COMMUNICATION RIGHTS FOR SOCIAL BOTS?

Options for the Governance of Automated Computer-Generated Online Identities

Stefano Pedrazzi and Franziska Oehmer

ABSTRACT

Social bots, automated agents operating in social networks, are suspected of influencing online debates, opinion-formation processes and thus, the outcome of elections and votes. They do so by contributing to the dissemination of illegal content and disinformation and by jeopardizing an accurate perception of the relevance and popularity of persons, topics, or positions, through their potentially unlimited communication and networking activities, all under the false pretense of human identity. This paper identifies and discusses preventive and repressive governance options for dealing with social bots on state, organizational, and individual levels respecting the constitutional provisions on free expression and opinion-formation.

Keywords: Social bots, governance, social media, freedom of expression, platform regulation

The influence of social bots on the results of elections and votes as well as on debates about issues such as vaccinations or migration is attracting more and more public and political attention. It is argued that social bots, automated agents operating in social networks, may contribute to the spread of defamations, disinformation, and conspiracy theories and distort opinions by feigning false urgency and popularity of issues and persons.[1]

1. Yücel.

Stefano Pedrazzi: Department of Communication Science and Media Research, University of Fribourg
Franziska Oehmer: Research Center for the Public Sphere and Society, University of Zurich
DOI: 10.5325/jinfopoli.10.2020.0549

There are also concerns that, with social bots becoming more similar to human users in appearance and function in the future, they will be even more difficult to identify, and their manipulative potential will continue to increase.

The scientific debate about social bots also demonstrates largely critical findings. Signs of interference in political discourses, opinion-formation, and decision-making through the use of social bots of both domestic and foreign origin were found in several cases, including in elections in the United States,[2] Japan,[3] Germany,[4] and France;[5] in referendums in Great Britain,[6] Switzerland,[7] and Catalonia;[8] in conflict situations such as in Ukraine,[9] Syria,[10] and Mexico;[11] or in connection with controversial issues such as vaccinations.[12] A study by Varol et al. estimated that 9% to 15% of all profiles on Twitter were bots.[13] Values of up to 20% were calculated for the 2016 US election campaign.[14] Simulations also showed that within a highly polarized setting, even a small number of social bots can be sufficient to tip the opinion climate over.[15]

When systematizing the challenges in connection with social bots discussed in public, political, and scientific debates, three specific problem areas can be identified[16]:

1. The dissemination of illegal content and disinformation
2. Under the false pretense of human identity
3. By means of a potentially unlimited number of communication and networking activities

The causes of these problems lie not only with the initiators and programmers of social bots, they can also be viewed as the consequence of

2. Bessi and Ferrara.
3. Schäfer, Evert, and Heinrich.
4. Neudert, Kollanyi, and Howard; Keller and Klinger.
5. Ferrara.
6. Howard and Kollanyi.
7. Rauchfleisch and Vogler.
8. Stella, Ferrara, and De Domenico.
9. Hegelich and Janetzko.
10. Abokhodair, Yoo, and McDonald.
11. Suárez-Serrato et al.
12. Broniatowski et al.
13. Varol et al.
14. Bessi and Ferrara.
15. Ross et al.
16. Ferrara et al.; Woolley; Steinbach; Libertus.

and gaps in the allocation of responsibility for organizational or individual actions of various actors with asymmetrical resources and possibilities to implement measures and thus as a "problem of many hands."[17] These actors include legislators who consciously or unconsciously fail to adopt standards against malicious bot activities, operators of social networks (including their management, programmers, etc.) whose network infrastructure and terms of use allow or even promote malignant bot activities, and social network users who intentionally or negligently disseminate or endorse such activities.

Given that the responsibility for collective action cannot be assigned to a single actor, the networked nature of the adopted measures and that the areas in which social bots as well as their initiators operate are not confined by territorial or national borders,[18] the solutions for these problem areas can only be worked out by taking into account different levels of regulation and the interactions among the aforementioned actors. This article, then identifies and discusses from a governance perspective[19] preventive and repressive state, organizational, and individual measures against (potentially) harmful activities of social bots that have already been implemented or are conceivable. Due to a (so far) missing intergovernmental regulation of platforms in general and social bots specifically, the discussion on possible state interventions refers to national legislation. In particular, the article discusses the limits and possibilities of the Swiss Federal Constitution (Cst.), especially the provisions on free (political) expression and opinion-formation. Since these provisions also have constitutional status in many other democratic states, the following explanations are not limited to Switzerland, but can also be considered as *pars pro toto*, even though the legal basis for bot regulation is of special significance in Switzerland because of its numerous possibilities for direct democratic participation including popular votes several times a year. In addition, however, national regulations of other states or associations of states (e.g., United States, Germany, and European Union [EU]) are also addressed if they have taken or are discussing a measure explicitly targeted at social bots.

To this end, we will next briefly outline the key features of social bots. The potential impact of social bots with regard to opinion-forming

17. Thompson; Nissenbaum; van de Poel et al.
18. Woolley.
19. Rhodes; Puppis.

processes in social networks and the necessity of regulatory measures is discussed in the section Mechanisms of Social Bots. In the subsequent section we will then discuss options for the governance of social bots, taking into account state, organizational, and individual measures and actors.

Definition and Key Features of Social Bots

Social bots fall into the broader category of bots, which are generally defined as autonomous, reactive software agents capable of perceiving signals and performing actions in a specific online environment for the purpose of pursuing an agenda[20]—that is usually specified by a programmer. Social bots can be distinguished from other bots by the nature and intended effect of the activities they perform. These include *web robots* that collect information through crawling and scraping, *chatbots* that are used as human-machine communication systems mostly in commercial settings, or *spambots* that are deployed to direct users to compromised websites.[21] The literature further differentiates between social bots and hybrid forms such as *trolls* or *cyborgs*, which combine automation with human profile behavior and can pursue similar goals as social bots and are sometimes used in combination with them.[22]

Social bots are computer algorithms that automatically produce and distribute content in social networks or online forums and interact with human users, as well as other bots, thereby imitating human identity or behavior, in order to (possibly) influence opinions or behavior.[23] Social bots operate on behalf of individual or collective actors, for example, for political or commercial motives. They are often associated and investigated in connection with the deception of users, the spread of rumors, defamations, or disinformation, but they can also be employed for nonmalicious purposes when, for example, they automatically aggregate and disseminate content from different sources.[24]

Depending on how social bots are programmed, they can perform different types of operations to pursue their goals, ranging from liking

20. Franklin and Graesser; Tsvetkova et al.; Howard, Woolley, and Calo.

21. Gorwa and Guilbeault; Oentaryo et al.; Stieglitz et al., "Do Social Bots Dream of Electric Sheep?"

22. Gorwa and Guilbeault.

23. Ferrara et al.; Stieglitz et al., "Do Social Bots Dream of Electric Sheep?"

24. Ferrara et al.

accounts to identically distributing specific content to learning new behavior patterns. Despite advances in machine learning and artificial intelligence (AI) social bots cannot be considered "full ethical agents"[25] at present as they cannot (yet) be attributed consciousness, free will, and intentionality.[26] With the delegation of human agency within a predefined or expanding scope of action, social bots gain a limited autonomy that can affect the construction of reality and the behavior of users of social networks, leaving open questions about the responsibility and liability for actions performed by social bots.[27]

Due to the false pretense of human identity and imitation of human behavior, the identification of social bots poses a complex challenge. In addition, the existence of hybrid forms (e.g., trolls) as well as the lack of a consensus on from which degree of automation an account is to be considered a bot further complicate an unambiguous classification.[28] For this purpose, numerous scientists have developed classifications and recognition programs,[29] including the Botometer[30] that is frequently used in academic research. Using a variety of profile, network, and behavioral data and based on machine learning, it computes a probability value for any Twitter profile that indicates whether a particular user is a bot.[31]

However, doubts have been raised as to the validity of the results of studies that have used Botometer to detect social bots, which in many cases contain a not insignificant number of false positives (i.e., profiles identified as bots that are not actually bots) or false negatives (i.e., bots identified as human profiles).[32] Botometer is based on a machine learning process that uses a variety of characteristics of profiles from a training data set to identify patterns that distinguish bots from human users. Based on this, Botometer computes a probability that a profile is a fully automated account. This would include, for example, the account of a newsroom, which automatically distributes all published articles via Twitter.

25. Moor.

26. Moor; Mitcham.

27. Just and Latzer; Guilbeault; Klinger and Svensson.

28. Hegelich and Thieltges.

29. Subrahmanian et al.; Chu et al.; Ferrara et al.; Oentaryo et al.; Stieglitz et al., "Do Social Bots (Still)."

30. Davis et al.; Varol et al..

31. Davis et al.; Ferrara et al.

32. Rauchfleisch and Kaiser.

It is therefore not recommended to choose a random threshold value, but to determine and justify it depending on the data set used.[33]

Mechanisms of Social Bots

Social bots unfold their effect on the basis of their own features and capabilities and in interaction with characteristics of the network architecture and the algorithmic selection logic of platforms as well as the individual receptivity of social network users, which are briefly explained in the following.

Automation

Automation is one of the constituent features of social bots and allows their activities to be restricted only by physical limits to the transmission and computational processing of signals, and by the design and architecture of an online environment. For instance, a less sophisticated social bot automatically retweets any post that contains a specific hash tag. In doing so, it potentially expands the distribution of this particular contribution by its own network of followers. More complex social bots, on the other hand, can draw on a versatile repertoire of options for action and adaptation in order to create publics in social networks and influence the flow of information or the perception of people or topics. They can create and post original content, imitate time patterns of use, copy profile information and usernames from other users with slight changes, and seek interaction with other users.[34] Furthermore, social bots can, for example, specifically misuse "benign" bots in order to promote the spread of false messages, whereby these unintentionally enter into the service of the initiators of the "malignant" social bots. As a result, an unambiguous and temporally persistent categorization into "good" or "bad" bots is difficult, if not impossible, whereby the normative subdivision of this categorization is per se highly complex, especially since the detection and proof of deception, manipulation, or disinformation are not carried out value-free. Further developments in the field of AI suggest that the predictability of

33. See Botometer FAQ, available at: https://botometer.iuni.iu.edu/#!/faq (accessed June 21, 2020).

34. Aiello et al.; Ferrara et al.; Yang et al.

actions[35] as well as the recognition of social bots by humans or algorithmic systems will become even more difficult.[36]

Distortion of Popularity Indicators

By feigning human identity and due to the potential of unlimited communication and networking activities through automation, social bots can distort popularity indicators such as number of followers, likes or retweets of people, topics, or positions.[37] For example, on Twitter, they can make use of its asymmetrical network structure, which allows one profile to follow another without its explicit consent. At the individual user level, popularity indicators can serve as cue for the relevance or the credibility of a profile or (the source) of content and thus affect its processing.[38] According to the Elaboration Likelihood Model,[39] source characteristics can influence the extent of cognitive engagement with and the persuasiveness of content.[40] Further, popularity indicators can stimulate the selective exposure to and the use of content[41] as well as activities such as liking or sharing,[42] which can ultimately lead to a self-reinforcing effect.[43] In that popular profiles, content and positions can suggest reflecting the majority's position, they can also promote the effect of the "spiral of silence,"[44] thus discouraging or preventing supporters of the supposed minority opinion from publicly expressing their opinion for fear of social isolation.[45] This tendency has been observed among the users of social media affecting their willingness to express their opinion in both online and offline settings.[46] Simulations have shown that in a highly polarized setting even a small number of social bots can succeed in tipping the opinion climate over and, as postulated by the theory of the spiral of silence, suggesting supposed majorities that

35. Just and Latzer.

36. Cresci et al.

37. Porten-Cheé et al.; Ross et al.

38. Porten-Cheé et al.

39. Petty and Cacioppo, "Source Factors and the Elaboration"; Petty and Cacioppo, "The Elaboration Likelihood Model."

40. DeBono and Harnish.

41. Messing and Westwood; Knobloch-Westerwick et al.; Yang.

42. Shao et al.; Haim; Bobkowski.

43. Porten-Cheé et al.

44. Noelle-Neumann.

45. Porten-Cheé et al.; Keller and Klinger.

46. Hampton et al.

do not exist in reality.[47] At the technological level, increased metrics also affect the algorithmic selection logic of social networks: Contents or profiles with many search queries or high or rapidly growing popularity values are more likely to be recommended and displayed to users, which again can affect their reception and behavior.[48]

Reach Extension

Social bots can not only distort popularity indicators and thus affect how people, positions, and content are perceived, their activities can also lead to an increase in their reach within and possibly also outside social networks, for example, through interpersonal offline communication activities of users.[49] In this context, different practices and strategies have been identified that make use of diffusion factors such as time and the centrality of nodes in a network. Shao et al., for example, were able to demonstrate through an analysis of Twitter data that the diffusion of false messages from low credibility sources was facilitated by the retweet activity of several social bots within seconds after publication (and thus before any review by fact-checking organizations, media organizations, or other users).[50] Considering that the perception of the correctness of implausible news or even of news explicitly marked as false grows with increasing exposure frequency,[51] this strategy appears all the more successful. Especially since there is empirical evidence that human users are as likely to share contents provided by social bots as by human users[52] and that complex contagion processes (i.e., the propagation of information through activities such as retweets, shares, or likes by several users), promote the diffusion of messages.[53] In addition, this distribution mechanism is supported by social bots targeting influential users (with many followers), for example, by using the response and mention functions, which at the same time provide a link to a false message.[54]

47. Ross et al.; Cheng, Luo, and Yu.

48. Papakyriakopoulos, Serrano, and Hegelich; Just and Latzer; Lazer et al.; van Dijck, Poell, and de Waal; Yang et al.

49. Shao et al.; Vosoughi, Roy, and Aral.

50. Shao et al.

51. Pennycook, Cannon, and Rand.

52. Vosoughi, Roy, and Aral; Varol et al.

53. Mønsted et al.

54. Shao et al.

Previous research suggests that the production of disinformation and the targeted use of social bots to promote its dissemination, especially during the last days and hours before an election or vote, are arranged and coordinated by the same initiators.[55] By deceiving and targeting human users and supported by recommendation algorithms, they can achieve a faster and more effective diffusion of content. The reasons why people interact with and also share content from social bots are not yet well researched, that is, whether they do so because they do not recognize social bots or whether they regard their nonhuman nature as irrelevant. Studies in the tradition of the "Computers Are Social Actors" approach,[56] which deal with whether people unconsciously apply the same social rules and heuristics in dealing with computers, or in this case social bots, as in dealing with people, can provide an indication of this: Edwards et al. showed in an experiment that bots are perceived as credible, attractive, and competent—qualities normally associated with people.[57] The participants also expressed an equally strong intention to interact with a bot or a human user on Twitter.[58] In addition, it has been demonstrated that users show a similar level of effort and motivation in information search and cognitive processing when they receive information from a bot or a human user.[59] People use the same processing strategies when interacting with social bots as they do when interacting with people, they retweet or like their posts and attribute credibility to them.

The Governance of Social Bots

As has been argued, social bots have the potential to influence the democratic opinion-forming and decision-making processes to an undesirable extent. The central challenges identified are: (1) the dissemination of false and/or unlawful content, (2) under the false pretense of human identity, and (3) by means of automated and thus principally unrestricted activity.

Their potential to do so cannot only be attributed to the initiators and programmers of social bots. It can also be viewed as the consequence of organizational or individual decisions and actions of various other actors—including legislators, operators of social networks (including management,

55. Shao et al.; Ferrara; Stella, Ferrara, and De Domenico; Bastos and Mercea.
56. Reeves and Nass.
57. Edwards et al., "Is That a Bot Running the Social Media Feed?"
58. Ibid.
59. Edwards et al., "Differences in Perceptions of Communication Quality."

programmers, etc.) and their users—as well as gaps in the distribution of responsibility among these actors.[60] This phenomenon, referred to as the "problem of many hands,"[61] can be illustrated by the example of social bots increasing the reach of disinformation: If a false message spread by a social bot is presented to a human or computer-generated user, this is the result of a missing or unsuccessful examination of the content for correctness as well as an algorithmic recommendation by the platform, which itself is based on characteristics of the message as well as the user's own and other users' behavior. If a user likes, comments, or shares this message—regardless of any knowledge about the message being false—he or she will firstly increase its reach and popularity indicators, secondly influence the perception of the message by other users, and thirdly change the basis for algorithm-based operations such as recommendations for other users. The platforms primarily design and use recommendation algorithms to pursue economically motivated goals such as generating user data and increasing the number of interactions[62] rather than to promote the dissemination of verified content. This in turn can also be understood as the consequence of a US-influenced values and legal system that, with *47 U.S.C. § 230*, exempts platforms from any liability claims and attributions of responsibility for the distribution of third-party content, and on which platforms based in the United States mainly orientated their business models and practices until legal disputes with state authorities of other western countries arose.[63]

Since the responsibility for collective action cannot be assigned to a single actor[64] and the areas in which social bots and their initiators operate are not confined by territorial or national borders,[65] and given the networked nature of the measures adopted by the different actors, solutions for these problems can only be developed by taking into account different levels of regulation and the interactions among different actors such as legislators, platform operators, and users.[66] In this context, particular attention needs to be paid to the asymmetrical resources, coordination possibilities,

60. Thompson; Nissenbaum; van de Poel et al.

61. Thompson.

62. van Dijck, Poell, and de Waal.

63. Jin; van Dijck, Poell, and de Waal; Gillespie, "Platforms are not Intermediaries"; Pasquale, "Platform Neutrality."

64. Nissenbaum; van de Poel et al.

65. Woolley.

66. Napoli; Helberger, Pierson, and Poell; van Dijck, Poell, and de Waal.

and enforcement capacities the different actors have. The options include forms of self-help by users, internal or collective self-organization by companies and sectors, co-regulation, and state interventions.[67]

Governance options for the three problem areas outlined are presented and discussed below from a *STATE*, *ORGANIZATIONAL*, and *USER* perspective. Potential measures can be classified according to the time of the intervention into preventive and repressive as well as already implemented and conceivable. From a public interest perspective, governance should serve to strengthen benefits and minimize risks.[68] In the context of the activities and the influence of social bots, this specifically means that governance options should be used to guarantee and promote fundamental rights such as freedom of expression and unbiased decision-making, and to curb the potential for censorship and repression as well as for manipulation of elections and socially relevant discourses.

For the discussion of state options in dealing with social bots, reference is made to Swiss constitutional law, in particular to the provisions on freedom of expression and formation of opinion (*Articles 16 and 34 Cst.*), which are consistent with other nation-state provisions in western democracies (e.g., *the First Amendment to the Constitution of the United States of America, Article 5 of the German Basic Law, and Article 10 of the European Convention on Human Rights*), and can therefore claim generalizability for these provisions.

Governance Options Addressing Illegal Content

In Switzerland, illegal content includes, among other things, statements that make false harmful factual claims, contain discriminatory, violence-glorifying or pornographic content or infringe the personal rights of third parties. *STATE:* Governance options for this problem area must consider the right to freedom of expression established in *Article 16 Cst. paragraph 2* and the conditions for the restriction of fundamental rights formulated in *Article 36 Cst.*: The fundamental right to freedom of expression, which is concretized in *Article 16 Cst. paragraph 2*, serves to protect the needs and interests of people, which are of fundamental importance for their personality.[69] The article serves two primary functions: First,

67. Saurwein, Just, and Latzer; Puppis.
68. Black; Saurwein, Just, and Latzer.
69. Rhinow and Schefer.

it should ensure the free exchange of views and perspectives, which is relevant for democratic processes. Second, it enables citizens to express criticism of (perceived) political and social grievances.[70] In particular, it also establishes a right of defense against interventions by the state or third parties, especially in connection with censorship or sanctions. A restriction of the fundamental right to freedom of opinion and expression, according to *Article 36 Cst.*, requires a legal basis, is subject to the principle of proportionality and is only possible if it is in the public interest or justified by the protection of fundamental rights of third parties.[71] The core essence of the fundamental right must not be subject to any restriction.[72] This includes, for example, a prohibition of pre-censorship in the sense of a systematic control of intended expressions of opinion.[73] A restriction of the freedom of expression can be justified if the civil and criminal legal protection of the personality of private third parties is affected and in cases of discrimination, defamation, or the deliberate dissemination of false facts.[74] If contents posted by a social bot infringe the fundamental rights of third parties or applicable law, the freedom of expression can be restricted. The same repressive state defense mechanisms, for example, regarding liability and injunction, are applied as for other communication channels. Often, however, it is not always possible to distinguish between content that is actually illegal or merely undesirable. Moreover, there are different international regulations. For example, denying the Holocaust is not considered illegitimate in all countries. It should be noted that in the United States, false allegations are generally protected by the First Amendment.[75]

Since social bots do not (yet) have their own legal personality, the programmer or the initiator is regarded as the owner of the legal interest against which an intervention can be asserted and as the addressee of sanction claims,[76] whereby the possibilities for sanctioning also depend in part on international treaties and cooperation with foreign authorities in the prosecution and combating of criminal activities. In the course of technological development in the field of robot technology and AI, an

70. Kley and Tophinke.

71. Rhinow and Schefer.

72. Kley and Tophinke.

73. Ibid.

74. Biaggini; Schweizerisches Bundesgericht, "Bundesgerichtsurteil 6B_119/2017"; Sprecher et al.; Schweizerisches Bundesgericht, "Bundesgerichtsurteil 6B_267/2018"; Rhinow and Schefer.

75. Wood.

76. Oehmer.

increasingly autonomous behavior of social bots is assumed. In this context, it is questionable whether illegal content distributed by autonomous social bots can still be legally charged to the programmer in the future. In this context, e-personhood is discussed as a new legal entity.[77] However, there are no concrete implementations so far.

ORGANIZATIONS: The responsibility of platforms in the distribution of illegal or deliberately false content must also be reflected: Due to their network architecture as well as the algorithmic selection, recommendation and curation of content, platforms not only act as transmitters, but also allow social bots to exploit these algorithmic mechanisms in order to pursue their goals. Accordingly, from a governance perspective, the coresponsibility of platforms for the distribution of content by social bots must also be taken into account. At the level of self-regulation, the operators of platforms such as Twitter or Facebook point out in their usage guidelines[78] that the publication and distribution of such content via their networks is prohibited. In order to prevent this, certain platforms—for example, Facebook in connection with terrorist, violence-glorifying, and pornographic content[79]—check whether content is illegal or violates their guidelines during the upload process by means of algorithmic filters and human editors. Preventive content moderation is, however, not entirely unproblematic from a constitutional point of view, especially since algorithms are used for such tasks that are based on human programming and machine learning and therefore operate neither value-neutral nor objectively.[80] This bears a potential risk of systematic bias due to incorrect results,[81] particularly if the law also provides for sanctions against platform operators for the distribution of such content.

Repressively, the platforms can remove illegal content and links to such (or even entire profiles), either by order of a court or through a reporting procedure in which users partly assume the work of content moderation. In the latter case, too, the platform operators must ultimately decide whether or not a content infringes the guidelines. Depending on how

77. Fanti.

78. See https://help.twitter.com/en/rules-and-policies/twitter-rules; https://www.facebook.com/legal/terms (accessed June 21, 2020).

79. See https://newsroom.fb.com/news/2019/05/enforcing-our-community-standards-3/ (accessed June 21, 2020).

80. Gillespie, "The Relevance of Algorithms"; Gillespie, "Governance of and by Platforms."

81. The same risk applies to human content moderators. In addition, they are confronted with highly problematic working conditions. See Gillespie, "Platforms are not Intermediaries."

the reporting procedure is designed, it also opens up room for manipulation by social bots themselves: If, for example, a user account is blocked by reporting an (alleged) violation, bots can be programmed to identify, report, and thus block (influential) accounts that support opposing positions.[82] Whereas the detection of illegal content already poses a substantial challenge to platform operators, the challenge is even greater in connection with the dissemination of intentionally false facts. While in practice it is often nearly impossible to draw a clear line between false and true content, it is even more difficult to make a judgment about the intention of an author of a piece of content, all the more if it originates from a social bot, to which no awareness or ability of intentionality can currently (yet) be attributed.[83] Preventive as well as repressive content moderation on the part of the platform operators thus harbor a potential risk for the freedom of expression and the free formation of opinion in the form of misjudgments and systematic precensorship. In addition, they also pose a risk to the commercial success of platforms by requiring the use of considerable resources and potentially frustrating users whose content is deleted or accounts (temporarily) are blocked.[84]

In response to the dissemination of deliberately false content, measures can be implemented that help to promote the quality of information in a social network in general[85] and thus indirectly address the harmful potential. These include, for example, visual cues that users can use to orient themselves with regard to the quality and credibility of content, sources and authors or an adaptation of recommendation algorithms in favor of high-quality content.[86] Again, it should be noted that such adaptations can be a source of possible bias. Research on the use of warnings in connection with mis- and disinformation have shown that they can reduce the perceived credibility of such content and the probability of sharing it, but unintended effects have also been observed.[87] Particularly problematic is the fact that the existence of warning labels can have a detrimental effect on the credibility attribution of other (true) content as well as question the perceived correctness of existing beliefs.[88] Furthermore, false content

82. Gollmer.
83. Mitcham.
84. West.
85. Napoli.
86. Clayton et al.; Pasquale, "The Automated Public Sphere."
87. Chan et al.; Mena; Walter and Murphy; Walter and Tukachinsky.
88. Carey et al.; Clayton et al.

that is not accompanied by a warning is implicitly validated.[89] Given these difficulties, it is important from a democratic perspective that governance be based on a common understanding among key stakeholders, what is seen as a real threat to third-party rights and public security.[90] In order to meet this objective, it would be conceivable at the co- and self-regulatory level to participate in, support, and finance expert panels as well as bot detection and fact-checking agencies jointly operated by citizens, academics, journalists, platform operators, and state actors.[91]

As far as *USERS* are concerned, their behavior also contributes to how content spreads in social networks, for example, by liking, sharing, or reporting content as offensive if a social network has a corresponding feedback function, or by refraining from doing so.[92] It is however important to highlight that the options available to users are largely determined by the platforms. The reports of particularly credible users, so-called "Trusted Flaggers,"[93] are given privileged treatment by some platforms and the content that has been flagged is removed more quickly. For the average user, however, it is an extremely demanding task to judge whether content is illegal, deliberately misleading, or comes from a reputable source. Such tasks require a certain level of competence in the critical use of sources and content as well as in the functioning of social networks, especially given that activities such as liking and sharing can be classified as independent expressions of opinion and are also subject to possible claims for sanctions.[94]

In contrast to state actors or platform operators, users do not have an organization and resources that would allow them to act in a goal-oriented and strategic way. Correspondingly, platforms have a major responsibility for and play a particularly important role in how users deal with illegal or deliberately false information disseminated by social bots: in the design of the terms of use, in the transparency of their operations, by equipping users with true and effective options to organize and enforce their interests vis-à-vis the platforms and thus be able to exercise a supervisory and custodian role,[95] and in providing the necessary skills in cooperation with

89. Pennycook et al.
90. Helberger, Pierson, and Poell.
91. Wardle and Derakhshan.
92. Helberger, Pierson, and Poell.
93. In Switzerland, the Federal Office of Police fedpol holds this status. See: Bundesrat.
94. Koltay.
95. Gillespie, *Custodians of the Internet*.

and complementary to state institutions such as educational institutions. In this context, voluntary or mandatory tests (e.g., for users exceeding a certain number of followers or influence level), similar to a driving exam, would help to raise the users' awareness of illegal content and untrustworthy sources, inform them about reporting possibilities and thus enable them to use a social network responsibly.[96] Finally, users are free to turn away from certain social networks if the quality of the content or the benefits no longer meet their expectations. However, it is often network effects or a lack of equivalent alternatives that make a change difficult.[97]

Table 1 provides an overview of the governance options to protect against the distribution of illegal content by social bots.

TABLE 1 Governance Options Addressing Illegal Content Distributed by Social Bots

	State Actors	Organizations (Platforms)	Users
Preventive	[no intervention desired]	Content moderation via upload filter[a] Indications for quality of content[a] Algorithmic recommendation of "quality content"[b] Transparency obligations with regard to operations[a]	Profile settings Proof of competence[c]
Repressive	Sanctions against the author Removal order	Content moderation	Reporting of illegal content (also by trusted flaggers) Switching or leaving a network
	Expert panels; bot detection, and fact-checking agencies[a]		

Notes: Although actors and potential risks arising from the activity of social bots for democratic processes and societies do not act nor manifest in isolation, they are assigned as far as possible to the three levels for the purpose of clarity. This also applies to Tables 2 and 3. No mark means that the measure is already largely implemented;

[a] partially implemented;

[b] unknown whether partially implemented;

[c] measure conceivable.

96. Examples of such voluntary offers are "Troll Factory" provided by Finnish public service broadcasting company YLE that teaches how information operations work on social media (see https://trollfactory.yle.fi/, accessed June 21, 2020) and "Interland," an initiative of Google, which wants to make children familiar with potential dangers on the Internet (see https://beinternetawesome.withgoogle.com, accessed June 21, 2020).

97. Barwise and Watkins; Saurwein, Just, and Latzer.

Governance Options Addressing the False Pretense of Human Identity

The freedom of expression guaranteed by *Article 16 Cst.* applies equally to natural and legal persons and also includes statements made anonymously or using a pseudonym. Under certain conditions, this can serve to protect against reprisals, hostility, or personal disadvantage and can be conducive to the expression of opinions that would otherwise not be voiced.[98] Accordingly, this also means that a requirement to communicate under clear name is not intended by the constitution. The validity of the scope of protection is independent of the choice of the communication channel.

STATE: It can therefore be argued that the basic right cannot be denied even to a social bot or to the programmer or initiator of a social bot who uses it, for example, to protect his or her identity.[99] This argument could be countered by saying that a social bot is not merely a communication channel that enables anonymous communication, especially when it is deliberately used to manipulate opinions or popularity indicators under the false pretense of human identity.[100] Under these circumstances, freedom of expression as a premise and freedom of opinion as a conclusion come into a seemingly irreconcilable conflict. However, no right to restrict freedom of expression can be derived from the mere pretense of human identity. The right of a programmer or user to freely express himself or herself under anonymous conditions must be given higher weight than the users' expectation to interact with humans in social networks.[101] This becomes even more evident when one considers the example of a programmer in an authoritarian regime who, for fear of reprisals, uses social bots to spread critical positions under the protection of anonymity. On the one hand, this requirement can contribute to a climate of opinion that is conducive to the free formation of opinion and is characterized by diversity of content and pluralistic interests. At the same time, however, it also conceals the danger that users may be deliberately and unnoticeably deceived, for example, by social bots copying profile information and usernames with slight changes or by imitating usage patterns.[102] A potential combination of the need for disclosure of nonhuman profiles and anonymous communication by social bots is offered by bot disclosure legislation, such as

98. Milker.
99. Steinbach.
100. Milker.
101. Oehmer.
102. Yang et al.

that envisaged in the draft of the U.S. *Bot Disclosure and Accountability Act of 2019* or the *Medienstaatsvertrag der Länder* in Germany. These laws impose an obligation to label social bots, but do not prohibit them. This preserves the possibility of using social bots to make anonymous statements. A repressive approach at the state level, on the other hand, is undesirable, as the potential danger to freedom of expression would outweigh the benefits.

ORGANIZATIONS: Implementation of the labeling obligation would require, as mentioned earlier, a very difficult and in particular valid and unambiguous identification of social bots and thus a strong cooperation with operators of social networks or other forums in which social bots can be active. Although the terms of use of Twitter or Facebook, for example, prohibit operating under a false identity, these two platforms do not check the authenticity of an identity when creating a profile.[103] However, mandatory verification, for example, by means of an official identity document, would not be entirely uncritical for reasons of data protection and privacy. The verification process would have to be designed to protect the anonymity of users (even in authoritarian regimes), especially with regard to possible repression, but also to ensure that no discrimination in access, for example, due to missing documents, is implemented. A possible solution for this might be a blockchain-based self-sovereign identity approach,[104] which leaves the sovereignty over one's own data to the user and allows him or her to decide which service provider gets access to what data.[105] For example, a social media platform would simply receive information from a certified digital identity (e-ID) that an applicant is a natural or legal person and whether a possible quota of permitted profiles has already been exhausted. In addition, further elements such as name or picture could be verified at the request of a user, which could be of particular importance for persons in public life. In addition to simply indicating that a profile is verified, this information could further be used to display the ratio between verified and nonverified followers or friends in the case of profiles

103. Verification for persons of public interest can, for example on Twitter or Facebook, be performed at a later date: https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts; https://www.facebook.com/help/1288173394636262?helpref=faq_content (accessed June 21, 2020).

104. After the e-residency was introduced in Estonia, the implementation of e-ID solutions is being discussed in several countries, including Switzerland, see https://www.bj.admin.ch/bj/de/home/staat/gesetzgebung/e-id.html (accessed June 21, 2020).

105. Mühle et al.; Sullivan and Burger.

or, in the case of content, the relationship between verified and nonverified profiles that retweet or like it. This would at the same time increase transparency with regard to popularity indicators. However, it can be assumed that the acceptance and effect of such indications would be based on the perception of their validity, which in turn would have consequences, for example, with regard to the selection and credibility attribution of sources and content, similar to those discussed for warnings in connection with false content.

Authentication of human users can also be done technologically using challenge–response techniques such as captcha tests, although progress in machine learning and AI has resulted in such tests, as currently designed, being increasingly mastered by software.[106] Such authentication could be limited, for example, to profiles that prefer to remain anonymous or unverified or where the determined value for the automation probability exceeds a certain value. This would prevent users from having to face such tests permanently. In addition, verified profiles could be assigned more relevance for recommendations, rankings or the identification of emerging issues by means of algorithmic procedures.[107] However, it should also be noted in this context that the visibility of opinions and interests, which are anonymous for fear of reprisals, could suffer as a result of this, and that algorithms, again, have the potential for bias due to nonneutral programming and selection decisions.

On the *USERS* side, the possibilities for dealing with social bots that feign a false identity are limited primarily to detecting them and preventing deception. User behavior contributes to the connections and reach that social bots can have within a network. Depending on the platform characteristics, appropriate profile settings and critical behavior can facilitate, for example, checking requests from unknown or nonverified accounts and, if necessary, cutting connections. Again, this requires a certain competence and literacy in the use of social networks. In this context, operators of social networks play a crucial role in designing the conditions (functions for and assistance in identification, operation, configuration, interaction, etc.) that enable individual users to meet their responsibilities.[108] This also applies to teaching the necessary skills in cooperation with and complementary to state institutions such as educational institutions.

106. Stark et al.
107. Lazer et al.; Grinberg et al.
108. Helberger, Pierson, and Poell.

Table 2 provides an overview of governance options to protect users from social bots feigning a false human identity.

Governance Options Addressing Automated and Potentially Unlimited Activities

Article 16 Cst. on freedom of expression in general and *Article 34 Cst.* on free political decision-making in particular do not only guarantee freedom of expression, they also oblige the Swiss *STATE* to ensure a functioning communication system that enables free and undistorted political opinion-formation and participation.[109] If the integrity of an election or vote is at risk, the state is required to intervene. However, such intervention must not lead to a situation in which freedom of expression is no longer possible under anonymous conditions.[110] Two potential threats to freedom of expression and opinion-formation can be identified as a result of automated and thus principally unlimited activities of social bots and their (hitherto) nonexistent sense of awareness, which impedes argumentation and understanding based on reflection. On the one hand, the adequate representation of pluralistic interests, on the other hand, an accurate perception of the relevance and popularity of political personalities or positions can be

TABLE 2 Governance Options Addressing the Pretense of Human Identity by Social Bots

	State Actors	Organizations (Platforms)	Users
Preventive	Bot disclosure laws (mandatory labeling for social bots)[a]	Identification obligation for users (e.g., via e-ID)[b] Challenge response tests[a]	Profile settings Proof of competence[c]
Repressive	[no intervention desired]	Profile deletion Labeling of profiles[a] Display of parameters (e.g., verification ratio)[c] Algorithmic promotion of verified profiles[b]	Termination of connections Switching or leaving a network

Notes: No mark means that the measure is already largely implemented;

[a] partially implemented;

[b] unknown whether partially implemented;

[c] measure conceivable.

109. Biaggini.

110. Oehmer.

endangered and even promote a spiral of silence.[111] A preventive intervention at state level could consist of the legislator imposing restrictions that apply for specific periods of time, actors, and channels. For example, a general prohibition of the use of social bots for parties and other political and social groups on information and opinion platforms could be considered, as envisaged in the US *Bot Disclosure and Accountability Act of 2019*. This could, for example, be extended to all users during the final weeks before an election or vote and would directly address the increased use of social bots shortly before votes or elections as identified by studies[112] and counteract the danger of artificially generated changes of opinion in highly polarized settings.[113] Such legislation would again require the cooperation of platform operators and coordination with all involved actors.

In extreme cases, repressive options at state level can go as far as declaring the results of an election or vote invalid.[114] However, it should be remembered that while the detection of social bots already presents a complex challenge, quantifying the effectively manipulative influence of social bots on the communication order, the formation of opinions, and the actual outcome of an election or vote is even more difficult. There is also a risk that such an instrument could be misused in the case of undesired election or voting results, especially in authoritarian regimes. Correspondingly, repressive measures aimed at restoring the communication order and thus enabling the free formation of opinion are normally preferable. However, this would first require comprehensive cooperative efforts on the part of the platform operators for the purpose of a continuous monitoring of discourses in terms of content and quantitative indicators. Given the asymmetrical access to data, this task would have to be demanded by state actors

111. Noelle-Neumann; Porten-Cheé et al.; Ross et al.; Keller and Klinger; Oehmer.

112. Ferrara; Stella, Ferrara, and De Domenico; Bastos and Mercea.

113. Ross et al.; Cheng, Luo, and Yu.

114. The case is known in Switzerland in which incomplete and nontransparent information from the Federal Council led the Federal Supreme Court to annul the result of the 2016 referendum on the popular initiative " Für Ehe und Familie – gegen die Heiratsstrafe [For marriage and family - against the marriage penalty]." The Federal Supreme Court justified the decision on the grounds that the incorrect information provided by the Federal Council, which was disseminated by political players and mass media, violated the right of voters to objective and transparent information, with the consequence that they were unable to form and express their opinion correctly. Given the tight outcome (50.8% no votes), the result of the vote could have been different. See https://www.bger.ch/files/live/sites/bger/files/pdf/de/1C_315_2018_yyyy_mm_dd_T_d_13_11_34.pdf (accessed June 21, 2020). If erroneous communication by the Federal Council can lead to the annulment of a voting result, this seems imaginable as well in the case of an online discourse biased by social bots in the run-up to an election or vote.

from the platform operators and could be supervised, for example, by independent expert panels and bot detection and fact-checking agencies, which have already been discussed earlier. From a democratic perspective, it is important also in this context that the implementation of governance measures reflects cultural and legal traditions and is based on a consensus among the main political and social actors on what is regarded as a threat to the communication order and public security, which circumstances justify an intervention (e.g., exceeding certain thresholds for popularity indicators with the help of the activity of social bots) and in which form such an intervention could take place (e.g., flagging or blocking of profiles and/or content).

ORGANIZATIONS: At the self-regulatory level, platform operators also dispose of preventive technological options with which they could, for example, limit the activities of a profile in order to make it more difficult to use social bots in a scalable and manipulative way. These could be based, among other things, on measuring the frequency of posted messages, likes, friend requests, and so on, per defined time unit and profile or on analyzing the diffusion of content, for example, with regard to reaction time, retweet ratio, to name a few.[115] Contents or profiles identified accordingly could then be downgraded or ignored in the algorithmic recommendations and when measuring trends.[116] It should be noted, however, that accounts incorrectly identified as bots would also be affected by such measures.

Automated activities of social bots, which are becoming more sophisticated and complex, make it more difficult for *USERS* to detect them and thus increase the likelihood that they will be deceived and ultimately, unconsciously and unintentionally, become assistants of the initiators or programmers of social bots in pursuing their goals. In this context, the promotion of knowledge and skills in the critical use of (social) media, their content, their (algorithm-based) functioning and their potential to influence opinions and political preferences would be helpful. Further, information from platforms that help to recognize the profiles of social bots or indicate the ratio of social bots within popularity indicators of profiles or content would be of help.[117] This would enable users to understand how their actions contribute to the way information is spread on social networks and how social bots can systematically exploit this. It would also

115. Shao et al.; Mønsted et al.
116. Lazer et al.
117. Ibid.

empower them to better identify and, where appropriate, report suspicious activity, content, or profiles. Again, the difficulty of validly identifying such profiles should be pointed out, which could affect the acceptance and impact of such indications.

Table 3 provides a summary of governance options that address the potentially unlimited activities of social bots.

Conclusion

The potentially harmful role of social bots in the opinion-formation and decision-making processes before votes, elections, or within controversial debates is increasingly being discussed in public, political, and scientific forums. Critical issues are (1) the dissemination of illegal content and disinformation, (2) the false pretense of human identity, and (3) the potentially unlimited number of communication and networking activities that can deceive about relevance and actual majorities.

The aim of this article was to present and discuss (existing and possible) governance options for addressing these issues. In order to take into account the characteristics of the cross-border and platform-based activities of social bots, (collaborative) options for action by state, organizational,

TABLE 3 Governance Options Addressing the Potentially Unlimited Activities of Social Bots

	State Actors	Organizations (Platforms)	Users
Preventive	Restrictions on the use or prohibition of social bots for certain time periods/ channels/actors[c]	Limitation of communication activities (per profile and time)[b]	Proof of competence[c]
Repressive	Declaration of the invalidity of a result of an election or vote (possible repetition order)[c]	Adjustment of trends and recommendations for the activities of social bots[b]	Switching or leaving a network
	Supervision by expert panels, bot detection, and fact-checking agencies[a]		

Notes: No mark means that the measure is already largely implemented;

[a] partially implemented;

[b] unknown whether partially implemented;

[c] measure conceivable.

and individual actors were considered.[118] For the fact that social networks, which present themselves as facilitators of global networking and open exchange of ideas, have proven to be particularly susceptible to manipulation through the use of social bots and thus tend to contribute more to undermining a normatively desirable exchange,[119] can be understood both as a consequence of their emergence, (technological, economic, and networking) functional logic and gaps in the distribution of responsibility for actions in social networks and their consequences.[120]

In the following, the governance options will be presented in the form of theses, grouped by problem areas:

- Illegal content distributed by social bots, which, for example, contains defamation or discrimination, can be subject to the same legal sanctions as content distributed through analogue channels. Respecting the core content of the constitutional provision on freedom of expression the state can only act repressively, never preventively, as this would constitute censorship.[121] Platforms, on the other hand, can take both preventive and repressive action against illegal content within the legal framework. Through the use of algorithmic filters and manual checks by platform employees, they can prevent its publication, downgrade it when making recommendations and measuring trends, label it with a quality indication, or remove it completely.[122] However, it should be noted that these forms of content selection, moderation, and curation offer room for systematic bias.[123] In addition, the assessment of the illegality or the factuality and quality of content is often not possible without doubt and can also vary in different legislations. For this reason, the participation of various stakeholders in the development of selection and weighting rules would be desirable. Platforms also react to notifications from users who report illegal content via the feedback options implemented by platforms and, through their actions, also contribute to whether and how such content spreads.[124] Consequently, it is also the duty of the

118. Gillespie, "Governance of and by Platforms"; Helberger, Pierson, and Poell; Napoli; Puppis; Saurwein, Just, and Latzer.

119. Tucker et al.; Deibert; Jarren.

120. Gillespie, "Governance of and by Platforms"; van Dijck, Poell, and de Waal; van de Poel et al.

121. Kley and Tophinke.

122. Lazer et al.; Pasquale, "The Automated Public Sphere."

123. Gillespie, "The Relevance of Algorithms"; Gillespie, "Governance of and by Platforms."

124. Helberger, Pierson, and Poell.

platforms, together with state institutions, to provide users with the skills for a responsible use of social networks with regard to content and sources and the knowledge of the functioning of algorithms in order to reduce the potential for deception.

- Since social bots can also serve as an instrument for anonymously sharing opinions and perspectives, a legal prohibition of the anonymous use of social bots should be rejected. However, a governmental or organizational obligation to label bots (as part of bot disclosure legislation) would be possible provided a—so far not realizable—unambiguous identification. Platforms further have the possibility of linking the creation of a user account to the use and verification of a legal or digital identity and to secure activities by means of authentication tests. Beyond mere labeling, such information can also be used to give greater weight to verified profiles in recommendations and to provide users with clues regarding automated agents' participation in online discourse. It cannot be ruled out, however, that such indications may also have unintended consequences in terms of reception, for example, in the selection and credibility attribution of sources and content. Depending on the settings options of a platform, users can minimize the possibility of getting in contact with unknown other users (including social bots).
- The provisions of the Swiss Federal Constitution on the protection of freedom of expression also establish a right to protection against a biased online discourse before elections, votes, and in controversial debates due to the potentially unlimited activity of social bots. Therefore, drastic instruments such as a prohibition of bot activities during certain periods, through certain channels or by particularly relevant and resource-rich actors would be possible. Less invasive would be a limitation of activities per user account and defined time unit.

It is important to note, however, that the governance measures described earlier may not only address the respective problem area but can also affect the other challenges at the same time: For example, limiting communication activities also minimizes the damaging potential of illegal content and vice versa. The success of the governance instruments discussed depends essentially on the extent to which the following challenges can be met:

- *National legislation:* Since activities of social bots are not confined by national borders, a national regulation seems somewhat anachronistic. For example, a prohibition of the use of social bots by Swiss actors before elections could possibly be circumvented by social bots programmed by German actors and operated from servers located in

the Ukraine. An inter- or supranational solution that unites different cultural and legal traditions (as far as possible and desirable) and different stakeholders and gathers the necessary technical know-how would be desirable in this context.

- *Balancing of legal interests:* The governance of social bots is always a balancing of different legal interests: On the one hand, the right of the individual to spread opinions and perspectives anonymously while protecting his or her identity via a social bot, on the other hand, the right of the public not to be exposed to a biased online discourse. In the choice of governance measures, both of these goods must be taken into account. Furthermore, it must also be kept in mind that governance measures introduced in states with functioning democratic systems should not serve to legitimize and expand the possibilities for censorship and repression in autocratic regimes.[125]
- *Unambiguous identification*: Governance options of social bots can only work if they can be clearly identified and false positives and negatives can be excluded. Currently, however, identification is associated with considerable uncertainty. This is mainly due to the fact that the boundaries between different profile forms are fluid and identification is like a cat-and-mouse game between programmers of social bots and programmers of recognition software.
- *Technological development:* The regulation of technologies, such as the activities of social bots but also algorithmic filters and detection programs, can only react *ex post* to technological possibilities and innovations. Future developments, their potential dangers, and the corresponding need for regulation are usually not foreseeable. Therefore, it is not possible to conclusively assess whether the options outlined earlier would also be viable in the future.
- *Economic maxim of the platforms*: The implementation of the governance options is largely dependent on platform operators. This is because legal regulations require technical and organizational implementation by the platform operators and are therefore subject to their willingness to cooperate. Platforms provide the operational framework[126] by enabling or preventing the existence of and at the same time providing settings for protection against and for social bots on their networks. A variety of the governance options outlined change the way platforms and users

125. Deibert; Tucker et al.

126. Gillespie, "Governance of and by Platforms."; Jarren.

interact, limiting the extent or intensity of communication activities on social networks and thus the basis of the platforms' business model, which is based on user data and their interactions.[127] The question as to whether platform operators, for economic considerations, are more likely to adopt a reactive stance in the development and implementation of governance options, or whether they are committed to the ideal they communicate in their marketing activities as promoters of global networking and open exchange of ideas, remains open.

BIBLIOGRAPHY

Abokhodair, Norah, Daisy Yoo, and David W McDonald. "Dissecting a Social Botnet: Growth, Content and Influence in Twitter." Paper presented at the Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, Vancouver, BC, 2015.

Aiello, Luca Maria, Martina Deplano, Rossano Schifanella, and Giancarlo Ruffo. "People Are Strange When You're a Stranger: Impact and Influence of Bots on Social Networks." Paper presented at the Sixth International AAAI Conference on Weblogs and Social Media, Dublin, Ireland, 2012.

Barwise, Patrick, and Leo Watkins. "The Evolution of Digital Dominance." In *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*, edited by Martin Moore and Damian Tambini, 21–49. New York: Oxford University Press, 2018.

Bastos, Marco T., and Dan Mercea. "The Brexit Botnet and User-Generated Hyperpartisan News." *Social Science Computer Review* 37, no. 1 (2019): 38–54. doi:10.1177/0894439317734157.

Bessi, Alessandro, and Emilio Ferrara. "Social Bots Distort the 2016 Us Presidential Election Online Discussion." *First Monday* 21, no. 11 (2016). doi:10.5210/fm.v21i11.7090.

Biaggini, Giovanni. *Bv Kommentar: Bundesverfassung Der Schweizerischen Eidgenossenschaft.* 2nd ed. Zurich, Switzerland: Orell Füssli, 2017.

Black, Julia. "Risk-Based Regulation: Choices, Practices and Lessons Being Learnt." In *Risk and Regulatory Policy: Improving the Governance of Risk*, edited by OECD, 185–224. Paris, France: OECD Publishing, 2010.

Bobkowski, Piotr S. "Sharing the News:Effects of Informational Utility and Opinion Leadership on Online News Sharing." *Journalism & Mass Communication Quarterly* 92, no. 2 (2015): 320–45. doi:10.1177/1077699015573194.

Broniatowski, David A., Amelia M. Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C. Quinn, and Mark Dredze. "Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate." *American Journal of Public Health* 108, no. 10 (2018): 1378–84. doi:10.2105/ajph.2018.304567.

Carey, John M., Victoria Chi, D. J. Flynn, Brendan Nyhan, and Thomas Zeitzoff. "The Effects of Corrective Information About Disease Epidemics and Outbreaks: Evidence from Zika and Yellow Fever in Brazil." *Science Advances* 6, no. 5 (2020): eaaw7449. doi:10.1126/sciadv.aaw7449.

127. van Dijck and Poell; Jarren.

Chan, Man-pui Sally, Christopher R. Jones, Kathleen Hall Jamieson, and Dolores Albarracín. "Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation." *Psychological Science* 28, no. 11 (November 1, 2017): 1531–46. doi:10.1177/0956797617714579.

Cheng, Chun, Yun Luo, and Changbin Yu. "Dynamic Mechanism of Social Bots Interfering with Public Opinion in Network." *Physica A: Statistical Mechanics and its Applications* 551 (January 15, 2020): 124163. doi:10.1016/j.physa.2020.124163.

Chu, Zi, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?" *IEEE Transactions on Dependable and Secure Computing* 9, no. 6 (2012): 811–24.

Clayton, Katherine, Spencer Blair, Jonathan A. Busam, Samuel Forstner, John Glance, Guy Green, Anna Kawata, et al. "Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media." *Political Behavior* (February 11, 2019). doi:10.1007/s11109-019-09533-0.

Cresci, Stefano, Marinella Petrocchi, Angelo Spognardi, and Stefano Tognazzi. "Better Safe Than Sorry: An Adversarial Approach to Improve Social Bot Detection." *arXiv preprint arXiv:1904.05132* (2019).

Davis, Clayton Allen, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. "Botornot: A System to Evaluate Social Bots." Proceedings of the 25th International Conference Companion on World Wide Web, Montréal, QC, 2016.

DeBono, Kenneth G., and Richard J. Harnish. "Source Expertise, Source Attractiveness, and the Processing of Persuasive Information: A Functional Approach." *Journal of Personality and Social Psychology* 55, no. 4 (1988): 541–46. doi:10.1037/0022-3514.55.4.541.

Deibert, Ronald J. "The Road to Digital Unfreedom: Three Painful Truths About Social Media." [In English]. *Journal of Democracy* 30, no. 1 (2019): 25–39. doi:10.1353/jod.2019.0002.

Der Bundesrat. Rechtliche Basis Für Social Media: Erneute Standortbestimmung. Nachfolgebericht Des Bundesrates Zum Postulatsbericht Amherd 11.3912 "Rechtliche Basis Für Social Media." October 5, 2017. https://www.bakom.admin.ch/dam/bakom/de/dokumente/2013/10/rechtliche_basisfuersocialmediaberichtdesbundesrates.pdf.download.pdf/rechtliche_basisfuersocialmediaberichtdesbundesrates.pdf (accessed June 21, 2020).

Edwards, Chad, Austin J. Beattie, Autumn Edwards, and Patric R. Spence. "Differences in Perceptions of Communication Quality between a Twitterbot and Human Agent for Information Seeking and Learning." *Computers in Human Behavior* 65 (December 1, 2016): 666–71. doi:10.1016/j.chb.2016.07.003.

Edwards, Chad, Autumn Edwards, Patric R. Spence, and Ashleigh K. Shelton. "Is That a Bot Running the Social Media Feed? Testing the Differences in Perceptions of Communication Quality for a Human Agent and a Bot Agent on Twitter." *Computers in Human Behavior* 33 (2014): 372–76. doi:10.1016/j.chb.2013.08.013.

Fanti, Sébastien. "Switzerland." In *Comparative Handbook: Robotic Technologies Law*, edited by Alain Bensoussan and Jérémy Bensoussan, 293–308. Brussels, Belgium: Larcier, 2016.

Ferrara, Emilio. "Disinformation and Social Bot Operations in the Run up to the 2017 French Presidential Election." *First Monday* 22, no. 8 (2017). doi:10.5210/fm.v22i8.8005.

Ferrara, Emilio, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. "The Rise of Social Bots." *Commun. ACM* 59, no. 7 (2016): 96–104. doi:10.1145/2818717.

Franklin, Stan, and Art Graesser. "Is It an Agent, or Just a Program?: A Taxonomy for Autonomous Agents." In *Intelligent Agents Iii Agent Theories, Architectures, and Languages*, edited by Jörg P. Müller, Michael J. Wooldridge and Nicholas R. Jennings, 21–35. Berlin, Heidelberg, Germany: Springer, 1997.

Gillespie, Tarleton. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media.* New Haven & London: Yale University Press, 2018.

———. "Governance of and by Platforms." In *The Sage Handbook of Social Media*, edited by J. Burgess, A. Marwick and T. Poell, 254–78. London: Sage, 2018.

———. "Platforms Are Not Intermediaries." *Georgetown Law Technology Review* 2, no. 2 (2018): 198–216.

———. "The Relevance of Algorithms." In *Media Technologies: Essays on Communication, Materiality, and Society*, edited by Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot, 167–94. Cambridge, MA & London: MIT Press, 2014.

Gollmer, Philipp. "Twitter Sperrt Lieber Zu Viel Als Zu Wenig." *NZZ Online*, May 17, 2019. https://www.nzz.ch/feuilleton/medien/twittersperrt-neue-meldefunktion-sorgt-fuer-kritik-ld.1482317 (accessed June 21, 2020).

Gorwa, Robert, and Douglas Guilbeault. "Unpacking the Social Media Bot: A Typology to Guide Research and Policy." *Policy & Internet* 12, no. 2 (2020): 225–48. doi:10.1002/poi3.184.

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. "Fake News on Twitter During the 2016 Us Presidential Election." *Science* 363, no. 6425 (2019): 374–78.

Guilbeault, Douglas. "Growing Bot Security: An Ecological View of Bot Agency." *International Journal of Communication* 10 (2016): 19.

Haim, Mario, Anna Sophie Kümpel, and Hans-Bernd Brosius. "Popularity Cues in Online Media: A Review of Conceptualizations, Operationalizations, and General Effects." [In de]. *SCM Studies in Communication and Media* 7, no. 2 (2018): 186–207. doi:10.5771/2192-4007-2018-2-58.

Hampton, Keith N, Harrison Rainie, Weixu Lu, Maria Dwyer, Inyoung Shin, and Kristen Purcell. *Social Media and the ‚Spiral of Silence'.* Washington, DC: Pew Research Center, 2014. https://www.pewresearch.org/internet/2014/08/26/social-media-and-the-spiral-of-silence/ (accessed June 21, 2020).

Hegelich, Simon, and Dietmar Janetzko. "Are Social Bots on Twitter Political Actors? Empirical Evidence from a Ukrainian Social Botnet." Paper presented at the Tenth International AAAI Conference on Web and Social Media, Cologne, Germany, 2016.

Hegelich, Simon, and Andree Thieltges. "Desinformation Und Manipulation." *aktuelle analysen* 71 (2019): 97–109.

Helberger, Natali, Jo Pierson, and Thomas Poell. "Governing Online Platforms: From Contested to Cooperative Responsibility." *The Information Society* 34, no. 1 (January 1, 2018): 1–14. doi:10.1080/01972243.2017.1391913.

Howard, Philip N., and Bence Kollanyi. "Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum." *SSRN* (June 20, 2016). doi:10.2139/ssrn.2798311.

Howard, Philip N., Samuel Woolley, and Ryan Calo. "Algorithms, Bots, and Political Communication in the Us 2016 Election: The Challenge of Automated Political Communication for Election Law and Administration." *Journal of Information Technology & Politics* 15, no. 2 (April 3, 2018): 81–93. doi:10.1080/19331681.2018.1448735.

Jarren, Otfried. "Fundamentale Institutionalisierung: Social Media Als Neue Globale Kommunikationsinfrastruktur." *Publizistik* 64, no. 2 (May 1, 2019): 163–79. doi:10.1007/s11616-019-00503-4.

Jin, Dal Yong. *Digital Platforms, Imperialism and Political Culture.* New York: Taylor & Francis, 2015. doi:10.4324/9781315717128.

Just, Natascha, and Michael Latzer. "Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet." [In English]. *Media Culture & Society* 39, no. 2 (March 2017): 238–58. doi:10.1177/0163443716643157.

Keller, Tobias R., and Ulrike Klinger. "Social Bots in Election Campaigns: Theoretical, Empirical, and Methodological Implications." *Political Communication* 36, no. 1 (January 2, 2019): 171–89. doi:10.1080/10584609.2018.1526238.

Kley, Andreas, and Esther Tophinke. "Art. 16 Meinungs- Und Informationsfreiheit." In *Die Schweizerische Bundesverfassung: Kommentar*, edited by Bernhard Ehrenzeller, Philippe Mastronardi, Rainer J. Schweizer, and Klaus A. Vallender. Zürich/St. Gallen: Dike/Schulthess Juristische Medien AG, 2008.

Klinger, Ulrike, and Jakob Svensson. "The End of Media Logics? On Algorithms and Agency." *New Media & Society* 20, no. 12 (2018): 4653–70. doi:10.1177/1461444818779750.

Knobloch-Westerwick, Silvia, Nikhil Sharma, Derek L. Hansen, and Scott Alter. "Impact of Popularity Indications on Readers' Selective Exposure to Online News." *Journal of Broadcasting & Electronic Media* 49, no. 3 (2005/09/01 2005): 296–313. doi:10.1207/s15506878jobem4903_3.

Koltay, Andràs. *New Media and Freedom of Expression: Rethinking the Constitutional Foundations of the Public Sphere.* Oxford, UK: Bloomsbury Publishing, 2019.

Lazer, David M. J., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, *et al.* "The Science of Fake News." [In English]. *Science* 359, no. 6380 (March 9, 2018): 1094–96. doi:10.1126/science.aao2998.

Libertus, Michael. "Rechtliche Aspekte Des Einsatzes Von Social Bots De Lege Lata Und De Lege Feranda." *Zeitschrift für Urheber- und Medienrecht* 62, no. 1 (2018): 20–26.

Mena, Paul. "Cleaning Up Social Media: The Effect of Warning Labels on Likelihood of Sharing False News on Facebook." *Policy & Internet* 12, no. 2 (2020): 165–83. doi:10.1002/poi3.214.

Messing, Solomon, and Sean J. Westwood. "Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online." *Communication Research* 41, no. 8 (2014): 1042–63. doi:10.1177/0093650212466406.

Milker, Jens. ""Social-Bots" Im Meinungskampf. Wie Maschinen Die Öffentliche Meinung Beeinflussen Und Was Wir Dagegen Unternehmen Können." *Zeitschrift für Urheber- und Medienrecht* 61, no. 3 (2017): 216–22.

Mitcham, Carl. "Agency in Humans and in Artifacts: A Contested Discourse." In *The Moral Status of Technical Artefacts*, edited by Peter Kroes and Peter-Paul Verbeek, 11–29. Dordrecht, Netherlands: Springer Netherlands, 2014.

Mønsted, Bjarke, Piotr Sapieżyński, Emilio Ferrara, and Sune Lehmann. "Evidence of Complex Contagion of Information in Social Media: An Experiment Using Twitter Bots." *PLOS ONE* 12, no. 9 (2017): e0184148. doi:10.1371/journal.pone.0184148.

Moor, James H. "The Nature, Importance, and Difficulty of Machine Ethics." *IEEE Intelligent Systems* 21, no. 4 (2006): 18–21.

Mühle, Alexander, Andreas Grüner, Tatiana Gayvoronskaya, and Christoph Meinel. "A Survey on Essential Components of a Self-Sovereign Identity." *Computer Science Review* 30 (November 1, 2018): 80–86. doi:10.1016/j.cosrev.2018.10.002.

Myers West, Sarah. "Censored, Suspended, Shadowbanned: User Interpretations of Content Moderation on Social Media Platforms." *New Media & Society* 20, no. 11 (2018): 4366–83. doi:10.1177/1461444818773059.

Napoli, Philip M. "Social Media and the Public Interest: Governance of News Platforms in the Realm of Individual and Algorithmic Gatekeepers." [In English]. *Telecommunications Policy* 39, no. 9 (October 2015): 751–60. doi:10.1016/j.telpol.2014.12.003.

Neudert, Lisa-Maria, Bence Kollanyi, and Philip N. Howard. *Junk News and Bots During the German Federal Presidency Election: What Were German Voters Sharing over Twitter?*

Oxford, UK: Project on Computational Propaganda, March 27, 2017. http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/03/German-What-Were-German-Voters-Sharing-Over-Twitter-v9.pdf (accessed June 21, 2020).

Nissenbaum, Helen. "Accountability in a Computerized Society." *Science and Engineering Ethics* 2, no. 1 (March 1, 1996): 25–42. doi:10.1007/bf02639315.

Noelle-Neumann, Elisabeth. *The Spiral of Silence: Public Opinion--Our Social Skin.* 2nd ed. Chicago, IL: University of Chicago Press, 1993.

Oehmer, Franziska. "Meinungsfreiheit Für Social Bots?" *Jusletter*, April 4, 2019. https://jusletter.weblaw.ch/juslissues/2019/977/meinungsfreiheit-fur_0e2c3de5fc.html (accessed June 21, 2020).

Oentaryo, Richard Jayadi, Arinto Murdopo, Philips Kokoh Prasetyo, and Ee-Peng Lim. "On Profiling Bots in Social Media." *arXiv e-prints* (September 1, 2016). https://ui.adsabs.harvard.edu/abs/2016arXiv160900543J (accessed June 21, 2020).

Papakyriakopoulos, Orestis, Juan Carlos Medina Serrano, and Simon Hegelich. "Political Communication on Social Media: A Tale of Hyperactive Users and Bias in Recommender Systems." *Online Social Networks and Media* 15 (January 1, 2020): 100058. doi:10.1016/j.osnem.2019.100058.

Pasquale, Frank. "The Automated Public Sphere." *SSRN Scholarly Paper* (November 8, 2017). U of Maryland Legal Studies Research Paper No. 2017-31. https://ssrn.com/abstract=3067552 (accessed June 21, 2020).

———. "Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power." *Theoretical Inquiries in Law* 17, no. 2 (2016): 487–513.

Pennycook, Gordon, Adam Bear, Evan T. Collins, and David G. Rand. "The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines without Warnings." *Management Science* (2020). doi:10.1287/mnsc.2019.3478.

Pennycook, Gordon, Tyrone D. Cannon, and David G. Rand. "Prior Exposure Increases Perceived Accuracy of Fake News." *Journal of Experimental Psychology: General* 147, no. 12 (2018): 1865–80.

Petty, Richard E., and Pablo Briñol. "Attitude Change." In *Advanced Social Psychology: The State of the Science.*, edited by R. F. Baumeister and E. J. Finkel, 217–59. New York: Oxford University Press, 2010.

Petty, Richard E., and John T. Cacioppo. "The Elaboration Likelihood Model of Persuasion." In *Communication and Persuasion: Central and Peripheral Routes to Attitude Change.*, edited by R. E. Petty and J. T. Caccioppo, 1–24: New York: Springer, 1986.

———. "Source Factors and the Elaboration Likelihood Model of Persuasion." *ACR North American Advances* 11 (1984): 668-72.

Porten-Cheé, Pablo, Jörg Haßler, Pablo Jost, Christiane Eilders, and Marcus Maurer. "Popularity Cues in Online Media: Theoretical and Methodological Perspectives." [In de]. *SCM Studies in Communication and Media* 7, no. 2 (2018): 208–30. doi:10.5771/2192-4007-2018-2-80.

Puppis, Manuel. "Media Governance: A New Concept for the Analysis of Media Policy and Regulation." *Communication, Culture & Critique* 3, no. 2 (2010): 134–49. doi:10.1111/j.1753-9137.2010.01063.x.

Rauchfleisch, Adrian, and Jonas Kaiser. "The False Positive Problem of Automatic Bot Detection in Social Science Research." *Social Science Research* (March 2020). Berkman Klein Center Research Publication No. 2020–3. doi:10.2139/ssrn.3565233.

Rauchfleisch, Adrian, and Daniel Vogler. #Nobillag Auf Twitter: Grabenkämpfe Zwischen Gegnern Und Befürwortern. February 20, 2018. https://www.foeg.uzh.ch/dam/jcr:7b9901f5-2942-43e3-b3b3-e1345ae6a62b/%23NoBillag%20auf%20Twitter.pdf (accessed June 21, 2020).

Reeves, Byron, and Clifford Ivar Nass. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* Cambridge, UK: Cambridge University Press, 1996.

Rhinow, René, and Markus Schefer. *Schweizerisches Verfassungsrecht.* 2nd ed. Basel, Switzerland: Helbing Lichtenhahn, 2009.

Rhodes, Roderick A. W. "The New Governance: Governing without Government." *Political Studies* 44, no. 4 (1996): 652–67. doi:10.1111/j.1467-9248.1996.tb01747.x.

Ross, Björn, Laura Pilz, Benjamin Cabrera, Florian Brachten, German Neubaum, and Stefan Stieglitz. "Are Social Bots a Real Threat? An Agent-Based Model of the Spiral of Silence to Analyse the Impact of Manipulative Actors in Social Networks." *European Journal of Information Systems* 28 (2019): 394–412. doi:10.1080/0960085X.2018.1560920.

Saurwein, Florian, Natascha Just, and Michael Latzer. "Governance of Algorithms: Options and Limitations." *Info* 17, no. 6 (2015): 35–49. doi:10.1108/info-05-2015-0025.

Schäfer, Fabian, Stefan Evert, and Philipp Heinrich. "Japan's 2014 General Election: Political Bots, Right-Wing Internet Activism, and Prime Minister Shinzō Abe's Hidden Nationalist Agenda." *Big Data* 5, no. 4 (December 1, 2017): 294–309. doi:10.1089/big.2017.0049.

Schweizerisches Bundesgericht. "Bundesgerichtsurteil 6b_119/2017." 2017. https://www.bger.ch/ext/eurospider/live/de/php/aza/http/index.php?highlight_docid=aza%3A%2F%2Faza://12-12-2017-6B_119-2017&lang=de&zoom=&type=show_document (accessed June 21, 2020).

———. "Bundesgerichtsurteil 6b_267/2018." 2018. https://www.bger.ch/ext/eurospider/live/de/php/aza/http/index.php?highlight_docid=aza%3A%2F%2Faza://17-05-2018-6B_267-2018&lang=de&zoom=&type=show_document (accessed June 21, 2020).

Shao, Chengcheng, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. "The Spread of Low-Credibility Content by Social Bots." *Nature Communications* 9, no. 1 (November 20, 2018): 4787. doi:10.1038/s41467-018-06930-7.

Sprecher, Franziska, Andreas Lienhard, Pierre Tschannen, Axel Tschentscher, and Franz Zeller. "Die Staatsrechtliche Rechtsprechung Des Bundesgerichts in Den Jahren 2017 Und 2018." *Zeitschrift des Bernischen Juristenvereins* 154, no. 10 (2018): 641–707.

Stark, Fabian, Caner Hazırbas, Rudolph Triebel, and Daniel Cremers. "Captcha Recognition with Active Deep Learning." Paper presented at the Workshop New Challenges in Neural Computation, Aachen, Germany, 2015.

Steinbach, Armin. "Social Bots Im Wahlkampf." *Zeitschrift für Rechtspolitik* 50, no. 4 (2017): 101–05.

Stella, Massimo, Emilio Ferrara, and Manlio De Domenico. "Bots Increase Exposure to Negative and Inflammatory Content in Online Social Systems." *Proceedings of the National Academy of Sciences* 115, no. 49 (2018): 12435–40. doi:10.1073/pnas.1803470115.

Stieglitz, Stefan, Florian Brachten, Davina Berthelé, Mira Schlaus, Chrissoula Venetopoulou, and Daniel Veutgen. "Do Social Bots (Still) Act Different to Humans?–Comparing Metrics of Social Bots with Those of Humans." Paper presented at the International Conference on Social Computing and Social Media, Vancouver, BC, 2017.

Stieglitz, Stefan, Florian Brachten, Björn Ross, and Anna-Katharina Jung. "Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts." *arXiv preprint arXiv:1710.04044* (2017).

Suárez-Serrato, Pablo, Margaret E. Roberts, Clayton Davis, and Filippo Menczer. "On the Influence of Social Bots in Online Protests." In *Social Informatics: SocInfo2016, Lecture Notes in Computer Science*, edited by E. Spiro and Y.-Y. Ahn, 269-78. Cham, Switzerland: Springer International Publishing, 2016. doi: 10.1007/978-3-319-47874-6_19.

Subrahmanian, Venkatramanan S., Amos Azaria, Skylar Durst, Vadim Kagan, Aram Galstyan, Kristina Lerman, Linhong Zhu, *et al.* "The Darpa Twitter Bot Challenge." *Computer* 49, no. 6 (2016): 38–46.

Sullivan, Clare, and Eric Burger. "E-Residency and Blockchain." *Computer Law & Security Review* 33, no. 4 (August 1, 2017): 470–81. doi:10.1016/j.clsr.2017.03.016.

Thompson, Dennis F. "Moral Responsibility of Public Officials: The Problem of Many Hands." *American Political Science Review* 74, no. 4 (1980): 905–16.

Tsvetkova, Milena, Ruth García-Gavilanes, Luciano Floridi, and Taha Yasseri. "Even Good Bots Fight: The Case of Wikipedia." *PLOS ONE* 12, no. 2 (2017): e0171774. doi:10.1371/journal.pone.0171774.

Tucker, Joshua A., Yannis Theocharis, Margaret E. Roberts, and Pablo Barberá. "From Liberation to Turmoil: Social Media and Democracy." *Journal of democracy* 28, no. 4 (2017): 46–59.

van de Poel, Ibo, Jessica Nihlén Fahlquist, Neelke Doorn, Sjoerd Zwart, and Lambèr Royakkers. "The Problem of Many Hands: Climate Change as an Example." journal article. *Science and Engineering Ethics* 18, no. 1 (March 1, 2012): 49–67. doi:10.1007/s11948-011-9276-0.

van Dijck, José, and Thomas Poell. "Understanding Social Media Logic." *Media and Communication* 1, no. 1 (2013): 2–14.

van Dijck, José, Thomas Poell, and Martijn de Waal. *The Platform Society: Public Values in a Connective World.* New York: Oxford University Press, 2018.

Varol, Onur, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. "Online Human-Bot Interactions: Detection, Estimation, and Characterization." Paper presented at the Eleventh International AAAI Conference on Web and Social Media, Montréal, QC, 2017.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. "The Spread of True and False News Online." *Science* 359, no. 6380 (2018): 1146–51. doi:10.1126/science.aap9559.

Walter, Nathan, and Sheila T. Murphy. "How to Unring the Bell: A Meta-Analytic Approach to Correction of Misinformation." *Communication Monographs* 85, no. 3 (July 3, 2018): 423–41. doi:10.1080/03637751.2018.1467564.

Walter, Nathan, and Riva Tukachinsky. "A Meta-Analytic Examination of the Continued Influence of Misinformation in the Face of Correction: How Powerful is it, Why Does it Happen, and How to Stop it?" *Communication Research* 47, no. 2 (March 1, 2019): 155–77. doi:10.1177/0093650219854600.

Wardle, Claire, and Hossein Derakhshan. "Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making." *Council of Europe Report* 27 (2017): 1–107.

Wood, Julia K. "Truth, Lies, and Stolen Valor: A Case for Protecting False Statements of Fact under the First Amendment." *Duke Law Journal* 61, no. 2 (2011): 469–510.

Woolley, Samuel C. "Automating Power: Social Bot Interference in Global Politics." *First Monday* 21, no. 4 (2016). doi:10.5210/fm.v21i4.6161.

Yang, JungAe. "Effects of Popularity-Based News Recommendations ("Most-Viewed") on Users' Exposure to Online News." *Media Psychology* 19, no. 2 (April 2, 2016): 243–71. doi:10.1080/15213269.2015.1006333.

Yang, Kai-Cheng, Onur Varol, Clayton A. Davis, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. "Arming the Public with Artificial Intelligence to Counter Social Bots." *Human Behavior and Emerging Technologies* 1, no. 1 (2019): 48–61.

Yücel, Dennis. "Gefährliche Lautsprecher." *Der Tagesspiegel Online*, April 29, 2019. https://www.tagesspiegel.de/themen/freie-universitaet-berlin/social-bots-gefaehrliche-lautsprecher/24258248.html (accessed June 21, 2020).