

行政治理中的情感计算技术应用及其风险预防

黄雪英

内容提要:情感计算系基于目前的人工智能等技术,对人类此前未用科学描述过的情绪、感情等“主观”内容作出相对理性、精确分析乃至预测的技术。该技术的应用毫无疑问让行政治理迭代、高效,突出表现在催生跨物理空间的新治理场景、实现高精度的治理对象侧写、转型智慧型的行政治理模式等。但是,行政治理的无序利用将会产生至少三个层面的风险:因技术限度造成了无法产生对于治理来说所需要的精确性、因技术本身的要求产生对信息权等基本人权侵犯的高度风险以及因技术的不当应用造成的公权力扩张之风险。有鉴于此,应当基于法治精神实现情感计算技术之于行政治理应用中的内部和外部规制。内部规制上,将个人信息权益等作为保障核心并融入伦理原则形成情感计算的行动方案;外部规制上,搭建风险预防框架并确立规范、价值上的引导路径。

关键词:行政治理 情感计算 信息权益 法治精神

黄雪英,广西财经学院马克思主义学院副教授。

一 问题的提出

党的十九大提出了“智慧社会”这一概念,以顺应互联网、区块链、大数据、人工智能等高新技术的快速发展。^[1] 在国家和政府治理领域中,建立“数字政府”或“数字法治政府”的呼声也愈发高涨。顾名思义,“数字法治政府”意味着政府在提供公共服务、开展行政治理过程中的数字化、智慧化、法治化建设。在此种智慧化、智能化的转型趋势下,情感计算已然成为数字行政治理的关键性技术,行政治理的智慧化转型需要依赖于情感治理的数字化发展。从意志要素角度看,在弱人工智能时代,以 ChatGPT 为代表的新一代生成式人工智能即便已经初具强人工智能的某些特征,^[2]但其仍无法完全满足强人工智能

[1] 参见帅奕男著:《智慧社会的司法范式转型》,知识产权出版社 2021 年版,第 1 页。

[2] 参见令小雄、王鼎民、袁健:《ChatGPT 爆火后关于科技伦理及学术伦理的冷思考》,《新疆师范大学学报(哲学社会科学版)》2023 年第 4 期,第 123-136 页。

自我意识、自主学习、自主决策的突出表征,^[3]更不必说情感过程的真正“智慧化”。从情感要素角度看,当前的人工智能体对情感领域的涉足较为贫乏,多集中于人类思维、操作过程的拟人化、智能化模拟,无法实现各方主体间的真正共情。特别是在法治国家、法治政府的建设视域下,程序化、机械化、固定化的行政治理往往因为缺乏对行政相对人的情感共鸣以及对社会公众的需求响应而与整体目标相背离。在此基础上,将情感计算嵌入数字法治政府建设领域能够一定程度上回应公众的微观需求和国家的宏观需要,以行政治理的智慧化转型推动国家治理能力和治理体系的系统性优化。^[4]

在传统行政治理过程中,感知、判断特定主体的情感状态十分耗费精力,对执法人员的能力有极高的要求;而在审讯或长期的治安秩序维持过程中,执法工作效率更会随着时间跨度的延伸而逐渐降低。情感计算技术能够将上述负面影响降至最低。它不仅能够大幅降低执法人员的时间和精力损耗、提高行政效率,还可以借助相关技术产品提供的科学分析提升行政执法活动的合法性与合理性。同时,现代的人工智能与情感计算技术能够通过深度学习不断增强自身的独立分析和判断能力,这也能够弥补人工智能体和人类智能体之间在决策能力上的差距,充分发挥人工智能之于人类智识的效能优势,使行政执法技术工具从传统的“收集信息型”升级为“智能分析型”。

然而,情感计算在推动数字行政治理发展的同时也带来了一些新的挑战和问题。一方面,情感计算在行政法中的应用是一个前沿领域。实际上,我国是世界上情感计算主要的研发和应用国家之一,目前关于情感计算的应用已分布多个行业,包括医疗、^[5]教育、^[6]商业、就业、执法等等。^[7]但遗憾的是,从法律视角切入对情感计算的主要风险与规制路径的研究较少,并不能为实践应用提供充分的理论供给,这使对情感计算的规制与约束不足,并由此引发了一系列问题:包括危及人的自由与尊严、造成人的客体化、权力行使溢出侵犯公民隐私权等等。那么,如何保障情感计算的恰当、合理、准确应用,就是亟待理论予以解答的问题。另一方面,情感计算在行政法中的应用要求对一些传统概念进行反思。譬如,情感信息到底是敏感信息还是一般信息,其所分属的信息层级不同,行政治理的使用要求和程序也相应不同;又如,情感计算在行政治理当中的运用会因伦理、技术等方面的困境而折损公众对政府的信任度,如何维护和强化行政治理的权威性、公信力也是亟待重视的问题。

基于此,需要在明确情感计算技术的运行逻辑基础之上,就其在数字行政治理中的风险点进行归纳、分析或者预测,并在行政法治原则之下结合技术手段规避其可能存在的风险,以确保数字行政治理中情感计算能够在法治轨道上运行。

[3] 参见莫宏伟:《强人工智能与弱人工智能的伦理问题思考》,《科学与社会》2018年第1期,第14-24页。

[4] 参见杜振雷、刘金婷、史金鹏:《ChatGPT及其核心技术在科技名词规范化中的应用潜力与挑战》,《中国科技话语》2023年第4期,第45-54页。

[5] 参见岳远雷、徐着雨:《医疗情感计算应用的多维困境及其法治实现》,《医学与哲学》2022年第19期,第39-44页。

[6] 参见龚善要:《教育场景中情感计算的应用风险及其法律规制》,《复旦教育论坛》2022年第6期,第40-46页。

[7] 参见王禄生:《情感计算的应用困境及其法律规制》,《东方法学》2021年第4期,第49-60页。

二 情感计算与行政治理的交融实证

自“情感计算”(Affective Computing)这一术语在上世纪被皮卡德(R. W. Picard)提出以来,^[8]有关该项技术的理论研究与实践应用便不断涌现。可以将情感计算列为人工智能的分支,而且有一定的适用范围。近年来,情感计算的应用已经十分广泛。同样地,情感计算技术在嵌入数字行政治理的过程中,不仅能通过技术逻辑提升数字行政的高效化、智能化水平,还能通过情感交互促进数字行政的智慧化、亲民化转型。

(一)跨物理空间的新治理场景

情感计算技术的嵌入能够实现行政治理场景的虚拟化建构,将散乱的数字化行政治理统一于完整的治理场域中。目前,情感计算在虚拟场景中的应用主要存在于司法活动中。虽然在这种拟真式的智慧诉讼服务中,情感计算只是为合成个性化语音和完成人脸表情建模服务的,其本身并不直接面向用户感知与分析其情感状态,但类似技术的兴起却意味着情感维度已经可以通过情感计算被纳入虚拟场景搭建的过程之中。在可预见的未来,实时性、具象化、沉浸式的法律交往和协作已经呼之欲出:参与者的各种生理信号都可以通过情感计算技术进行特征提取与情绪展示,使得虚拟场景中的“数字分身”具备与现实相一致的情绪生成和表达能力。

基于此,情感计算的嵌入,将有可能使得行政治理更契合情感治理的目标要求,以情感识别和情感分析为行政治理提供内生性动力,以更好地发挥信息技术的工具性价值。其一,情感计算技术能够更好地发挥人工智能体的中介作用,实现行政管理者与行政相对人的情感互通与情感交互,以行政治理虚拟场景的高度拟真性增强行政治理者对行政相对人的情感识别与情感回应能力,最终借助行政行为或制度调整回应公共服务需求。其二,情感计算技术的嵌入能够在实现情感治理的同时提升行政服务效能,以个性化的情感识别和情感分析为方式,结合行政相对人的情感状态和情感变化提供针对性、具体性的行政服务,在提升行政治理效能的同时也提升了政府的公信力。此外,在理想的虚拟交互场景中,情感计算是针对治理者和治理对象双方同时进行的,这使得数字行政治理不仅能够满足各方的拟真性需求,还能完成对传统“面对面”式行政治理模式的超越。以往的情感治理聚焦于面向个体或群体的治理向度,基于“情理法合一”的治理需要,情感计算更能契合面向国家或整体社会的情感治理需要,通过产生、整合情感治理资源构建国家权力的情感基础和社会基础。^[9]基于此,情感治理与数字行政都能够在某种程度上超越现有的体系。对“人”及其情感的关切不仅要求治理者密切关注治理对象的情感状态,还要求被治理者能够适当理解治理者所肩负的责任。在此过程中,情感计算不仅为搭建一个和谐的虚拟治理场景提供可能性,还为数字行政治理活动设定了限制框架:情感计算使治理者

[8] See Rosalind W. Picard, *Affective Computing*, The MIT Press, 1997, p. 42.

[9] 参见向德平、向凯:《情感治理:驻村帮扶如何连接国家与社会》,《南开学报(哲学社会科学版)》2020年第6期,第84-93页。

消极的情感反应无所遁形,而为建构一个健康的行政治理面貌,各治理主体必须抱持诚恳的态度,否则便无法在情感计算中获得积极反馈。

(二)高精度的治理对象侧写

情感计算技术能够推动实现行政执法活动的情感识别与情感量化,降低主观因素的理解偏差,全面提升行政执法活动的整体效能。在传统的行政治理活动中,智能化工具更侧重于从便捷度、高速度、低成本等维度提升行政执法效率,往往忽视了人民群众的情感需求和真实需要。然而,由于情感具有一定“神秘性”,^[10]借助情感计算适配行政执法技术工具升级实为必要。近年来,已经有相关企业推出了一系列情感计算技术产品,并且已经在北京、广州、杭州等地完成了项目试点测试。该系列产品的核心技术是基于“前庭情感反射理论”,通过视频动态分析精神生理参数来监测潜在可疑危险人员,即通过摄像头采集的人体面部视频流,分析头部和颈部微小肌肉振动的频率和振幅,计算出个人压力、侵略性和焦虑度等参数,甄别有自体原发性焦虑紧张状态的潜在可疑危险人员,并将受测人员的状态区分为“休息或安静状态”“正常或平和状态”“焦虑或紧张状态”和“激动或攻击状态”等,以辅助安检人员进行排查。与此类似的技术或产品还包括用来掌握被讯问人心理和生理反应的“灵视多模态情绪研判系统”以及用来进行情绪监控和分析的“审讯场景的无感知情绪检测分析系统”等。

基于此,情感计算的嵌入能够实现行政机关与行政相对人的多模态情感交互,借助智能体的情绪表现力实现对于行政执法资源的全面整合与准确分析。相较于单模态的情感交互,多模态的情感计算更能够结合特定对象的背景信息实现高效、准确的情感识别与分析。^[11]一般而言,情感计算的输入信号包括视频信号(或图像信号,包括肢体动作、面部表情、姿态特征等)、音频信号(或声音信号,包括音色、语调等)、文本信号(包括词性、词频、语词位置等)和生理信号(包括体温、脑电、皮肤电、心电、呼吸频率、眼动等)。在单模态条件下,情感计算的基础来源于不同数据对象的情感特征,分析主体往往借此明晰数据对象的情绪表达和个体立场。但是,单模态的条件下,情感特性存在一定的语义缺失,再加上情感特性的获取方式单一,难免会导致情感计算的结果出现偏差或失误。当然,多模态与单模态并非对立排斥的关系,多模态的情感计算主要包括基于单模态的数据处理和情感计算以及多模态之间的情感融合两个环节。^[12]基于此,在行政执法过程中,情感计算技术能够依据视频、音频、文本和生理信号等信息进行单模态识别,结合多模态数据分析,通过模拟人脑对于情感数据的融合与处理,实现执法对象内心情感特征和心理活动的拟真化复刻,从而提高行政执法的针对性和准确性。

(三)智慧型的行政治理模式

情感计算技术通过在行政治理活动中嵌入情感要素,为机械化、固定化、冰冷化的行

[10] 参见王峰:《人工智能的情感计算如何可能》,《探索与争鸣》2019年第6期,第89-100页。

[11] See Sidney K. D' mello, Jacqueline Kory, A Review and Meta-Analysis of Multimodal Affect Detection Systems, 47 *ACM Computing Surveys* 1-36 (2015).

[12] 参见饶元、吴连伟、王一鸣等:《基于语义分析的情感计算技术研究进展》,《软件学报》2018年第8期,第2397-2426页。

政治治理赋予情感温度,推动行政治理的智慧化转型。一方面,情感计算技术的嵌入有效提升了行政治理活动的针对性和交互性,更加契合“建设人民满意的政府”的目标。在行政治理活动中,拥有公权力的行政机关与行政相对人的地位实质上存在一定差距,虽然行政机关借助公权力、强制力在一定程度上也能实现政绩要求和政治目标,但其可能由于忽视与社会公众的情感联结而与真正的治理目标相背离。在行政治理活动中,人民是当之无愧的评价主体,人民群众的真实感受是行政治理活动的重要评价指标。而情感计算技术的嵌入则能够提高行政机关关切政务服务对象情感、态度和立场的能力,通过治理主体与治理对象的情感交互与情绪表达,帮助治理主体真切地了解治理对象对于行政行为的情感状态与立场倾向,为治理主体贯彻高效便民原则提供方向指引和资料来源,从而提高行政治理活动的可行性与针对性。例如,政府部门可以通过情感分析技术了解民众在社交媒体等渠道上的情感反馈,进而了解民众对某些政策的态度和看法。毫无疑问,这种具有针对性、具体化的技术路径,能够在极大程度上满足治理对象的异质化需求,从而将情感治理的基层叙事延展至国家治理的宏观维度,最大限度满足提升人民群众的获得感、幸福感、安全感的宏观治理需要。

另一方面,情感计算将行政治理的场域由外在行为规制拓展到内在情感共鸣,有利于推动行政治理的情感治理转向。从朴素的法理角度看,法律的调整对象存在一定的局限性,无法对人的思想进行规制,仅能对人的外在行为进行规范。在传统的行政治理活动中,无论是方法的革新还是技术的升级,行政治理活动的调整对象始终是人的外在行为,一定程度上也会将行政治理置于形式主义的尴尬境地。但是,情感计算技术能够将行政治理的场域延展至治理主体与治理对象的内在情感共鸣,通过对情感特征的识别、处理、模拟和影响,实现治理活动外部效益和内在结果的统一,从实质角度提升人民群众的满意度。例如,有学校曾经推出“AI 双师课堂”,通过“慧眼系统”对每个学生的面部进行情绪识别,既能发现学生是否认真听讲,还能辨别出学生的“高兴、悲哀、惊讶、正常”等情绪,基于这些内心的反馈效果,老师可以更有针对性地进行教学设计。此道理亦适用于行政治理当中,通过情感的反馈判断行政相对人的内心想法,进而选择最为合适和具有可接受度的行政行为,以确保行政治理实现标本兼治之效果。更为重要的是,行政治理活动的开展基础和决策内容拥有更为可靠的数据来源,社会公众的情感需求和情绪表达能够从实质上约束行政治理主体的职能膨胀和权力扩张,防止行政公权力的滥用。

总而言之,以情感计算为代表的人工智能等新兴信息技术对我国的行政治理体系已经且必将持续产生深刻的影响。面对数字时代下的智慧化转型及其隐含的风险,治理者必须与时俱进,正确对待技术发展所带来的正面和负面影响,及时调整行政治理观念,革新行政治理方略,完善行政治理手段,如此方能在时代的浪潮中稳步推进国家治理体系和治理能力现代化。

三 情感计算对行政治理的负面影响：现实与可能

在风险社会视域下,技术性风险与制度化风险同生共存。^[13] 基于此,情感计算在嵌入数字行政治理的过程中,虽然能够带来一定的技术红利,但同时也展现出技术的负面性与风险的不确定性等特征,对治理结果的准确度、情感信息的安全性以及行政权力的公共性产生不利影响。

(一) 技术限度难以达到理想高度

行政治理的有效性和准确性应是情感计算在数字治理过程中的重要目标。一般来讲,数字治理包括“基于数字化的治理”与“对数字化的治理”两个维度,^[14]前者是将数字化作为工具和手段,后者是将数字化作为对象和内容。情感计算作为数字技术,虽然能够提升数字治理的效能,但也会引发一系列的数字化治理问题。一方面,情感计算过程中的情感信息与生物信息存在范围交叉,极易通过收集和分析环节影响情感计算结果的准确性。“情绪识别系统”是以自然人的生物特征数据为基础对自然人的情绪或意图进行识别或推断的人工智能。其中,“生物特征数据”属于“生物信息”中的“生物反馈信息”,^[15]而“自然人的情绪和意图”则属于“情感信息”的范畴,只有对两者进行有效界定,才能厘清应重点收集和在输入端输入何种信息,继而进一步展开情感计算。同样地,也有研究指出,在进行文本数据、语音数据、视觉数据、生理数据的收集和分析过程中,也会涉及到人脸信息、声纹语调等生物信息。^[16] 从外在表现上看,情感信息与生物信息的范围存在交叉,二者的表现介质也存在一定重合,从而加大了廓清二者范围界限的难度。

现阶段情感计算的第一步是通过各类传感器实现对自然人各类生物信息的“感知”,感知阶段数据来源多样,包括声音、表情、肢体语言、姿势等各类生物数据。在情感计算过程中,“感知”是“识别”的前置性条件。在“感知”阶段,情感计算智能体要收集声学特征、面部表情、生理信号等生物信息,而后借助情感识别模型对丰富的数据源进行识别和分析,形成“输入生物信息——输出情感状态”的技术闭环。基于此,在情感计算过程中,生物信息是情感状态的数据来源和数据基础,进而导致生物信息的属性地位出现模糊,在信息收集和信

[13] 参见吴汉东:《人工智能时代的制度安排与法律规制》,《法律科学(西北政法大学学报)》2017年第5期,第128-136页。

[14] 参见蔡翠红:《数字治理的概念辨析与善治逻辑》,《中国社会科学报》2022年10月13日第5版。

[15] 生物反馈信息的收集与处理是为了识别信息主体的情感状况,具有一定的主观性和难以辨别性。参见张会云、黄鹤鸣:《基于异构并行神经网络的语音情感识别》,《计算机工程》2022年第4期,第113-118页。

[16] 参见之江实验室、德勤中国、中国科学院文献情报中心等:《情感计算白皮书(2022年版)》,上海科学技术出版社2022年版,第21-23页。

死循环。在信息处理环节,数据源的不准确可能会导致针对性信息的遗漏,不仅加重了信息处理的压力,还可能导致情感计算结果的错误和偏差,继而影响行政决策的科学性和准确性。

另一方面,现有技术的发展程度与准确性要求的实现尚存在距离。文本情感计算、语言情感计算、面部表情情感计算是情感计算三种具有典型性的应用方式。具体而言,其一,文本情感计算包括文本信息采集、情感特征提取以及情感信息分类三个部分,具体实现路径为“文本信息采集模块通过文本抓取工具获得情感评论文本,并传递到下一个情感特征提取模块,然后对文本中自然语言文本转化成为计算机能够识别和处理的形式,并通过情感信息分类模块得到计算结果”。^[17] 其二,语言情感计算则是指通过对语音信号的测量、分解、分析、合成等方法来进行情感方面的计算,从而使计算机具备一定程度的情感能力。如,在医疗系统中探索适用的情感聊天机器人便是一种典型的语言情感计算,其本质是一种包含信号分析和自动语言识别、语义分析和对话策略、响应生成和语音合成组建的自主系统,可以使用情感计算技术与病人进行交互,以检测情感,确定并模拟情感答案。^[18] 其三,面部表情情感计算与语言情感计算的运行机理基本相似,只是将所收集分析的对象变成了人脸表情模态,通过对面部表情的分析来辨别其蕴含的情绪。如,浙江某学校便探索适用了全国首个智慧课堂行为管理系统,“通过摄像头捕捉同学们的面部表情和动作,可以对课堂上学生的行为、表情等进行统计分析,并对异常行为及时反馈”。^[19]

通过对上述运行机理的剖析,可以发现其核心技术逻辑在于“通过情感测量和建模的过程,从定量的方面将情感客观化。”^[20] 要想将此技术落地,需要满足以下三个前提:(1)一个人的内在情感需要能够通过外在表现予以展示;(2)上述内在情感的外在表现需要在提取上具有普遍性;(3)这种提取需要足够的可靠性能够支撑作出决策。遗憾的是,情感本身的主观性、自主性更强,一个生理信息与某一情感之间并不具有唯一性的联系,可能会承载多重感情。整体而言,心理生理数据非常复杂,难以证明其与情绪之间的因果关系。^[21] 而且,不同的文化传统与社会背景,个人的成长环境和社会经历,甚至在仅有细微差别或相同的情况之下都可能对人的感情产生影响。现有技术的不成熟不仅难以有效界定情感信息与生物信息,还无法准确地对多样化、各有差异的情感进行准确标记。

(二) 缺乏限制造成人权侵犯

理性与情感是影响用户作出决策的两类因素。在数字时代,科学技术对人的操纵性

[17] 张瑞:《文本情感计算研究综述》,《企业与工商管理》2017年第13期,第28页。

[18] See Joachim von Braun, Margaret S. Archer, Gregory M. Reichberg & Marcelo Sánchez Sorondo, *Robotics, AI, and Humanity*, Springer, 2021, pp. 205-206.

[19] 纪驭亚:《智慧课堂行为管理系统上线 教室“慧眼”锁定你》,《浙江日报》2018年2月17日第9版。

[20] 朱敏:《情感计算展现强大应用潜力》,《社会科学报》2023年1月19日第4版。

[21] See R. D. Ward & P. H. Marsden, *Affective Computing: Problems, Reactions and Intentions*, 16 *Interacting with Computers* 707, 707-713 (2004).

和支配力逐渐强化,甚至异化了人的主体地位。在情感计算的视域下,情感操纵是指技术平台基于情感识别完成对特定对象的信息投放,通过利用、控制或影响其情绪的手段操纵特定对象的情感决策而获得收益。^[22]由此可见,情感操纵的前置基础在于情感信息的收集,其本质上是将被特定对象困于“情感茧房”中。在此过程中,无论是情感信息的获取,还是情感信息的计算都极易侵犯主体的隐私权。

首先,情感信息一定程度上可以落入隐私权的保护范围内。情感作为人的一种内心感受,并不愿意为外人所察觉或知悉,自然人在社会生活中甚至会有意识地对个人情绪进行伪装。《民法典》第1032条第2款规定:“隐私是自然人的私人生活安宁和不愿为他人知晓的私密空间、私密活动、私密信息。”那么,当特定主体的某一种情感是不愿意被他人所察觉或知悉的,应肯定该种情感具有一定的秘密性和敏感性。《民法典》第1034条规定个人信息中的私密信息适用隐私权的规定。《个人信息保护法》第28条规定:“敏感个人信息是一旦泄露或者非法使用,容易导致自然人的人格尊严受到侵害或者人身、财产安全受到危害的个人信息,包括生物识别、宗教信仰、特定身份、医疗健康、金融账户、行踪轨迹等信息,以及不满十四周岁未成年人的个人信息。”也就是说,当个人情感信息满足上述界定时便属于“隐私”的一种。然而,“情感计算应用的重要假设是可以推断内心的真实情感,并应当以此为基础作出决策。”^[23]在数字行政治理过程中亦是如此,通过对自然人所流露出来的情感信息进行大范围收集,具有对隐私权构成侵犯的较高概率。尤其是当通过一个人的细微生理行为如语言行为、面部表情等便可分析出此人的内心感受、想法与偏好时,便犹如使人处于一个单向透明的“全景监狱”之中,永远处于监视者的目光之下。

其次,情感信息的识别计算过程增加了其因暴露于公共空间而被侵犯的风险。对于主体直接显露在脸上的表情特征,虽然某种意义上达到了“公开”标准,但其仍处于主体的个人控制之下。而在行政机关收集情感信息进行情感计算的过程中,主体的表情数据、人脸信息以及生理信号便被完全置于“公共空间”,此时情感信息的使用方式和公开范围会脱离信息主体的控制,一定程度上构成对个人隐私的侵犯。例如,在审讯过程中,侦查人员能够通过非接触式数据采集等方式收集犯罪嫌疑人的面部特征信息,进而借助情绪识别技术全面直观地了解询问对象的情绪状态和心理活动,通过询问对象的压迫感和焦虑度等因素辅助侦查人员判断其供述的真实性和客观性。在此过程中,即便询问对象的压迫感和焦虑度等因素能够辅助侦查人员提升侦查实效,但上述因素明显具备私密性,难免会加剧提升行政治理效能与保护犯罪嫌疑人合法权益之间的张力,一定程度上产生侵犯特定主体隐私权的风险。

再次,主体失去对情感信息收集方式和收集范围的控制力。为适应信息技术的发展,芬恩等人将人机交互中的隐私分为身体隐私、行为隐私、信息隐私、数据和图像隐私、思想

[22] 参见褚婧一:《情感计算的信息隐私法律风险及其应对》,《交大法学》2023年第3期,第66页。

[23] 王禄生:《情感计算的应用困境及其法律规制》,《东方法学》2021年第4期,第49-60页。

和感情隐私、空间隐私、团体隐私。^[24] 基于此,面部图像、情感状态、行为信息等均属于隐私的规制保护范围内。美国著名的隐私法学者索洛夫(Daniel Solove)也曾提出,个人能够自主选择信息的收集方式和收集范围是隐私保护制度的中心内容,在此制度下,主体对其隐私信息的使用应当享有充分而自由的选择权。^[25] 但是,在情感操纵的情境下,平台能够通过情感识别实现对于特定对象的信息投放,使得特定对象一定程度上丧失公开信息的控制力和决断力,无法控制何种信息将被以数据的形式公开呈现,更无法控制上述信息公开后将用于何种用途,情感信息的私密性在一定程度上的丧失会导致行政治理活动存在潜在的侵犯隐私权风险。

最后,为准确进行情感计算大量收集的信息也加剧了侵害隐私权的风险。人类的情感既是与生俱来的,又可能受后天环境的影响。根据纽约研究所的研究,情景、社会和文化因素都会对情感表达产生重要影响。那么,遗漏某一信息都可能对最后结果的准确性产生消极影响。因此,在进行情感计算之前,需要大范围、大数量地采集关于情景、环境与背景等方面的信息。无疑,当收集的信息越多,可能揭露与展示个人隐私的深度与范围就越大。可以说,这个过程会进一步加剧侵犯个人信息与隐私的风险。

(三) 手段滥用产生权力膨胀

情感计算在给行政治理带来便利的同时,也伴随着一系列公权力不断扩张的风险。具体而言,权力的扩张主要来源于以下三个方面:

其一,“算法黑箱”的专业壁垒会加剧算法公权力的滥用风险。算法本身的技术复杂性和高度专业性以及技术提供方公开算法内生动力不足等因素加剧了“算法黑箱”的形成。^[26] 伴随智能算法的迭代升级,其不透明性和自主性等特征表明算法逐渐脱离了工具化的范畴。^[27] 再加上立法层面对于算法解释权的限制和弱化,在情感计算过程中,情感计算算法的不透明和不公开难免会为行政公权力的扩张提供环境基础。尤其是在情感信息与生物信息存在交叉的情形下,行政机关缺乏透明公开的履职制约,极易产生过度收集个人信息、肆意使用个人信息等行为,加剧算法公权力的滥用风险。

其二,情感计算的“监控”特质会加剧行政治理活动中的主体不平等。私主体相较于行政机关在行政治理活动中往往处于被动的弱势地位。而情感计算在应用的过程中,无疑会将私主体置于行政机关的监控视域下,加剧双方主体地位的失衡。并且,由于“算法黑箱”的存在,多数私主体自身的算法素养缺乏,难以准确识别情感计算的“监控”方式和“监控”范围。情感计算的“监控”特质虽然能够为行政机关开展行政执法、进行行政决策提供数据来源和数据支撑,但治理过程的不透明以及治理主体地位的失衡等同时也表明了算法公权力的肆意膨胀,甚至呈现出异化的趋势。

[24] See Rachel L. Finn, David Wright & Michael Friedewald, *European Data Protection: Coming of Age*, Springer, 2013, pp. 3-32.

[25] 参见[美]丹尼尔·索洛夫:《隐私权与权力:计算机数据库与信息性隐私权隐喻》,孙言译,载张民安主编《信息性隐私权研究:信息性隐私权的产生、发展、适用范围和争议》,中山大学出版社2014年版,第176页。

[26] 参见谭九生、范晓韵:《算法“黑箱”的成因、风险及其治理》,《湖南科技大学学报(社会科学版)》2020年第6期,第92-99页。

[27] 参见张凌寒:《算法权力的兴起、异化及法律规制》,《法商研究》2019年第4期,第63-75页。

其三,情感计算可能会扩张算法提供者的权力进而导致算法监管失灵。在情感计算技术的嵌入过程中,行政机关囿于技术限制和资金缺口,往往采取与第三方技术公司合作的方式。但是,算法提供者作为中介主体,在借助情感计算技术嵌入行政治理活动的过程中,可能会影响行政公权力的实现方式。一方面,算法提供者可能会基于功利主义立场,将自己的主观感觉和主观倾向嵌入情感计算活动中。^[28] 并且,生理信号与情感信息之间并非一一对应的关系,而是存在交叉、重合、分离的多种样态。情感计算算法的治理内容极其复杂,行政机关可能会因为技术限制等因素无法开展有效监管。另一方面,情感计算为算法提供者介入行政治理活动提供了技术可能,决策权力所有者与决策内容作出者可能出现分离。从规范依据上看,行政决策的作出主体为行政机关,但算法技术的嵌入会使算法提供者无形中成为行政决策的实际作出者,行政机关的决策权力一定程度上被虚置,行政主体的算法监管失灵,甚至影响政府的公信力和权威性。

值得一提的是,权力扩张与权利侵害是事物的正反两面。在情感计算造成权力扩张的情况下必然会对行政相对人带来除上文隐私权侵害之外的其他不合理、不合法或不成比例的权利侵害。比如,情感计算的应用对个人控制其情绪状态的自主权造成了威胁,但无论情感计算技术的发展程度如何,都不能将行政相对人的辩解机械化;又如,多模态拟合使算法规则难以理解和解释造成全流程监管困难,从而损害行政相对人的知情权;再如,自动化决策和算法争议带来责任风险等等。

(四) 无序应用催生信任危机

促进政府权威性的树立是数字政府构建过程的重要目标之一。情感计算的应用却可能对政府的权威性产生消极影响。情感计算算法可能偏袒或歧视某个个体和团体。通过分析某区域的情感状态得出该区域居民的特点,导致对该区域居民的偏见歧视。此歧视结果不可避免地会影响行政决策的公正性,成为冲击人的主体性、造成公民质疑行政行为合理性的重要依据之一。实际上,在行政治理过程中运用情感计算更应该以实现人的价值为基本目标。也就是说,任何制度都应当以实现人的主体性为最终价值。当人的主体性都得不到尊重时,政府的权威性更无从论及。此外,情感计算还可能引发盲信技术和懒政等风险。在一些地区,行政机关可能对技术的积极作用、准确性和成熟度的认识不足,导致高估情感计算而忽视行政人员在治理过程中的重要性,产生“以数据技术完善为核心”的错误认知。在此影响之下,可能有的地方政府会习惯性地会将众多行政任务交给情感计算技术来完成,自己成为了“甩手掌柜”,但囿于情感计算有限的准确性,情感计算所得出的结果并不能够得到切实保障,这必然将会减损行政机关的权威和公信力,并不利于法治政府形象在民众心中的确立。

四 情感计算在行政治理中的恰当应用:内部规制与外部框架

在数字行政治理活动中,情感计算是通过情感计算技术对人类情感状态的感知、识

[28] 参见徐琳:《人工智能推算技术中的平等权问题之探讨》,《法学评论》2019年第3期,第152-161页。

别、模拟和影响实现的。^[29] 情感计算技术的嵌入,能够实现行政治理的情感化和柔性化转型,提高行政治理的整体效能,但与此同时也存在一定的风险隐忧。因此,我们应当识别并防范数字行政治理活动中的情感计算风险,在发挥情感计算技术优势的同时保障行政相对人的合法权益。

(一) 内部保障:构建信息权益为核心的规制手段

通过情感计算技术原理可知,情感信息的收集和使用是情感计算的中心环节,是为情感模拟和情感影响提供数据支撑、为行政治理活动提供数据来源的前置性环节。基于此,为防止因数据收集使用而侵犯隐私权,同时保障数据的准确性,应首先明确数据收集要求。

第一,在情感数据收集过程中,行政机关要坚持合法原则。当前,我国已经形成了以《个人信息保护法》《数据安全法》《网络安全法》为基础的个人信息保护规范体系。若情感信息具备可识别性且未经匿名化处理,则属于个人信息的范畴,适用个人信息保护的法律规定。例如,人脸情绪识别算法作为人脸识别的特殊变体以及情感计算的重要构成,^[30]其所对应的人脸情绪本身具备可识别性,并与生物体一一对应,当然属于个人信息的保护范畴,行政机关应当在贯彻“知情—同意”原则的基础上合法、合规地收集、存储、加工、传输人脸情绪数据等。

第二,在情感数据使用过程中,行政机关要落实合法性审查。行政机关要通过强化情感数据滥用的内部责任,进行动态的情感数据使用合法性控制。一方面,对于情感算法设计者和提供者,行政机关要审查上述主体的相关资质,对其存在的技术风险和违规行为进行纠查,最大限度上提升情感数据使用的规范程度。另一方面,对于情感算法的使用者和监管者,即行政机关,要适当公开情感数据的使用方式和使用范围等内容,涉及国家秘密、商业秘密和个人隐私等内容的除外,并对履职的监督管理机关进行问责。在情感数据使用的监管活动中,可以借助电子监察管理系统进行全程监管,对数据的异常使用和违规使用进行预警,^[31]以此来确保数字行政符合行政法治原理。

第三,行政机关应当完善情感数据保护措施,具体包括:其一,情感数据的存储和传输应采取加密和安全传输协议,防止未经授权的访问和篡改;其二,情感数据的访问权限应限制在必要的人员范围内,避免非授权人员的访问和使用;其三,在情感数据的处理过程中,不仅需要采取匿名化和去标识化的方法,还应该严格限制情感数据的收集和使用范围,以降低个人信息被识别、追踪、泄露以及侵害的风险。

(二) 内部实现:融入人机交互伦理原则的技术发展

人机情感协同是克服情感计算风险的必经之路。目前,情感计算在行政治理中存在人机脱节的障碍,^[32]而且就其发展水平而言,所扮演只的是辅助性、工具性的角色,

[29] 参见王禄生:《情感计算的应用困境及其法律规制》,《东方法学》2021年第4期,第49-60页。

[30] 参见包康赞:《人脸情绪识别算法的风险与规制》,《北方法学》2022年第1期,第44-45页。

[31] 参见查云飞:《行政裁量自动化的学理基础与功能定位》,《行政法学研究》2021年第3期,第114-124页。

[32] 参见石磊:《智能司法裁判的人机协同机制探索》,《行政与法》2022年第4期,第75页。

故必须走向人机协同。人机协同的本质是以人为本,由人、专家和智能机器三个子系统构成。^[33]

一是实现从提高算法的可解释性向建构可解释的算法模型的转变。人机协同的高度交互和理想状态离不开主体对于算法的理解和解释,为此,行政机关应当制定相关政策使公众得以了解情感计算的运行机制和结果生成过程。具体而言,其一,情感计算的适用者应当提供清晰明确的情感计算过程,包括情感数据的收集、情感计算模型的构建和训练、情感计算结果的生成等。其二,情感计算使用者应当使用可解释的情感计算模型,使用者和利益相关者能够理解情感计算结果的生成过程和依据。其三,情感计算使用者应当确保数据主体能够参与决策和管理自己的情感数据,这可以通过隐私设置、数据共享选项和用户参与机制来实现。其四,情感计算使用者应当公开向用户和公众说明情感数据的收集目的、使用范围和共享方式。值得注意的是,算法透明也存在一个度,但是总体而言,面向行政机关要比面向个人和社会的透明程度高得多。^[34]

二是实现人机情感协同治理的共商共建共享。人机情感协同治理主体主要包括技术开发者、应用提供商、政府监管机构、学术界、用户等各个相关方。各方应当共同参与制定情感计算的标准和规范,并在技术研发、应用设计和监管执法等方面各司其职、密切协作。分别就各个主体而言:第一,情感计算技术开发者应当注重对于可解释算法模型的研发,提升情感计算算法的透明性;第二,情感计算应用提供商应当积极履行算法备案制度,在情感数据告知书中以明显的方式提示用户关于情感数据的使用方式、使用范围、使用主体等内容,确保“知情——同意”原则的贯彻实施;第三,行政机关应当加强对于情感计算算法的事前审查和事后监管,在事前阶段通过数据抽样和算法备案等途径强化对于情感计算算法的审查,在事后阶段通过行政监管和安全评估等方式强化对于情感计算算法的管理;第四,学术界应当加强与技术人员的交流,在明确情感计算技术基础和运行逻辑的基础上,结合行政治理的一般原理,探索可行性的发展建议;第五,用户应当努力提升自身的算法素养,在降低“算法黑箱”的不利影响的同时,增强自身在情感计算过程中的参与感,切实维护自身的情感数据安全等等。

还要提升情感计算算法的道德判断能力。虽然人类的道德要求是蕴含了价值判断的自然语言,难以完全转化为代码,但仍然可以加强情感计算算法自身对道德要求的深度学习和机器训练。主要包括:一是可以通过数据收集和学习提高道德判断能力。即,数据必须具有多样性和代表性。只有当数据来自于不同的文化、社会群体和背景,才能避免行政治理过程中的偏见和歧视,让行政相对人感受到公平公正与温暖。二是可以通过伦理模型和规则更新提高道德判断能力。伦理模型是一种用于指导道德决策的框架,可以基于人类伦理学理论和道德准则进行构建。通过不断迭代和改进这些模型,情感计算算法可以更好地理解和应用伦理原则。举例而言,已有研究基于罗尔斯的“公平的机会平等”理

[33] 参见钱学森、于景元、戴汝为:《一个科学新领域——开放的复杂巨系统及其方法论》,《自然杂志》1990年第1期,第4-5页。

[34] 参见姜野:《算法的法律规制研究》,吉林大学2020年博士学位论文,第81页。

论,引入“歧视指数”的概念,提出设计公平算法的方法,并证明其开发的“公平”算法没有表现出结构化的歧视行为。^[35] 可以此模型为镜鉴将其进一步推广。三是可以利用人类专家的反馈来改进道德判断能力。即,可推动算法与人类专家的交互,使算法接受专家的评估和指导,以此帮助算法纠正错误判断和偏见,并提供更全面、深入和准确的道德分析。

(三)外部框架:行政治理应用技术手段的边界和限度

鉴于情感计算应用利弊兼具的特点,并不是在行政治理的任何场景、任何环节中皆能一概适用情感计算算法。基于此,需要进一步明确情感计算在行政治理中的基本要求以及在不同场景中的应用限度。

第一,行政机关应当建立全面的情感计算适用标准。从宏观层面看,行政机关可以将公开透明确定为情感计算的应用原则,从而为增强情感计算算法的可解释性、接受多元主体的共同监督提供原则基础。从微观层面看,其一,在事前阶段,行政机关应当确定情感计算的适用范围和适用条件,防止算法行政权力的膨胀和滥用;其二,在事中阶段,行政机关应当明确情感数据的收集、存储、分析和传输等规则,在开展行政治理活动的同时保护主体的信息安全;其三,在事后阶段,行政机关应当明确监管主体和检举渠道,在监督情感计算算法规范运行的同时防止行政公权力的过度膨胀。

第二,行政机关须具有使用情感计算的正当目的。情感计算适用的正当目的是为政府决策提供参考和改善公共服务,在行政治理过程中超越此目的或假借此目的之名使用情感计算都应当被视为非正当的使用。如,利用情感计算来测试行政相对人对决策结果的满意程度,从而制定措施实现对相对人的控制或影响便是较为典型的不具有正当目的的滥用。再如,只是为了迅速、高效完成原本难以完成的行政任务,而不是为了更好地实现行政治理效果、全方位保障人民利益而使用情感计算手段,也在一定层度上存在目的偏颇。

第三,行政机关在使用情感计算时不得违背比例原则。比例原则作为行政法中的重要原则,其逻辑起点是人权保障,本质是基于基本公共需要而限制公民基本权利的国家权力的限制。为防止情感计算应用冲击比例原则,若能够通过其他手段替代情感计算对行政治理的积极作用,则应当优先使用其他手段。此外,在技术层面,收集一万人的情感数据与收集十万人情感数据在成本上可能出入不大,后者甚至更有利于情感计算结论的准确性,但公民权利却因数据收集数量的扩大而面临更大的侵害可能,此亦应当被视为对比例原则的违背。

(四)外部内容:伦理、价值和规范上的边缘明确

伴随数字技术的迭代升级,技术资本逻辑的嵌入使得算法的工具属性被过度放大,甚至异化,此时,就需要坚持工具理性与价值理性相统一的原则规范算法的运行。基于此,为最大限度发挥情感计算算法的工具价值,应当将伦理性因素嵌入算法治理过程中,为情感计算应用划清价值底线与伦理禁区。

[35] 参见姜野:《算法的法律规制研究》,吉林大学2020年博士学位论文,第81页。

一方面,在事前阶段,强化对于情感计算应用的价值评估和伦理审查。行政机关应当及时采取措施防止突破价值底线和伦理禁区的问题的发生。具体而言,主要包括以下方面:其一,要对情感计算的算法模型开展价值评估,确保个人的主体性和自治性,^[36]充分保护主体的情感数据安全和平等地位;其二,尊重情感计算应用过程中的主体异质化表达,增强情感计算模型的可解释性和透明性,重视多元化的主体参与;其三,完善情感计算的伦理审查反馈渠道,囿于“算法黑箱”的限制,应当重视专业人员和技术人员的力量,鼓励社会主体对于情感计算伦理审查的广泛参与,以促进情感计算的持续改进和优化。

另一方面,在事后阶段,完善对于情感计算应用的伦理监督。其一,成立专门的伦理委员会由其负责对情感计算项目的伦理监督。该委员会应由跨学科的专家组成,包括伦理学、计算机科学、法律和社会科学等领域的专家,以确保监督的全面性和客观性。其二,加强行政监督和执法。加强对情感计算应用的监督和执法力度,确保情感计算的合规合法,专门设立有效的投诉和举报机制,公众和利益相关者可以通过正式渠道向其他行政机关或独立第三方机构报告涉嫌违规或不当使用情感计算的行为。相关机构应及时处理举报,对涉嫌违规的情感计算应用进行调查和处理。其三,定期审查和评估。这可以通过定期的审核、现场检查和独立评估来实现对情感计算的监督。其四,开展特定教育和培训。通过举办研讨会、培训课程、媒体宣传等形式,帮助公众更好地理解情感计算的概念、原理和应用。^[37]

五 结 语

科技的进步促进了人们技术应用能力的提升,但同时科技的应用也可能带来以往所不能预见的一些问题。情感计算本质上是人工智能算法的一种,相比于以客观行为数据为质料而进行运算的智能算法,情感计算在技术上试图通过一系列数据信息以观测、分析甚至影响人们的情感。这对于建设服务型政府,讲求情理法并重的现代化中国具有重要意义。可以推断,情感计算技术未来必将广泛应用于数字政府的各项服务当中,但在此过程内还需要规避该技术运用的诸多风险。一方面防范技术本身带来相对人权益损害风险,另一方面防范技术运用者的权力扩张风险。故需要采取有针对性的措施防止此类情感计算风险的产生,以防止突破保护公民权利的底线,实现情感计算积极效能的最大化发挥。总而言之,“人始终是目的而非手段”以及“将权力关进制度的笼子里”,应当是一切人工智能技术与政府治理相结合应当遵循的基本原则。

[感谢上海政法学院张西恒副教授在本文修改过程中提供的宝贵帮助。]

[36] 参见张欣:《算法解释权与算法治理路径研究》,《中外法学》2019年第6期,第1425-1445页。

[37] 参见贾开:《人工智能与算法治理研究》,《中国行政管理》2019年第1期,第19-21页;曹建峰、方龄曼:《欧盟人工智能伦理与治理的路径及启示》,《人工智能》2019年第4期,第44页。

The Application of Affective Computing Technology in Administrative Governance and Its Risk Prevention

[**Abstract**] Affective computing has a broad application space in the process of establishing digital administrative governance; firstly, constructing a virtual scene for service administrative governance, so that administrative governance activities can directly penetrate into the emotional or affective dimensions of subjects' personality; secondly, adapting to the upgrading of administrative law enforcement technology tools, so that subjects of digital administrative governance can integrate and analyze multi-modal information and maximally reproduce the inner emotional tendency and position of law enforcement objects; and thirdly, promoting the intelligent transformation of administrative governance, so that administrative governance actions become more concrete and touch people's heart and soul more directly. However, because of the lag of legal research and the complexity of affective computing, it is difficult to regulate and restrict affective computing, leading to a number of risks. Firstly, it is difficult for affective computing to fully guarantee the accuracy of governance results. Secondly, the intrinsic characteristics of affective information make it possible to violate the right to privacy. And thirdly, the application of affective computing can easily lead to the over-expansion of algorithmic public power. In view of the above risks, the following paths need to be constructed to prevent the risks; firstly, to protect information rights and interests by regulating emotional data, that is, to ensure that the use of emotional data complies with the current legal requirements, improve emotional data protection measures, and establish a data internal legitimacy review mechanism; secondly, to embed the logic of emotional computing in the human-machine emotional synergy of the training dimension, that is, to not only improve the transparency and interpretability of affective computing algorithms and establish effective mechanisms for the collection and use of affective data but also promote the establishment of a human-computer interaction system for affective computing with the participation of multiple parties; thirdly, to set standards and limits for the application of affective computing in digital administrative governance, that is, administrative organs should set up standards for affective computing through administrative legislation and policy formulation, and clarify the purpose and limits of affective computing application; fourthly, to delineate the value bottom line and theoretical forbidden zone for the application of affective computing, that is, to determine the principles of value bottom line and ethical forbidden zone, strengthen the ethical assessment and legitimacy review of affective computing, and establish the application supervision mechanism; and fifthly, to add rationality and legitimacy requirements for the iteration of affective computing, that is, to enhance ethical judgment capacity through data collection and learning and through ethical modeling and rule updating and to improve ethical judgment by using feedback from human experts.

(责任编辑:支振锋)