

【实践研究】

网络健康谣言的传播阻断和治理框架研究

陈燕方 周晓英 裴俊良

【摘要】[目的/意义]文章设计针对网络健康谣言的多维度、多层次的综合性治理框架来抑制网络健康谣言的流传和误用。[方法/过程]文章通过抽象网络健康谣言传播过程及各个传播阶段的阻断方式,进而总结分析网络健康谣言治理过程的主要阶段及关键点;然后,按照健康谣言生成与传播特性,详细分析各个阶段的治理目标、治理对象和治理手段,并据此归纳分析得出闭环治理框架。[结果/结论]本治理框架主要包括一个核心技术模块(内容判别技术模型)和三个中心平台模块(网络媒体运营平台、辟谣专家工作平台和辟谣信息共享平台),基本涵盖了当前网络健康谣言治理需要考虑的核心问题,可作为健康谣言治理活动的顶层设计基础和行动指南。

【关键词】健康谣言传播;健康谣言治理;防谣;辟谣;止谣

【作者简介】陈燕方(1992-),女,中国人民大学图书馆馆员;周晓英(通讯作者)(1963-),女,中国人民大学信息资源管理学院教授,博士生导师;裴俊良(1996-),男,中国人民大学信息资源管理学院博士研究生(北京100872)。

【原文出处】《情报资料工作》(京),2024.1.35~45

【基金项目】本文系“中国人民大学公共健康与疾病预防控制文理交叉重大创新平台‘中央高校建设世界一流大学(学科)和特色发展引导专项资金’”、国家社会科学基金重点项目“全媒体语境下的信息流行病学理论与实践研究”(批准号:20AZD132)的研究成果之一。

人口老龄化的加速、食品安全等问题的频发、各类慢性病的层出不穷^[1],使我国公众的健康诉求逐渐开始从温饱问题转向健康生活品质问题,与此伴随的健康信息需求也呈现出空前高涨的形势。同时,随着信息技术的飞速发展,微信、微博等社交媒体成为人们获取健康信息的主要渠道之一。然而,网络的草根性、匿名性等特点,直接或间接地导致了近年来网络健康谣言的井喷式爆发。中国互联网联合辟谣平台发布的2022年度网络谣言盘点中,医疗健康与食品安全仍是网络谣言频发的集中领域^[2]。

为了应对网络健康谣言,政府和社会各界均采取了相应的措施,包括建立各种辟谣平台,如“中国互联网联合辟谣平台”“新浪微博辟谣”“微信辟谣中心”“微信辟谣助手”小程序、新华网联合食品行业协会推出的“中国食品辟谣联盟”等。但是,上述以专家或大众鉴别为主的辟谣工作仅能在谣言扩散后缓

解其给社会带来的危害,更为有效的、全局的、自动化的手段亟待完善。

为了应对当前网络健康谣言泛滥的问题,社会公众、媒体、政府等多元主体需要共同对网络健康谣言从产生到消退的每个节点进行把关,建立针对网络健康谣言的多维度、多层次的综合性治理框架来抑制网络健康谣言的流传和误用,从而保证大众获取更加真实可靠、优质的健康信息,免受健康谣言的侵害和误导;同时有效提升我国网络健康信息服务水平,推动“健康中国”战略目标的早日实现。

1 相关研究

网络健康谣言是与健康主题相关,未经官方或当下健康领域专业共同体证实,且在网络上产生了一定影响的信息。根据健康谣言的特征分析,本文从治理视角出发将网络健康谣言分为两类:第一,被已有科学研究证明为捏造、子虚乌有的论断;第二,

在业界、科学共同体内部尚未得到认可,明确缺乏支持依据的信息。

1.1 研究现状

笔者在中国知网、万方、维普等中文数据库以及 Web of Science、Google Scholar 等外文数据库检索网络健康谣言相关研究,通过对检出的文献进行筛选、剔除并精读,发现当前研究重点主要分布在以下几个方面:

(1)网络健康谣言特征及检测研究。包括基于质性研究方法的文本特征分析^[3-5],基于对比分析的网络传播特征分析^[6],以及基于混淆评估工具的多维特征分析^[7]。特征分析的目的在于两个方面:一是形成特征列表,帮助用户提升健康信息素养以及鉴别健康谣言的能力;二是为健康谣言检测奠定基础,网络健康谣言检测研究大多基于提取的各类特征及机器学习文本分类算法来探索精准、高效的检测方法^[8]。近些年,随着卷积神经网络方法被广泛应用,深度学习方法也开始被运用在网络健康谣言检测中^[9]。

(2)影响网络健康谣言信任感知、传播和分享等行为的因素分析。研究方法以在线问卷调查和在线模拟实验为主,此外也包括精细加工可能性模型(ELM)、结构方程模型、文本挖掘技术等其他方法。已有研究发现人口统计学变量^[10](性别、年龄、受教育水平、收入等),信息线索^[11](质量线索、吸引力线索、情感极性),附加信息^[12],健康素养^[13],科学素养^[14],个人情绪状态与谣言引发情绪的一致性^[15],情感类型^[16](恐惧或希望)等因素都正向或负向地影响着用户对网络健康谣言的信任感知和传播分享行为。

(3)网络健康谣言治理研究。主要集中讨论网络健康谣言的辟谣策略,通过对比实验比较幽默式辟谣与逻辑式辟谣^[17]、简单反驳式辟谣与事实阐述式辟谣^[18]、图片辟谣^[19]等不同辟谣方式的效果,比较普通用户与疾病预防控制中心(CDC)^[20]、强社会关系与弱社会关系^[21]等不同来源信息的辟谣效果,以及比较新闻机构与同行等不同辟谣对象的辟谣难度^[22]。与已有的网络谣言辟谣研究方法(如影响因素分析^[23-24]、真实网络数据分析^[25]、模拟仿真^[26]等)相比,网络健康谣言辟谣研究方法还比较单一。此

外,还有部分研究致力于探讨政府^[27]、医疗行业^[28]、公众^[29]等不同主体如何积极参与到网络健康谣言治理实践中。

1.2 尚需进一步研究的内容

(1)结合健康谣言的特征体系开展治理的研究。与一般谣言不同,健康谣言的生成方式与传播机制具备一定的独特性。首先,健康谣言在网络谣言中占比高,类型多,单一的预警或辟谣方式难以有效遏制健康谣言传播,需要采用批量的、全局的、系统的治理方式。第二,从生成方式来看,健康谣言内容具备较强的衍生性^[30]与关联性^[31],是其被自动识别的重要突破口^[32]。第三,健康谣言传播具备较强的周期性和反复性^[33],这充分凸显了健康谣言事前治理(如识别、预警)以及事后治理(信息公开与辟谣信息共享)的重要性。第四,健康信息的知识性和科学性特征决定了健康谣言是可识别且可辟谣的,同时,不同于其他谣言的观点表达(如社会事件谣言),健康谣言通常以谣言信息与真实信息穿插叙事的中长文本为主,以上特征决定了健康谣言在事中治理(辟谣)的可靠性与权威性原则,以及对健康领域专业共同体的充分依赖。由此可见,对健康谣言的治理需密切结合其特征体系来进行,才能具备更好的准确性和针对性。

(2)开展健康谣言治理顶层设计方面的研究。目前,对于网络谣言治理主要集中在技术视角下的识别预警、传播学视角下的辟谣或自消解、公共管理学视角下的多元共治等几个维度,识别针对谣言潜伏期、辟谣针对谣言增长期和高潮期、多元共治针对谣言的全生命周期。一方面,跨学科视角拓展了网络谣言治理的研究外延,丰富了网络谣言治理的路径;但另一方面,研究对象的泛化,研究目标的窄化(仅聚焦于某一阶段的治理方法)和虚化(扩大到全阶段策略的泛泛而谈),均导致网络谣言治理研究难以落地。研究者们试图找到一个普适的治理方案,但谣言是一个极其复杂的社会现象,谣言的主题、传播媒介、社会环境都会影响治理效果。对此,需要尝试聚焦某类典型代表,如网络健康谣言,针对其个性化特征,在治理战略的顶层设计和各阶段的具体治理战术上进行突破。

(3)开展面向健康谣言传播全生命周期治理的研究。网络健康谣言的治理是一个面向全生命周期,贯穿形成、发展到消退等各阶段的长线过程,包括对网络健康谣言的监控预警、应急、反馈等一系列治理手段和方法。然而,当前网络健康谣言治理实践过程较为关注高频健康谣言的应急式辟谣,而忽视了面向潜伏期的网络健康谣言监控预警,以及面向消退期的健康谣言传播规律特征复盘和数据归档沉淀,导致同质化的健康谣言反复传播,网络健康谣言的治理也始终处于被动地位。面向传播阶段的应急式辟谣是单一的阶段式治理方式,缺乏系统性与整体性,唯有面向健康谣言传播全生命周期的闭环治理才能更高效地缓解健康谣言泛滥的困境。

综上,当前网络健康谣言治理可以从网络健康谣言的多维特征入手,在现有多元主体联动治理方式的基础上,构建一个贯穿网络健康谣言全生命周期、群体智慧与人工智能相结合的多层次、多维度、有重点的治理框架。面向网络健康谣言传播全过程的治理框架立足于网络健康谣言关键特征与实际治理需求中的要点,着力阐明网络健康谣言多元治理主体在网络健康谣言治理过程中的具体分工。本文将从网络健康谣言传播的形成阶段和阻断方式分析开始,提出防谣、辟谣、止谣三阶段传播阻断方法,进

而提出网络健康谣言治理框架。

2 网络健康谣言的传播及阻断方式分析

网络健康谣言治理的根本要点在于对网络健康谣言内容的预警、判别和处置。具体来说,针对网络健康谣言内容的治理需要从其形成和传播过程入手,综合判断需要在什么传播阶段,针对什么对象,使用什么方法,进行什么类型的处置。要回答以上问题,首先需要对网络健康谣言传播的关键共性链路进行分析,找到共性链路的核心着力点,并针对关键位置/对象进行精准治理,从而实现不同阶段的网络健康谣言治理。

对此,笔者将网络健康谣言的治理分为三阶段传播阻断方法:第一阶段为防谣(源头阻断,防微杜渐),第二阶段为辟谣(受众教导,影响削弱),第三阶段为止谣(成果共享,信息复用)。如图1所示,抽象出了当前网络健康谣言传播的基本链路,以及针对网络健康谣言传播路径所划定的三道防线。

网络健康谣言传播基本链路的流程是:由网络健康谣言生产者发布健康谣言内容,经由传播媒介(如微博、微信等平台),触达网络健康谣言受众。在这一传播过程中,涉及了三个基本节点(健康谣言生产者、传播媒介平台、健康谣言受众)和两条基本链路(健康谣言生产者→传播媒介平台,传播媒介平

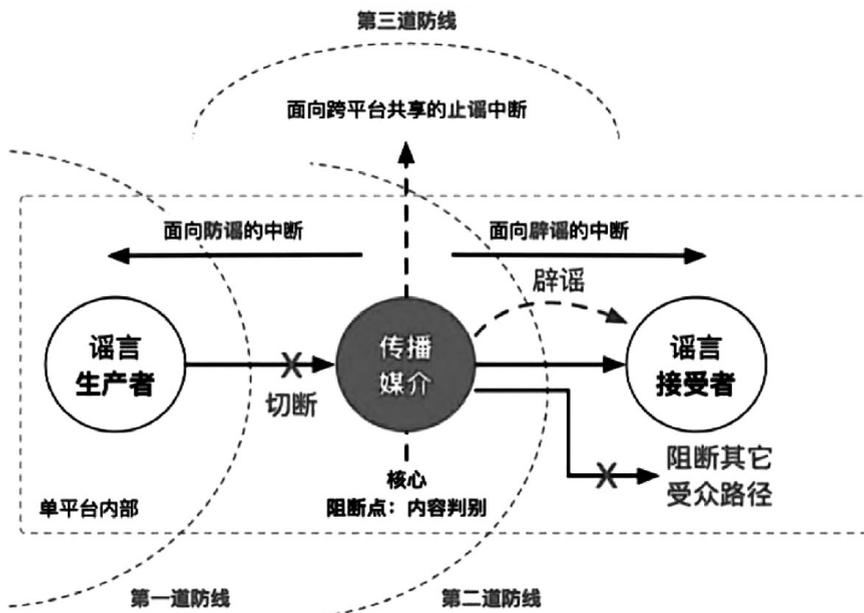


图1 网络健康谣言传播路径及其阻断方式

台→受众)。于此,可以非常明确地看出:对于单个平台内部的网络健康谣言治理而言,其核心要点在于切断两条基本链路,并以此构建两道基本的传谣防线;对于跨平台的网络健康谣言治理而言,其核心要点在于综合多个单平台的谣言内容和辟谣成果形成知识库,通过知识库的开放共享阻断网络健康谣言从当前平台向下一平台跨界传播,以此构建进阶版的第三道传谣防线。三道防线的作用目标、作用时期、作用地点、作用对象、主要任务如表1所示。

2.1 防谣:源头阻断,防微杜渐

第一道防线:防谣(源头阻断,防微杜渐),即在发布端阻断网络健康谣言生产者到传播媒介平台这一基础链路。从效率上来说,该防线是最直接,也是最快速的阻断方式,通过判别和过滤网络健康谣言内容,降低网络健康谣言流入公共信息池的风险。因此,这一阶段的根本要点在于对即将流向媒体平台的网络健康谣言进行事前处置,防止健康谣言的流入,从源头上阻断网络健康谣言的传播。在第一道防线中,网络媒体运营平台是最直接的处置中心,并且是第一责任主体,平台基于文本内容特征对用户传播内容进行判别和处置,能够在第一时间阻断该信息内容在本平台上的传播,从而完成第一道防线的任务。需要说明的是,在发布端识别网络健康谣言并阻断其信息流,能依赖的特征属性只有网络健康谣言的文本内容特征以及少量的用户静态属性特征(如年龄、性别、学历等)。此外,从网络健康谣言的周期性传播规律和文本内容要素的衍生性来看^[31],在发布端识别并阻断的对象主要以周期性/衍生性网络健康谣言为主。综上,第一道防线的作用时期是网络健康谣言传播的潜伏期,作用地点是信息发

布端,作用对象是周期性/衍生性网络健康谣言,主要任务是对健康谣言进行实时识别、标记或删除,显著提升网络健康谣言治理效率。

2.2 辟谣:受众教导,影响削弱

第二道防线:辟谣(受众教导,影响削弱),即在网络健康谣言形成并进入传播后,在传播端尽早阻断以防止网络健康谣言进一步扩散,同时对已收到网络健康谣言的用户进行及时辟谣。在网络健康谣言治理实践中,部分新型健康谣言因其新颖特征未被预警识别,从而突破第一道防线,成功发布到平台,并被用户阅读和感知,此时为了尽量减小健康谣言的危害性,需要在传播端及时采取行动。具体而言,在传播端阻断谣言的关键是:针对已形成并进入传播阶段的新型网络健康谣言,通过引入早期传播链路特征和用户动态行为特征,采用“随机采样+定点采样”的方式进行判别,同时结合优质用户的举报对高危疑似健康谣言进行聚合,并将该内容及时传递给辟谣专家进行核实并解读,及时阻断新型网络健康谣言的继续传播,同时对已触达的用户及时输入纠正性信息,也即辟谣。综上,第二道防线是针对第一道防线失效采取的“补救”性治理措施,作用时期是网络健康谣言的扩散期,作用地点是信息传播端,作用对象是新型网络健康谣言,主要任务是通过自动与人工相结合的手段进行早期预警,经专家核实后及时阻断传播并反馈辟谣。在这一防线中,辟谣专家工作平台是核心的责任人,通过对新型网络健康谣言快速高效的解读,降低该类网络健康谣言带来的社会影响。

2.3 止谣:成果共享,信息复用

第一道防线和第二道防线的设置,能够有效阻止网络健康谣言的发布与传播并降低其影响。然

表1 三道防线对比分析

三道防线	第一道防线	第二道防线	第三道防线
作用目标	防谣	辟谣	止谣
作用时期	潜伏期	扩散期	消退期
作用地点	发布端	传播端	沉淀端
作用对象	周期性/衍生性网络健康谣言	新型网络健康谣言	已识别并处置过的网络健康谣言
主要任务	实时识别、标记或删除	早期预警与及时辟谣	共享与复用

而,当前媒体平台存在多样性,一个用户通常活跃于多个不同的媒体平台,加之网络健康谣言搬运工的普遍存在,导致同一则健康谣言会在不同时期/平台反复传播,即所谓的旧谣新传现象。因此,在以往的单平台、单周期治理过程中,常常出现辟谣信息共享不及时导致重复检测、遗漏检测和影响扩散的现象。基于此,笔者针对单时空治理可能存在的信息不对称问题,构建了第三道防线:止谣(成果共享,信息复用)。第三道防线(止谣)的根本目标在于提升单时空的治理效果,促进跨平台治理信息与成果的实时共享,防止在单一平台上已被处置的健康谣言从横向继续扩散到其他平台,从纵向上经过一段周期后反复传播,从而最大程度降低信息差带来的高治理成本,增加造谣或传谣难度。例如,某个用户在多个媒体平台发布了某则新型网络健康谣言内容,但该谣言在单个平台上的传播特征并不明显,此时可以将多个媒体平台的传播数据进行聚合,以此识别谣言的传播共性,从而能够在最短的时间内进行辟谣和处置,提升网络健康谣言治理的敏感性和效率。综上,第三道防线的作用时期是网络健康谣言的消退期,作用地点是信息沉淀端,作用对象是经第一道防线与第二道防线识别并处置过的网络健康谣言,主要任务是对辟谣成果进行共享与复用,主要目标是防止已处置过的健康谣言再生。

3 网络健康谣言治理框架

从网络健康谣言传播的形成及阻断方式分析结果来看,网络健康谣言治理的要点是:在网络健康谣言传播的发布端、传播端以及沉淀端分别设置三道防线进行阻断,以网络健康谣言的判别为起点,构筑起以网络媒体运营平台为中心的防谣、以辟谣专家工作平台为中心的辟谣、以辟谣信息共享平台为中心的止谣三道基础防线。

据此思路,本文提出了网络健康谣言治理框架,如图2所示,旨在厘清网络健康谣言的核心传播链路,并通过畅通信息协作与共享的渠道,来实现网络健康谣言的有效治理。框架图中的主要模块构成可以分为一个核心技术模块(内容判别技术模型)和三个中心平台模块(网络媒体运营平台、辟谣专家工作平台和辟谣信息共享平台),接下来将分别针对上述模块的基本构成进行概述。

3.1 内容判别技术模块

内容判别技术模块是整个网络健康谣言治理框架的起点,即谣言治理先从识别入手,然后根据其传播阶段、类型特点,选择不同的处置方式。对于第一道防线,需要在发布端基于文本内容特征对网络健康谣言进行识别(主要面向周期性/衍生性网络健康谣言),进而进行删除或标记等处置。前已述及,已有研究从理论和模型实证上均已充分证明,健康谣

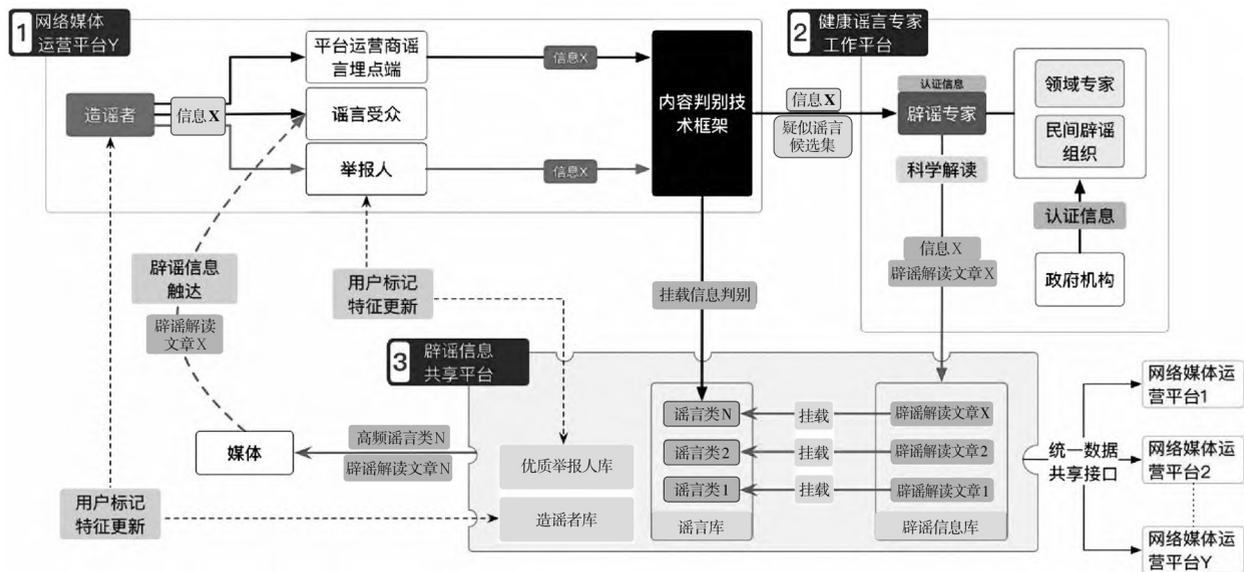


图2 网络健康谣言治理框架

言的衍生性与关联性特征是其被识别的重要标记特征。在当前的健康谣言事前治理中,关键词或主题特征也是用来识别重复性健康谣言的常见手段。由此可见,充分利用健康谣言的衍生性文本特征和周期性传播特征,可以顺利实现第一道防线的治理目标。第二道防线,需要在传播端基于文本内容特征、传播链路特征、用户行为特性对网络健康谣言(主要面向新型网络健康谣言)进行早期识别预警,以便筛选出高危疑似集,减轻专家审核解读的工作量,提高辟谣效率。基于多维特征(语言统计特征、网络特征、交互特征等)的谣言早期预警一直是谣言识别研究中的重要关注点,也形成了一系列的机器学习识别模型^[34]和深度学习识别模型^[9]。虽然在大规模范围上的识别精确度有待进一步验证,但可以肯定的是,作为高危疑似健康谣言辅助筛选手段,可大大提高健康谣言治理效率。第三道防线,基于已识别出并处置过的网络健康谣言的模型与特征信息进行共享复用。由此可见,内容判别模块是健康谣言治理的起点,判别结果的质量直接影响着治理的效率和效果。

3.2 网络媒体运营平台

网络媒体运营平台主要指多种在线网络媒体运营平台,如微博、微信、今日头条、小红书等,该类平台的主要特点包括内容的丰富性、巨量性和高时效性等,同时,平台中所传播的大量内容通常由各类用户自主生成(User-Generated Content, UGC),其论述内容的真实性通常需要进一步验证。由于网络媒体运营平台是网络健康谣言信息流转的载体,因此整个治理框架中防谣功能的实现是由网络媒体运营平台在发布端通过实时判别并处置来实现的。但同时,网络媒体运营平台还承担着待辟谣的可疑健康谣言筛选、辟谣信息定向触达等任务,以辅助下一步辟谣工作的顺利展开。

3.3 网络健康信息辟谣专家工作平台

网络健康信息辟谣专家工作平台是一个以辟谣专家为中心的、基于政府权威机构认证的、疑似健康谣言解读与分析平台。该平台主要承担疑似网络健康谣言的科学解读、专家认证、专家奖励和科普宣讲工作,其数据的输入来源为网络媒体运营平台的代

表性疑似谣言集,数据的输出结果为与该健康谣言对应的判别结果和科学解读信息。该工作平台可以看作疑似网络健康谣言的“审判中心”,通过将相关健康谣言分发给对应领域的权威专家进行拆解分析,并结合投票模型给出最终判别结果。

3.4 辟谣信息共享平台

辟谣信息共享平台是对与网络健康谣言相关的信息进行特征抽取、特征融合和数据共享的服务平台。通过采取三统一(标记统一、特征统一、接口统一)的策略,形成全域共享与共治的健康谣言数据链接渠道,最大程度降低跨平台之间的信息传输与融合成本,同时通过融合不同平台之间的数据特征,进一步提升单一平台和场景下的治理效果。例如,对于周期性或衍生性健康谣言而言,跨平台特征共享,使得造谣者无法利用平台信息差(特征差)来规避检测,能最大程度降低该类健康谣言的传播概率;对于新型健康谣言而言,虽然在起始阶段的单一平台中传播量较小导致难以检测,但是通过对多个平台的数据进行融合检测,有助于将该细微特征进行放大,从而能够在传播早期得到及时有效的检测结果,以此降低健康谣言带来的影响。

4 网络健康谣言治理框架模块功能分析

治理框架清晰阐述了网络健康谣治理各阶段的治理要点、治理主体、治理对象及治理需求,为保证各阶段治理目标的顺利实现,下文将进一步细化治理框架中各模块的功能架构及其实现路径。

4.1 以网络媒体运营平台为中心的防谣模块功能分析

作为网络健康谣言形成与传播的核心模块,网络媒体运营平台(如微博平台、微信平台、搜狐平台等)对于网络健康谣言信息流转的基本链路(形成、传播、管理、监控、特征复用等)具有重要的媒介与处置作用。因此,从数据的逻辑流转层来说,平台运营商是领域专家辟谣模块、跨域共享模块等核心关联模型的重要数据转接口。同时,从技术的前沿性和落地的有效性来说,由于平台处于网络健康谣言传播的第一线,因此其数据真实性和实时性都能够得到有效保障。由此,对于每一个平台运营商而言,其内部的核心逻辑功能主要包括:数据特征抽取、内容判

别、人群/内容标记和发布管控。据此,笔者设计出了网络媒体运营平台防谣功能架构,如图3所示,从右至左包括发布端的管控与隔离功能,传播端的聚合、判别、挂载与共享功能,数据端的标记复用与共享功能三大功能模块。

(1)发布端功能分析。发布端在网络媒体运营平台治理过程中承担沙箱隔离与管控的重要作用。首先,对于周期性/衍生性网络健康谣言,通过基于文本内容的识别模型,在最大限度和最快时效内对其检测识别并进行标记或删除等处置。其次,对于多次被检测出发布健康谣言的用户,当其发布次数达到特定阈值后,可对其进行禁言处置。再次,对于多次阅读或转发健康谣言的易感人群,进行重点保护,重点推送辟谣信息及优质健康信息。总之,发布端主要目标是将“危险消灭在源头”,重点解决网络健康谣言中的绝大部分周期性/衍生性谣言。

(2)传播端功能分析。传播端的核心功能在于对新型网络健康谣言进行判别预警与疑似聚合。首先,结合谣言传播特征、谣言发布用户特征以及用户举报特征等对冲破第一道防线的新型网络健康谣言进行判别和疑似性分析。其次,将高疑似健康谣言进行聚合,经过筛选后形成有限的候选集合,进而传输给相应领域的网络健康谣言辟谣专家进行科学解读和分析。最后,将反馈的辟谣信息回流给已知的

受影响用户,从而形成完整的数据链路闭环,大幅提升网络健康谣言的处置效果。

在传播端,除了实现内容判别模型中新型网络健康谣言的判别预警外,还需要着力优化当前的网络健康谣言举报渠道与方式,充分调动优质用户参与健康谣言把关的积极性。现有网络辟谣平台举报渠道以电话、邮箱为主,少数平台提供了简单的表单提交渠道,这些渠道既无跟踪机制也无反馈机制,在一定程度上消磨了用户举报的积极性。因此,传播端需要从举报方式、举报管理与分析机制、举报反馈机制和举报激励机制等多方面进一步优化。

(3)数据端功能分析。数据端的核心功能在于通过协定的通用数据形式和接口进行跨平台数据共享,将自身平台的数据特征输出,并将外部共享的数据特征接入,从而进一步提升当前模型的数据效果。例如,通过协定的加密和表示方案对用户ID进行表示,使得每个用户都拥有全网唯一的标记,形成“一处造谣、处处受限”的联动机制,抬升用户的造谣成本。

4.2 以辟谣专家工作平台为中心的辟谣模块功能分析

对于突破第一道防线的新型网络健康谣言,需要在第二道防线进行处置。作为参与网络健康谣言辟谣的重要角色,科学辟谣专家的认证、解读与激励

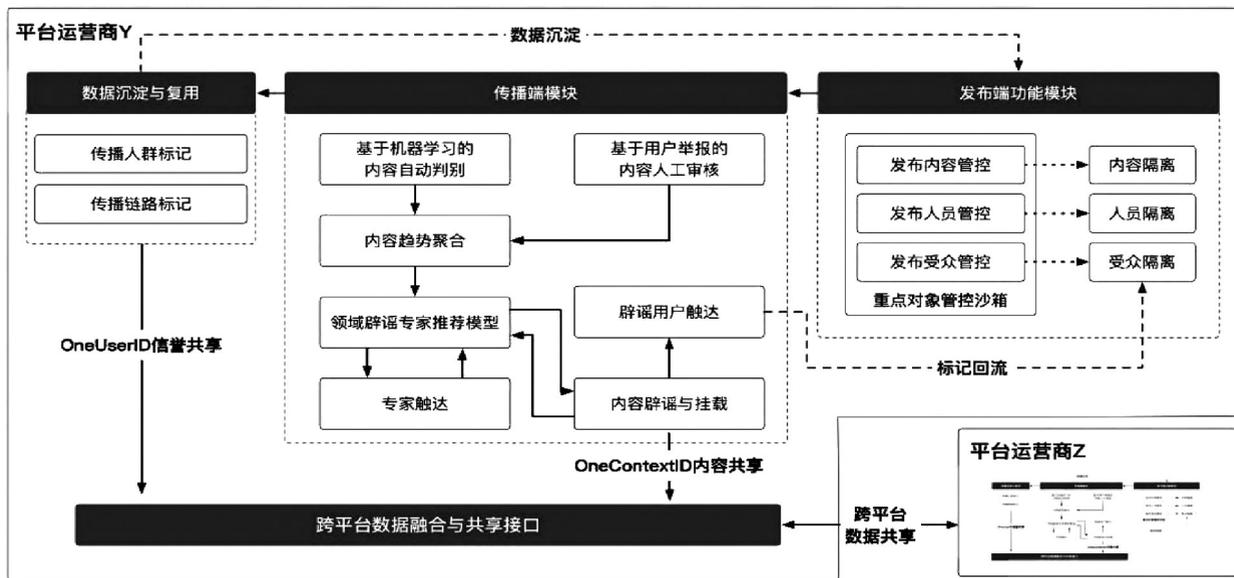


图3 网络媒体运营平台防谣功能架构

关系直接影响着网络健康谣言的辟谣效率。因此,笔者设计出了辟谣专家工作平台辟谣功能架构,如图4所示,包含三个核心部分:专家认证模块、内容分发与解读模块、效果评估与激励模块。

专家认证模块的核心功能在于将相应领域的专家以政府授信的形式引入专家工作体系。具体内容如下:首先,专家推荐,来自高校、科研院所、医院等官方机构的专家以及来自商业医疗机构或民间公益组织的民间科学家,只要符合招募领域范围,均可采用自荐或他荐模式申请加入网络健康谣言的研判、分析和审核的专家库中。然后,政府授信,政府机构需对自荐或他荐的专家进行初始授信(认证),以卫健委为代表的政府机构可根据自荐或他荐专家的学历、从事工作、科研成果、社会威望等多个维度对其进行认证。接着,领域评估,在认证基础上对专家擅长的辟谣领域及领域权威值进行评估,并对专家擅长的领域进行数字化、特征化筛选,方便后续待辟谣内容的分发与解读。

内容分发与解读模块的核心功能在于将候选的健康谣言分配给对应领域的权威专家进行解读和审查,并返回结果。具体内容如下:首先,在接收到网络媒体运营平台针对高危可疑健康谣言集合发出的

研判审核请求后,辟谣专家工作平台根据领域评估结果逐一分发给各位领域专家(高信誉值专家)。其次,领域专家针对接收到的疑似谣言进行研判、分析、审核。然后,针对多位专家的判定与解读结果,借助专家投票模型产出最终的评估审查结果。最后,将最终的评估审查结果反馈至跨平台数据融合与共享接口。

效果评估与激励模块的核心功能在于对专家辟谣工作的效果、效率以及影响进行评估,并据此以适当形式给予专家激励,如采用津贴或者荣誉值的方式。该模块旨在调动专家参与辟谣的积极性,同时保证网络健康谣言研判的及时性与可靠性,从而最终形成“健康谣言接入—辟谣解读—辟谣激励”的完整闭环。

4.3 以辟谣信息共享平台为中心的止谣模块功能分析

辟谣信息共享平台的核心功能在于对网络健康谣言治理过程中产生的一切数据进行统一规划、统一调配和统一使用,降低不同健康谣言治理平台之间的数据共享成本,避免网络健康谣言跨平台的共时或历时的周期性传播,从而提升网络健康谣言的治理效率。笔者设计出了辟谣信息共享平台止谣功

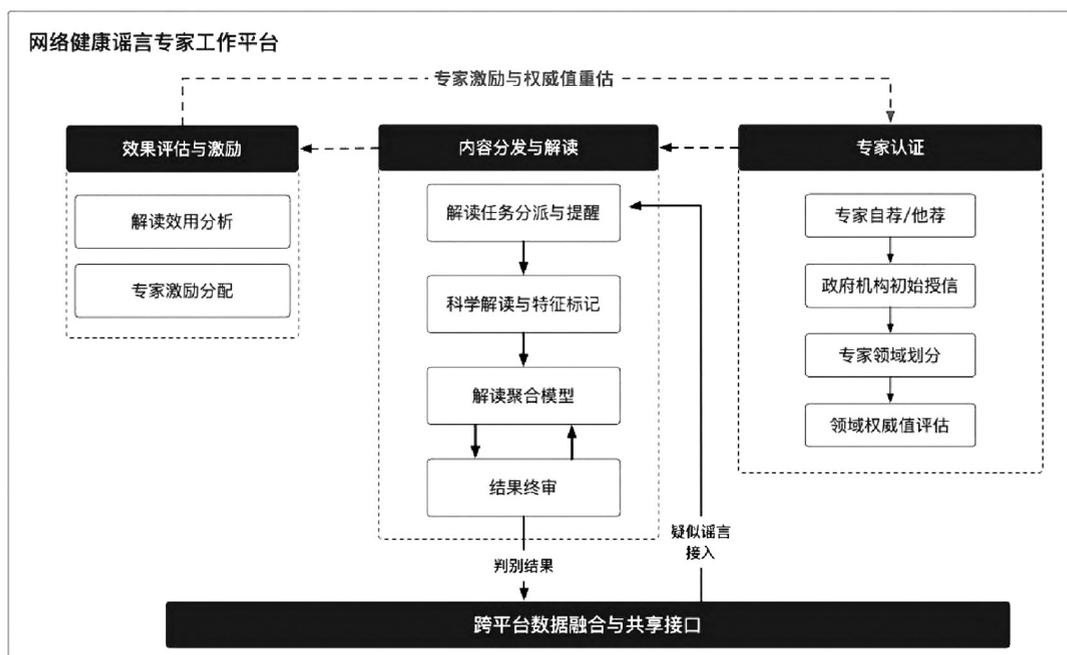


图4 辟谣专家工作平台辟谣功能架构

能架构,如图5所示,可以分为三个层面:数据资源接口层、数据模型层和数据应用层。从逻辑关联上来说,上述层次存在递进关系。

首先是数据资源接口层。该层的核心功能是对网络健康谣言形成、传播和辟谣过程中涉及的核心数据进行统一定义和类型规范。具体内容包括:统一文档接口(Unified Document Interface, UDI),其典型的数据来源包括健康谣言内容文档、辟谣解读文档、用户评价语料等;统一用户接口(Unified User Interface, UUI),其典型的数据来源包括健康谣言传播者、健康谣言举报者、辟谣专家等;统一特征接口(Unified Feature Interface, UFI),包括健康谣言传播者的用户画像特征、健康谣言传播趋势特征等。在数据资源接口层定义UDI、UUI和UFI,有助于提升数据跨平台传输与融合的准确性和有效性,减少由数据污染带来的模型效果降低的现象,从而形成统一的数据资源池,提高网络健康谣言的治理效率。

其次是数据模型层。该层的核心功能是对网络健康谣言治理过程中涉及的核心问题进行数据特征抽取和建模。具体内容包括:人群跨域标记模型、谣言—解读挂载模型、谣言—人群挂载模型、专家—领域挂载模型、人群统一信誉评估模型、网络健康谣言趋势聚合模型以及统一特征融合模型等。该层可以被视为对底层输入数据在模型层面的初步加工与处理,通过特征分析、抽取与融合,形成统一的特征备

选框架,使其能够适用于多个平台的多种应用,从而扩展底层数据的应用范围。例如,人群统一信誉评估模型旨在依据用户跨平台的网络健康谣言行为,对用户进行信誉评估,为后续健康谣言传播判别、热点趋势分析等典型应用提供特征参考。

最后是数据应用层。该层处在辟谣信息共享平台的最顶层,其核心功能是对网络健康谣言传播中需聚合多个平台才能得出的数据进行分析。具体内容包括:网络健康谣言热点趋势榜、网络健康谣言预警榜以及网络健康谣言辟谣机器人。其中,网络健康谣言热点趋势榜旨在对多个平台正在传播的热门网络健康谣言进行聚合分析,给出热门健康谣言排序,而其他新闻媒体平台则可以通过应用接口将该榜单分发至自身的媒体频道,从而扩大止谣效果。网络健康谣言预警榜旨在通过聚合不同平台上正在传播的微小量级的疑似网络健康谣言,对可能形成传播趋势的网络健康谣言内容进行预警,将预警结果反向传回各网络媒体运营平台,提升网络预警的时效性,减小网络健康谣言的影响面。

因此,以辟谣信息共享平台为中心的止谣模块的关键在于构建全面、高效、可持续的治理数据融合与共享体系,该体系以“处置结果”为中心,包含“治理对象信息”“数据模型信息”“对象特征信息”“处置结果信息”等要点。短期内止谣模块的建设需考虑以下两点:其一,在政府的协调与指导下(如国家网

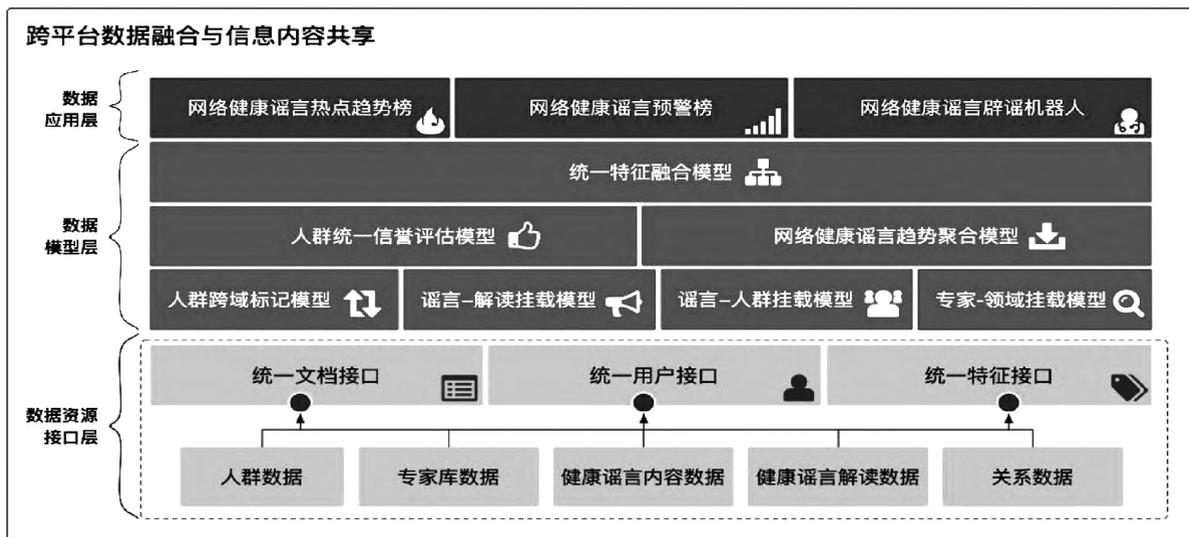


图5 辟谣信息共享平台止谣功能架构

信办等),各网络媒体传播平台就统一接口、统一协议、统一特征的体系框架进行协商与构建,明确政府组织和不同平台组织之间信息共享的权利与界限,弥合辟谣信息共享平台的实体鸿沟,畅通信息共享的政策渠道,提升信息共享的对接与实施效率。其二,针对不同平台的数据类型与数据特征集合,构建通用的平台数据价值体系方案,增强大平台在网络健康谣言治理任务中的“责任感”和小平台的“参与感”,从而提升网络健康谣言治理的广度与深度。

综上,止谣的根本目的是借助统一的数据与特征资源接口构建一个辟谣信息共享平台,为健康谣言相关信息的有效共享与复用奠定基础,从而提升社会整体资源的利用效率。

5 研究总结

本文重点分析讨论网络健康谣言从形成到消退的传播过程中,各个关键节点的治理逻辑与方法,提出了防谣—辟谣—止谣的网络健康谣言阻断方式,在此基础形成了一套相对完整、系统的网络健康谣言治理框架,具体的理论价值与实践意义如下。

5.1 理论价值

与常见的健康谣言阶段式、离散式治理策略不同,本研究从全局性、整体性的角度构建了网络健康谣言闭环治理框架,确定了健康谣言治理的主要阶段及目标,以及各阶段治理的相互关系及手段。整个框架的设计立足实际并层层递进,具体表现为:

首先,立足于健康谣言泛滥的实际问题,面向网络健康谣言自发布到消退的全生命周期,通过抽象其实际传播方式及各个传播阶段的阻断方式,进而总结分析网络健康谣言治理过程的主要阶段及关键点,即切断三条基本链路,构筑三道防线。

然后,严格按照健康谣言生成与传播特性,详细分析了各个阶段的治理目标、治理对象和治理手段,并据此归纳分析得出闭环治理框架。在潜伏期,要充分利用健康谣言内容的关联性,识别衍生性与周期性健康谣言,同时尽可能的预警新型健康谣言。在扩散期,要充分利用健康信息的知识性与科学性,高效地匹配领域专家以便及时辟谣。在消退期,要充分关注健康谣言的周期性,及时沉淀单平台、单周期内的健康谣言相关数据,防止健康谣言的跨周期

与跨平台传播,实现高效的闭环治理。

最后,以各阶段治理目标为导向,聚焦各阶段的治理需求点,详细揭示框架各功能模块架构及其实现路径,以保证潜伏期的防谣、扩散期的辟谣和消退期的止谣的顺利实现。

5.2 实践意义

本研究所提出的治理框架并非空中楼阁,其中许多局部的功能模块在现实实践中早有雏形,涉及的健康谣言特征提取技术及识别技术,在当前理论研究中也取得了大量的研究成果。由此可见,无论是实践层面还是技术层面,本研究提出的治理框架均具备一定的可行性。

对于网络媒体运营平台,微信平台目前推出的“谣言过滤器”公众号,可以接受用户的谣言举报,同时推送给经平台认证的第三方辟谣机构进行核实,一旦核实为谣言,平台将对该文章进行标记,并附上辟谣公众号的名称、辟谣结论以及与辟谣结论相关的科普文章。微信平台作为典型的网络媒体运营平台,基本实现了治理框架中用户举报—专家辟谣—辟谣推送—谣言标记的基本工作流程。当然,在周期性谣言的技术治理等方面的实践应用还有待进一步加强。

对于辟谣专家工作平台,当前,由中央网信办指导,中国科协、卫健委等部门主办的科学辟谣平台^[35]已开始尝试聚集领域专家开展科学谣言的解读工作(目前已入驻医学、食品安全、营养健康等相关领域的1881名专家学者),为辟谣专家工作平台的构建提供了一个很好的原型。为了有效发挥辟谣模块的功能,辟谣专家工作平台还应进一步细化专家的认证、匹配与解读以及激励等方面的实施要点,充分集中专家的力量,降低信息核实成本。

对于辟谣信息共享平台,健康谣言治理过程中相关数据信息的沉淀与复用是整个治理框架中的重要部分,也是实现闭环治理的关键所在,需要多方治理主体的充分协作,更需要国家层面的政策引导,这也为辟谣信息共享平台功能的充分实现增加了难度。目前,各独立平台内部的辟谣信息共享已基本实现,如各大平台每月公布的谣言榜,但平台之间的辟谣信息共享还面临一些阻力。伴随着2023年3月

国家数据局的成立^[36],重要信息资源开发利用与共享、信息资源跨行业跨部门互联互通以及公共服务与社会治理信息化被提上日程。相信在不久的将来,在数字中国建设进程中,平台间的信息壁垒终将被打破。

综上所述,本研究所提出的治理框架基本涵盖了当前网络健康谣言治理需要考虑的核心问题,既可以作为健康谣言治理活动的顶层设计基础,也可作为查漏补缺的行动指南。

参考文献:

[1]华颖. 健康中国建设:战略意义、当前形势与推进关键[J]. 国家行政学院学报, 2017(6): 105-111, 163.

[2]中国互联网联合辟谣平台 2022 年度网络谣言盘点[EB/OL].[2023-02-12].<https://www.piyao.org.cn/20230130/17e16497990946e9b350fd80d04be6a5/c.html>.

[3]ZHENG X, WU S, NIE D. Online health misinformation and corrective messages in China: a comparison of message features[J]. *Communication Studies*, 2021(5): 1-16.

[4]YLA B, ZFA B, XY C, et al. Recognizing fake information through a developed feature scheme: a user study of health misinformation on social media in China[J]. *Information Processing and Management*, 2022(59): 1-14.

[5]李月琳,张秀,王姗姗. 社交媒体健康信息质量研究:基于真伪健康信息特征的分析[J]. 情报学报, 2018, 37(3): 294-304.

[6]BAGAVATHI A. Contrasting misinformation and real-information dissemination network structures on social media during a health emergency[J]. *American Journal of Public Health*, 2020, 110(S3): S340-S347.

[7]PU G, JIN L, XIAO H, et al. Systematic evaluation of COVID-19 related Internet health rumors during the breaking out period of COVID-19 in China[J]. *Health Promotion Perspectives*, 2021, 11(3): 288-298.

[8]LIU Y, YU K, WU X, et al. Analysis and detection of health-related misinformation on Chinese social media[J]. *IEEE Access*, 2019(7): 154480-154489.

[9]赵月华,朱思成,苏新宁. 面向网络虚假医疗信息的识别模型构建研究——一种基于预训练的BERT模型[J]. 情报科学, 2021, 39(12): 165-173.

[10]PAN W, LIU D, FANG J. An examination of factors contributing to the acceptance of online health misinformation[J].

Frontiers in Psychology, 2021, 12(3): 1-11.

[11]SONG X, ZHAO Y C, SONG S, et al. The role of information cues on users' perceived credibility of online health rumors [J]. *Proceedings of the Association for Information Science and Technology*, 2019, 56(1): 760-761.

[12]邓胜利,付少雄. 社交媒体附加信息对用户信任与分享健康类谣言的影响分析[J]. 情报科学, 2018, 36(3): 51-57.

[13]SONG S, ZHAO YC, SONG X, et al. The role of health literacy on credibility judgment of online health misinformation [C]. 2019 IEEE International Conference on Healthcare Informatics, Xi'an, China, 2019: 1-3.

[14]HE L, CHEN Y, XIONG X, et al. Does science literacy guarantee resistance to health rumors? The moderating effect of self-efficacy of science literacy in the relationship between science literacy and rumor belief[J]. *International Journal of Environmental Research and Public Health*, 2021, 18(5): 1-10.

[15]NA K, GARRETT R K, SLATER M D. Rumor acceptance during public health crises: testing the emotional congruence hypothesis[J]. *Journal of Health Communication*, 2018, 23(8): 791-799.

[16]SOON JIA J, BANERJEE S, CHUA AYK. Analyzing medical personnel's perceptions of online health rumors[J]. *Lecture Notes in Engineering and Computer Science*, 2017(2227): 457-460.

[17]VRAGA E K, KIM S C, COOK J. Testing logic-based and humor-based corrections for science, health, and political misinformation on social media[J]. *Journal of Broadcasting & Electronic Media*, 2019, 63(3): 393-414.

[18]TONI GLA, VAN DER MEER, JIN Y. Seeking formula for misinformation treatment in public health crises: the effects of corrective information type and source[J]. *Health Communication*, 2020, 35(5): 560-575.

[19]VRAGA E K, BODE L. Addressing covid-19 misinformation on social media preemptively and responsively[J]. *Emerging Infectious Diseases*, 2021, 27(2): 396-403.

[20]VRAGA E K, BODE L. Using expert sources to correct health misinformation in social media[J]. *Science Communication*, 2017, 39(5): 621-645.

[21]CHOUDHURY, N. Ageinag, health misinformation and mobile messaging apps[J]. *Catalan Journal of Communication and Cultural Studies*, 2021, 13(2): 233-247.

[22]WALTER N, BROOKS J J, SAUCIER C J, et al. Evaluating the impact of attempts to correct health misinformation on

social media: ameta-analysis[J]. Health Communication, 2020(4): 1-9.

[23] BORDIA P, DIFONZO N, HAINES R, et al. Rumors denials as persuasive messages: effects of personal relevance, source, and message characteristics[J]. Journal of Applied Social Psychology, 2005, 35(6): 1301-1331.

[24] 熊炎. 辟谣信息构成要素: 一种整合框架——二战以来西方辟谣实证研究回顾[J]. 广东行政学院学报, 2015, 27(1): 5-12.

[25] 宗干进, 黄子风. 社交媒体辟谣信息的扩散机理研究——基于辟谣信息演化网络的纵向分析[J]. 现代情报, 2017, 37(9): 44-52, 121.

[26] 滕婕, 夏志杰, 占欣. 基于改进 CA 模型的群体辟谣信息扩散效果预测[J]. 计算机工程与应用, 2020, 56(6): 51-57.

[27] 卡斯·R·桑斯坦. 谣言[M]. 张楠迪扬, 译. 北京: 中信出版社, 2010: 57.

[28] HERMANSYAH A, SUKORINI A I, RAHAYU T P, et al. Exploring pharmacist experience and acceptance for debunking health misinformation in the social media[J]. Pharmacy Education, 2021, 21(2): 42-47.

[29] 黄崑, 郭森, 郝希嘉, 等. 公共健康危机事件下健康信息素养文献综述[J]. 图书馆杂志, 2020, 39(7): 59-69, 82.

[30] 陈燕方, 周晓英. 基于文本特征融合的衍生性网络健康谣言识别模型研究[J]. 图书情报工作, 2023, 67(14): 73-84.

[31] 黄森, 黄佩. 基于知识关联特征的网络内容识别——以健康谣言为重点[J]. 北京邮电大学学报(社会科学版), 2020, 22(1): 1-6, 13.

[32] 谭励, 王舸, 周丽娜, 等. 基于多示例学习的食品健康领域长文谣言检测[J]. 计算机工程与设计, 2022, 43(11): 3101-3107.

[33] 刘鹏飞, 周悦. 食品安全谣言的法律处置[J]. 中国报业, 2015(13): 52-53.

[34] SAEED F, YAFOOZ W, AL-SAREM M, et al. Detecting health-related rumors on Twitter using machine learning methods [J]. International Journal of Advanced Computer Science and Applications, 2020, 11(8): 324-332.

[35] 中国科协. 科学辟谣[EB/OL]. [2023-08-25]. <https://piyao.kepuchina.cn/>.

[36] 新华网. 组建国家数据局[EB/OL]. [2023-08-25]. http://www.news.cn/politics/20231h/2023-03/07/c_1129419141.htm.

Research on the Spread Blocking and Governance Framework of Online Health Rumors

Chen Yanfang Zhou Xiaoying Pei Junliang

Abstract: [Purpose/significance] Design a multi-dimensional and multi-level comprehensive governance framework for online health rumors to curb the spread and misuse of online health rumors. [Method/process] By abstracting the dissemination process of online health rumors and the blocking methods at each stage of dissemination, the main stages and key points of online health rumors governance are summarized and analyzed. Then, following the characteristics of health rumors generation and dissemination, each stage's governance objectives, governance targets, and governance measures are analyzed in detail, and a closed-loop governance framework is derived accordingly. [Result/conclusion] The governance framework mainly includes one core technology module (content identification technology model) and three central platforms modules (the platform for the online media, the platform for the rumor-refuting experts and the platform for sharing rumor-refuting information), which basically covers the core issues that need to be considered in current online health rumors governance. It can serve as the top-level design foundation and action guide for health rumors governance activities.

Key words: Health rumors transmission; Health rumors governance; Preventing rumors; Refuting rumors; Stopping rumors