海外智库观点要览
北京大学国际战略研究院
2018年7月5日 第40期

假消息之战

克里斯·梅瑟罗尔、阿丽娜·波利亚科娃[1]

编者按：随着人工智能、去中心化应用程序等新技术的兴起，世界政治中的虚假消息变得更加难以检测和监管。本期摘译推荐美国布鲁金斯学会专家对这一问题的研究。她们认为，政府和企业只有尽快联手，把对人工智能等新技术的研究和反假消息的措施结合以来，才能有效应对技术进步带来的新威胁。

对于即将到来的人工智能引发的“深度假象”浪潮，美国和欧洲都没准备好。

俄罗斯的虚假消息已经成为欧洲政府面对的一个难题。在过去两年里，克林姆林宫已经传播了诸多假消息，如声称法国总统马克龙得到“男同性恋者游说组织”的支持，编造关于一名俄裔德国籍女孩被阿拉伯移民性侵的事件，散布一系列关于加泰罗尼亚独立公投的阴谋论等等。

欧洲最终采取了行动。今年一月，德国的《网络执行法》生效。这部法律旨在限制网络上的仇恨言论和虚假消息。法国和西班牙也考虑制定它们的反假消息法律。更重要的是，欧盟在四月公布了一项应对网络虚假消息的新策略。欧盟的计划着重于几种理性的反应：提升媒体素养，资助第三方事实核查服务，敦促脸

[1] 克里斯·梅瑟罗尔（Chris Meserole）是布鲁金斯学会中东政策中心的研究员，是研究新兴技术和网络极端主义的专家。阿丽娜·波利亚科娃（Alina Polyakova）是布鲁金斯学会美国和欧盟中心的大卫·M·鲁宾斯坦研究员，是研究俄罗斯政治战和新兴威胁的专家。本文英文原文发表在外交政策（Foreign Policy）网站：http://foreignpolicy.com/2018/05/25/disinformation-wars/#。此为中文摘译版。

书(Facebook)和其他社交网络公司采取措施突出由可信媒体发布的新闻。尽管这项计划还缺乏监管措施，但欧盟官员已暗示监管可能即将到来。脸书首席执行官扎克伯格在出席欧盟听证会时，试图回避关于假新闻和极端主义内容的问题，但立法者提醒扎克伯格，他们具有监管权力。

最近欧洲采取的行动是重要的第一步。但这些已经公布的法律或策略是不够的。问题在于，技术的进步远远快于政府制定政策。欧盟的措施只是针对昨日而非明日的假消息。为了解决这个问题，欧洲和美国的决策者应关注即将到来的具有颠覆性的技术浪潮。由人工智能和去中心化计算所推动的下一代虚假信息势必会更加复杂，更难以检测。

立法者应关注四种新兴威胁：人工智能的民主化，社交网络的演变，去中心化应用程序的兴起，以及假消息的“后端”。

由于更大的数据、更好的算法和定制硬件的出现，在未来几年，世界各地的人们将可以逐渐使用尖端的人工智能。从医疗到交通，人工智能的民主化具有巨大的前景。

然而，与任何双重用途的技术一样，人工智能的扩散也带来重大的风险。**人工智能使人人都能制作虚假的印刷品、音频和视频故事。**虽然计算机一直允许对数字内容进行操控，但在过去这种操控几乎总是可以被检测到的：假图像无法说明光的微小变化，被篡改的语音无法完全掌控节奏和音调。但是，深度学习和生成式对抗网络可以很好地篡改图像和视频，以至于难以区分伪造的文件和真实的文件。此外，由于有像 FakeApp 换脸软件和 Lyrebird 变声软件等应用程序，这种所谓的“深度假象”可以被任何人用电脑或手机制作出来。今年早些时候，一种能让用户轻松换掉视频中的人脸并伪造名人色情片的工具在推特(Twitter)和色情影片网站 Pornhub 上疯狂传播。

对政府和社会来说，有效应对深度假象和虚假信息的民主化是有难度的。由于生成假象的算法不断获得更有效地复制现实的能力，深度假象便不能轻易地被其他算法检测到——在生成式对抗网络中，算法通过变得真正善于欺骗它自己而起作用。因此，为了应对虚假信息的民主化，政府、社会和技术部门不能只依靠算法，而是需要投资社交验证的新模型。

在人工智能和其他新兴技术成熟的同时，传统平台将继续在制作和传播网络信息上发挥重要大作用，例如假消息在谷歌、脸书和推特上的扩散。

越来越多搜索引擎优化(SEO)操控的山寨行业为希望上升谷歌排名的客户提供服务。虽然在大多数情况下，谷歌能通过不断调整领先于那些操控算法，但是**搜索引擎优化操控者正变得更精明，能让包括假消息在内的内容出现在搜索结果的顶部**。

举个例子，在三月份英国神经毒剂袭击事件和四月份叙利亚化学武器袭击事件发生之后，今日俄罗斯(RT)和俄罗斯卫星通讯社(Sputnik)——俄罗斯政府的宣传媒体——的报道出现在谷歌搜索的首页。类似的，YouTube（谷歌旗下的视频网站）有一种算法，能优先考虑用户观看内容的时间量，将其作为确定哪些内容首先出现在搜索结果中的关键指标。这种算法的偏好性导致虚假的、极端主义的和不可靠的信息出现在搜索结果的顶端，这反过来意味着这些内容更频繁地被用户查看，被认为更可靠。搜索引擎优化操控行业的收入预计达到数十亿美元。

在脸书上，假消息通过共享内容和付费广告两种方式出现。脸书公司已试图减少各种来源的假消息，但到目前为止都徒劳无功。最著名的例子是，脸书引入了一种“有争议的标志”，用以表示可能的假消息——结果是这一标志没有减少反而增加了用户与这类内容互动的可能性。还有个不那么出名的例子，在加拿大，脸书公司尝试让包括针对一小部分用户的微型广告在内的所有广告可供审查，以增加付费广告的透明度。然而，这种努力的效果是有限的：广告赞助商经常是隐藏的，需要用户去做耗时的研究，而且脸书为广告设置的存档不是一个永久的数据库，只是显示活跃的广告。**脸书早期的努力没有很好地预见到：外来行为者能够继续利用脸书的动态消息和广告产品传递假消息——包括制作针对特定个人或群体的深度假象。**

虽然推特已经采取措施抗击其平台上的机器人魔怪和软件机器人的泛滥，但它面对假消息仍非常脆弱，因为该平台上的账号未经过验证，它的应用程序编程接口(API)仍然可以让错误内容在平台上轻易地生成和传播。**即便推特采取进一步措施打击信息滥用行为，它的检测算法也可以像谷歌的搜索算法那样被反向操作。**如果不对其应用程序编程接口和交互设计进行根本性的改变，推特仍会充斥着假消息。

举个例子，当美军在四月份袭击叙利亚化学武器设施时——正是在推特最新改革措施实施之后——五角大楼报告称袭击发生后几小时内俄罗斯假消息大量增加。这些推文似乎发自合法账号，而且没有办法将它们报告为错误信息。

区块链技术和其他分布式账本技术以支持比特币(bitcoin)和以太坊(ethereum)等加密货币而著称。但它们最大的影响可能在于改变了互联网的运作方式。随着越来越多的去中心化应用程序上线，网络将逐渐受到抵制脸书和其他网络公司所享有的集中控制的那些服务和协议的支持。例如，用户可以在 DTube 而不是 YouTube 上浏览视频，在 Blockstack 而不是 Safari 上浏览网页，使用 IPFS，即点对点文件系统，而不是 Dropbox 或 Google Docs 来存储文件。当然，去中心化应用程序生态系统仍处于萌生阶段，需要时间来成熟和解决故障。但是，随着时间的推移，安全性不断提高和底层网络体系结构不断被修复，分布式账本技术有望让网络变得更加安全，并脱离主要公司和国家的控制。

如果线上活动转移到去中心化应用程序上，它们所提供的安全性和去中心化对于隐私倡导者和人权异议人士来说是福音。但这对于恶意行动者来说也是天赐之物。这些服务大部分都由匿名性和公钥密码体制所支撑，因此难以追踪到现实中的个人或组织。此外，一旦信息被提交给去中心化应用程序，就几乎不可能取消。例如，IPFS 没有删除的方法——用户只能添加而不能删除内容。

对于政府、社会和个人而言，去中心化应用程序将带来前所未有的挑战，因为当前用以应对和干扰假消息活动的方法将不再适用。政府和社会最终可以呼吁推特首席执行官杰克·多西阻止或删除在推特上的恶意用户或有问题的内容，但对于去中心化应用程序，它们不能求助于任何人。如果曼彻斯特爆炸袭击者是在去中心化应用程序而不是在 YouTube 上看到制造爆炸的指导视频，那么谁应该或可以采取措施阻止这类内容仍不清晰。

在过去三年，对俄罗斯假消息的重新关注引发了越来越多非营利组织、政府、记者和活动家的研究和行动。目前，这些努力集中在记录假消息活动中涉及的机制和参与者——追踪僵尸网络，识别巨魔账号，监控媒体叙述以及追踪假消息内容的传播，还包括政府为落实数据保护和隐私政策所做的努力，例如，欧盟《通用数据保护条例》以及为网上广告空间引入更多透明度和问责制的立法提案。

虽然这些努力对提高公众和决策者的认识确实是有价值的，但它们关注最终

产品（内容），几乎没有深入研究驱动假消息活动的潜藏基础设施和广告市场。**需要对假消息的“后端”进行更深入的检查和评估。算法和产业——网上的广告市场、搜索引擎优化操控市场和数据经纪人——隐藏在产品背后。**与机器学习配合的自动化程度的提高也将改变这一空间。

为了应对这些新出现的威胁，欧洲和美国应考虑以下几种政策反应。

首先，欧盟和美国应当投入大量资金，用于人工智能和信息战的交叉点的研究和开发。今年四月，欧盟委员会呼吁到 2020 年将至少拨出 200 亿欧元用于研究人工智能，优先关注健康、农业和交通领域。这些资金都没有被专门用于研究假消息。同时，当前欧洲反假消息的措施优先在教育和事实核查上，忽略了人工智能和其他新技术。

只要技术研究和反假消息的努力仍各自为战，在应对新兴威胁方面就难有进展。在美国，政府一直不愿参与推动技术研究，硅谷的创新很少受到监管。2016 年奥巴马政府关于人工智能未来的报告并未分配资金，特朗普政府尚未发布它的战略。随着俄罗斯操控数字平台的行为持续被揭露，政府显然需要与私营部门共同识别脆弱性和国家安全威胁。

此外，欧盟和美国政府也应当快速行动，阻止去中心化应用程序上错误信息的增加。去中心化应用程序的出现为决策者提供难得的第二次机会：当十年前社交网络刚被建立起来的时候，立法者没能预见到恶意行为者可以利用社交网络。现在去中心化应用程序还是一个萌发的市场，决策者可以在它达到全球规模之前做出反应。政府应当构建新的公私伙伴关系，帮助开发者确保下一代网络不被假消息活动利用。联合国的反恐技术项目可以作为一个典范，该项目与小型科技公司紧密合作，帮助他们从头开始设计自己的平台，用以防范恐怖主义利用。

最后，立法者应当继续推动数字广告业的改革。随着人工智能继续改变着这个行业，假消息内容将更加精准地针对特定受众。人工智能将让恶意行为者和合法广告商都能更容易地追踪用户的线上行为，识别并瞄准潜在的新用户，收集用户态度、观念、偏好等信息。2014 年，美国联邦贸易委员会发布一份报告，呼吁数据经纪行业的透明度和责任心。这份报告呼吁国会考虑立法，通过向个人提供途径和信息，让他们知道自己的数据是如何在线上被收集和利用的，从而使这

些公司的活动透明化。欧盟的保护法规在帮助用户掌控其数据方面发挥了重要作用，并限制了社交媒体平台为实现广告定位而处理用户数据的方式。脸书也在争议投票前试图阻止外国广告销售。但是，数字广告行业作为一个整体仍是一个黑匣子，在限制技术挖掘和管理网上政治广告方面需要做的还很多。

有效追踪和定位上述各个领域并不容易。然而，决策者需要从现在开始关注它们。如果欧盟反假消息的新努力和其他相关政策不能跟上不断发展的技术，那么这些政策在被引入之前就可能过时了。

（黄琳摘译，归泳涛校）