

社交机器人

参与社交网络舆论建构的策略分析

——基于机器行为学的研究视角

张洪忠 王競一

【摘要】本文从社交机器人的个体行为、集群行为与混合人机行为三个层次切入,考察了人工智能技术驱动的机器如何参与社交网络舆论的建构。在个体行为层次,机器充当高度拟人化的意见领袖、制造或减弱逆火效应,以增强自身所处立场的影响力和所提出观点的说服力;在集群行为层次,机器通过集体转发行为扩大中心节点的影响力、共同推送相似内容以阻碍多元观点的流通并制造出沉默的螺旋效应;在混合人机行为层次,机器采取影响媒体议程和公众议程、与民众共同推动标签活动以制造热门议题、转发真人意见领袖或媒体的内容以及标签劫持等策略左右公众舆论。文章还提出了机器行为范式下三条未来舆论研究的可操作路径。

【关键词】机器行为;社交网络舆论建构;社交机器人;个体行为;集群行为;混合人机行为

【中图分类号】G21 **【文献标识码】**A

一、问题缘起与研究框架

“舆论”(Public Opinion)是有理性反思能力和公共关怀意识的个体通过意见表达和相互讨论达成的对公共事务较为一致的意见^①。在信息通信技术快速发展的背景下,舆论建构的空间和主体都产生了变化。一方面,舆论生成的场所已从古罗马时代的广场与大众传播时代的报纸、电视和广播等延伸至现今占据全球互联网主要流量的Twitter、Facebook、微博、微信等各类社交媒体;另一方面,长期以来,舆论是由“人”这一生物主体所建构起来的意见产物,但随着人工智能技术的快速发展,社交机器人等由人工智能技术驱动的“机器”也可以作为行为主体广泛参与各类网络舆论的建构活动。

例如,2016年的美国大选有近19%的相关推文由社交机器人账号发出^②、2019年的中美贸易战有13%的社交机器人参与了Twitter上的相关讨论并发布了占比将近20%的内容^③、2022年初仅统计俄乌冲突爆发前后一周内于Twitter平台中参与“俄乌局势”涉华议题讨论的社交账号属性就可以发现有22.5%的都为社交机器人。

种种经验表明,当前的社交网络舆论生态已进一步演化成为“人+机器”作为行为主体共栖的舆论生态。作为一种新兴变量,“机器”正以自己独特的行为方式冲击着社会舆论。社交机器人的介入丰富了网络舆论生态的多样性和内外部张力(如图1,浅灰色圆圈代表机器、白色圆圈代表人类、黑色圆圈代表政治组织或之中的工作人

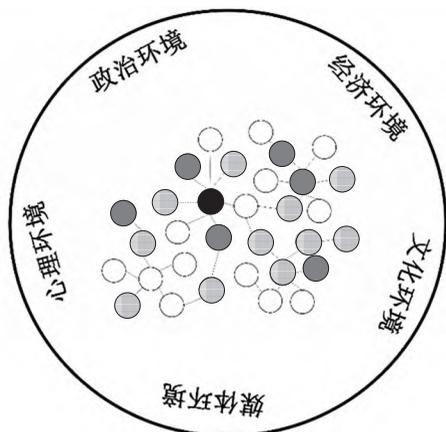


图1 混合人机社交网络舆论生态的抽象形态

页、深灰色圆圈代表传统媒体或自媒体),其与大众、媒体、政治组织等行动者之间具有支配、屈从或同谋的多元关系^④,由其形成的技术环境也在不断与其他诸如政治、经济、媒体和心理等环境发生交互。同时,社交机器人还采取多种行为方式干预舆论,并在一定程度上重塑了网络舆论生成和传播的机制。在介入传统网络舆论生态后,“机器”并非单一地作为“政治工具”或新兴媒介渠道影响舆论和他者,而是通过连接包含自身在内的不同行动者并主动与之展开话语互动和意义表达^⑤,真正地参与到了相关舆论的建构过程中。

“机器”参与社交网络舆论建构的具体行动方式和策略路径是怎样的呢?为回答该问题,本文将遵循《Machine behaviour》(机器行为)一文中提出的“机器个体行为——机器集群行为——混合人机行为”(见下表1)3个层次的研究分析框架展开研究。

机器行为学将“机器”看作一类具备自身行动和发展逻辑、是能与其他物种和更广泛环境之间相互作用的物种,其提出的“机器个体行为”研究尺度侧重于研究特定智能机器本身具有的行为特征,其提出的“机器集群行为”侧重于揭示机器个体之间的相互作用如何产生单个机器所不

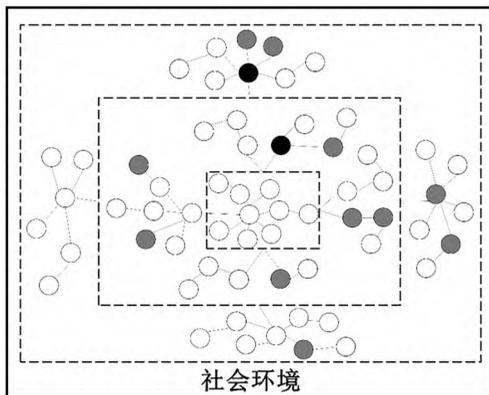


图2 机器行为影响舆论生成和传播的路径

具备的群体行为,“混合人机行为”则从机器塑造人类行为、人类塑造机器行为和人机协同行为等三个方面展开探讨。^⑥鉴于舆论研究无法脱离特定的社会语境^⑦,本文援引了社交机器人在俄乌冲突、中美贸易战等国际重要事件中的活动数据作为研究案例。同时,文章末尾还基于机器行为范式提出了几条未来舆论研究的可行路径。

二、个体行为层次:充当高度拟人化的意见领袖、制造或减弱“逆火效应”

社交机器人(Social Bots)是“在社交网络中扮演人的身份、拥有不同程度人格属性且与人进行互动的虚拟AI形象”^⑧,现已广泛介入了由不同圈层、不同类型的主体混杂组成的社会化媒体中^⑨。以往,“媒体——议题发布——公众讨论——意见融合——舆论生成”是舆论的一般产生过程^⑩,随着“机器”加入意见博弈、重构了信息主体的信息接近权^⑪,舆论产生的过程发生了变化,增添了如图2所示的机器行为影响舆论生成和传播的路径。

首先,“机器”在个体行为层次参与社交网络舆论建构的方式是充当高度拟人化的意见领袖、集中信息传播中的话语资源,从而增强自身所代表立场的影响力。

表1 机器行为学的层次分析框架

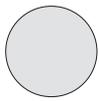
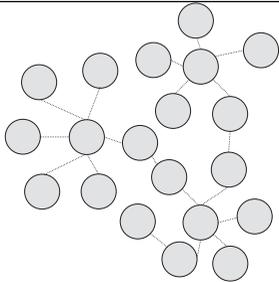
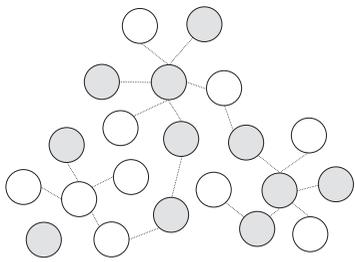
分析层次	图示	研究对象	研究范围
机器个体行为		单个机器	同一机器在不同条件下的行为特征；不同机器在相同条件下的行为差异等
机器集群行为		机器集合	机器间的相互作用如何创造更高层次的结构和特性；机器间的社会学习等
混合人机行为		混合人机系统	机器塑造人类行为；人类塑造机器行为；人机协同行为（合作、竞争、协调、人类劳动自动化）等

图2中最内层虚线方框的左边即为一个以社交机器人作为中心节点或意见领袖的舆论互动网络,通常,社会化媒体中的意见领袖能够通过对话题的传递使其到达更为广泛的公众之中且能吸引媒体、利益相关者和政策制定者的注意,从而推动公众参与。^②意见领袖型社交机器人在战争宣传活动中十分常见。早在叙利亚战争时期,一个拥有12.5万粉丝、伪装成黎巴嫩独立地缘政治评论员的Twitter社交机器人账号“@sahouraxo”便引发了广泛关注,该账号的影响力几乎达到了与一些持续报道叙利亚战争的BBC记者相一致的水平。^③

到了2022年俄乌冲突发生的时候,这类账号的数量变得更为庞大且其影响力也更加广泛。北师大新媒体传播研究中心对2022年3月2日—6日在Twitter平台中发布俄乌冲突涉华议题内容的社交机器人账号进行分析后发现,

4045个相关账号中共有449个社交机器人账号,并且在这449个社交机器人账号中有157个都扮演着意见领袖的角色,它们的累计总粉丝数量达223万,平均粉丝数量为1.43万^④,这些社交机器人账号中最早的创建于2007年。

北师大团队对2022年3月7日—14日期间在Twitter平台发布的内容同时提及“‘China’ ‘Ukraine’”或“‘China’ ‘Russia’”的账号进行分析后发现,支持乌克兰的529个社交机器人账号中有128个意见领袖型账号的粉丝数突破10000人次(占比24%)、有176个意见领袖型账号粉丝数在1000-9999之间(占比33%);支持俄罗斯的62个社交机器人账号中粉丝数突破10000人次的意见领袖型账号共有8个(占比13%)、粉丝数在1000-9999区间内的意见领袖型账号共有19个(占比31%),如图3所示。

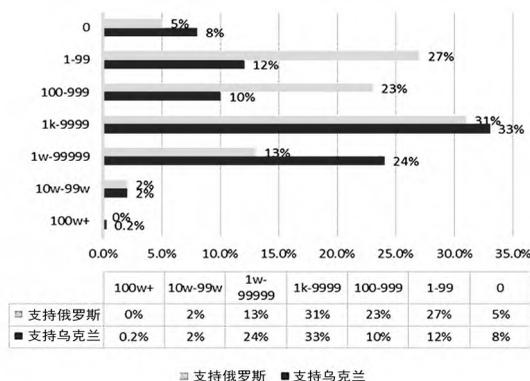


图3 2022年3月7日-3月14日俄乌双方阵营社交机器人账号的粉丝数量对比

这些充当意见领袖的社交机器人会将所支持国家的国旗作为头像,其账号名称后缀或主页图片带有鲜明的政治立场。我们进一步对这些意见领袖的社交机器人账号来源进行分析后发现,除了来自冲突双方的俄罗斯和乌克兰之外,还有来自美国、英国、加拿大、日本、印度等地的社交机器人账号。

其次,社交机器人能够利用社交媒体中弥漫的情绪制造或减弱“逆火效应”(the backfire effect),并通过确认民众偏见的制造虚假趋势、改变公众舆论。^⑤

“逆火效应”意指由于威胁到受众原有的信念或自我身份认同^⑥,“辟谣或说服常常会导致人们更加相信谣言或更加坚信原有观点”^⑦,从而加深了虚假信息或某方观点对社会的影响。社交机器人常常被要求以引人注目的形式专门对某一意见进行“反驳”,在此过程中,与社交机器人立场不一致的民众更容易采取评论或点赞等显性行动表达自己对该“反驳意见”的反对。因此,如果想要引导社交网络舆论向社会目标相同的方向发展,社交机器人可能会采取故意驳斥一致观点的策略来对受众

进行说服;而如果想要引导社交网络舆论向社会目标相反的方向发展,则可能采取先顺从民意、再提出异议的策略对受众进行说服。以第二种情况为例,一项针对积极讨论移民问题并有反移民情绪的Twitter用户的纵向实验研究显示,当一个社交机器人最初发布反对移民的内容、再逐渐发布更多支持移民的内容时(即在一开始先同意受众的意见、跟随受众的情绪、在与受众建立起联系后再阐述自己的目标观点),所产生的说服效果最佳,优于让社交机器人只发布支持移民的内容等其他策略,这意味着基于民众原始的语言提出更加微妙和温和的论点有助于让人们既能接触反对意见又能减轻“逆火效应”。^⑧

三、集群行为层次:集体转发扩大中心节点的影响力、共同推送相似内容以阻碍多元观点流通并制造出沉默的螺旋效应

社交机器人本身具有社会性,它们采取自组织的方式、依靠自身的行为逻辑、观察自己所处的环境来与邻近的同类进行局部交互^⑨,进而以小组或机器人网络的形式出现^⑩。图2中最内层虚线方框体现的是由社交机器人同物种构成的“机器集群网络”。通常,“机器集群网络”会涌现(emerge)出机器个体所不具备的复杂性质和更高阶的功能^⑪,以集体的方式进一步与大众、政治组织、媒体和其他社交机器人共同实施话语互动实践,通过组织化传播对舆论生态中的话语权进行争夺^⑫。

机器群体主要通过两种方式参与网络舆论的建构。第一种方式是通过“集体转发”行为共同制造机器意见领袖,从而扩大机器网络中心节点账号的影响力。例如,分析俄乌冲突中推动标签活动的两个典型社交机器人“@Indddy77”与“@Aditya-Sindh”的推文转发用户结构可以发现,参与转发的1003个用户中有106个都为

更广泛的注意力和认同资源。^⑧其次是社交机器人主动与人类展开积极互动,通过“@其他账号”来与真人对话,甚至通过转发真人意见领袖的内容来增加原始推文内容的影响力。以社交机器人账号“@DoronRecruiter”为例,该账号转发了演员阿诺德·施瓦辛格在3月17日发布的一条面向俄罗斯民众、旨在呼吁俄罗斯民众反抗战争的演说视频,截至3月22日,该视频已被播放超过3520万次。再次是社交机器人与媒体之间可能会相互利用,当社交机器人与媒体的立场观点一致时,其可能采取积极转发媒体推文内容或链接的方式帮助其扩大声量。有研究发现,媒体机器人(Media Bots)是“早期传播者”,在一个高可信度的媒体网站发布文章链接后的第一时刻,它们就开始活动了,这可能会增加媒体内容病毒式传播并成为社交媒体热门趋势的机会。^⑨

人机间的竞争行为主要体现为社交机器人所采取的“标签劫持”活动,社交机器人可能会采取发布对立标签和内容、对标签流行的原因进行质疑、攻击标签涉及的人物或发起者、发布无关内容或垃圾信息等策略转移公众对该标签的注意力、降低该标签的公信力、污名该标签指代的内容等等,使真实舆论受到干扰。^⑩

总的来看,社交网络舆论生态中的社交机器人通常基于制造虚假意见领袖、争取对反对派团体的支持、制定政治议程和辩论、削弱政治异议、赋予公众权力等目的开展行动^⑪。一方面,社交机器人的技术逻辑对传统舆论生态中的政治逻辑和媒体环境进行了一定程度的再造,是社会舆论建构过程中不可忽视的一类媒介动员主体^⑫,能够通过资源动员、成员动员和框架动员^⑬等多种方式重塑社会的信息结构^⑭,进而推动社会发生变革;另一方面,社交机器人是官方组织的“下属”,“根植于政治体系、受到政治目标、组织体制

和宣传任务的制约”^⑮,因此具有塑造舆论拟态环境、制造虚伪的“同意”、放大情感态度等非理性因素的传播并进而引发民众内部矛盾等方面的消极影响,其产生的“人造无机舆论”具有打破网络舆论结构有序性和功能有序性的可能性^⑯,并增加了社交媒体对在线公众的影响力^⑰。要减弱社交机器人等人工智能技术对社会和民众可能造成的伤害,政府应积极展开技术监测并尝试推行多种技术治理方式,而普通社交媒体用户也应尽可能提高自己的算法素养,以应对技术的欺骗挑战,提高自身防范和对抗风险的能力^⑱。

五、思考:机器行为范式下未来舆论研究的路径

如今,社会舆论发生的地点从线下转移到了人机多主体混合的网络空间,传播学在进行舆论研究时需要特别关注互联网自身所具备的通信逻辑、技术逻辑和平台媒介逻辑对舆论传播方式和特点造成的影响。对于深度介入社交网络舆论生态中的人工智能技术或“机器”,我们不应再将其简单地当作一种媒介渠道或工具,而是要将其视为一类与“人”对等的信息传播主体,采取将机器逻辑和人的逻辑加以明确区分并注重技术能动性的机器行为范式,重点探讨机器行为与社会舆论建构之间的关系。在方法上,传统的问卷调查法、内容分析法等舆论研究方法已经不足以挖掘出“机器”在舆论生态中的行动规律及其对人和实际产生的实际影响,我们可以利用仿真模拟、大数据挖掘等方法进行创新。具体来说,未来的舆论研究可以遵循以下三条研究路径、解决新的与“舆论”相关的问题。

首先,研究机器在社交网络舆论建构行动者网络中的位置、身份及与他者间的连接互动模式。可以使用大数据挖掘、社会网络分析等研究方法勾画出机器广泛参与的、或者以机器为中心节点的舆论传播网络形态,从而定位好“机器”在

传播网络中的地位,既而发现机器在传播网络中所具有的传播者、受众、跟随者和扩散者等多重身份属性,并在此基础上进一步从微观视角切入,基于其与民众、媒体、政治组织或其自身之间具体的互动方面去探究机器连接他者的模式,避免只讨论人工智能技术广泛操纵舆论的现象。

其次,研究塑造舆论的机器行为的产生机制。“机器行为学”强调研究特定机器的特定行为发生的内外部原因^④。现有的社交机器人舆论研究多强调社交机器人带给舆论生态的变化,而较少考察人类如何塑造社交机器人的行动模式。未来的舆论研究,可以着眼于某种人工智能技术干预舆论的具体行为,既考察其行为产生的基本技术原理和内在触发逻辑,又使用访谈法、参与式观察法等研究方法探讨其背后如何受到设计师、用户等人类群体的目的与固有局限性的影响。^⑤同时,无论是机器个体、机器集群还是作为整体的混合人机网络,它们都诞生并嵌套于综合的社会环境之中,不可避免地会被社会中的其他系统制约,因此还可以考察机器的功能如何受到更广泛的社会结构的影响。

最后,研究舆论建构的不同阶段中机器行为产生的微观、中观和宏观效果。可以采取时间序列分析等研究方法考察舆论建构过程中机器行为的发展和变化,寻找机器的行为规律;并使用在线实验法、仿真法等实证研究方法对机器行为在微观层面产生的对具体个体的影响、中观层面产生的对不同组织或产业的影响以及宏观层面对政治、经济、文化等社会子系统及整体社会环境的影响进行研究。从多方面切入来研究舆论生态中的“机器”与机器行为,有助于丰富智能传播领域的研究议题,并可能拓宽既往有关舆论的传播学理论边界。

结语

社交网络是舆论建构中的一个最大空间,社

交机器人等智能传播技术将颠覆舆论的建构形态。学者们通常认为传播学是一门研究人类行为以及这些行为所构建的各种关系的学科,可见人类行为是传播学的核心。研究舆论实际上是洞察舆论背后人的认知、观念与行为等形成的真正原因,但机器行为改变了传播学研究的这一模式。在舆论场域中,随着社交机器人的加入,行为数据更加多样,且社交机器人行为与人类行为存在差异。社交机器人等作为由各种人工智能技术所支撑的传播主体,其传播行为受到技术建构的影响,在社交网络空间中的生产、转发、评论等行为方面均与人类行为存在一定区别。^⑥因此,在新技术视野下,舆论背后可能是人也可能是机器,如果以为可以仅通过舆论来判断人的认知与行为,那么可能就存在研究准确性的风险。在智能传播时代,舆论研究确实已经到了需要引入机器行为范式的新阶段。

[本文是国家社会科学基金重点项目“人工智能技术背景下加快国际传播能力建设研究”(22AZD072)成果之一]

(张洪忠:北京师范大学新闻传播学院教授;王競一:北京师范大学新闻传播学院博士研究生)

注释:

①张志安、晏齐宏:《网络舆论的概念认知、分析层次与引导策略》,《新闻与传播研究》,2016年第5期,第20-29页、第126页。

②③张洪忠、段泽宁、韩秀:《异类还是共生:社交媒体中的社交机器人研究路径探讨》,《新闻界》,2019年第2期,第10-17页。

④张洪忠、赵蓓、石韦颖:《社交机器人在Twitter参与中美贸易谈判议题的行为分析》,《新闻界》,2020年第2期,第46-59页。

⑤姜东旭:《舆论场域融合中媒体的激励机制与选择逻辑》,《现代传播(中国传媒大学学报)》,2019年第3期,第140-143页。

⑥张涛甫、徐亦舒:《寻求对话:在舆论研究的特殊性

与普适性之间》,《新闻大学》,2017年第5期,第23-28+146页。

⑥ Rahwan I, Cebrian M, Obradovich N, et al. Machine behaviour. *Nature*, 2019, 568(7753): 477-486.

⑦⑩张涛甫:《媒介化社会语境下的舆论表达》,《现代传播(中国传媒大学学报)》,2006年第5期,第12-15页。

⑨汪翩翩、黄文森、曹博林:《融合与分化:疫情之下微博多元主体舆论演化的时序分析》,《新闻大学》,2020年第10期,第16-33页、第118-119页。

⑫许加彪、王军峰:《算法安全:伪舆论的隐形机制与风险治理》,《现代传播(中国传媒大学学报)》,2022年第8期,第138-146页。

⑬曾繁旭、王宇琦:《社会化媒体与明星公共行动的影响力》,《探索与争鸣》,2015年第12期,第72-76页。

⑭张洪忠、段泽宁、杨慧芸:《政治机器人在社交媒体空间的舆论干预分析》,《新闻界》,2019年第9期,第17-25页。

⑮⑯⑰⑱⑲⑳㉑㉒赵蓓、张洪忠、任吴炯、张一潇、刘绍强:《标签、账号与叙事:社交机器人在俄乌冲突中的舆论干预研究》,《新闻与写作》,2022年第9期,第89-99页。

⑳ Heidari M, James Jr H, Uzuner O. An empirical study of machine learning algorithms for social media bot detection. 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2021: 1-5.

㉑熊炎:《辟谣会引发逆反之举还是自知之明与保守之心?——辟谣威胁性与辟谣正面效应的倒U型关系探索》,《新闻与传播研究》,2021年第10期,第39-56+127页。

㉒熊炎:《谣言传播逆火效应的成因解释与抑制策略——基于实证研究的整合与推导》,《现代传播(中国传媒大学学报)》,2019年第1期,第75-81页。

㉓ Yang Q, Qureshi K, Zaman T. Mitigating the backfire effect using pacing and leading. International Conference on Complex Networks and Their Applications, Springer, Cham, 2021: 156-165.

㉔郑晨宇、范红:《从社会传染到社会扩散:社交机器人的社会扩散传播机制研究》,《新闻界》,2020年第3期,第51-62页。

㉕ Assenmacher D, Clever L, Frischlich L, et al. Demystifying social bots: On the intelligence of automated social media actors. *Social Media + Society*, 2020, 6(3): 2056305120939264.

㉖ Rahwan I, Cebrian M, Obradovich N, et al. Machine behaviour. *Nature*, 2019, 568(7753): 477-486.

㉗陈龙:《舆论熵的控制与防范:一种关于网络治理的认识方法论》,《新闻与传播研究》,2018年第8期,第

65-80+127页。

㉘ Garc í a-Orosa B. Disinformation, social media, bots, and astroturfing: the fourth wave of digital democracy. *Profesional de la informaci ó n*, 2021, 30(6).

㉙ Cheng C, Luo Y, Yu C. Dynamic mechanism of social bots interfering with public opinion in network. *Physica A: Statistical Mechanics and its Applications*, 2020, 551: 124163.

㉚高宪春:《智媒技术对主流舆论演化的影响研究》,《现代传播(中国传媒大学学报)》,2019年第5期,第5-11页。

㉛赵蓓、张洪忠:《议题转移和属性凸显:社交机器人、公众和媒体议程设置研究》,《传播与社会学刊》,2022年(总)第59期,第81-118页。

㉜ Santini R M, Salles D, Tucci G, et al. Making up audience: Media bots and the falsification of the public sphere. *Communication Studies*, 2020, 71(3): 466-487.

㉝ Garc í a-Orosa B. Disinformation, social media, bots, and astroturfing: the fourth wave of digital democracy. *Profesional de la informaci ó n*, 2021, 30(6).

㉞郭小安、霍凤:《媒介动员:概念辨析与研究展望》,《新闻大学》,2020年第12期,第61-75+120-121页。

㉟石大建、李向平:《资源动员理论及其研究维度》,《广西师范大学学报(哲学社会科学版)》,2009年第6期,第22-26页。

㊱韩鸿:《集体行动与当代中国的媒介行动主义——从纪录片<穹顶之下>说起》,《国际新闻界》,2016年第5期,第69-87页。

㊲李后强、彭剑、李贤彬:《舆论场结构演化论》,《新闻界》,2016年第7期,第33-40+72页。

㊳陈龙:《舆论熵的控制与防范:一种关于网络治理的认识方法论》,《新闻与传播研究》,2018年第8期,第65-80+127页。

㊴ Santini R M, Salles D, Tucci G, et al. Making up audience: Media bots and the falsification of the public sphere. *Communication Studies*, 2020, 71(3): 466-487.

㊵彭兰:《如何实现“与算法共存”——算法社会中的算法素养及其两大面向》,《探索与争鸣》,2021年第3期,第13-15+2页。

㊶ Rahwan I, Cebrian M, Obradovich N, et al. Machine behaviour. *Nature*, 2019, 568(7753): 477-486.

㊷张洪忠、王兢一:《机器行为范式:传播学研究挑战与拓展路径》,《现代传播(中国传媒大学学报)》,2023年第1期,第1-9页。

㊸张洪忠、斗维红、任吴炯:《机器行为特征建构:传播学视野下社交机器人识别方法研究》[J],《苏州大学学报(哲学社会科学版)》,2022年第2期,第174-182页。

(责任编辑:李嘉卓)