLs 嵌入数字政府建设，容易引发歧视性后果。虽然 PTMLs 嵌入数字政府建设能够借助数字技术减少主观因素对治理过程的不当干预，但是由于数字技术本身嵌入了开发者的个人偏好，故而其应用过程和输出内容在本质上仍是人的主观意志的表达。PTMLs 帮助数字政府表达行政意见、做出行政决策，从本质上看体现了 PTMLs 设计者和开发者的主观意志。由于 PTMLs 仍处于弱人工智能的发展阶段，其算法程序和数据基础仍存在不足，生成的结果可能会存在歧视性、差错性等特征，从而影响用户行为。

另一方面，PTMLs 嵌入数字政府建设，极易导致数字技术资本的入侵，加剧行政权力公共属性的解构风险。“公共性”是数字政府的本质特征之一，也是数字时代下行政基本伦理原则的应有之义与合理延伸。在大数据时代下，数字技术资本中的技术力量相较传统经济资本中的技术力量而言，能产生更多的政治价值、经济价值和社会价值，能更好地帮助政府进行决策^[21]。虽然 PTMLs 能够推进数字政府的科技革新、加快数字政府的建设步伐，但也不可避免地加大了政府行政权力公共属性的解构风险。在私权力的嵌入下，PTMLs 甚至可能侵夺政府等公权力机关的公共职能，并享有某种“数据权

力”。技术资本在数据规模和质量、算法和模型优化、决策效率和响应速度以及国际和跨领域合作等方面的优势,使其在某些方面可能拥有比公权力机关更强的控制力和支配权。进而言之,在人工智能时代,PTMLs 极易借助其技术优势、资本优势支配公民行为,使行政法律关系的主体由“政府—相对人”逐渐演变为“PTMLs—相对人”。而且,资本的逐利性本质会引发“数字方式”的异化,并对政府公共秩序造成负面影响。技术的合理运用要通过服务于人的自由价值才能得以实现^[22]。一旦生成式人工智能借助数字政府的公共权力为资本力量服务,那么数字政府建设将可能成为资本“俘获”国家权力的纽带,资本可能会以“数字资本”的形式走到政府公共性的对立面,资本的价值也将取代人的价值^[23]。在此意义上,数字政府的公共属性及其行政伦理将面临着严峻挑战。

(三) 行政责任主体模糊不明

PTMLs 嵌入数字政府建设,极易导致数字政府治理主体的责任缺失。如果 PTMLs 等技术一开始就被确定,是对政府决策正确性的有效保障,那么特定行政程序的结果或特定的政策措施从一开始就已经明确了结果,并依此执行。此时,决策失误导致的后果是否应当由行政主体来承担,由哪一个行政主体来承担的问题就变得模糊不清了^[10]。而盲目迷信数据的态度也就演变成了一种数字时代的决定论,这将把公共事务的决策权从行政机关工作人员手中转移至负责编程、输入数据的人员手中,他们同时也享有了路径选择权和计算结果的合法解释权^[24]。

在传统的政府治理模式下,治理主体是人,治理过程和治理结果都体现着人的意志,责任主体也是人。在数字政府治理模式下,PTMLs 嵌入数字政府建设,能够在一定程度上影响数字政府的行政行为。这极易导致数字政府所做出的行政方案、决策中存在机器意志,从而侵蚀数字行政行为的合法性基础,也可能模糊行政责任主体的范围边界。比如,大型语言模型拥有强大的 NLU 和生成能力,但其往往缺乏辨别真伪的能力,从而可能做出不正确的回应。如果 PTMLs 扮演的智能客服在与公民交流时出现服务错误,此时如何界定相关责任、如何进行损失赔付等都没有相关依据,这将在很大程度上损害政府的行政权威。

四、风险应对目标下的数字政府建设完善进路

在 PTMLs 嵌入数字政府建设的过程中,需要从个体权利、伦理风险、权责明晰等角度对其加以规制,从而发挥其技术红利的正向价值,使用开源模型保护相关主体的隐私,提高相关信息的透明度和可信度^[25]。

(一) 以权利保障为核心落实相关法律法规

将 PTMLs 嵌入数字政府建设,虽然能够提升政府的治理效能,但也会为公民、法人和其他组织的法定权利带来现实或潜在威胁。《生成式人工智能服务管理暂行办法》第 9 条、第 11 条明确了生成式人工智能服务提供者的网络信息内容生产者责任和个人信息处理者责任,从而明确了对个人信息进行保护的具体要求。由此,为了坚持行政关系中从行政相对人出发的基本理念、实现人文数字政府建设的根本目标,在公民权利与生成式人工智能技术带来的利益产生冲突时,必须坚持公民权利的优先性,保障公民权利。

一方面,在 PTMLs 辅助数字政府建设的过程性行政行为时,要有效保障公民、法人或其他组织等行政相对人的实体性信息权利。这一过程的信息收集和利用范围既包括行政机关的内部数据,也包括公民的个人数据。因为算法的透明度和可解释性均较低,PTMLs 在辅助决策过程中使用的信息范围不特定,所以公共数据、企业商业秘密和个人敏感数据极易在未获得明确授权的情形下被使用,进而导致相关主体权益被侵犯。因此,在 PTMLs 嵌入数字政府建设的过程中,使用者要严格遵守《中华人民共和国个人信息保护法》《中华人民共和国数据安全法》的相关规定,保护公民和组织的合法权

益。其一，要贯彻知情同意原则，切实履行个人信息处理告知义务。在 PTMLs 嵌入数字政府建设的过程中，不能以牺牲个人合法权益为代价，也不能不履行法定的告知义务。就具体内容而言，个人合法权益包括个人拥有对自己的数据的控制权、知情权、访问权、更正权、删除权等。其二，要以法定规范为依据，严格遵守个人信息处理规则。PTMLs 的程序运行和结果生成依赖于海量数据，但是基于合法行政原则、比例原则等规则，不能为了追求生成内容的准确性和高效性而超越法律法规规定的限度。结合数字技术的发展趋势，未来应当围绕涉及 PTMLs 的人工智能技术等进行专门立法，制定保护个人信息的专门规范，可以参考欧盟《人工智能法》的规定，从风险管理、风险分级出发，针对不同类型、不同场景中 PTMLs 对个人数据权利可能产生的影响或侵害，设置相应的风险评估、限制使用等行为规范。相关的民众调查结果显示，大多受访者支持制定数据隐私法规，在制定个人信息保护规范方面，存在着广泛共识^[26]。

另一方面，在 PTMLs 嵌入数字政府建设的过程中，要保障各类主体的自主决定权。第一，要尊重各类主体的自主选择权。行政机关在提供公共服务的过程中，不能出于高效便民的目的，强制各类主体必须使用 PTMLs，也不能在各类主体未表示同意的情况下强迫其使用 PTMLs。第二，在行政机关运用 PTMLs 提供公共服务、进行行政决策时，要履行释明告知义务。行政机关要及时提示和告知 PTMLs 运行时可能使用的信息内容，例如通信信息、登录信息、设备信息等，从而有效保障使用者的知情同意权。第三，要完善各类主体的救济机制，保障其权利实现路径。在 PTMLs 嵌入数字政府建设的过程中，要保障各类主体的删除权和撤销权。《生成式人工智能服务管理暂行办法》第 15 条规定：“提供者应当建立健全投诉、举报机制，设置便捷的投诉、举报入口，公布处理流程和反馈时限，及时受理、处理公众投诉举报并反馈处理结果。”行政机关借助 PTMLs 提供公共服务、做出行政决策后，若权利主体在合理合法的范围内主张行使删除权或撤销权，行政机关应当履行配合义务。若在技术上难以实现，行政机关应当及时告知权利主体，并要求服务提供者履行相应的配合义务，不再对相关数据进行分析处理和扩大化应用。

（二）以价值理性为取向构建伦理保障机制

目前，关于 PTMLs 对数字政府建设功用的讨论主要集中在提升政府效率的领域，具体包括决策效率、治理效率、服务效率和监督效率的提升等内容^[27]。对于在行政行为做出过程中因模型技术参与而可能产生的伦理风险，需要用价值理性弥补工具理性的局限，以制度和规范强化技术伦理，最大限度地发挥 PTMLs 的工具性作用，以此助力数字政府的智慧性、高效性转型。

针对 PTMLs 嵌入数字政府建设引发的行政偏见、行政歧视等不公平问题，需要保证人的价值的优先性，要通过适当的制度构建以确保行政行为做出过程中的伦理正当。当前，由于大模型的出现，生成式人工智能的智能水平急速提升，但其“智能为何能够涌现，其涌现的相变点受哪些因素影响，目前均为未知”^[10]。在涌现原理未知、算法具有不可解释性的情况下，“算法黑箱”“算法偏差”等问题对人的价值的威胁一直存在。因此，应当以“以人为本”的价值理念作为价值引领，并完善相应的治理方案。

第一，构建人工智能伦理框架，完善数据伦理规范体系，以价值理性指导立法。从根本上看，要依法对 PTMLs 的应用进行有效的规制与监督。但从 21 世纪的法律规制与科技发展的互动经验来看，立法存在明显的滞后性。目前，就 ChatGPT、“文心一言”等产品而言，相关行业主要遵循自我监管模式，例如通过联合倡议和自愿承诺等方式进行自我监管。从比较法的角度来看，欧盟制定的《人工智能法》意在规制人工智能系统的具体利用和相关风险，对我国有参考意义。该法律通过明确具有技术中立性的人工智能系统的定义，并“基于风险的方法”对人工智能系统进行分类，为相关主体设定了义务^[28]。该法律将人工智能系统可能引发的风险分为四类：不可接受的风险、高风险、有限风险和轻微风险，并对每类风险规定了不同的义务。此外，该法律还涉及通用人工智能模型的使用，并对违

反法律的公司设定了罚款标准。欧盟的其他相关人士也在积极敦促国家投入更多资源以监管人工智能技术的发展,从而避免私营部门和公共部门之间的信息不对称问题和应对措施缺乏问题^[29]。有鉴于此,我国可以尝试建立相应的算法社会影响评估制度,在引入 PTMLs 做出行政行为时,进行不同层级的社会影响评估,评估行政行为对不同群体及利益相关者可能产生的影响,并做出相应的保障措施以减少负面影响。此外,国家数据局等相关部门也应该牵头制定 PTMLs 技术在数据利用、开发、应用等方面的行业标准,从而将社会公共价值融入其中,确立相应的规范。

第二,提高 PTMLs 应用的透明度以消除潜在的伦理风险。为避免因便捷性而产生的算法依赖,应保证算法解释的实现,理解并监督算法决策,以此规避“以算法为中心”的公共服务结构,防止行政相对人的知情权、参与权、异议权和救济权等重要权利被“虚置”^[30]。对于政府部门及模型技术提供商而言,需要充分保障行政主体、行政相对人的算法解释权,以实现公共部门在行政过程中使用模型技术的透明化。具体来说,要保证透明度、坚持可解释性原则,通过公开算法细节、公开设计用户界面、进行算法影响评估、提供培训和教育、建立申诉机制、制定相关法律法规等方式,充分保障行政主体和行政相对人的算法解释权。此外,要建立相应的外部审计机构,通过公民或管理员报告制度,进一步以公众参与的方式实现技术透明的多向闭环,并赋予监督人在必要时要求相关主体采取改正措施的权利^[31]。

(三) 以权责清晰为基点优化责任结构关系

遵循权责一致原则,公权力的行使必然伴随着责任的规范承担。虽然 PTMLs 会深度嵌入行政治理活动,但由于其暂时不具备独立的责任主体身份,所以在生成式人工智能造成损害时,应当对其开发者、生产者、使用者和监管者的责任范围和责任比例进行合理界定。

一方面,应当以主体类型为标准,明确各方的责任范围和责任内容。为完善责任承担机制,实现法律责任对法律主体的强制制约,应当对责任类型、责任内容、责任承担方式等方面做出具体规定。首先,要明确具体的责任类型。要以现行的部门法律为依据,明确不同义务的责任属性,如民事责任、行政责任和刑事责任。应当明确开发者、生产者、使用者、监管者等不同主体的责任范围,完善不同责任类型的构成要件,为责任界定和责任追究提供规范指引。其次,要细化不同责任类型的责任内容。例如,在民事责任方面,可以设定赔偿损失、恢复原状、消除影响等具体的责任承担方式;在行政责任方面,可以设定罚款、吊销执照、责令停产停业等处罚措施;在刑事责任方面,则需要明确罪名和相应的刑罚。制定这些具体规定时,应充分考虑技术发展的特点及其可能带来的风险和挑战。同时,还要注重平衡各方利益,确保规定的公平性和合理性。

另一方面,在 PTMLs 嵌入数字政府建设的过程中,行政机关有着使用者和监管者的双重角色。从使用者的角度而言,应当明确第三方机构和行政机关的责任范围和责任比例,避免重复承担和责任真空。在技术赋能的行政决策活动中,若出现决策失误或片面决策等问题,甚至是对公民造成了权利减损和义务增加时,要增加并明确公民的救济渠道和救济方式。例如,可以借助区块链技术固定和存储证据,为后续的调查环节提供证据支撑,方便确定相关的责任主体,明确主体的责任类型和责任内容。从监管者的角度而言,行政机关应当对 PTMLs 嵌入数字政府建设的各个过程进行全面监督。虽然 PTMLs 提升了数字政府的运行效率,但行政机关仍需对价值理念和运行原则进行审查和监督。数字政府的运行不仅是利用数字技术进行治理,更是治理空间和治理场域的延伸^[32]。因此,在行政审查、技术应用、内容输出等环节,行政机关要进行意识形态把控,保证决策过程和决策内容在政治导向和政治立场上的正确性。

因此,未来在制定相关法律法规和政策时,应当关注数字技术嵌入引发的政府内部结构再造、业务流程重塑和政府公共服务方式变革,鼓励公民通过获得信息、积极表达和采取行动等方式提升参与能力^[33],同时要细化人工智能技术服务监管规则,保障 PTMLs 在数字政府建设中的健康应用与发展。

五、结语

本文聚焦行政关系中所涉及的行政主体、行政相对人、行政行为等，明确了 PTMLs 嵌入数字政府建设所可能带来的行政关系样态迭代。数字技术对传统行政关系的重塑必将带来新问题、新挑战。面对行政相对人的数据权利被侵扰、行政伦理基本原则面临挑战、行政责任主体模糊不明等情形时，行政主体应当以权利保障为核心落实相关法律法规、以价值理性为取向构建伦理保障机制、以权责清晰为基点优化责任结构关系。只有这样，数字政府在引入和应用 PTMLs 时，才能更好地兼顾效率、安全性和公平性，同时提高政府行政的透明度和公民对政府的信任度。

注释：

- ① 根据 OpenAI 的隐私政策，ChatGPT 模型有权随时收集用户的 IP 地址、浏览器类型及设置、用户与网站互动的数据。另外，它还会收集用户在不同时间以及不同网站上的浏览活动信息。

参考文献：

- [1] 杜孝珍, 王琳. 生成式人工智能技术嵌入数字政府建设: 形态、效应与调适[J]. 天津师范大学学报(社会科学版), 2024(4): 78-89.
- [2] 汪波, 牛朝文. 从 ChatGPT 到 GovGPT: 生成式人工智能驱动的政务服务生态系统构建[J]. 电子政务, 2023(9): 25-38.
- [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(4): 45-56.
- [4] 张涛. 生成式人工智能训练数据集的法律风险与包容审慎规制[J]. 比较法研究, 2024(4): 86-103.
- [5] 周智博. ChatGPT 模型引入我国数字政府建设: 功能、风险及其规制[J]. 山东大学学报(哲学社会科学版), 2023(3): 144-154.
- [6] 熊杰, 张晓彤. 生成式人工智能的数据风险及其合规治理: 以 ChatGPT 为样例[C]//《新兴权利》集刊 2024 年第 1 卷: 智慧法治学术共同体研究文集. 2024: 51-61.
- [7] 徐艳红, 罗敏. 人工智能的社会风险及其规避之道[J]. 湖南社会科学, 2024(4): 123-131.
- [8] 康骁. 行政法如何应对生成式人工智能: 基于算法、训练数据和内容的考察[J]. 云南社会科学, 2024(4): 80-88.
- [9] 秦瑞标. 大模型时代生成式人工智能治理的逻辑更新与回应路径: 以《生成式人工智能服务管理暂行办法》为视角[J]. 时代法学, 2024, 22(3): 28-41.
- [10] 张洪雷. 生成式人工智能参与数字政府建设的技术跃迁、目标导向与可行路径[J]. 南昌大学学报(人文社会科学版), 2023, 54(4): 100-109.
- [11] 张夏恒. ChatGPT 的逻辑解构、影响研判及政策建议[J]. 新疆师范大学学报(哲学社会科学版), 2023, 44(5): 113-123.
- [12] 於兴中, 郑戈, 丁晓东. 生成式人工智能与法律的六大议题: 以 ChatGPT 为例[J]. 中国法律评论, 2023(2): 1-20.
- [13] 王益民. 数字政府[M]. 北京: 中共中央党校出版社, 2020: 94.
- [14] 王浦劬, 季程远. 新时代国家治理的良政基准与善治标尺: 人民获得感的意蕴和量度[J]. 中国行政管理, 2018(1): 6-12.
- [15] HUANG K, WANG Y, ZHU F, et al. Beyond AI: ChatGPT, Web3, and the business landscape of tomorrow[M]. Cham: Springer Nature Switzerland, 2023: 280.
- [16] 毕文轩. 生成式人工智能的风险规制困境及其化解: 以 ChatGPT 的规制为视角[J]. 比较法研究, 2023(3): 155-172.
- [17] 何炼红, 朱曦青. 论人工智能数据公共领域深度共享机制的构建[J]. 中南大学学报(社会科学版), 2024, 30(6): 33-48.
- [18] ATTARD J, ORLANDI F, SCERRI S, et al. A systematic review of open government data initiatives[J]. Government Information Quarterly, 2015, 32(4): 399-418.
- [19] 谭九生, 范晓韵. 算法“黑箱”的成因、风险及其治理[J]. 湖南科技大学学报(社会科学版), 2020, 23(6): 92-99.
- [20] 罗德刚. 行政伦理的基础价值观: 公正和正义[J]. 社会科学研究, 2002(3): 11-14.
- [21] LOUKIS E, SAXENA S, RIZUN N, et al. ChatGPT application vis-a-vis Open Government Data (OGD): Capabilities, public values, issues and a research agenda, international conference on Electronic Government[M]. Cham: Springer Nature

- Switzerland, 2023: 95–110.
- [22] 孟飞, 冯明宇. 数字资本主义的技术批判与当代技术运用的合理界域[J]. 东北大学学报(社会科学版), 2022, 24(4): 1–8.
- [23] 孟庆国, 崔萌. 数字政府治理的伦理探寻: 基于马克思政治哲学的视角[J]. 中国行政管理, 2020(6): 51–56.
- [24] FADAVIAN B. Chancen und Grenzen der algorithmischen Verwaltung im demokratischen Verfassungsstaat[J]. HMDPraxis der Wirtschaftsinformatik, 2018: 294–314.
- [25] MAMALIS M E, KALAMPOKIS E, KARAMANOU A, et al. Can large language models revolutionize open government data portals? A case of using ChatGPT in statistics.gov.scot[C]// Proceedings of the 27th Pan-Hellenic Conference on Progress in Computing and Informatics. Lamia Greece. ACM, 2023: 53–59.
- [26] ONWUKA I A, OKEKE O K, MELUGBO D U, et al. Facilitators and barriers influencing the integration of ChatGPT-based models in enhancing the delivery of public services within government institutions in Nigeria[J]. Ianna Journal of Interdisciplinary Studies, 2023, 5(1): 220–230.
- [27] 王祯军. 算法应用于社会稳定风险评估的可行性、法律问题及对策[J]. 法治研究, 2022(2): 114–125.
- [28] 刘绍宇. 论数字时代个人信息保护的风险规制路径[J]. 西南政法大学学报, 2024, 26(4): 142–158.
- [29] 曾雄, 梁正, 张辉. 欧盟人工智能的规制路径及其对我国的启示: 以《人工智能法案》为分析对象[J]. 电子政务, 2022(9): 63–72.
- [30] 刘甜甜. 智能刑事司法中算法解释的实践原则与具体路径[J]. 法理——法哲学、法学方法论与人工智能, 2024, 10(1): 195–213, 382.
- [31] SELBST. An institutional view of algorithmic impact assessments[J]. Harvard Journal of Law & Technology, 2021: 155–161.
- [32] 杨清望, 唐乾. 生成式人工智能与个人信息保护法律规范的冲突及其协调[J]. 河南社会科学, 2024, 32(12): 81–93.
- [33] 孟天广. 政府数字化转型的要素、机制与路径: 兼论“技术赋能”与“技术赋权”的双向驱动[J]. 治理研究, 2021, 37(1): 5–14, 2.

Risk management of introducing pre-trained language models into digital government construction: From the perspective of administrative relationships

CAO Dongying

(College of History and Administration, Yunnan Normal University, Kunming 650500, China)

Abstract: From the perspective of the relationship between administrative subjects and objects, introducing pre-trained language models into the construction of digital government can promote the communication upgrading between administrative subjects and their counterparts, facilitate the multi-dimensional improvement of internal and external processes of administrative actions, and enhance the rational level of administrative decision-making. While facing up to the positive effects brought by pre-trained language models, it is necessary to clearly recognize the risks that they may cause, such as the infringement of the data rights of administrative counterparts, the challenge to the basic principles of administrative ethics, and the ambiguity of the administrative responsibility subjects. To deal with risks, administrative entities should implement relevant laws and regulations with rights protection as the core, construct ethical guarantee mechanisms with value rationality as the orientation, and optimize the responsibility structure relationship with clear rights and responsibilities as the starting point.

Key words: pre-trained language model; digital government; administrative relations; ChatGPT; digital governance

[编辑: 郑伟]