

doi:10.13582/j.cnki.1672-7835.2025.01.013

生成式人工智能传播的文化责任伦理： 理论阐释、风险识别及治理路径

王炎龙,黄婧

(四川大学文学与新闻学院,四川成都610064)

摘要:生成式人工智能的迅速发展,极大提升了社会民生便利性,也造成了诸如虚假信息、文化歧视、隐私泄露等伦理风险的叠加,凸显出现有治理政策与技术之间的不匹配。以“伦理先行”为理论框架,回顾生成式人工智能与科研伦理问题的现有讨论,分析其实际应用带来的伦理挑战。从公共文化责任意识角度,探讨在内容生成技术应用过程中,如何平衡技术创新带来的风险与伦理责任,揭示其在文化生成与认同之间的断层和歧视,以及认知安全引起的文化风险。对此,可通过公共文化服务政策设计中的智能化标准治理,确保数据使用合规,加强文化责任与伦理治理,构建完善的技术治理体系,从而实现全面有效的治理效果。

关键词:生成式AI;文化治理;科技伦理;文化责任;智能传播

中图分类号:G203

文献标志码:A

文章编号:1672-7835(2025)01-0101-09

生成式人工智能(以下简称生成式AI)作为新兴的智能技术,依托深度学习、大规模数据分析与优化算法,能够自主生成文本、图像、音乐及视频等内容,展示出极大的创新潜力。这一技术突破使得内容生成技术在艺术创作、科学研究及社会生产等多个领域得到广泛应用,并逐步渗透到人类日常生活的各个方面。与传统大模型相比,生成式模型不仅在效率上带来了革命性的提升,更凭借其独特的创造性,推动了从创意生成到知识传播等多维度的变革。尤其在公共文化传播领域,内容生成技术的介入已不仅仅是技术革新,更是对文化内容的生成、传播及接受方式的深刻变革。从传统媒体到社交平台,从内容创作到受众互动,生成式AI在文化传播中的功能不再局限于单向内容输出,这就需要重新对技术进行创新与社会责任的双重考量。在此背景下,生成式AI所引发的伦理问题不仅涉及技术使用层面的风险识别,还涉及文化认同的混淆、社会价值观的偏离,以及文化安全的维护。因此,本文旨在通过分析生成式AI在公共文化传播中如何平衡技术创新与伦理责任,在尽力推动文化和科技发展的同时注意规避文化歧视、认知偏差等风险问题。通过回溯“伦理先行”的理论来源,阐明其在现代文化治理,特别是在文化传播中的应用,分析生成式AI在内容生成、数据应用与文化传播过程中可能引发的文化离散、文化歧视以及认知偏差等多维度风险。基于这些风险特征与核心问题,本文将探讨如何通过智能化标准治理、数据资产管理与教育资源适配,防范和化解公共文化领域中的伦理风险,以推进文化责任意识的共建。

一、“伦理先行”与“文化治理”的阐释

科技向善,伦理为先。“伦理先行”是近年来科技治理领域广泛讨论的概念,在生成式AI这一新兴科技的背景下显得尤为重要。生成式AI作为当今科技领域的代表性技术,带来前所未有的创新机会,但其“双刃剑”效应也不可忽视。它在推动社会发展的同时,带来了滥用技术的风险,进而引发一系列

收稿日期:2024-09-21

基金项目:教育部人文社会科学研究规划基金项目(22YJA860006)

作者简介:王炎龙(1972—),男,湖南双峰人,博士,教授,博士生导师,主要从事网络与新媒体、出版理论与实务、公共传播等研究。

全球性社会、经济和文化挑战。因此,许多国家在应对生成式 AI 的技术风险时普遍强调“伦理先行”这一原则。在此背景下,深入理解“伦理”和“治理”这两个核心概念,对理解科技发展的未来走向至关重要。

(一)理论阐释

从哲学和伦理学的角度看,“伦理先行”是伦理思考和技术发展交织的结果,源于对技术与伦理关系的深入探讨。伦理学中的“善”是最为核心的概念之一。苏格拉底曾指出“善”无法直接被定义,但引入了一个概念,“善的儿子,就是那个看上去很像善的东西”^①即“太阳之喻”。他通过“太阳之喻”来说明,善的理念与可知世界的关系就如同太阳给予可感世界的光明一样。善的理念不仅是理解的前提,也是其他理念存在的基础。在这一哲学框架下,“伦理先行”的核心理念是,伦理应在技术发展之前就进行预见与规制,确保科技发展有益于人类社会,推动科技向善。俞可平在《治理与善治》一书中这样概括:“善治就是使公共利益最大化的社会管理过程。”^②“科技理性的膨胀导致人文失落,产生了一定程度的道德失范和信仰危机。”^③前沿领域的科技突破具有带动社会变革的可能,但也会引发价值观念上的冲突和日益复杂的伦理问题。因此,科技的发展不能脱离伦理规范,尤其是在生成式 AI 的应用上,必须优先考虑其对社会伦理、文化价值以及个体权利的潜在影响,避免技术滥用或负面后果的产生。

伦理通常被定义为人与人之间、人与社会和人与自然之间的相互关系应遵循的道理和准则。正如马克思所言:“我们的一切发明和进步,似乎结果是使物质力量成为有智慧的生命,而人的生命则化为愚钝的物质力量。”^④人经常依靠情感判断并抉择,所以人不可能具有绝对理性。反而,一个依赖形式逻辑的机器能够实现完全的理性,因此具备成为完全道德主体的潜力。在此语境下,“科技伦理则是开展科学研究、技术开发等科技活动需要遵循的价值理念和行为规范,是促进科技事业健康发展的重要保障。”^⑤“伦理先行”的理念最初用于医学领域,规范医疗实践与研究中的伦理问题,而后逐渐被引入到更广泛的科技治理领域。就文字表达而言,“伦理先行”即伦理“先行”,是开展科技活动前,主动进行科技伦理的风险评估,涉及科技伦理风险、达到科技伦理审查规范要求的,必须要及时开展审查^⑥。“伦理先行”是对伦理审视的呼吁,更是科技发展不可忽视的前提。就科技和伦理的关系而言,前者是对人类技艺的延伸及技术层的突破,其对社会、经济、文化的影响带来了伦理道德可能的后果;后者则是对行为和决策的规范和引导,以确保科技的应用不会超越社会道德的边界,保障公众利益与社会稳定。从这一关系看,“伦理先行”是把伦理道德融入生成式 AI 研发和应用的全生命周期,体现以人为本的理论内涵,实现对人格尊严的维护^⑦。具体到生成式 AI,伦理先行要求在技术研发与应用过程中,优先考虑其对社会伦理、文化价值、个体权利的影响,避免对社会结构、文化认同及公共伦理带来负面影响。

将文化构想成一个原初的“构造”过程,在形成各种社会问题和历史事件方面,其重要性不亚于经济和物质基础,不再单纯是事件发生以后对世界的反映^⑧。福柯强调通过自由显示出治理的“治理术”^⑨观念为“文化治理”理论发展提供了思路。当文化被视为治理对象时,关注点是“指涉着下层社会阶级的道德、礼仪和生活方式”的广义文化;而当文化被视为治理工具时,是“对道德、礼仪和行为符码等领域的管理进行干预和调节的手段”^⑩。本尼特强调了政府在审美和智性活动中的作用,以及如何通过管理和调节道德、礼仪和行为规范等来形塑公共政策和社会结构。因此文化治理是利用随着“现代”

①柏拉图:《理想国》,商务印书馆 2013 年版,第 175 页。

②俞可平:《治理与善治》,社会科学文献出版社 2000 年版,第 8 页。

③王学川:《科技伦理价值冲突及其化解》,浙江大学出版社 2016 年版,第 31 页。

④《马克思恩格斯文集(第二卷)》,人民出版社 2009 年版,第 580 页。

⑤《关于加强科技伦理治理的意见》,《人民日报》2022 年 3 月 21 日。

⑥杨舒:《科技向善,伦理先行》,《光明日报》2022 年 3 月 24 日。

⑦商建刚:《生成式人工智能风险治理元规则研究》,《东方法学》2023 年第 3 期。

⑧斯图尔特·霍尔:《表征——文化表征与意指实践》,商务印书馆 2003 年版,第 6 页。

⑨史蒂文·卢克斯:《权力:一种激进的观点》,彭斌译,江苏人民出版社 2008 年版,第 92 页。

⑩托尼·本尼特:《文化、治理与社会:托尼·本尼特自选集》,王强等译,东方出版中心 2016 年版,第 210 页。

阶段产生的知识和专门技术等具体文化形式的发展而产生的各种文化手段塑造人类总体^①,文化治理“作为一种全新的治理范式,既是国家治理的重要组成部分,也是推进国家治理体系和治理能力现代化的重要手段”^②,其本质是通过对文化现象和文化资源的有效管理和调控,推动文化繁荣发展,维持社会稳定。在当代文化治理的理论体系中,生成式 AI 技术作为一种新的治理工具,逐步渗透到文化管理的各个方面,尤其是在公共文化领域的运用,为文化治理提供了新的视角和实践路径,成为推动文化治理现代化的关键力量。生成模型和大数据的结合及普及,使文化治理的互动性和社会参与度得到大幅提升,社会公众可以更方便地参与文化活动、分享文化资源,实现文化治理的协同化、网络化,增强文化治理的全民性和共享性。文化治理与生成模型的结合,是对传统治理模式的补充和完善,更是对新时代文化治理需求的精准回应。文化治理不局限于传统的政策引导和资源配置,还融入了科技创新和社会参与,增强文化活动的影响力与可达性,为实现文化繁荣与社会治理现代化提供有力支持。

(二) 公共文化中的伦理先行

“伦理先行”是科技治理的核心原则,也是文化治理中的重要导向。在公共文化领域,随着社会开放与民主化进程的推进,公共文化已不仅仅局限于传统的文化教育和社会认同的塑造。通过普惠式的文化服务,公共文化在实现社会个体文化平等中发挥着重要作用。政府通过“农家书屋”、调配艺术教育等举措,试图消弭贫困或偏远地区资源匮乏、教育水平较低带来的文化鸿沟,让所有社会成员都能享有文化权益,推动文化普及和公平性,从而在社会文化层面实现了伦理价值。

所有社会活动都发生在一定的空间内,文化服务亦如此。网络的出现使公共文化的传播突破图书馆、博物馆、剧院等物理空间,连结报刊、杂志、电视、网络平台等数字空间共同构建公共文化服务体系,成为社会交往和价值引导的重要场所。然而,在近年的影视剧、新闻报道和网络文化产品中,出现了一些通过内容生成技术伪造假冒的历史人物视频或声线,误导公众对重大历史事件的理解,产生误导性、偏激性甚至破坏社会和谐的文化价值观。所以,在社会问题导向、历史文化造假等方面伦理审核需要更加严格,将社会伦理价值置于优先位置,确保文化内容符合社会共同价值观,帮助社会成员形成正向的历史认知和价值认同。生成式 AI 通过大数据分析和深度学习算法,能够根据用户偏好定制内容,生成文本、图像、音频等多媒体形式的文化产品,从而提高了文化产品的生产效率。然而,这种技术普及的同时也伴随着文化内容的单一性、过度标准化以及伦理风险。例如,社交媒体平台上广泛使用 AI 算法推荐内容,用户只能接触到与其兴趣相符的有限内容,产生“信息茧房”效应,这可能使文化产品趋同,缺乏多元性和创新性;而 AI 生成的广告、视频或文章可能不负责任地放大某些社会问题,助长极端、偏激的观点,给社会伦理和文化价值观带来冲击,甚至可能导致社会分裂、信任危机以及文化认同的混乱。因此,公共文化中的“伦理先行”需要关注文化设施的建设,重视网络平台、社交媒体等新兴文化传播媒介的伦理规范,确保技术工具在文化传播中的使用符合社会伦理标准,避免导致文化内容的机械化、标准化,确保文化表达符合社会多样性与伦理价值^③。生成式 AI 应在文化内容生产过程中,将社会责任和伦理审查作为其技术应用的前提,推动文化的传承和创新,有效防范技术滥用带来的负面影响。

二、公共文化中生成式 AI 传播的风险识别

生成式 AI 大规模生产文化内容,延伸人类技术的背后,是海量数据驱动选取,缺乏对文化背景、社会价值和伦理责任的深度理解。这一过程中,生成式 AI 技术与公共文化的深度融合,在推动文化发展、信息传播和社会互动上发挥着重要作用。然而生成式 AI 在优化文化服务和内容传播的同时,也带来了诸多潜在的风险,尤其是在文化认同、社会公平、认知安全等领域。

①金莉:《本尼特“文化治理性”的理论逻辑与价值意蕴——从“葛兰西方法”到“福柯思想”》,《文艺评论》2017年第2期。

②张森,顾海娥:《文化治理的理论源流、反思与现实路径》,《社会科学辑刊》2024年第1期。

③王海建,郝宇青:《社会治理数字化的伦理挑战及其应对》,《郑州大学学报(哲学社会科学版)》2024年第4期。

(一) 生成式 AI 与公共文化的互融性偏离

生成式 AI 的出现提高了信息内容的生产、传播和分发效率,推动文化生活便捷化,与公共文化的关系应是互利共生的。生成式 AI 作为文化创作与传播的技术工具,推动了公共文化的创新与变革。通过智能化的数据处理与创作算法,生成模型拓展了文化创作的形式与内容,提升了创作效率与多样性。在文学、艺术等领域,生成模型为创作提供创新支持,突破传统创作的边界,突破物理空间的限制,使公共文化能够通过数字平台传播。个性化推荐系统使文化产品能精确匹配受众需求,提升文化传播的效果与覆盖面。因此,生成式 AI 与公共文化的互融不仅推动了文化创作与传播的智能化,也为公共文化的多元化开辟新的路径。

而实际应用中,随着生成式 AI 的广泛渗透,公共文化的“技术大众化”和“伦理先行”之间的张力逐渐显现,尤其在文化认同建构上,内容生成技术带来的偏离现象愈加明显。生成式 AI 改变了文化的生产方式。文化生产过程中,技术和信息资源大众化使文化生产不再局限于精英或专业创作者。据 Open AI 披露,GPT-3.5 的文本语料多达 45TB,相当于 472 万套中国四大名著,而 GPT-4 在此基础上又增加了多模态数据^①。内容生成技术帮助公众拓展了文化创作的范围,增加了文化内容和文化主体的多样性。然而,这种普及化的创作方式也带来了一定的风险,即生成的文化内容在一定程度上趋向同质化。由于生成模型依赖已有的文化成果进行训练,它的创作方式有时显得“旧瓶装新酒”,缺乏独创性与创新性,但生成式 AI“狗尾续貂”的创作手法过度依赖训练数据,容易形成文化生产的“机械化”和“标准化”,进而抑制文化的多元性与创新性。

生成式 AI 在文化传播过程中所面临的偏离现象与文化认同失衡密切相关。传统文化传播主要依赖语言这一文化认同和社会价值传递的载体。生成式 AI 虽然可以模仿人类的表达方式,并生成语法上符合要求的内容,但其生成内容缺乏伦理框架的支持,容易产生价值观的失衡。例如,由于 AI 无法完全规避训练数据中的偏见,它可能会无意中放大社会中的性别歧视、种族偏见、地域歧视等负面信息,导致“语言污染”,影响用户的认知和判断力,进而加剧社会群体之间的对立与分化。在此背景下,确保生成式 AI 在语言生成过程中的“伦理先行”变得尤为重要,尤其要加强对内容的伦理审查和把关,防止因模型训练数据的不完全性或偏差而损害文化认同。

生成式 AI 技术通过算法推荐推动信息传播,这种基于历史行为数据的个性化推送虽然扩大了文化传播的速度和规模,却容易形成“信息茧房”。用户在这种环境下,接触到的文化内容往往局限于符合其认知背景和兴趣的领域,而难以接触到有挑战性的、跨文化的或跨领域的内容。其应用虽然推动了文化创作的技术化与高效化,但也带来了文化创作的去人性化现象。传统文化的生产 and 传播通常依赖个体创作者的独特视角、情感表达和社会背景,而生成式 AI 则通过历史数据与统计模型生成内容,缺乏人类创作者的主观性与情感联结。这种机器化创作方式使得文化内容变得单一和表面化,长期依赖 AI 生成内容可能导致公众对文化的认知与情感体验逐渐疏离,忽视文化的深度与复杂性。生成式 AI 在文化创作中的使用,容易忽视伦理与社会责任,可能产生误导性、偏激甚至虚假的文化信息。生成模型产出的新闻报道、历史叙述或社会问题讨论,缺乏人类判断和伦理审核,可能导致对历史的曲解或社会问题的过度放大,形成社会认知偏差。此外,文化的传递方式变得机械化和数据化,模型在生成内容时无法充分传达传统文化的社会性、情感性与道德伦理,使得文化传播的内涵与价值观出现缺失。从而使文化创作的个性和人性化特质受损,导致生成式 AI 与公共文化的互融偏离。生成式 AI 虽然带来了文化传播的高效性,但也使文化内容的创作和传播逐渐去人性化,削弱了文化的独特性与人文内涵。

(二) 文化歧视与技术歧视的交织性偏见

生成式 AI 的使用,推动了公共文化的生产与传播,但也暴露了其潜在的风险,特别是文化歧视与技术歧视的交织性偏见问题。在数据处理和信息传播的过程中,生成式 AI 的学习模型依赖海量数据输入,这些数据通常来源于社交媒体、网络平台和数字化服务。由于历史数据本身可能带有社会偏见,AI

^①李广建,潘佳立:《人工智能技术赋能情报工作的历程与当前思考》,《信息资源管理学报》2024年第2期。

在训练过程中不可避免地吸收并放大了这些偏见。

数据库中的固化显性偏见。生成式 AI 的算法设计,虽然看似中立,但实际上在某些情况下也携带了设计者的价值观。生成式 AI 在训练过程中需要分析大量的数据,往往无意识地吸收和延伸了数据中的偏见。这些数据不仅反映了某些社会群体的文化认知与行为习惯,还通常夹杂了种族、性别、地域等方面的偏见。AI 模型训练中的数据偏向可能会使得特定文化、语言或身份群体的需求和价值观被忽视或误解。当训练数据不断将“某一族群”和“负面标签”相关联时,AI 模型就容易输出带有偏见的内容,误导大众认知。例如,“Pro Publica 的一项调查发现,执法部门使用的人工智能工具显示黑人比白人更有可能犯罪”^①。输出信息内容和传播时,数字歧视加剧社会群体间的不平等,无形中加剧了文化歧视。

算法设计中的隐形偏见。生成式 AI 的算法设计,虽然看似中立,但实际上在某些情况下也携带了设计者的价值观。数据本身的不准确,算法规则代码的不透明,以及运行程序中道德规范的缺失可能衍生出算法歧视^②。算法歧视和文化偏见通过技术手段延伸放大,进而导致特定文化、语言或身份群体的认知与表达被技术化忽视,形成了某些群体在数字时代的“技术歧视”。例如 AI 模型通过学习历史数据中的典型模式来进行预测和推荐,而这些模式可能是特定文化背景下的“常态”。因此,技术本身并不具有天然的“公平性”,反而不可避免地嵌入某些文化假设和判断,这种“隐性偏见”在 AI 应用中形成了“技术歧视”。在技术的设计和优化过程中,数据的选择、特征的定义、模型的调优等环节,都会影响 AI 系统对不同文化群体间差异的处理,进而影响文化表达和文化认同。

隐性和显性歧视的交织。在生成式 AI 的应用中,文化歧视与技术歧视可能通过隐性方式影响公众对文化认同的理解和群体之间的互动。智能算法技术逻辑的本质就是根据特定的价值偏向进行信息排序,在信息过滤后将优先级更高的信息投喂给受众。“基于受众本位的算法机制,会通过大数据勾勒出的受众兴趣偏好而投送同质性信息,导致个体被包裹在同一价值偏向营造的‘气泡’性空间中”^③,使不同群体之间的文化交流和思想碰撞减少,加剧社交“信息茧房”的形成,影响多元文化的交流与融合,进一步强化了社会群体之间的认知分歧。例如,AI 系统可能倾向于生成更符合大众口味或主流文化偏好的内容,忽略少数民族、女性或其他边缘群体的文化需求,这种“文化边缘化”的现象在数字时代尤为突出。而文化表现的单一化与边缘化,又与社会不平等和技术鸿沟的边缘化群体重叠。老龄群体或少数民族群体等边缘群体,可能因为技术设备的不完善、信息技术的学习障碍或互联网的普及性问题,无法平等地享受生成式 AI 带来的文化创新成果。更重要的是,这种技术鸿沟可能进一步固化社会不平等,使数字时代的文化传播呈现出“二元化”的格局。

(三) 认知安全与社会风险的叠加性共振

在文化传播和生成式 AI 技术的场景下,除了文化歧视与技术歧视的交织性偏见问题外,生成式 AI 的潜在风险还在于其可能引发虚假信息的传播、侵犯隐私权以及引发文化危机等社会风险。

虚假信息的生成与传播是生成式 AI 认知域安全的外在风险。生成式 AI 在数据、算法和算力三要素的支撑下生成内容并传播信息,在处理大量数据过程中,可能会因过度拟合而产生 AI“幻觉(hallucination)”风险,即大语言模型常生成的内容对提供的源内容无意义或不忠实^④。受限于训练数据的有限范围、信息的时效性不足,以及可能潜藏的错误与偏见,生成式 AI 和输入法预测一样,生成的文本不总是真实的。其唯一“目的”似乎就是生成文本,因此即使“无法回答”某问题,生成式 AI 依旧会提供看上去正确但实际上错误的内容,从而扰乱公众判断、影响社会舆论。这些虚假内容通过社交平台 and 数字媒体迅速传播,极大影响公共文化的真实性和可信度。虚假信息的传播不仅误导公众对历史、文化、社会

①闫坤如:《人工智能的算法偏差及其规避》,《江海学刊》2020 年第 5 期。

②张琪,林佳怡,陈璐,等:《人工智能技术驱动下的人力资源管理:理论研究与实践应用》,《电子科技大学学报(社科版)》2023 年第 1 期。

③方正:《数智化赋能铸牢中华民族共同体意识的传播样态、潜在风险与路径建构》,《民族学刊》2023 年第 8 期。

④Ji ZW, Lee N, Frieske R, Yu TZ, Su D, Xu Y, Ishii E, Bang YJ, Madotto A, Fung P. Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, 2023, 55(12): 1-38.

问题的认知,还加剧了社会群体之间的撕裂。生成式 AI 的普及,使信息传播的速度和广度大幅提升,导致虚假信息的传播不仅仅局限于个别事件,甚至对整体文化认同造成长期损害,引导大众的认知偏差。

用户隐私侵犯是生成式 AI 应用过程的隐形风险。生成式 AI 算法离不开数据的收集,用户应用大模型的过程中,其个人基本信息、兴趣、情感,甚至是思维方式的行为数据不断被采集以支撑模型生成个性化内容,再推送回用户。隐私被信息技术重新包装,放入虚拟电子黑箱,隔绝于人的感官,“被‘稀释’处理之后,以‘中性化’的数据再现”^①,通过对用户的个人数据进行深度分析,影响用户的思想和决策,进而挑战个人认知的自主性,形成技术对个体思维的隐性操控。当 AI 技术缺乏监管时,用户的个人数据可能被未经授权地收集、存储和使用,导致隐私泄露。这不仅侵犯了个体隐私,还可能加剧公众对 AI 技术的不信任,进而影响到公共文化传播的效果和社会的认知稳定性。这种隐私侵犯的隐性风险与文化传播的开放性相冲突,容易造成技术与公众之间的疏离,甚至引发社会的不安和恐慌。

文化危机的共振是生成式 AI 发展路径上的“梦魇”。生成式 AI 的滥用不仅加剧了虚假信息的传播和隐私侵犯的风险,还可能引发更深层次的文化危机。当 AI 推荐系统和生成算法被某些极端观点或文化叙事操控时,AI 技术的传播可能充斥着极端政治思想、错误的历史叙事或其他有害的社会观念,这些内容在网络和社交媒体上迅速蔓延,进一步扭曲了公众的认知与文化认同。同时,当特定的文化认知和社会观念在多平台上得到广泛传播时,弱势群体的文化表达往往难以突破技术平台的算法壁垒,从而被忽视或扭曲。生成式 AI 通过选择性过滤、推荐机制操控文化话语和公共舆论,引发文化上的偏见和歧视,导致社会的文化断裂和社会身份的对立危机。认知安全的威胁与社会风险相互交织,形成了数字时代特有的社会风险链条,呈现出叠加性共振的特点,这一特点威胁到个体的认知自主权和社会群体的文化认同,扰乱公共秩序。因此,在面对生成式 AI 的风险挑战时,需要从多维度构建合理的治理机制,不仅要技术角度进行风险防范,更应从社会伦理和政策层面提出应对策略,以实现技术与社会的和谐共生。

三、生成式 AI 传播的文化治理路径

传统的文化治理方式依赖粗放型政策设计,但大数据、互联网技术的发展,使公共文化服务的治理更加精细和精准,具备动态调整的能力。公共文化服务作为社会文化体系的重要组成部分,应立足我国生成式 AI 发展阶段的客观规律,坚持“伦理先行”,提高文化服务效率,实现服务均衡。

(一) 公共文化服务政策设计中的智能化标准治理

公共智能化服务的个性化定制。在传统的文化服务模式,各区域、各群体的文化需求被混为一谈,出现城乡“二元化”等资源分配严重不均衡的问题。某些群体的文化需求得不到充分关注,但另一些则可能得到过度的资源倾斜。而生成式模型通过对大数据的实时分析和深度挖掘,能够精准捕捉到不同群体的文化需求变化,极大地提升资源配置的针对性。为此,政府应加大对偏远地区网络基础设施和智能硬件的建设投入,推动 5G 网络、光纤宽带等关键基础设施的普及。鼓励企业和社会资本参与公共文化服务智能化建设,尤其是通过公益性项目向边远地区提供便捷的文化服务平台,确保智能化技术的普惠性。再借助各地区大数据的反馈分析特定群体的需求特征,定制个性化和精准的文化服务策略,确保文化资源能够惠及各个社会层级,促进数字资源的公平分配。

智能化技术的标准化监管。在内容生成技术应用的过程中,文化服务的监督不仅要依赖技术手段,还需结合相关法律进行有效的制度化监督。在数据监控和算法透明度方面,政府可以通过法规要求所有使用人工智能算法的文化项目,在执行过程中公开其算法模型和决策过程,保障文化项目实施的透明度与公正性。维系社会秩序的两套规则中,法律比道德更具强制性和效力。“法律研究以伦理为先导”^②,《生成式人工智能服务管理暂行办法》从技术进步、服务规范化及过程监管等多维度入手,为应对

^①金观涛:《赛先生的梦魇:新技术革命二十讲》,东方出版社 2019 年版,第 93 页。

^②毛子骏,朱钰谦:《人工智能的国外社会科学研究热点综述》,《电子科技大学学报(社科版)》2023 年第 2 期。

短期内可能涌现的技术风险提供了政策保障和操作指引。针对内容生成的可信性与质量问题,中央网信办多次开展专项行动,就利用生成式 AI 技术制作发布涉未成年人低俗色情图片和有害信息^①、人工智能生成信息标注以及虚构摆拍内容标注等突出问题乱象,开展从严整治^②。生成式 AI 技术的快速变化和不可预测性,使得传统的事前监管规则和事后监管手段常常难以有效应对技术迭代带来的新风险,因此,结合“伦理先行”和现有监管机制,从数据安全保障、算法应用监管和信息内容治理等方面,构建涵盖生成式 AI 全生命周期的综合治理体系,以应对技术风险并及时进行政策调整。

智能化文化服务政策的标准化创新。在智能化治理模式下,公共文化服务政策需要融入更多的科学性与前瞻性,以应对日益变化的社会需求。通过智能化技术,建立一套符合现代需求的标准化体系,以确保文化服务的质量与效益。同时,文化服务的创新路径也呈现多元化趋势。文化服务不仅仅局限于传统的物理空间,还可以通过数字平台、虚拟空间等多样化载体进行传递。因此,制定适应这一新需求的技术标准与政策规范,推动文化服务创新,已经成为公共文化治理中不可忽视的重要内容。具体而言,“政府应积极接入生成式 AIGC,建构“技术—平台—生产者—使用者—生态”五位一体的政务智能媒体传播格局,实现精准传播,驱散政治迷雾,纠正用户认知偏误,促成社会共识”^③,提升文化服务的质量与效率,也为文化产业的未来发展提供了更加广阔的空间。

(二) 公共文化安全决策中的智能化数据资产管理

智能化数据资产管理是提高公共文化决策效率、确保文化安全的重要基础。在确保数据安全的同时,高效管理公共文化资源,利用智能化手段进行文化决策和安全管理,已成为亟待解决的问题。

技术研发,内容生成模型的数据可追溯。高新技术研发阶段,研发者的责任自觉、道德规约、行业自律极为关键。“这既是人类抵御技术风险与灾难的第一道单薄而脆弱的防线,也是技术治理的基础性环节。”^④为了避免技术滥用与文化歧视,技术设计者必须将“以人为本”的理念内化到技术开发的每一环节。具体而言,采取可解释性设计,以确保算法大模型的透明性与可追溯性,使得技术应用过程中的每一项决策都能接受社会的监督与法律的审查。同时,“需加大网络空间合法信息的保护,加强对各大主流媒体、自媒体、网络平台的培训和引导,将中华民族共同体意识培育作为算法运行的重要依据,避免‘算法偏好’形成信息推送偏差”^⑤。此外,内容生成技术在公共文化决策中的应用不应仅仅将技术效益作为唯一衡量标准,还必须充分考虑其对社会文化结构的潜在影响。尤其在文化传播领域,算法推荐系统必须避免单一价值观或文化霸权的倾向,而应尊重和保护的多样性与包容性。技术设计者通过可解释性设计来确保算法的透明度与可追溯性,使得技术应用过程中的每一步都可以接受公众和相关监督机构的审查,这对于构建可信赖的生成式 AI 服务体系至关重要。

训练过程,内容生成模型的数据可解释。训练过程是内容生成模型构建复杂知识体系与优化策略的关键环节,其内在机制和生成路径需通过多层次的解析加以揭示。训练过程的数据处理影响着算法模型面对不同输入时的反应特征及其对社会文化背景的敏感与适应性。因此,训练数据的选择和整理必须经过严格筛选与审定,以确保生成内容的合理性、合规性和文化适应性。在智能化数据资产管理的背景下,模型训练的每个环节都应基于明确的逻辑依据,且每一层次的生成模型都涉及复杂的参数调整与权重优化。为此,训练过程的透明性要求所有决策和操作均可追溯并具备清晰的解释,接受理性与逻辑的审视。生成内容的来源、构建过程及其生成路径等静态数据处理成为可解释性的核心,通过对生成过程的详细追溯和解析,识别潜在的误差或不合理输出,以及对特征权重、输入条件和参数调节的精细

①中央网信办秘书局:《关于开展“清朗·2023年暑期未成年人网络环境整治”专项行动的通知》,中国政府网,https://www.gov.cn/zhengce/zhengceku/202306/content_6888642.htm?share_token=8e860c69-0599-47fb-9898-46744faf54c6。

②《中央网信办部署开展2024年“清朗”系列专项行动》,中国网信网,https://www.cac.gov.cn/2024-03/15/c_1712088026696264.htm。

③曾润喜,秦维:《人工智能生成内容的认知风险:形成机理与治理》,《出版发行研究》2023年第8期。

④田永鸿,韩冬雪,吴满意,等:《人工智能与马克思主义理论学科交叉研究》笔谈,《西南交通大学学报(社科版)》2023年第4期。

⑤丁忠毅,程桂龙:《智能算法赋能铸牢中华民族共同体意识的机理与优化策略》,《民族学刊》2023年第10期。

化审视,确保生成输出符合预定的文化安全标准,帮助研究人员理解模型生成过程中的每一细节,确保其结果具有高度的透明性和可解释性。

隐私保护,内容生成模型的数据可控制。平衡创新与隐私、技术与伦理是智能化数据管理成功实施的关键。内容生成模型通过对海量数据的学习与分析,能够为决策提供高效精准的支持。但这项技术的应用同时伴随着隐私泄露、知识产权侵权等潜在风险,亟须加强监管与管理。“在算法模型设计层面降低人工智能生成内容对单一在先版权作品的依赖,强化生成内容与训练数据之间的差异性而非相似性,能够从源头上降低侵权风险。”^①在实际操作中,算法模型涉及大量个人信息与敏感数据,该过程可能存在侵犯用户隐私,引发数据滥用与泄露的法律风险。因此,技术企业需加强对数据使用合法性的监管,建立机制识别与规避潜在的侵权风险,确保数据的可控性,避免未经授权的数据收集与使用,最大限度降低数据侵权的风险。内容生成模型的应用不仅仅是技术层面的挑战,还是法律与伦理的复杂交汇点。加强对模型训练数据的审查、生成内容的规范化管理以及侵权风险的有效控制,已成为保障公共文化安全、提升决策公正性的必要手段。

(三) 公共文化责任意识共建的智能化教育资源适配

“进入知识经济社会以后,各民族之间的竞争从传统的军事‘硬实力’较量转向知识与科技的竞争,知识创新能力成为各国提升综合竞争力的核心关切。”^②生成式 AI 作为现代知识创作的载体之一,要求对公众进行普惠化智能教育,提升文化教育资源的适配性,同时为公众提供更为灵活和个性化的学习体验,使得文化责任教育能够更广泛地覆盖到社会的各个层面。

公共文化服务与责任意识的内生培育。利用智能化技术和大数据分析,精准识别并匹配不同群体的文化需求,是公共文化责任意识共建的关键。技术企业通过分析用户行为、偏好和反馈信息,动态捕捉并理解基层群众的潜在需求,进而提供个性化、定制化的文化教育资源。例如,通过大数据分析,技术企业识别到偏远地区居民对文化活动的参与度较低,或特定年龄段群体在文化学习方面的需求未被满足时,调整文化服务内容与形式,提供线上文化课程、定制化的文化教育资源。在满足用户即时需求的同时,挖掘出用户的潜在文化需求,帮助其更好地理解并承担文化责任。智能化技术和大数据分析不仅是文化服务精准供给的保障,更为公共文化责任意识的内生式培育提供了支持。群众在享受更为个性化、丰富的文化资源的同时,也逐步意识到自身在文化建设中的责任,实现文化服务与文化责任的内生互动。例如,利用大数据算法模型为不同群体设计符合其特点的文化学习及活动,激发其对文化建设的参与意识,形成文化责任感,使文化教育更具针对性和持久性,在社会各层面形成广泛的共识。

智能化技术与公共文化服务的外部驱动。生成式 AI 打破传统文化服务的时空和地域限制,推动文化资源的高效传播与跨区域共享。公众对技术的接受度与其对生成原理及潜在风险的理解密切相关,且应用是否符合社会价值观和个人利益,直接影响技术的社会认同与发展方向。强化全民网络安全与人工智能伦理教育,能够推动公众理性认识技术发展,减轻滥用风险,优化社会环境,为生成式 AI 的“向善”发展创造条件。这不仅是社会治理的必然要求,也对提升公众科技素养至关重要。此外,缩小“数字鸿沟”、支持数字弱势群体,尤其是老年人、残障人士等群体,成为生成式 AI 伦理治理的重要议题。技术发展应遵循普惠性和包容性的原则,推动社会公平与机会均等,需要平等接入技术创新,弥合群体间知识和技能上的差距。政府应加强对偏远地区基础设施建设和数字技能培训的支持,确保各群体平等享有技术进步带来的福利,实现生成式 AI 的普惠与包容,提升社会整体福祉。

结语

生成式 AI 的全面落地应用正蓄势待发,不同于传统信息技术,生成系统在研发和应用之初就会产

^①邵红红:《生成式人工智能版权侵权治理研究》,《出版发行研究》2023年第6期。

^②郑淳,刘长军:《“去理想化”:对西方国家高等教育国际化实践类型的再审视》,《西南交通大学学报(社会科学版)》2023年第2期。

生相关伦理问题,体现出明显的伦理始涉性,引发了关于技术创新与伦理责任的讨论与反思。在内容生成技术推动文化生产与传播效率提升的同时,数据偏见、算法不透明等问题却在潜移默化中塑造了文化认同的断裂与异化,进而加剧了社会群体间的不平等与歧视。在这一过程中,虚假信息扩散与隐私侵犯现象频发,突显出公共文化中的认知安全问题,且生成模型的应用往往引发更复杂的社会风险叠加效应。通过对生成式 AI 与科研伦理挑战进行梳理,发现生成模型的应用不仅仅是一个技术问题,也涉及数据治理、算法透明度、信息真实性等层面的伦理考量。因此,在技术研发阶段,必须注重建立相应的标准化规制与数据监控体系,保障公众隐私权和信息的真实可靠;在算法模型训练阶段,应加强算法的透明性与数据筛选,减少社会文化偏见和群体间不平等的可能性;在内容生成与传播过程中,应适当调控文化基础设施与资源,确保技术的普惠性和公平性。由生成模型发展带来的文化责任与伦理风险等问题,需要以“伦理先行”为导向,将伦理规范融入技术的设计与应用全过程,确保技术创新与文化责任并行推进。将技术播种在伦理的土地上,把公平、隐私、透明度、责任等伦理价值观嵌入其设计中,使生成式 AI 生长成“菩提树下的花”,将“道德滋养”贯穿始终,自然中向“善”而生。

尽管对生成模型与文化责任伦理的关系进行了深入分析,研究仍然存在一定局限性。现有分析主要聚焦于国内技术与伦理的平衡,不同文化背景下的内容生产技术应用会面临不同的伦理和治理挑战,全球化背景下跨文化的伦理规范值得进一步探索。生成式 AI 技术不断发展,其在文化传播中的应用将更加广泛,随之而来的伦理问题和社会风险将愈发复杂。因此,未来的研究应继续深入探讨生成式 AI 在文化领域中的伦理治理框架,强化技术全流程监管,推动智能技术在保障文化责任前提下健康发展。

Cultural Responsibility and Ethics of Generative Artificial Intelligence Communication: Theoretical Explanation, Risk Identification, and Governance Pathways

WANG Yanlong & HUANG Jing

(School of Literature and Journalism, Sichuan University, Chengdu 610064, China)

Abstract: The rapid development of generative artificial intelligence has not only greatly enhanced the convenience of social livelihoods, but also triggered a convergence of ethical risks, such as false information, cultural discrimination, and privacy breaches, highlighting the mismatch between existing governance policies and technologies. This paper adopts “ethics first” as its theoretical framework, reviewing current discussions on generative artificial intelligence and research ethics, and analyzing the ethical challenges it poses in practical applications. From the perspective of public cultural responsibility awareness, the study explores how to balance the risks brought by technological innovation with ethical responsibility in the process of content generation, revealing the gaps and discrimination between cultural generation and identity, as well as the cultural risks caused by cognitive security. Based on this, the paper proposes the design of intelligent standards governance in public cultural service policies to ensure compliance in data usage, strengthen cultural responsibility and ethical governance, and build a comprehensive technology governance system to achieve a fully effective governance outcome.

Key words: generative artificial intelligence; cultural governance; technology ethics; cultural responsibility; intelligent communication

(责任校对 曾小明)