MediaWell

Research Review

Algorithms, Automation, and Misinformation

September 16, 2020

Automation: Problems and potentials

Our daily routines and interactions are increasingly assisted by and related to social media. We use social media platforms not only to connect to our friends but also to read news, express our social and political views, and make purchases. But we do so without necessarily realizing the roles of the under-the-hood algorithms in everything we use these platforms for. Outside of the social media world, our lives are also assisted, monitored, and guided by automated devices and systems. Some are as simple as a timer on a toaster, whereas others—like Amazon's virtual assistant Alexa, Apple's Siri, or Google Assistant, to name a few—are extremely sophisticated and involve cutting-edge research at the intersection of modern computer science, data science, linguistics, human-computer interaction, psychology, and social science.

Although typically designed and developed with the idea of making our lives simpler and more comfortable, automated systems have not always achieved this goal. Instead, their advent has also been associated with new social, political, and moral challenges and risks. New algorithms can exacerbate existing racial, ethnic, gender, and other biases and prejudices. An example—discussed in more detail below—would be Facebook's online advertisement algorithms, which can enhance racial and gender biases by automatically targeting specific genders and racial groups based on the content of the ads (Ali et al. 2019a).

These biases can emerge inadvertently as a result of algorithmic learning from interactions with humans. When Microsoft launched its artificial intelligence chatbot Tay on Twitter in late March 2016, no one expected that the company would need to take it down in less than 24 hours because the system would automatically learn to swear and make racist comments based on available data and interactions with other Twitter users (Wakefield 2016).

Tay's failure was not the last time an automated system deployed on social media and referred to as a "bot" would make newspaper headlines around the globe. Academic research conducted around the 2016 US presidential election revealed that about 400,000

Twitter bots posted almost 4 million tweets about the election, making up almost 20 percent of tweets on this topic (Bessi and Ferrara 2016). The same research pointed out that social media bots are capable of further polarizing online political conversations and promoting misinformation.

The spread of digital technologies has created an array of ethical, social, economic, and political issues that have drawn attention from scholars, journalists, and policymakers. This review focuses on a subset of these issues related to the use of algorithms and automation on social media platforms, and particularly on the reasons, techniques, and consequences of deploying social media bots. What do we know about the scope and importance of this phenomenon? And what are feasible remedies for addressing existing problems and potential challenges presented by these automated technologies?

Bots: Definitions, mechanisms, and strategies

Despite ample research on social bots across academic disciplines, there is still some controversy about what a bot is. A recent study focused on the use of political bots in Venezuela defines bots as “computer-generated programs that post, tweet, or message of their own accord” (Forelle et al. 2015). Although bots can and often do engage in these types of activities, they can also do more than that. As Varol et al. (2017) pointed out, “social bots are accounts controlled by software, algorithmically generating content and establishing interactions.” Alternatively, according to Bastos and Mercea’s (2018) definition, “bots are automatic posting protocols used to relay content in a programmatic fashion.” The lack of a common definition makes it hard to directly compare findings from different research projects. It also highlights scholarly disagreement about the specific characteristics required to consider a social media account a bot. Nevertheless, most existing definitions agree on the automatic nature of bot accounts, whose activity is controlled by an algorithm that can perform a variety of actions (such as liking other accounts, sending friend requests, posting content that was automatically generated or borrowed, and even engaging in online discussions with other social media users).

The automatic nature of bots makes it possible to use them to amplify a particular group of users or specific messages in online discussions (Bastos and Mercea 2019). Given the relatively low costs of developing simple social media bots (Grimme et al. 2017), different actors might be interested in getting access to bots instead of or in addition to other, more pricey promotion options on social media. The availability of social media bots is especially simplified by social media platform architectures that often have publicly available application programming interfaces (APIs)—or sets of tools for automatically accessing and managing social media accounts—that allow a functionality similar to that of a human user

(Takacs and McCulloh 2019). APIs are available on Twitter, Facebook, Instagram, Viber, and other platforms. Although social media platforms differ in terms of what is possible to do and what information is available via API, in most cases users can post new content and change their profiles (Yasu 2019). Importantly, some APIs (e.g., on Twitter) also allow users to automatically access other accounts' information, which enables automatic data collection for commercial, political, or other purposes that we discuss below.

Social bots are often designed to "mimic human social media users" (Woolley 2016). Nevertheless, they can be dramatically different in terms of their complexity and sophistication. Some researchers argued that early bots on Twitter could be spotted using simple rules of thumb based on the number of tweets posted, though these strategies have been contentious in the research community. Oentaryo et al. (2016) searched for bots among the accounts that posted at least 15 tweets per month in 2014. Other simple rules based on account activity were also proposed in Metaxas and Mustafaraj (2010) and Howard and Kollanyi (2016). Alternative rules of thumb included the similarity of posts (Gao et al. 2010; Gao et al. 2012; Sanovich et al. 2018). However, abundant evidence suggests that social media bots have evolved over time (Luceri et al. 2019; Cresci 2020). Luceri et al. (2019) analyzed data about the activity of Twitter bots during the 2016 and 2018 electoral campaigns in the US and found clear differences between human and bot behavior in 2016 on the one hand, and striking similarities in their activity patterns in 2018 on the other. The authors concluded that "some bots have grown more sophisticated and have been refined to emulate human timing." This trend of growing bot sophistication has even been labeled as a "paradigm shift" in the development of social media bots, a shift that poses an important challenge to existing bot-detection tools and methodologies based on patterns of account activities, the content of posted messages, or the network structure of accounts' online interactions (Cresci et al. 2016).

Given the high diversity and rapid evolution of social bots, it is little surprise that scarce research has been conducted about human perception of or interactions with bots on social media platforms. An early study came from the communication field and used a randomized experiment to measure a number of different psychological reactions and attitudes of almost 250 students who were exposed to identical content on one of two mock Twitter pages. One was presented as a Centers for Disease Control Twitter account, and the other as a bot-controlled page (Edwards et al. 2014). Although this study did find statistically significant differences in reported measures of social attraction between groups of students exposed to nominally human and bot accounts, the results were not significant for many other measures, including source credibility. The authors concluded that a "Twitterbot is perceived as a credible source of information."

These findings were further confirmed in nonexperimental settings by observing human accounts engaging in online conversations with Twitter bots (Cresci et al. 2016) and sharing social media posts produced by bots to the same extent as those posted by human users (Shao et al. 2018). Shao et al. concluded that “collectively, people do not discriminate between low-credibility content shared by humans versus social bots.” These unsettling findings are not, however, in line with evidence from political discussion on Twitter during the 2016 US presidential election, in which human users were shown to reply to other humans much more than to Twitter bots (Bessi and Ferrara 2016). Overall, further research is required to better understand human perception of social media bots, as this might shed light on new strategies for countering malicious bot activities.

For Whom the Bot Tolls

Automatic activity on social media platforms is not necessarily evil, as bots are sometimes used for social good. For example, Savage et al. (2016) developed a platform that employs Twitter bots to engage with volunteers and activists who were willing to contribute to anticorruption activities in Latin America. Efforts have been made to develop guidelines for creating social bots that would connect otherwise ideologically segregated groups of social media users, thereby attenuating ideological polarization on social media platforms (Graham and Ackland 2016). Mass media can also deploy bots to facilitate the spread of information and keep their readers informed about the latest news (Lokot and Diakopoulos 2016; Diakopoulos 2019). However, this type of news bot can be misused to disseminate propaganda or falsehoods—a topic that constitutes a burgeoning area of research at the intersection of data science and the social sciences.

A popular mass media topic is the use of bots in Russia, a country whose government has been accused of being linked to promoting social media bots with nefarious political purposes internationally. Stukal et al. (2019) addressed this issue systematically using Twitter data from 2014 to 2017 and found that bots make up around half of all Twitter accounts that discuss Russian politics (Stukal et al. 2017)—far more than the 9–15 percent estimate for the proportion of bots in the English-language Twittersphere (Varol et al. 2017). The prevalence of Twitter bots in Russia can partly be explained by the modest popularity of Twitter, which has around 9 million users in the country, whereas Instagram is utilized by over 32 million people (Statista 2020). What is more surprising is that Russia-focused research also found that progovernment bots make up only one-third of all bots that operate on Russian-language Twitter; the rest were found to either promote antigovernment content or have no explicit political leaning at all. Similar results have also been reported using data from Venezuela, where the most active bots turn out to act on the side of Venezuela’s radical opposition (Forelle et al. 2015). However, this scholarship avoids directly attributing bots to specific political actors, as this type of attribution is typically infeasible based on publicly available data and might require intelligence or data leaks.

Overall, the deployment of automated social media accounts for propaganda and manipulation of public opinion across the political spectrum is well documented. Abokhodair et al. (2015) analyzed a network of interconnected bot accounts (also called a “botnet”) that tweeted in Arabic in 2012 and flooded hashtags about the Syrian civil war with non-war content, thereby distracting public attention from the conflict. Propaganda-focused Russian bots also demonstrate a variety of tactics and strategies to promote proregime messages (Sanovich 2018; Stukal et al. 2020). Finally, automated accounts can even work for terrorist organizations, as shown by Berger and Morgan (2015), who described Twitter bots posting tens of thousands of pro-ISIS tweets per day in 2014.

Actors can also attempt to manipulate public opinion with social bots. The bot strategy that has drawn particular scholarly attention in this context is the spread of false information online. Vosoughi et al. (2018) studied the dissemination of true and false news on Twitter from 2006 through 2017 and found that adding tweets and tweet threads posted by bots into the analysis did not change the result: “false news still spread farther, faster, deeper, and more broadly than the truth.” Another study pointed out that bots could be employed to amplify the spread of false articles via resharing links posted by other accounts (Shao et al. 2018). The key advantage of using bots for this purpose is their scalability, which poses a major threat to efforts to fight media manipulation (Luceri et al. 2019; Yang et al. 2019).

Another major topic in the study of how bots are employed for manipulating public opinion in democratic countries is their use during elections and referenda. Election campaigns in the US have been of special interest for researchers studying bots. Early examples of research on this topic date back to 2010, when an analysis of Twitter data about the 2010 special US Senate election in Massachusetts uncovered Twitter spam attacks aimed at promoting particular candidate-related topics into Twitter trends (Metaxas and Mustafaraj 2010). More recent analyses of automated Twitter accounts during the 2016 US presidential election campaign revealed an increase in the activity of these accounts from the first to the second candidate debates, producing roughly one quarter of all tweets about this topic (Kollanyi et al. 2016). Another estimate is that about one-fifth of all tweets about the presidential election were produced by automated Twitter accounts (Bessi and Ferrara 2016). Despite all recent efforts to cope with the political activity of social media bots, their use was also documented in the 2018 US midterm election, during which millions of dormant bots were found following either Democratic or Republican candidates (Takacs and McCulloh 2019).

Bots have also been deployed during electoral campaigns outside of the US. In Germany, up to 10 percent of users following parties’ Twitter accounts during the 2017 German federal election were reported to be bots (Keller and Klinger 2019), although this proportion varied across the political spectrum in somewhat unexpected ways. For example, it turned out that AfD, a German far-right political party, did not have more bot followers than its rivals. In Asia-Pacific, in particular in the Philippines and Indonesia, around one-fourth of Twitter accounts involved in political conversations about elections were labeled as bots (Uyheng and Carley 2019). Finally, automated traffic generated in relation to the Brexit referendum in the UK in 2016 was reported to make up around 30 percent of all tweets (Howard and Kollanyi 2016); these automated accounts were found to amplify hyperpartisan information, although not necessarily so-called fake news content (Bastos and Mercea 2019).

Finally, not every bot is necessarily designed to interact with and affect human users directly. Instead, social media bots can be deployed to manipulate other algorithms we use online and thereby affect human behavior indirectly. The first efforts to use automation to manipulate the information people see online—although different from today's social media bots—were detected as early as 2004–2006. Multiple studies detected the use of so-called Googlebombs as a political campaigning tool to help promote negative information about a political opponent in Google search rankings (McNichol 2004; Zeller 2006). Today, social media bots can be employed for similar purposes, as Elmas et al. (2019) illustrated in their study of how this works with Twitter's trending topics algorithm. Their research analyzed multiple attacks on trending topics on Twitter in Turkey and revealed that they were successful in promoting specific keywords to trending topics in 90 percent of the cases. This alarming finding highlights that the effects of bot deployment can be diverse and hard to gauge accurately.

Dealing with social bots

Addressing the wide range of social and political issues of automatic activity on social media platforms necessarily requires that social media users be able to detect algorithmically controlled accounts. Bot detection has been an active area of research not only in computer and data sciences, but also in media and communication studies and journalism. Investigative journalists, media professionals, and activists have come up with diverse guidelines for users about how to spot a social media bot (Nimmo 2017; Roberts 2020). This issue has also been studied in a more systematic way in computer science, starting with a seminal paper by Chu et al. (2010).

Most research in this area has been based on supervised learning; in other words, it requires that researchers not only collect information that could potentially be indicative of bots, but also have access to ground-truth labels (bots vs. legitimate users) for a subset of social media accounts. The research goal is then to develop a classifier that would use provided features and ground-truth labels to uncover patterns of features that distinguish bots from humans. Once these patterns are identified, the classifier can be applied to larger collections of social media data to detect bots automatically. Although a number of supervised classifiers have been proposed for bot detection (see Orabi et al. 2020 for an overview), the most commonly used supervised learning tool is the Botometer software, which was developed by a group of scholars from Indiana University (Davis et al. 2016; Varol et al. 2017). Botometer uses an ensemble of multiple classifiers known as *random forest* to find bot-identifying patterns among over 1,000 features related to accounts' characteristics, the content of their tweets, their network ties, and so on. For each tested account, Botometer produces a numerical score, with higher values indicating higher confidence in the automated nature of the account (Yang et al. 2019).

Although the use of Botometer is extremely popular in a social media research (Stieglitz et al. 2017; Vosoughi et al. 2018; Kitzie et al. 2018; Luceri et al. 2019), it faces the same challenges as all supervised learning algorithms, which require extensive high-quality labeled sets with ground-truth labels. In the case of Botometer, the labeled set includes over a dozen different sets of bot accounts collected by different researchers at different times (Bot Repository 2020). But there is also a growing understanding that a mere combination of different sets of detected bots for training a supervised learning bot-detection algorithm might not be an ideal strategy, due to important crossnational differences in bot development and the evolution of bot technologies over time (Cresci et al. 2017; Uyheng and Carley 2019; Rauchfleisch and Kaiser 2020).

An alternative approach that does not require extensive labeled sets is typically referred to as unsupervised learning. The key idea here is to identify anomalies in the behavior of different accounts and flag those as bots. This idea underlies DeBot, another popular and publicly available bot-detection tool, which detects groups of accounts with highly correlated activity over time (Chavoshi et al. 2016). Some approaches are based on identifying groups, or clusters, of similar social media accounts using a number of features (Wu et al. 2018; Khalil et al. 2020). Cresci et al. (2016) suggested a DNA-research-inspired approach that involves coding account activities over time with sequences of symbols, finding the longest common subsequence, and flagging as potential bots those accounts that share it. Finally, so-called semi-supervised approaches, which leverage the benefits of both supervised and unsupervised methods, have been recently proposed for detecting bots (Dorri et al. 2018; Shi et al. 2019).

A fundamental limitation of the described efforts to detect automated social media accounts is the underlying assumption that a bot should constantly remain algorithmically controlled. A recent study of attacks on Twitter's trending topics in Turkey has, however, revealed that algorithms can get access to humans' compromised accounts intermittently and perform unauthorized automated activities without being detected by bot detectors or even account owners (Elmas et al. 2019). Attempts to overcome this limitation include the development of systems that compare newly posted messages to the long-term behavioral profile of a Twitter account (Egele et al. 2017). These studies also highlight the importance of another strand of academic research that is mainly conducted in the field of human-computer interaction and aims to reverse-engineer bots' strategies of infiltrating social media platforms. Freitas et al. (2015) created over a hundred Twitter bots with different characteristics, monitored their success in interactions with human users, and found a positive correlation between automated accounts' activity level and their success in gaining human followers. Another study also revealed that bots that are equipped with better conversational skills are more successful in retaining followers (Savvopoulos et al. 2018). These studies indicate that human users often lack skills to identify bots (Cresci et al. 2017), which becomes especially challenging for hybrid accounts, also known as cyborgs, that exhibit both human-driven and automated behavior (Chu et al. 2010; Grimme et al. 2017). But overall, there is a limited understanding of the extent to which automated accounts are able to infiltrate social media communities or have effects on human users.

Beyond Twitter bots

Most research cited so far focuses on Twitter bots, although earlier studies also featured analysis of bot strategies on Facebook. For example, Boshmaf et al. (2011) deployed a network of over a hundred social bots and maintained it for several months on Facebook to study how bots infiltrate networks of human accounts. However, the changes in data-access policies that Facebook implemented in 2018 after the infamous Cambridge Analytica scandal (Confessore 2018) have made it particularly difficult to analyze automated activity on the platform. A recent study that proposed a set of features and an automatic classifier for detecting Facebook bots (Santia et al. 2019) also highlighted that new Facebook data-access policies prevent users from easily deploying this new bot-detection tool. The new Facebook policy requires that scholars who are interested in understanding the role of algorithms and automation on the platform either use pieces of data Facebook chooses to release via CrowdTangle, a social analytics tool that belongs to Facebook, or come up with inventive research designs and feasible research questions that take these data limitations into account (Freelon 2018). An example of the former approach is Giglietto et al.'s (2020) study of coordinated link-sharing behavior on Facebook that uses CrowdTangle. This study monitored over 300,000 Facebook links posted during recent electoral campaigns in Italy and found that links from media sources that were previously flagged as problematic by fact-checkers were roughly twice as likely to be shared in a coordinated campaign as in a noncoordinated way.

An example of the latter approach is Ali et al.’s research on Facebook advertisement algorithms. Ali et al. (2019a) use publicly available voter records from North Carolina and identify locations populated mostly by Black and white people to analyze racial skew in Facebook algorithms for housing-advertisement delivery. This research found “significant ad delivery skew along racial lines … with certain ads delivering to an audience of over 85% white users while others delivering [sic] to an audience of as little as 35% white users.” The authors admit that further research is required to understand the particular ad properties that drive this skew. The same study found significant gender skew in the delivery of Facebook ads about bodybuilding and cosmetics, and identified images associated with the ads as a major driver of the algorithmic skew. Another study by same group of scholars also found algorithmically induced echo-chamber effects in the delivery of political ads (Ali et al. 2019b). In particular, the latter study used an experimental design to show that over 65 percent of users who automatically received ads with content from Bernie Sanders’s 2016 campaign were registered Democrats, whereas the percentage of registered Democrats among the recipients of ads from Donald Trump’s campaign was under 40 percent. The authors emphasized that this algorithmic skew is driven by complex optimization and the bidding procedure used for ad delivery on Facebook.

Algorithmic biases are also a concern outside of social media platforms. Noble (2018) provides a book-length analysis of racial biases and expressions of racism in Google search outputs that range from showing photos of Black people for image search with the keyword “gorilla” to the Google Maps search on “N*gga House” showing the White House under Barack Obama’s administration in April 2016.

A separate line of research focuses on digital assistants (e.g., Alexa, Siri, or Google Assistant). It has been argued that these technologies are capable of tightening ideological echo-chambers by not providing an ideologically diverse news stream to their consumers, or even potentially assisting nondemocratic regimes in their censorship endeavours (Stucke and Ezrachi 2017). Another set of issues are potential algorithmic biases related to users’ gender, ethnicity, or language. Lima et al. (2019) conducted experiments asking participants from different parts of Brazil to read out loud a series of sentences in Portuguese to Siri or Google Assistant. The results indicated that these digital assistants were biased toward accents of individuals from the more developed part of the country. The issue of social and economic inequalities being amplified by new forms of digital inequality and exclusion is also highlighted by Park and Humphry (2019), who illustrated the problem with examples from Australia, where people in rural areas with intermittent internet access were severely affected by the government’s automated debt recovery system.

Research into algorithmic biases in automated technologies that have started permeating our physical reality beyond social media platforms constitutes an emerging field of study that is expected to shed light on the new aspects of algorithmic impact on our lives.

Works Cited

Abokhodair, Norah, Daisy Yoo, and David W. McDonald. 2015. “Dissecting a Social Botnet: Growth, Content and Influence in Twitter.” In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 839–851. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/2675133.2675208.

Ali, Muhammad, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019a. “Discrimination through Optimization: How Facebook’s Ad Delivery Can Lead to Biased Outcomes.” *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW): 199:1–199:30. https://doi.org/10.1145/3359301.

Ali, Muhammad, Piotr Sapiezynski, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019b. “Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging.” In *ArXiv*. http://arxiv.org/abs/1912.04255.

Bastos, Marco, and Dan Mercea. 2018. “The Public Accountability of Social Platforms: Lessons from a Study on Bots and Trolls in the Brexit Campaign.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376 (2128). https://doi.org/10.1098/rsta.2018.0003.

Bastos, Marco T., and Dan Mercea. 2019. “The Brexit Botnet and User-Generated Hyperpartisan News.” *Social Science Computer Review* 37 (1): 38–54. https://doi.org/10.1177/0894439317734157.

Berger, J.M., and Jonathon Morgan. 2015. *The ISIS Twitter Census: Defining and Describing the Population of ISIS Supporters on Twitter*. Brookings Institution. https://www.brookings.edu/research/the-isis-twitter-census-defining-and-describing-the-population-of-isis-supporters-on-twitter/.

Bessi, Alessandro, and Emilio Ferrara. 2016. “Social Bots Distort the 2016 U.S. Presidential Election Online Discussion.” *First Monday* 21 (11). https://doi.org/10.5210/fm.v21i11.7090.

Boshmaf, Yazan, Ildar Muslukhov, Konstantin Beznosov, and Matei Ripeanu. 2011. “The Socialbot Network: When Bots Socialize for Fame and Money.” In *Proceedings of the 27th Annual Computer Security Applications Conference*, 93–102. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/2076732.2076746.

Bot Repository. 2020. “Bot Repository.”

https://botometer.iuni.iu.edu/bot-repository/index.html.

Chavoshi, Nikan, Hossein Hamooni, and Abdullah Mueen. 2016. "DeBot: Twitter Bot Detection via Warped Correlation." In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, 817–22. https://doi.org/10.1109/ICDM.2016.0096.

Chu, Zi, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. 2010. "Who Is Tweeting on Twitter: Human, Bot, or Cyborg?" In *Proceedings of the 26th Annual Computer Security Applications Conference*, 21–30. ACSAC '10. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/1920261.1920265.

Confessore, Nicholas. 2018. "Cambridge Analytica and Facebook: The Scandal and the Fallout So Far." *New York Times*, April 4, 2018. https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html.

Cresci, Stefano. 2020. "Detecting Malicious Social Bots: Story of a Never-Ending Clash." In *Disinformation in Open Online Media*, edited by Christian Grimme, Mike Preuss, Frank W. Takes, and Annie Waldherr, 77–88. Lecture Notes in Computer Science. Springer International Publishing. https://doi.org/10.1007/978-3-030-39627-5_7.

Cresci, Stefano, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi. 2016. "DNA-Inspired Online Behavioral Modeling and Its Application to Spambot Detection." *IEEE Intelligent Systems* 31 (5): 58–64. https://doi.org/10.1109/MIS.2016.29.

———. 2017. "The Paradigm-Shift of Social Spambots: Evidence, Theories, and Tools for the Arms Race." In *Proceedings of the 26th International Conference on World Wide Web Companion*, 963–972. WWW '17 Companion. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3041021.3055135.

Davis, Clayton Allen, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2016. "BotOrNot: A System to Evaluate Social Bots." In *Proceedings of the 25th International Conference Companion on World Wide Web*, 273–274. WWW '16 Companion. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/2872518.2889302.

Diakopoulos, Nicholas. 2019. *Automating the News: How Algorithms Are Rewriting the Media*. Harvard University Press.

Dorri, Ali, Mahdi Abadi, and Mahila Dadfarnia. 2018. "SocialBotHunter: Botnet Detection in Twitter-Like Social Networking Services Using Semi-Supervised Collective Classification." In *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl*

Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), 496–503. https://doi.org/10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00097.

Edwards, Chad, Autumn Edwards, Patric R. Spence, and Ashleigh K. Shelton. 2014. “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 (April): 372–76. https://doi.org/10.1016/j.chb.2013.08.013.

Egele, Manuel, Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna. 2017. “Towards Detecting Compromised Accounts on Social Networks.” *IEEE Transactions on Dependable and Secure Computing* 14 (4): 447–60. https://doi.org/10.1109/TDSC.2015.2479616.

Elmas, Tuğrulcan, Rebekah Overdorf, Ahmed Furkan Ozkalay, and Karl Aberer. 2019. “Lateral Astroturfing Attacks on Twitter Trending Topics.” In *ResearchGate*. https://www.researchgate.net/publication/336638958_Lateral_Astroturfing_Attacks_on_Twitter_Trending_Topics.

Forelle, Michelle, Phil Howard, Andrés Monroy-Hernández, and Saiph Savage. 2015. “Political Bots and the Manipulation of Public Opinion in Venezuela.” In SSRN. https://ssrn.com/abstract=2635800.

Freelon, Deen. 2018. “Computational Research in the Post-API Age.” *Political Communication* 35 (4): 665–68. https://doi.org/10.1080/10584609.2018.1477506.

Freitas, Carlos, Fabricio Benevenuto, Saptarshi Ghosh, and Adriano Veloso. 2015. “Reverse Engineering Socialbot Infiltration Strategies in Twitter.” In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 25–32. https://doi.org/10.1145/2808797.2809292.

Gao, Hongyu, Yan Chen, Kathy Lee, Diana Palsetia, and Alok Choudhary. 2012. “Towards Online Spam Filtering in Social Networks.” Paper presented at the 19th Annual Network & Distributed System Security Symposium, San Diego, CA,

https://www.ndss-symposium.org/wp-content/uploads/2017/09/02_3.pdf.

Gao, Hongyu, Jun Hu, Christo Wilson, Zhichun Li, Yan Chen, and Ben Y. Zhao. 2010. “Detecting and Characterizing Social Spam Campaigns.” In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, 35–47. IMC ’10. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/1879141.1879147.

Giglietto, Fabio, Nicola Righetti, Luca Rossi, and Giada Marino. 2020. “It Takes a Village to Manipulate the Media: Coordinated Link Sharing Behavior during 2018 and 2019 Italian Elections.” *Information, Communication & Society* 23 (6): 867–91. https://doi.org/10.1080/1369118X.2020.1739732.

Graham, Timothy, and Robert Ackland. 2016. “Do Socialbots Dream of Popping the Filter Bubble? The Role of Socialbots in Promoting Deliberative Democracy in Social Media.” In *Socialbots and Their Friends*, edited by Robert W. Gehl and Maria Bakardjieva, 203–22. Taylor & Francis. https://doi.org/10.4324/9781315637228-18.

Grimme, Christian, Mike Preuss, Lena Adam, and Heike Trautmann. 2017. “Social Bots: Human-Like by Means of Human Control?” *Big Data* 5 (4): 279–93. https://doi.org/10.1089/big.2017.0044.

Howard, Philip N., and Bence Kollanyi. 2016. “Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum.” SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2798311.

Keller, Tobias R., and Ulrike Klinger. 2019. “Social Bots in Election Campaigns: Theoretical, Empirical, and Methodological Implications.” *Political Communication* 36 (1): 171–89. https://doi.org/10.1080/10584609.2018.1526238.

Khalil, Hunia, Muhammad U. S Khan, and Mazhar Ali. 2020. “Feature Selection for Unsupervised Bot Detection.” In *2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 1–7. https://doi.org/10.1109/iCoMET48670.2020.9074131.

Kitzie, Vanessa L., Ehsan Mohammadi, and Amir Karami. 2018. “‘Life Never Matters in the DEMOCRATS MIND’: Examining Strategies of Retweeted Social Bots During a Mass Shooting Event.” In *ArXiv*. http://arxiv.org/abs/1808.09325v1.

Kollanyi, Bence, Philip N. Howard, and Samuel C. Woolley. 2016. “Bots and Automation over Twitter during the Second U.S. Presidential Debate.” Oxford Internet Institute. https://comprop.oii.ox.ac.uk/research/bots-and-automation-over-twitter-during-the-second-u-s-presidential-debate/.

Lima, Lanna, Vasco Furtado, Elizabeth Furtado, and Virgilio Almeida. 2019. “Empirical Analysis of Bias in Voice-Based Personal Assistants.” In *Companion Proceedings of The 2019 World Wide Web Conference*, 533–38. WWW ’19. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3308560.3317597.

Lokot, Tetyana, and Nicholas Diakopoulos. 2016. “News Bots – Automating News and Information Dissemination on Twitter.” *Digital Journalism* 4 (6): 682–99. https://doi.org/10.1080/21670811.2015.1081822.

Luceri, Luca, Ashok Deb, Silvia Giordano, and Emilio Ferrara. 2019. “Evolution of Bot and Human Behavior during Elections.” *First Monday* 24 (9). https://doi.org/10.5210/fm.v24i9.10213.

McNichol, Tom. 2004. “Your Message Here.” *The New York Times*, January 22, 2004, sec. Technology. https://www.nytimes.com/2004/01/22/technology/your-message-here.html.

Metaxas, Panagiotis Takis, and Eni Mustafaraj. 2010. “From Obscurity to Prominence in Minutes: Political Speech and Real-Time Search.” Paper presented at WebSci10: Extending the Frontiers of Society On-Line, Raleigh, NC, http://cs.wellesley.edu/~pmetaxas/Metaxas-Obscurity-to-prominence.pdf.

Nimmo, Ben. 2017. “#BotSpot: Twelve Ways to Spot a Bot.” *Medium* (blog). September 15, 2017. https://medium.com/dfrlab/botspot-twelve-ways-to-spot-a-bot-aedc7d9c110c.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press. https://nyupress.org/9781479837243/algorithms-of-oppression/.

Oentaryo, Richard J., Arinto Murdopo, Philips K. Prasetyo, and Ee-Peng Lim. 2016. “On Profiling Bots in Social Media.” In *Social Informatics*, edited by Emma Spiro and Yong-Yeol Ahn, 92–109. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-47880-7_6.

Orabi, Mariam, Djedjiga Mouheb, Zaher Al Aghbari, and Ibrahim Kamel. 2020. “Detection of Bots in Social Media: A Systematic Review.” *Information Processing & Management* 57 (4): 102250. https://doi.org/10.1016/j.ipm.2020.102250.

Park, Sora, and Justine Humphry. 2019. “Exclusion by Design: Intersections of Social, Digital and Data Exclusion.” *Information, Communication & Society* 22 (7): 934–53. https://doi.org/10.1080/1369118X.2019.1606266.

Rauchfleisch, Adrian, and Jonas Kaiser. 2020. “The False Positive Problem of Automatic Bot Detection in Social Science Research.” SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.3565233.

Roberts, Siobhan. 2020. “Who’s a Bot? Who’s Not?” *New York Times*, June 16, 2020. https://www.nytimes.com/2020/06/16/science/social-media-bots-kazemi.html.

Sanovich, Sergey. 2017. “Computational Propaganda in Russia: The Origins of Digital Misinformation.” Computational Propaganda Research Project Report 2017.3. Oxford, UK: Oxford Internet Institute. https://www.oii.ox.ac.uk/blog/computational-propaganda-in-russia-the-origins-of-digital-misinformation/.

Sanovich, Sergey, Denis Stukal, and Joshua A. Tucker. 2018. “Turning the Virtual Tables: Government Strategies for Addressing Online Opposition with an Application to Russia.” *Comparative Politics* 50 (3): 435–82. https://doi.org/info:doi/10.5129/001041518822704890.

Santia, Giovanni C., Munif Ishad Mujib, and Jake Ryland Williams. 2019. “Detecting Social Bots on Facebook in an Information Veracity Context.” In *Proceedings of the International AAAI Conference on Web and Social Media*, 13:463–72. https://www.aaai.org/ojs/index.php/ICWSM/article/view/3244.

Savage, Saiph, Andres Monroy-Hernandez, and Tobias Höllerer. 2016. “Botivist: Calling Volunteers to Action Using Online Bots.” In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing – CSCW ’16*, 811–20. San Francisco, CA: ACM Press. https://doi.org/10.1145/2818048.2819985.

Savvopoulos, Alkiviadis, Pantelis Vikatos, and Fabricio Benevenuto. 2018. “Socialbots’ First Words: Can Automatic Chatting Improve Influence in Twitter?” In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 190–93. Barcelona: IEEE. https://doi.org/10.1109/ASONAM.2018.8508786.

Shao, Chengcheng, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. “The Spread of Low-Credibility Content by Social Bots.” *Nature Communications* 9 (1). https://doi.org/10.1038/s41467-018-06930-7.

Shi, Peining, Zhiyong Zhang, and Kim-Kwang Raymond Choo. 2019. “Detecting Malicious Social Bots Based on Clickstream Sequences.” *IEEE Access* 7: 28855–62. https://doi.org/10.1109/ACCESS.2019.2901864.

Statista. 2020. “Russia: Social Media Audience by Platform 2019.” Statista. 2020. https://www.statista.com/statistics/1110977/russia-social-media-audience-by-platform/.

Stieglitz, Stefan, Florian Brachten, Björn Ross, and Anna Jung. 2017. “Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts.” *ACIS 2017 Proceedings*, January. https://aisel.aisnet.org/acis2017/89.

Stucke, Maurice E., and Ariel Ezrachi. 2017. “How Digital Assistants Can Harm Our Economy, Privacy, and Democracy.” SSRN Scholarly Paper. Rochester, NY: Social Science

Research Network. https://doi.org/10.2139/ssrn.2957960.

Stukal, Denis, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker. 2017. “Detecting Bots on Russian Political Twitter.” *Big Data* 5 (4): 310–24. https://doi.org/10.1089/big.2017.0038.

Stukal, Denis, Sergey Sanovich, Richard Bonneau, and Joshua A Tucker. n.d. “Bots for Autocrats: How Pro-Government Bots Fight Opposition in Russia.” Working paper. http://www.denisstukal.com/uploads/8/4/7/0/84708866/stukal_et_al__2020__bots_for_autocrats.pdf.

Stukal, Denis, Sergey Sanovich, Joshua A. Tucker, and Richard Bonneau. 2019. “For Whom the Bot Tolls: A Neural Networks Approach to Measuring Political Orientation of Twitter Bots in Russia.” *SAGE Open* 9 (2). https://doi.org/10.1177/2158244019827715.

Takacs, Richard, and Ian McCulloh. 2019. “Dormant Bots in Social Media: Twitter and the 2018 U.S. Senate Election.” In *Proceedings of 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 5. Vancouver, Canada. https://doi.org/10.1145/3341161.3343852.

Uyheng, Joshua, and Kathleen M. Carley. 2019. “Characterizing Bot Networks on Twitter: An Empirical Analysis of Contentious Issues in the Asia-Pacific.” In *Social, Cultural, and Behavioral Modeling*, edited by Robert Thomson, Halil Bisgin, Christopher Dancy, and Ayaz Hyder, 153–62. Lecture Notes in Computer Science. Springer International Publishing. https://doi.org/10.1007/978-3-030-21741-9_16.

Varol, Onur, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. 2017. “Online Human-Bot Interactions: Detection, Estimation, and Characterization.” In *Eleventh International AAAI Conference on Web and Social Media*. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. “The Spread of True and False News Online.” *Science* 359 (6380): 1146–51. https://doi.org/10.1126/science.aap9559.

Wakefield, Jane. 2016. “Microsoft Chatbot Is Taught to Swear on Twitter.” BBC News, March 24, 2016. https://www.bbc.com/news/technology-35890188.

Woolley, Samuel C. 2016. “Automating Power: Social Bot Interference in Global Politics.” *First Monday* 21 (4). https://doi.org/10.5210/fm.v21i4.6161.

Wu, Wei, Jaime Alvarez, Chengcheng Liu, and Hung-Min Sun. 2018. “Bot Detection Using Unsupervised Machine Learning.” *Microsystem Technologies* 24 (1): 209–17.

https://doi.org/10.1007/s00542-016-3237-0.

Yang, Kai-Cheng, Onur Varol, Clayton A. Davis, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2019. “Arming the Public with Artificial Intelligence to Counter Social Bots.” *Human Behavior and Emerging Technologies* 1 (1): 48–61. https://doi.org/10.1002/hbe2.115.

Yasu. 2019. “Top 10 Social Media APIs: Twitter, Facebook, Instagram, and Many More.” *Medium* (blog). February 7, 2019. https://medium.com/rakuten-rapidapi/top-10-social-media-apis-twitter-facebook-instagram-and-many-more-5c13262c61fe.

Zeller, Tom Jr. 2006. “Gaming the Search Engine, in a Political Season.” *New York Times*, November 6, 2006. https://www.nytimes.com/2006/11/06/business/media/06link.html.