

# 人工智能时代政治谣言的算法治理: 逻辑、限度与应对<sup>[\*]</sup>

孙会岩

(上海大学 政党治理与社会发展研究中心, 上海 200244)

[摘要]在人工智能时代,政治谣言通过各类社交平台不断呈几何级数式增长和传播,很容易扰乱公共秩序和社会情绪。近年来,单纯以政府为主导的政治谣言治理方式难以遏制其蔓延,正日益转向依托算法技术的治理模式。但由于程序设计、数据挖掘、机器学习等算法技术固有的双刃剑特点,致使“算法治理”还存在短期内难以克服的“利维坦”局限。因此,既要总结传统依托政府主导和算法治理政治谣言的效能存量,更要加强政府顶层设计、企业协同合作、主流媒体真相引导以及民众担当伦理责任,共同推动“算法—利益相关者”多元协同的治理模式,这样才能形成上下联动、多元参与、深度融合的治理路径,最大限度提升人工智能时代政治谣言治理效果。

[关键词]人工智能时代;社交平台;政治谣言;算法;协同治理

DOI:10.3969/j.issn.1002-1698.2021.02.007

当今世界,半导体、处理器、神经网络等引发了广泛的社会技术变革,互联网与大数据共同推动快速度、多途径、广范围的人工智能时代到来。这不仅改变了民众的交往方式,也助推了“政治谣言”在社交平台加速传播,进而给社会带来诸多不稳定因素。特别是在当前面临百年未有之大变局背景下,各种与政治相关的谣言不断增多,对主流意识形态安全带来很大挑战。面对政治谣言治理难题,一些地方已经突破传统的政府主导治理模式,日趋走向依托社交平台的算法治理,并取得一定效果。但由于算法“内部—外

部”逻辑固有的缺陷,也滋生出一些新问题。因此,准确把握人工智能时代社交平台政治谣言发展的新趋势,从算法维度探讨政治谣言的治理逻辑及其限度,进而从技术与社会协同维度全面审视治理模式,可以为社交平台政治谣言治理提供新思路。

## 一、政治谣言及其在人工智能时代的发展

纵观其演进历程,政治谣言几乎与人类政治文明史同步发展;在当今各种智能媒介加持下,更是呈海量增长趋势。在各类社交平台作用下,

作者简介:孙会岩,政治学博士,副教授,上海大学政党治理与社会发展研究中心副主任,研究方向为技术进步与政党政治。

[\*]本文系国家社会科学基金青年项目“互联网时代中国共产党的执政安全研究”(18CDJ004)的阶段性成果。

政治谣言对原始社会舆情起到推波助澜作用,严重扰乱社会情绪和公共秩序,进而给主流意识形态安全和社会治理提出了新要求。

广义而言,学术界大都认同让-诺埃尔·卡普费雷关于谣言的释义,他认为“谣言是在社会中出现并流传的未经官方公开证实或者已经被官方辟谣的信息”,<sup>[1]</sup>可见,谣言是信息世界中一种没有可靠证实标准的语言表达形式,它赋予与现实有关的某人或某事一些竭力使人信服的新要素。在政治领域,用奥尔波特和波斯特曼的等式来看,谣言=(问题的)重要性×(事实的)模糊性。当民众对信息的需求量与官方对信息的供给量不对等时,政治谣言就会恰逢其时地满足民众的信息需求,既可以传递民众想要的消息,也可以树立或毁坏名声,甚至会导致桑斯坦在《谣言》中所言及的蔑视、恐惧、仇恨和暴力。回顾政治史,不管是清朝乾隆年间的“叫魂妖术”,<sup>[2]</sup>还是法国大革命时期的“大恐慌”,<sup>[3]</sup>政治谣言扩散都给社会造成很大冲击;进入19世纪,民众开始沉浸在报刊、电报等媒介带来的信息便利中,马克思深刻意识到这些“在一天当中所制造的神话,比以前一个世纪之内所能制造的还要多”,<sup>[4]</sup>在商人和政客的推动下,很多媒体以“电讯”之名编织谣言,一度进行“新闻敲诈”;<sup>[5]</sup>到20世纪,随着经济社会的快速发展,“社会越来越服从或依赖于媒体及其运行逻辑,而媒体则成为社会发展不可分割的一部分,甚至成为社会本身”。<sup>[6]</sup>媒体发展不仅没有消灭政治谣言,反而助推其传播得更加专业与广泛。

20世纪中后期,科学技术高速发展,互联网和大数据成为人类社会基本结构的重要部分,不管是媒介产品的供给,还是民众的信息消费,都变得越来越便利,以至于麦克卢汉将这种技术进步带来的普遍性交往范式描述为“地球村”。截至2019年底,我国网民已达8.5亿,手机网民已突破网民总数的99%,智能手机、应用程序和社交媒体等已经把大众带入一个传统媒体与新媒体传播语境交融共生的人工智能时代,各种社交

平台不断争取日渐稀缺的注意力资源,在提供政治信息的同时更是催生政治谣言。虽然把关机制和议程设置也发挥了作用,但由于社交平台的“强关系”属性,以及低度自我纠错能力,加之资本运作“按点击次数付费”逻辑,共同催生了大批造假群体,为谣言极速网状式扩散提供了便利。数据显示,虚假内容在Twitter上传播给1500人的速度要比真相快6倍;在我国,近60%假新闻首发于微博、微信朋友圈等社交媒体。<sup>[7]</sup>可见,社交平台越来越在政治谣言的传播中起到关键作用,不仅使传播速度、传播周期发生改变,也改变了谣言的传播形式与影响范围,这中间有些政治谣言成为影响决策的重要因素,甚至会酿成社会治理危机。

如今人类已经进入人工智能时代,由于主流媒体提供信息的渠道单一、内容有限以及民众对一些地方政府的信任度低,使得通过微博、微信、直播平台等发布的政治谣言一度火热,短时间内形成了强大的舆论效果。这些社交平台中的政治谣言根据对象不同主要分为针对政治人物、政治事件、政治体制三类,政治谣言在“有限注意空间”和“占主导地位的社交媒介融合”影响下极易蛊惑民众情绪,日益成为“叛逆传播的必要手段”。加之民众对政治常识、国家政策、官方消息等不了解,政治谣言在社交平台快速扩散,这些谣言以“被揭露”“被证实”等字眼刻意抹黑党的领袖人物、否定我国政治制度、抨击社会治理方式等,严重损害了党的形象和政府的公信力。因此,人工智能时代社交平台政治谣言治理日益成为学界研究的重要命题,智能技术进步不仅是媒介走向现代化的重要生产力,也是政治谣言治理走向智能化的关键驱动力。发展至今,算法(Algorithm)日益成为新时代的重要关键词,“算法治理”为社交平台政治谣言治理提供了新范式。然而,算法“黑箱”、难以审查等特性一定程度上也导致谣言治理的“技术异化”,如何辨别政治谣言,及时传递真相,已成为当下算法驱动的社交平台面临的重要问题。

综上所述,智能技术在社会治理中的权重正在日益增加,因此,本文在评估算法在政治谣言治理的作用的基础上,着重分析算法治理存在的公正缺失、技术垄断、责任模糊等问题,进而探讨算法与利益相关者相结合的治理模式,以期为综合治理政治谣言提供具有针对性的正面策略。

## 二、解构神话:社交平台政治谣言算法治理的生成逻辑

随着政治谣言在社交平台的快速发展,催生了谣言粉碎机、较真查证、一点资讯辟谣等依托算法技术的谣言治理平台。通过分析发现,这些算法平台在政治谣言治理中,主要存在依托算法介入谣言生产环节的内部治理逻辑,以及依托算法介入谣言流通环节的外部治理逻辑两种方式。

### 1. 政治谣言生产环节治理:语义分析与情绪判定

信息社会的固有悖论告诉我们,“信息的富裕造成注意力的匮乏,因此我们需要在丰富的信息源中有效配置注意力”。<sup>[8]</sup>社交平台会依托算法来判定用户的关注程度,如果算法认定该信息不受欢迎,则就不会具备“爆款”的可能。<sup>[9]</sup>因此,政治谣言在社交平台生成后,并不会立即进入快速传播阶段,这为实行算法内容监测和治理提供了契机。算法介入谣言生产环节,也就是算法内部治理,主要依托算法不断采集和分析平台数据库,对用户发布的信息进行真实性评估,各大辟谣平台主要采取的方式如下:

一是通过政治谣言识别变量分析信息中的关键词。人工智能时代的造谣者在发布信息时,通常用一些夺人眼球、强烈暗示的权威修辞来发泄对政治体制、政府、政党的不满,并在谣言中使用感叹号、问号等符号以造成同情或恐慌。社交平台通过“算法化”与“技术化”监测发布者的信息风险,经算法推定,谣言中高频词汇有“专家、神秘、真相、曝光、震惊、竟然、警惕、必看”等,在标题中为夺人眼球常用“绝对、揭秘、惊呆了、没想到”等诱导性词汇。因此,通过算法对这些信息

进行预警和风险提示,自动对谣言进行预警,很多政治谣言就是在被算法监测到后对其可信度进行评估,并对低可信度信息进行“降权”处理的。<sup>[10]</sup>可见,依托算法的关键词分析减少了来自不实信息源的推送机会,一定程度上能抑制政治谣言进入流通环节。

二是通过机器学习模拟信息可能带来的社会情绪。奥尔波特把谣言带来的情绪归纳为“不安、恐惧、好奇、愿望、憎恨”等,政治谣言为了提高传播量通常大量使用第一人称和第二人称,以叙述真实经历展现内心独白,具有比真实信息更容易引起民众情绪波动的特征,进而带来病毒式传播。基于社会情绪判定理论,平台算法对谣言的内部审查已经从目录、标题至引擎发展到全文搜索模式,具备深度学习功能的计算机可以借助模仿人脑的人工神经网络处理大量的数据集,并完成情绪模拟,进而识别政治谣言。随着数据采集与运算技术进一步被应用,语法结构、词汇选择、标点符号等复杂性分析算法,很大程度上会提升谣言治理的精准度。

### 2. 政治谣言流通环节治理:“人工+算法”合力辟谣

资料显示,多数政治谣言是在海量传播后被发现的,所以算法外部治理显得尤为重要,也就是算法介入政治谣言流通环节。在这一阶段,社交平台通过增加人力资源与算法相结合对用户发布的信息进行复核,分析谣言的文本特征,并采用“人工+算法”对谣言易感群体进行精准辟谣。

一方面,以“人工+算法”复核信息真伪,及时发布官方信息。根据政治谣言从核心向外围人群传播的时间差,平台通过组建数据挖掘技术团队承担政治信息“内容审查”“事实核查”工作,24小时对社交平台信息进行动态监控,以便及时发现政治谣言的苗头,通过厘清政治谣言“数据化”的特征,再设计一套有较高识别准确率的算法及时切断传播路径。在2020年的新冠疫情期间,“一个武汉女人终结了法国大罢工,法国人都去抢口罩了”等谣言就是通过技术人员进

行人工复核后生成风险报告而进行治理的;在此基础上,还要积极联合官方媒体共同构筑辟谣主体,如相关平台在复核出“湖北解封3小时超千人进长沙”“韩国人到青岛不用强制隔离”等谣言时,及时联合@长沙发布、@青岛发布等发布辟谣信息,斩断“造谣黑链”,将危害降至最低。

另一方面,为政治谣言易感群体画像,对受众进行精准辟谣。虽然技术已经大幅提升民众看待世界的程度和范围,但个人在信息搜索上所花费的时间以及甄别信息真伪的能力,会受到“信息茧房”的限制,社交平台政治谣言治理也会遇到“信息茧房”,致使政治谣言的阅读量远高于辟谣信息的阅读量,形成传谣与辟谣的信息不对称现象。鉴于此,很多社交平台开始对账号进行实名制身份认证。由于很多政治谣言是通过普通民众进行线上传播,而易感群体却无法进行可信用度查证,误以为平台发布的都是官方信息。因此,微博的加V认证、微信的实名认证,能在很大程度上规范民众的言论责任。同时分析政治谣言的传播网络结构,对谣言易感群体定向推送辟谣信息。有些社交平台对在政治谣言链条中的角色进行有效识别,对谣言易感群体进行精准画像,把辟谣信息及时推送到民众手中。由此观之,虽然算法技术的进步提高了信息的可视化程度,但仍需要人类智慧对其中的各要素加以甄别,“人工+算法”相结合一定程度上实现了与谣言的指数级传播相抗衡。

### 三、技术的桎梏:算法偏见及其在政治谣言治理中的限度

通过复盘政治谣言治理的实践逻辑发现,算法治理在遏制谣言传播中取得了很大成效。但由于程序设计、数据挖掘、机器学习等固有的“双刃剑”特点,算法技术还存在短期内难以克服的“利维坦”风险。因此,系统分析这些潜在工具理性风险有助于克服算法治理的单向度思维。

#### 1. 平台算法的资本逻辑鼓励政治谣言

“信息技术发展和使用推动了主体的数字化

存在,并不断促使技术在思想世界里扮演着核心角色”,<sup>[11]</sup>这推动了资本加持的“工具理性”“技术理性”在研究和实践领域的发展。社交平台政治谣言不断增长,很大程度上是由平台的资本运作和算法偏见决定的。一是资本本身的趋利性属性。技术异化的根源在于把技术卷入资本的运转体系中,技术企业以及技术资本是媒介发展的“服务商”,“资本的力量正在逐渐支配市场,商家和企业依靠技术加持来获得更大利润,很大程度上会加剧恶性竞争”。<sup>[12]</sup>在“流量至上”的市场模式下,为了争夺注意力资源,不实的观点、情绪在社交平台上短时间汇聚,进而扩散为高点击量的政治谣言,广告商与平台则依靠民众对谣言的点击量获利。个人和企业为追求商业利益最大化,鼓励和默许在平台上编造与发布适应平台算法推送系统的信息,为夺人眼球甚至修改官方媒体与政府发布的信息。在资本运作逻辑下,民众对技术创造的参与越来越不具有自主性,信息也日益缺乏公共价值理性,更多是反映资本操纵者的价值观念和预期目标,资本化的算法会破坏政治话语的公平和质量。二是算法固有偏见带来的流量“马太效应”。所谓的“算法中立”更多是对数据和技术乌托邦式的构想,信息技术加剧了信息超载,各大社交平台的算法本身追求信息与用户间的相关度和快速、准确匹配,促进信息的“回声室”发展,这种算法固有的偏见会导致流量的“马太效应”。一些传播者为博取关注将不相关的内容移花接木,进一步加剧认知茧房、圈层固化;此外,算法推送中的流量激励机制制造了大量的同质化信息,不断压缩民众的认知空间,易导致官方权威媒体信息被淹没,阻断官方媒体与民众的有效交流,打破了官方媒体对信息传播的主导。

#### 2. 弱人工智能时代算法治理本身的局限

信息社会的风险更多体现在对技术把握的不确定性上,甚至很多学者认为人类将生活在“技术与知识的囚室”中。<sup>[13]</sup>当前人工智能还处于弱人工智能阶段,算法只能处理特定任务,很

难产生独立意识。笔者分析发现,很多政治谣言内容复杂,需要依靠专业知识、社会经验、话题语境等多重技能才能加以分辨,这给算法治理带来诸多挑战。一是很难将所有信息都“指标化”。有些政治谣言折射出民众对权威信息的期望,即使是人类也难以对其界定,而平台算法想通过数据、模型来进行鉴别,更是难上加难。此外,当算法面对政治谣言传播中涉及宗教、哲学等抽象的内容时,更是无法像人类一样会意或辨识其可信度。二是谣言制造者会转变手法应对算法治理。一些平台造谣者只要在原来谣言的基础上稍作改变,算法治理的适用性就成为问题。如用写作风格作为识别谣言手段的算法被各平台广泛采用,造谣者就会适当改变写作手法以蒙混过关。三是算法尚不能理解人类语言内涵。虽然算法治理正努力将复杂的政治谣言纳入技术处理的框架,但由于人工智能对人类语言理解的“中文房间”限度,当前的智能计算提供的更多的是可重复验证的、符合逻辑分析的结果,却很难感知到复杂的人类情绪以及情绪背后的语言。因此,算法的语言模式和符号处理可能会将真、假信息混为一谈。

### 3. 政府机关的算法依赖易导致治理误区

根据“守门人”和“议程设置”等理论,政府应在政治谣言治理过程中扮演着重要角色。面对强大的治理压力,越来越多的政府部门开始依赖算法治理;然而,算法治理往往与官僚主义具有天然的亲和性,这也一定程度上带来了政治谣言治理的误区。一是有些政府部门存在算法形式主义思维。传统的谣言治理主要以政府自上而下的“问题—约谈—整改”为主,随着算法治理的加入,有的地方注重新技术的“显成绩”,却忽视了“潜成绩”,投入专门人财物力,特别是为了完成指标性的要求,把主要精力放在算法治理的标准化上,“互相攀比购置计算机设备的先进程度,把主要财力投入到技术设备上”,<sup>[14]</sup>而缺乏结合不同时间空间的信息传播实际情况,以至于谣言治理算法体系建成后很难做到妥善维护

和有效运用。二是有些政府没有及时做到信息公开。如博尼法斯所言,“造假者之所以有市场,是因为他们顺应了民众思想,说出了民众准备听的话”。<sup>[15]</sup>广大民众倾向于将自己听到的政治谣言与公开的信息作对比,由于政府发布信息不及时、不透明,使得民众对地方政府信息源的信任大打折扣,从而让政治谣言在快速滚动的传播链条中不断发酵。三是数字鸿沟的存在导致治理盲区。数字鸿沟这个词描述的是技术产品使用者之间存在的平等,新技术快速发展会带来“在线”与“不在线”之间的数字鸿沟,算法更是加剧了这种鸿沟。中老年人属于谣言易感群体,而有关辟谣信息多以文字形态为主,并且辅之以专业科学阐释,这种传播形态难以在中老年群体中取得较好的传播效果。

## 四、协同治理:构建多元互动的政治谣言治理模式

随着普适计算日益大众化,虚拟和现实空间相互融合让社会生活焕发出新的活力,<sup>[16]</sup>这不仅需要人类智慧,也需要融合新技术的智慧。在未来的谣言治理中,我们既要总结传统政府、算法为主的治理政治谣言的技术存量,更要探讨“算法—利益相关者”多元协同的治理模式,进而形成上下联动、多元参与、深度融合的政治谣言治理路径。

### 1. 政府加强顶层设计承担协同治理的主导责任

创新人工智能时代社交平台政治谣言治理方式不是文字游戏,而是要深刻把握技术所承载的权力内核,这就需要通过顶层设计,实现跨功能、跨权限、跨层级的整体规划,才能形成真正的政治谣言治理格局。一方面,政府要充当规则制定者、过程管理者和底线把控者。在多元协同治理模式中,政府仍然处于中心位置,并推动利益相关方实现积极互动。<sup>[17]</sup>在顶层设计中,处于核心地位的是治理结构,政府通过主导智能技术对社会进行数字化呈现,促成有效的治理网络,使传统二元化的“人际交互”发展到多元化的“人

机和谐共生”场景。中共十九届四中全会《决定》明确指出,要“建立健全运用互联网、大数据、人工智能等技术手段进行行政管理的制度规则。推进数字政府建设,加强数据有序共享,依法保护个人信息”。<sup>[18]</sup>在政治谣言治理中,政府应设立谣言预警机制,相关制度文件也应对政治谣言进行清晰界定,及时公开相关信息以满足民众的信息诉求,真正做到从源头抑制谣言生成。有些地方政府在机构改革中设立了“大数据管理部门”等机构,以期担负算法治理的主导责任。此外,还需要突出政府的协调者角色,协调处于强势地位的社交平台 and 处于弱势地位的民众之间的关系,通过运用大数据预判谣言治理的空白点,通过优化顶层设计来压缩社交平台政治谣言治理层级冗余。

另一方面,积极构建政治谣言治理的制度供给体系。“社会数字化”使原有的制度滞后问题更加凸显,倒逼各国政府积极推动现存的信息披露、监督、使用等方面法律制度的变革。2019年4月,美国有立法者提出“算法问责”法案;<sup>[19]</sup>2020年3月1日,国家网信办发布的《互联网信息服务内容生态治理规定》正式实施,明确强调“依法严惩造谣传谣犯罪”,这为制造和传播政治谣言的责任人受到法律惩处提供了重要依据。未来还需要在政府主导下对算法治理的相关制度继续完善:一是建立完备的数据复核制度,我们不能“唯数据论”,需要通过人类自身的智慧增强数据的可靠度,促使技术的工具理性服务人的价值理性。二是建立健全数据管理制度,不断完善谣言治理资源数字化的规则和流程,使之与纯粹的技术结构和程序相适应,加大对篡改数据信息行为的惩戒力度。三是建立健全数据运用制度,借鉴英国的“网络观察基金会”、美国的“社交媒体监控中心”等经验,从源头规范信息保存和使用问题,防范数据泄露风险。

## 2. 社交平台承担研发与合作的主体责任

人工智能时代,在线公共存储库和商业数据库不断增加,这使政府很难从细节上全面把握政

治谣言的治理过程,而是转向对合作行动的支持,这为企业创造更多机会参与到算法治理中。这就需要各社交平台发挥企业的社会责任组成辟谣联盟,通过企业自治与行业组织等对网络社交平台进行管控,共同实现优质信息供给以及政治谣言的协同共治。一方面,优化算法设计,避免决策偏差。社交平台是政治谣言传播的主要媒介,所以要从平台算法本身入手,提升平台算法的治理水平。当前已在这方面取得一些技术性突破,如研究人员基于因果推理开发的捕捉谣言技术模型,通过区块链技术创造能被公众审查的信息算法等。<sup>[20]</sup>在这些研究基础上,未来平台企业还需要建立具有权威性的谣言研究和控制中心,通过大数据算法、信息挖掘、自然语言理解等技术精准定位虚假信息,进而实施数据监测。

另一方面,变各自为政为共同协作,共建负责任算法。“网络权力理论”认为所有的媒介都通过特定网络进行工作,并且权力不断集中于大型传媒集团手中。与此同时,政治谣言治理涉及计算机、新闻传播、经济、政治等多个学科和领域,单纯依赖某一个企业很难达到治理预期,企业间的合作治理尤为重要。如支付宝花重金研发“谣言粉碎机”、腾讯开发的“较真查证”等第三方辟谣机构,正在不断打破壁垒加深合作,对平台中的谣言进行联合辟谣。可见,企业间的不断开放为行业合作平台的建立提供了契机,为民众提供了一个关于现代技术如何影响公共生活的更为广泛的社会责任框架,消除技术的资本印记,使之成为治理政治谣言现实可行的解决方案。

## 3. 主流媒体做好真相传播的引导责任

人工智能时代,价值信仰依然是意识形态的精神内核,主流媒体具有引导真相传播的正当性。然而,主流媒体不能单纯正面宣传和歌颂赞美,更需要通过监督、批评建立自身的公信力,而公信力恰恰是引导力的前提。其一,主流媒体要做到第一时间发布官方信息。通过复盘政治谣言传播逻辑发现,谣言难治理既有技术因素也有

人为因素,但官方信息不能及时公开透明是谣言盛行的根本原因。官方信息的诚意、透明度及迅速发布是避免谣言滋生的重要途径,信息封锁不仅不会遏制社会恐慌,而且可能导致民众情绪或意见进一步极化。<sup>[21]</sup>技术发展需要坚持正确的价值导向,为极易产生泛娱乐化、去中心化、数字鸿沟等的算法产品加入干预和提醒机制,正如习近平强调的,要“用主流价值导向驾驭‘算法’”,如此才能实现价值导向与技术发展的有机平衡。比如,人民日报的“人民号”以算法为运作逻辑,实现智能算法技术与主流价值观的有机融合,从而打造了兼具主流意识形态和创新鲜活的媒体新生态。<sup>[22]</sup>因此,主流媒体应及时向公众提供与政治相关的信息,让新闻发言人和发布会制度日常化,及时针对政治最新发展与民众最关心的问题作出解答。

其二,顺应传播规律创新正能量表达方式。当下已有越来越多的民众认识到“泛娱乐化”信息的副作用,也厌倦对娱乐化信息的过度接收,正日益产生新的主流媒体诉求。因此,主流媒体在正能量传播中需要变宏大叙事为日常生活叙事,需要以平等的态度进入与各种媒介的互动过程中,将那些晦涩难懂的理论以简洁明了、通俗化的语言表达出来,通过日复一日不断积累的方式逐渐树立起“接地气”的形象;同时变系统化传播为微传播,例如善于借助微信、知乎等新媒体分享功能,在抖音、快手等平台发布信息,搭建专业的网上专家平台,为市民提供系统的专门服务,最终搭建中心化的权威信息平台与官方辟谣平台,确保公众获取信息的权威性与一体化,提高政府的公信力,最大化减少政治谣言滋生的土壤,确保稳定的秩序和安全的“政治逻辑”不受冲击。

#### 4. 广大民众承担政治谣言治理的伦理责任

《荀子·大略》有句名言是“流言止于智者”。因此,广大民众在信息扩散中的责任是辨别谣言与传播真相,并对自己的行为负责。一方面,激活民众个体,多渠道、多方式贡献“微力”。

在人工智能时代,虽然单个用户在社交平台的力量看似微小,但如果能激活这些信息传播中的节点,将具有无比强大的力量。一是社交平台要提供用户协作“打假”的“基础设施”,积极为用户提供一个有关谣言的数据库,便于核查、追溯和识别,鼓励民众自觉地核实查验,在转发时进行澄清;与此同时,算法也要持续关注用户日常生活中的见解、分析、判断,对其及时推送辟谣信息。二是需要平台和网络意见领袖建立合作机制,通过“@名人”拉高辟谣文章的热度。三是提升广大民众的媒介素养。通过教育,提高民众的媒介素养至关重要,广大民众应该自觉提高自己的知识储备量,自觉遵守互联网的准入标准,提升自我修养和网络素质。

另一方面,多方合作共建人工智能时代的社会安全伦理。人类对自己制造的智能技术比较了解,由于算法中存在“中文房间”等难题,智能算法技术固有的局限决定了“算法依赖”不足以推动政治谣言的妥善治理。社会安全是政治谣言治理的落脚点,我们正在想尽办法不断完善阿西莫夫的“机器人三大定律”,以使技术掌握在人类手中;更为重要的是,“我们需要明确技术只是达成目标的手段,它所带来的利弊取决于人类如何使用它”,<sup>[23]</sup>因此,通过多方共同努力明确社会伦理价值高于技术或商业需要的目标,共建一种置于个人和社会之上的公共伦理准则,才能提升广大民众在技术运用中的安全感。如国外一些高校设计了类似“道德机器”的模拟平台,我国很多高校的“人工智能”专业,还开设人工智能时代的社会伦理等相应课程,很多地方也积极成立科技伦理委员会,旨在收集人们对由机器智能作出道德决策的看法。随着社会合力的形成,呈现出“一核多元、协商共治”的政治谣言治理格局,最终形成超越所有个别组成部分的共同愿望或期待。

## 五、结 语

“混乱—秩序”循环的波动会影响到个人所

作的决策,而且将成为数字化信息社会的一个长期主题。<sup>[24]</sup>我们今天探讨政治谣言的算法治理逻辑理路,不仅要反思技术对治理的异化,更多的是要总结技术对人类能力的延伸的经验,为构建利益相关者与算法互动的谣言治理模式提供一些启示。一方面,对算法的合理运用是提升政治谣言治理水平的基础。数字技术和社会文化趋势在政治空间相关性日益显现,智能化的“去中心化”发展促进劳动者从韦伯式的“中心化”科层权力结构中解放出来,取而代之的是个体之间、团队之间的协作平台,这就需要在不断优化算法的过程中积极嵌入信息社会。另一方面,对算法的过度依赖会导致人的本质异化,现代世界充斥着大量奇异的发明和创新,人类需要不断提升驾驭能力,以免受到其自身发明产物之力量的束缚。因此,技术必须成为伦理问题,“技术治理需要在民主化路径中来获得与其他治理体系的协作与融合”,<sup>[25]</sup>只有各种社会力量形成合力,对于伴随技术进步而来的弊病,才能有明智的认知,并采取有效措施来解决。总之,只有在智能技术的智慧、法律制度的规范以及社会伦理的理性交相互作用下,政府、企业、民众共同协调现有价值规范和算法标准,才是人工智能时代政治谣言“算法—利益相关者”治理的关键。

注释:

[1][法]让-若埃尔·卡普费雷:《谣言:世界最古老的传媒》,郑若麟译,上海:上海人民出版社,2008年,第15页。  
 [2][美]孔飞力:《叫魂:1768年中国妖术大恐慌》,陈兼、刘昶译,上海:上海三联书店,1999年。  
 [3][法]乔治·勒费弗尔:《1789年大恐慌:法国大革命前夜的谣言、恐慌和反叛》,周思成译、高毅审校,太原:山西人民出版社,2019年。  
 [4]《马克思恩格斯全集》(第33卷),北京:人民出版社,1973年,第258页。  
 [5]《马克思恩格斯全集》(第15卷),北京:人民出版社,1965年,第458页。  
 [6]Hjarvard,S.,“The mediatization of society:A theory of the media as agents of social and cultural change”,*Nordicom Review*, 2008,29(2),pp.105-134.

[7]张超:《社交平台假新闻的算法治理:逻辑、局限与协同治理模式》,《新闻界》2019年第11期。  
 [8][美]詹姆斯·韦伯斯特:《注意力市场:如何吸引数字时代的受众》,郭石磊译,北京:中国人民大学出版社,2017年,第7页。  
 [9]闫译华:《内容算法:把内容变成价值的效率系统》,北京:中信出版集团,2018年,第60页。  
 [10]方师师:《哈佛大学肯尼迪政治学院:打击假新闻的研究议程与行动方针(编译)》,《汕头大学学报(人文社会科学版)》2017年第9期。  
 [11][美]尼尔·波兹坦:《技术垄断》,何道宽译,北京:中信出版集团,2019年,第29页。  
 [12][美]杰瑞·卡普兰:《人工智能时代》,李盼译,杭州:浙江人民出版社,2016年,第96页。  
 [13][德]乌尔里希·贝克、[英]安东尼·吉登斯、[英]斯科特·拉什:《自反性现代化》,赵文书译,北京:商务印书馆,2001年,第74页。  
 [14]孙会岩:《人工智能时代基层党组织组织力创新研究》,《探索》2020年第1期。  
 [15][法]博尼法斯:《造假的知识分子》,河清译,北京:商务印书馆,2013年,第17页。  
 [16][美]卡洛·拉蒂、马修·克劳德尔:《智能城市》,赵磊译,北京:中信出版社,2019年,第164页。  
 [17]刘小泉、朱德米:《协作治理:复杂公共问题治理新模式》,《上海行政学院学报》2016年第4期。  
 [18]《中共中央关于坚持和完善中国特色社会主义制度推进国家治理体系和治理能力现代化若干重大问题的决定》,《人民日报》2019年11月6日。  
 [19]Karen Ila. Congress Wants to Protect You from Biased Algorithms, Deepfakes, and Other Bad AI. 2019-04-15, <https://www.technologyreview.com/s/613310>.  
 [20]吴果中、李泰儒:《用区块链技术打击虚假新闻——Userfeeds与PressCoin模式介绍》,《新闻战线》2018年第13期。  
 [21]习近平:《加快推动媒体融合发展构建全媒体传播格局》,《求是》2019年第6期。  
 [22]王贤卿:《社会主义意识形态面对技术异化挑战——基于智能算法推送的信息传播效应》,《毛泽东邓小平理论研究》2020年第6期。  
 [23][美]温纳:《自主性技术:作为政治思想主题的失控技术》,杨海燕译,北京:北京大学出版社,2014年,第21页。  
 [24][美]杜布拉瓦茨:《数字命运:新数据时代如何颠覆我们的工作、生活和沟通方式》,姜昊鸾、李德坤、徐琳琪译,北京:电子工业出版社,2015年,第86页。  
 [25]向玉琼:《论技术治理的民主化:基于对风险治理技术化的反思》,《学术界》2020年第12期。

[责任编辑:刘 鑫]