

名人名家热点说之二 | 社交机器人

社交机器人是一种自主机器人,能够根据自己身份相称的社交行为和规范,与人类或其他实体进行互动与沟通。社交机器人的出现,对人类社会将产生怎样的影响?面对社交机器人泛滥,人类应该怎么办?带着一系列问题,华东周刊连线有关专家学者,作了访谈。

■ 华东周刊 马晓才 / 文

刘博

安恒信息首席科学家、高级副总裁

一、关于人机交互,虚拟账号收割虚假性流量。

现象:虚假流量最初出现在网络上播出的部分影视作品的点击量、评论量等数据中。除了影视圈,娱乐圈内各明星涉及流量造假的现象更为严重。2018年8月,当红偶像明星蔡徐坤发布了一条包含其新歌MV的新浪微博,该微博仅仅在九天之内转发量就突破了一亿大关。当年新浪微博的用户总人数为3.410亿,一首新歌的转发量过亿明显不符合当时的社会逻辑。

治理:作为流量造假的高发生地,各社交媒体平台当仁不让应当成为自身管控的第一主体。在平台中一旦发现流量造假的情况发生,要第一时间追责造假者。社交媒体平台自身管控无疑有着重要作用,但其作为企业势必有利盈利的目的,对于虚假流量的治理动力必然会不足,这就使得平台不可能成为唯一的治理主体。需要第三方的主管单位和技术公司联合对此行为进行综合治理。特别是当引入ChatGPT类型的机器人之后,虚拟人机交互的成本会更低,技术上也会更难甄别。

二、关于批量复制,信息污染降低社会能见度。

现象:大众传播中的虚假、伪劣信息具有欺骗性和误导性,是影响最坏的、受众意见最大的信息污染。这些信息批量复制,充斥网络,会导致用群众获取有用知识难度加大。政府在宣传上需要花费更多的精力才能让民众看到其宣传内容。

治理:政府和主流媒体及搜索引擎公司合作,构建专业的信息发布渠道,并治理信

息污染,让民众不论是被动还是主动,都能有官方渠道获取高质量信息。同时可以考虑通过低成本区块链等技术对重要新闻信息进行指纹信息上链,非上链重要信息不与发布,上链信息如果出现问题可以很容易的进行溯源。

三、关于算法联动,僵尸网络影响涉政判断力。

现象:英国脱欧期间不到1%的社交机器人生产了Twitter中32%的脱欧内容。在新冠肺炎疫情肆虐的背景下,社交机器人传播新冠病毒来源、种族仇恨、政治阴谋论等相关内容,与人类用户关注公共健康和福利形成鲜明对比。

治理:社交平台建立便于政府机构搜集和管控的信息流机制来抑制算法联动。在信息发布前,用AI+人工的方式对信息进行分类,如果有出现涉政判断力内容,限制发布量和转发量,比如非权威账号不能转发。不过有些信息的确很难事前分类,那么社交平台要具备能力在发生之后能够及时屏蔽或者撤回的功能,由于信息量很大,利用AI算法辅助识别不可避免。

金小刚

浙江大学计算机科学与技术学院教授

目前学界对社交机器人(Social bots)没有统一的定义,但普遍认为社交机器人是一种智能化算法程序,通过模仿人类语言或行为,与人类用户进行互动。社交机器人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。社交机器人按其功能的不同可分为四类:支持远程操控的恶意僵尸网络(malicious botnets)、用于爬取网络数据的调研机器人(research bots)、用于编辑网络信息的编辑机器人(editing bots)以及对用户基础请求进行回应的聊天机器人



(chat bots)。

社交机器人“伪装”成人类,隐秘地渗透进公共话语空间,低成本、高效率地传播特定信息,为开展“计算宣传”提供了可能。计算宣传(computational propaganda)是由Woolley和Howard于2016年提出的,指“在社交媒体网络上故意散布误导信息为目的,使用算法、自动化和大数据分析等方式操纵公众舆论”。

在人工智能时代,社交机器人可实现自动化大量转发特定信息,放大某一公共话题的影响力,改变公众心中对于议题的重要性排序。机器人通过推送大量同一主题的信息,形成话题热度假象,在沉默的螺旋作用下该话题被不断强化,成为主流的舆论风向。转发支持某一倾向的信息,传播垃圾信息混淆视听,干扰舆论。通过推送大量与议

题不相关信息来转移话题,即烟雾遮蔽(smoke screening),也叫标签劫持(hijacking the hashtag)。互联网为多元化的声音提供了讨论与对话的平台,不同的观点与意见在这里相遇、碰撞与融合。社交机器人发布极端的情绪化的信息,加剧了极端化情绪与观念的对抗,破坏了网络生态的理性讨论氛围,让舆论随时有失控的风险。

随着社交机器人干预舆论的程度越来越深,策略越来越多样化,社交机器人正在成为影响舆论的重要因素,通过传播虚假信息、营造虚假气氛、与更多人建立关系等手段操纵舆论,甚至对用户线下认知与观念产生影响。当社交机器人参与了网络后,媒体生态发生改变,从完全由“人”主导变为“人+社交机器人”的共生状态,社交机器人逐渐消解了人类用户在社交媒体中的唯一主体地位,成为信息内容生产、观念传播和意义表达的重要参与者。

网络舆论场本来就是一个被制造的空间,但当生成式AI和聊天机器人等人工智能技术的发展,让社交机器人参与到网络舆论中,干预舆论的程度越来越深、策略越来越多样化,一些假新闻在网络中被迅速传开,虚假标签活动被推至热搜,网络舆论场真实信息的价值密度会越来越低。社交机器人等技术应用会形成支持某一观点或某一方的强势音量,民众会产生与多数意见不一样的错觉,进而保持沉默,人工智能技术制造的沉默螺旋现象会显现。民众希望从网络舆论空间获得真实信息的难度加大,发声的意愿也会降低。

人工智能时代,社交机器人主导的计算宣传的介入使得网络舆论生态更为复杂。对其治理也开始从传统的以政府为中心的方式转向多主体共存的全球治理模式。治理体系的构建需要开展国际化合作,也需要强化社交媒体平台的自律性措施,提升公民“素养”。机器人对舆论生态产生的危害值得我们警惕,但另一方面,我们也应摒弃简单的

“技术善恶论”,在充分认识社交机器人阴暗面的同时,也应挖掘其对于社会发展的积极作用,让技术向善,服务于人类社会的良性发展。

在治理体现建设方面,政府作为国家治理者,需要针对社交机器人制定完善的应用规范,出台相应政策或法律确保使用目的、程序开发、应用场景的正当性,从源头上防止社交机器人被恶意使用。社交媒体平台是社交机器人得以存在和生长的基础设施,同时也是掌握算法和用户数据最为全面责任主体。因此要发挥企业的社会责任,提升发现恶意社交机器人的识别能力,加强对异常行为账号的管理和处罚力度。在个体层面,要培养公众在多元信息背景下甄别有效信息的能力,避免那些导致认知失败的认知陷阱,提升对于虚假信息的免疫力,让民众免受计算宣传的危害。

叶翠微

浙江省政府特约研究员、海亮教育集团总校长

社交机器人泛滥,是AI时代的一个过程。任何一种新技术的出现,总有洪水猛兽的一面。作为教育工作者,应该在乱象众生中学会雾里看花。

我们要提升的是一种对现代信息的甄别能力,这个甄别能力既要用到机器人,更要超越机器人;我们要通过深度阅读,经典阅读,不断提升对这种泛滥信息的甄别能力;我们要有良好的社交圈子,因为良好的社交圈子或社交网络,能够给我们带来的更多正确的信息。

作为教育工作者,很重要的一点就是要培养学生什么是信心,什么是认知,什么是切合实力的认知,并通过像“世界咖啡馆”这样一种形式,来进行完善和提升。

名人名家热点说之三 | 群体性安全事故

不时传来群体性安全事故导致人员伤亡的消息。在群体性安全事故突发情况下,群众的情绪安抚中存在什么难点?目前的故事应急安全教育存在哪些不足?安全教育不足会对事故救援造成哪些不利影响?带着一系列问题,华东周刊连线有关专家学者,作了访谈。

■ 华东周刊 马晓才 / 文

叶翠微

海亮教育集团总校长

以下难点,特别应该引起我们重视:

一、群体情绪在发生后存在过度渲染、飞快传播、集体卷入的特点,所以在群众情绪安抚时,群体情绪极易被负面情绪裹挟,造成对社会支持力量的不信任,给安抚造成难度。

二、主要表现在对安全形势的预判和对不安全事件的处理方式以及身边支持系统的有效建立。其不利影响表现为群众对于安全形势的判断会直接影响他是否参与其中,从众心理是克服的重点。身边支持系统建立不完善,受害者无法第一时间发送求救信号,也是难点之一。

俞国娣

崇文教育集团总校长

学校安全教育现在都很重视,演习很多。但到真实场景中,学生只知道一个大概,因为真实性场景欠缺,体验感不够。尤其缺乏有时效性的、能有代入感的实际案例。

安全教育中能很好地用于教育的案例没有好好开发,往往是事故处理后,就算解决了。还有“不便公开”等原因,我们无从了

解事情真相和细节,家长和老师也无法做有说服力的分析。如上海动车事故、各种交通事故等,权威性信息不公开,很难进行有针对性的、有震撼力的教育。

杨雪锋

浙江财经大学地方政府与城乡治理研究院副院长

大型集会导致的群体性踩踏伤亡事件,历史上多次发生。它存在的一个难点,就是由于在狭小的空间人群的高度集聚,一旦出现有人倒地,在后续前行推力下导致更多人被挤压,在惯性作用下前面就会多人被踩踏,场面难以控制。在缺乏有限的信息传播和导流作用时,接近受害人的人群会想办法挤开人群,惊慌失措,慌不择路,那么后面那些跟进的人群会造成混乱,加剧拥挤和踩踏,而外围人群并不知道前面发生的情况,所以人流还会持续地涌入,最后导致灾难性事故。这种情况下,如果没有硬隔离和外围的强制性分流,短时间内是无法控制群体性慌乱局面的。

目前应急安全教育方面存在不少短板。一是要改变那种喜欢凑热闹的习惯。越是大型集会,越是要有这种安全防范的意识,越是人群聚集的地方,越是要远离;二是一旦发现险情,要及时报警。像群体性的踩踏事件,往往受害者比较多的是女性和小孩、老人这种弱势群体。因为他们的身高通常会低

于大多数男性,而且他们的力量小、身体瘦弱,反应速度也比较慢。所以在遇到这种推搡和拥挤的情况下,非常容易受到挤压。由于他们身高偏低,在挤压状态下,他们的胸腔由于过度挤压而导致呼吸困难,窒息而亡。所以,特别提醒这类人群更不宜参与无组织的大型聚会活动。

韩梦龙

杭州市萧山区委宣传部常务副部长

我认为,应该避免出现以下两种现象。一是蛋糕现象。当奶油蛋糕掉下来时,总是有奶油的一面先落地。就网络舆论而言,也体现出这种负面性倾向。网络舆论呈现批评的多、肯定的少,甚至“为批评而批评”。负面舆论场一旦形成,加上大多数网民选择沉默,就造成了批评性意见一边倒的态势。

二是雪球现象。当一个事件和议题产生后,舆论就会如同滚雪球般不断衍生、聚合、裂变、扩散,其传播的速度、波及的广度及影响的力度几乎呈几何级数增长。

一旦遇到群体伤亡突发事件,建议五步走:

第一步,要团结媒体,第一时间建立新闻记者接待中心。迅速建立突发事件处置工作班子,在成立的工作班子中,一定要有负责媒体沟通的领导,并成立相应的媒体接待小组,确立单位的新闻发言人。

第二步,要组织新闻发布,24小时内滚动发布事件最新动态。要依据“快讲事实,重讲态度,诚讲措施,慎讲结论”的原则,第一时间向社会发布新闻信息。考虑到直接面对媒体和网络对新闻发言人本人可能会有影响,建议原则上尽量使用网站、微博、微信等发布平台发布准备好的新闻通稿。

第三步,要全媒展现,互联网随时更新,广电滚动直播。

第四步,要上下协调,利益攸关者持续沟通、一致对外。要妥善处理负面舆情的现实问题,事件情况及时报上级有关部门备案。

第五步,要关注舆情,实时舆情监控,不断预测问题,周密制定口径。

另外很重要的一点,凡是较大规模的群体活动,地方政府务必提前做好应急预案,尤其是对人员的控制,场地实景推演,应急救援演练等工作,防患于未然。

邹菲

江西省庐山市人民政府副市长

我认为,在突发的事件中,一个果断坚决的指挥者尤为重要。犹疑不决,群众完全没办法信任,就谈不上服从,情绪更得不到安抚。此外,安全教育在学校里开展比较频繁和规范,但在其他层面的人群中缺乏宣传教育,或者流于形式,这一点非常值得我们重视。

叶峰

浙江广电知名评论员

群体性安全事故突发时,群众情绪不稳定的缘由来自于事发原因不明、事件现状不明、发展态势不明、处置进展不明……核心问题是信息不明。

从政府的角度来说,应该灵活地根据实际情况,以事先制定的预案为蓝本进行操作。

具体的难点,主要表现在以下几个方面:

一是预案时考虑不周全,在具体内容、类别和处置手段上有缺失。

二是受救援紧迫性的制约,无法或不能有效地抚慰事发地相关人员的情绪、无法或不能有效地掌控信息传播内容,进而出现与稳定社会情绪相悖的言论,影响公众心境。

三是现场处置的最高机构或决策者因受主观或客观因素的影响,无法或不能有效地组织权威媒体(包括有影响力的自媒体)迅速并连贯地向公众通报简要情况、及时并持续地向公众发布处置进展,继而出现蔓延性的群体恐慌。

此外,面向全民的应急安全教育严重缺失,绝大多数群众对突发的公共安全事件缺乏了解和科学的应急常识。就事发地而言,公众会盲目地“逃跑”“自救”,增加了现场的混乱,影响了救援的质量。