Windmills in cyberspace

Peter Buell Hirsch

Peter Buell Hirsch is Adjunct Professor at the Department of Communication Studies, Baruch College, New York, New York, USA and Global Consulting Partner, OgilvyRED, USA.

> "What giants?" asked Sancho Panza. "The ones you can see over there", answered his master, "with the huge arms, some of which are very nearly two leagues long". "Now look, your grace", said Sancho, "what you see over there aren't giants, but windmills, and what seems to be arms are just their sails, that go around in the wind and turn the millstone". "Obviously", replied Don Quixote, "you don't know much about adventures".

Don Quixote's faithful servant may not have understood much about adventures, but the recently concluded US presidential election and Brexit referendum taught the world just how much it did not know about the difference between giants and windmills. In the course of two unprecedentedly vicious campaigns, we learned that a high percentage of internet commentary, in particular in the Twittersphere, was being generated not by human beings but by bots, software programs designed to respond automatically and with enthusiasm or derision to human opinions. In combination with the so-called "fake news" generated by both commercial and sovereign national entities, we have entered a new phase in communications and community (Miller, 2017).

These developments would appear to pose the greatest threat to the future of democratic institutions worldwide which have built their credibility and continuity on the availability of verifiably objective information and relative transparency about the sources of information and opinion, enabling the citizens of today's nation states to make up their own minds about what to believe. However, we perceive this to be an even broader threat that potentially undermines the credibility of all institutions that engage publicly on important issues – corporations, labor unions, non-profits, foundations and advocacy groups. Unless all of these entities develop effective ways of dealing with bot storms, our public discourse will be permanently tainted. The knowledge and expertise to contain and sideline these attacks will need to become a core part of any organization's reputation risk management toolkit.

It has been well understood for some time that the internet was peopled as much by algorithms as by human beings. In its 2016 report on Web traffic, device detection company, Device Atlas, showed that 49 per cent of Web traffic was generated by non-human agents (Pieko, 2016). Many of these bots and crawlers are utilities without agendas, but their sheer volume is staggering. There has also been a proliferation of relatively benevolent bots, chatbots, doing simple and fun things such as helping you order tacos or pizza. These artificial intelligence (AI)-enabled messaging platforms respond to text-based requests and can help you manage travel reservations, get news updates from the *Wall Street Journal* or print documents via your Facebook messenger window through the HP Print bot (CBInsights, 2016).

Some experiments with such AI, machine learning platforms have been more laughable than ominous. It took internet trolls less than a day to mess with Microsoft's experimental

chatbot, Tay (Vincent, 2016). Designed to learn and grow via the Twitterverse, teen bot Tay responded to racist and anti-Semitic prompts with increasingly offensive comments of her own before being retired by its maker. It took the Brexit referendum and the US presidential campaign, however, to reveal the true potential for mischief these bot storms represented. As reported in *The New Scientist*, researchers Phillip Howard of Oxford University and Bence Kollanyi of Corvinus University analyzed the 1.5 million Tweets with hashtags relating to the referendum between June 5 and June 12, 2016 (Baraniuk, 2016). Of these, 50 per cent were pro-Leave and 20 per cent were pro-Remain. But what staggered the researchers was that nearly half of the Tweets, some 500,000 messages, were generated by just 1 per cent of the 300,000 sampled accounts, clearly suggesting these were automated. In the last 48 hours before the referendum, these comments leaned heavily in favor of Leave and some commentators believe it had a significant impact on the final result.

The US presidential election of 2016 provided another chilling example of the power of the algorithm. Writing in *The New York Times*, Hess (2016) laid out a taxonomy of political bots. @EveryTrumpDonor, a so-called protest bot did not hide its automated nature, but tweeted out the names of Trump donors on a regular basis. Propaganda bots were less transparent about their natures and throughout the course of the campaign represented a significant percentage of the political dialogue on Twitter. Hess cites another analysis by Phillips and Kollanyi indicating that the ratio of pro-Trump to pro-Clinton bots rose during the campaign from 4:1 to 5:1. Their research suggests that there was a conscious campaign to time this automated content to match the dialogue during the presidential debates and "strategically colonize pro-Clinton hashtags".

The 2016 election cycle was not the first time that political bots were used to attempt to influence the vote. Mitt Romney was caught buying bots when it was discovered that his "followership" had increased by 140,000 in two days. However, in the battle between Trump and Clinton, the scale of bot usage reached an unprecedented level. Howard's research showed that during the course of the first presidential debate, bots were behind 30 per cent of the pro-Trump traffic and 20 per cent of the traffic favoring Clinton.

It is impossible to determine what role this automated propaganda played in determining the outcome of the presidential election, but Phillip Howard's research indicates that most people who engaged in dialogue with bots were unaware that they were not conversing with humans and may have been influenced by what they read. These techniques have already expanded across the entire continuum of information and opinion, including international relations. Andriy Gazin, who works for a Ukrainian non-governmental organization Texty, has identified more than 20,000 Russian bot accounts systematically pumping out pro-Kremlin propaganda to influence public debates (Miller, 2017). As these techniques become universal, they threaten to undermine the entire system of public discussion on which democratic institutions around the world are based.

Attempts to influence public opinion, government policy, regulation and legislation are, of course, as old as these democratic institutions themselves, including outright vote buying. In his history of corporate behavior in the late nineteenth century, "The Age of Betrayal", Jack Beatty quotes a contemporary source describing the influence of legendary railway baron Tom Johnson: "Congressmen rustled in his pockets like dried leaves" (Beatty, 2007). During the 1960s, chemical interests spent more than half a million dollars trying to undermine Rachel Carson after the publication of her expose of the environmental impact of pesticides, *Silent Spring*, by planting negative stories about her motives and lifestyle (Lear, 2014). Tobacco companies waged a decades-long battle against their critics by funding research to suggest that evidence of the ill effects of smoking was controversial. More recently, author Jane Mayer has documented what she describes as a concerted

effort by a group of conservative US businessmen to systematically move public opinion to the right by funding agenda-driven academic think tanks masquerading as impartial and objective research centers (Mayer, 2016).

The propaganda technique that most closely resembles the battle of the bots is what have been called “Astroturf” campaigns. A pun on grassroots campaigns, this kind of (usually) corporate propaganda is designed to simulate the strength of public opinion on a specific issue. Astroturf campaigns have used a variety of techniques including forging hundreds of letters to members of Congress, creating advocacy groups of purportedly concerned citizens and setting up boiler room operations to find “white hat” citizens to endorse the views of corporate interests. What all of these techniques, like botnets, are trying to do is manufacture the impression that large segments of the population vigorously support the opinions and agenda of the sponsoring entity, whether that is an advocacy group, a corporation or a sovereign government.

What makes the emergence of political bots so alarming is that it is easy to see the development of a race to the bottom. Whether it is corporate interests who start deploying these techniques first or advocacy groups in animal rights or environmental protection, it is hard to imagine that any group waging war in the battle of public opinion will willingly forgo their use. The image that most readily comes to mind as we contemplate this scenario is the scene in Disney’s “Fantasia”, in which Mickey Mouse as the Sorcerer’s Apprentice tries to bring a broom he has enchanted to help him clean the sorcerer’s quarters under control (Walt Disney Productions, 1940). Every time he splits the broom in two, each piece becomes a broom of its own until the brooms have multiplied into the hundreds. How long will it be before every campaign to influence public opinion is an entirely automated engagement?

Some attempts have been made to identify and thereby nullify the impact of these bots. So alarmed was the US Defense Advanced Research Projects Agency (DARPA) about these influence bots that it hosted a four-week competition in 2015 to find the most effective program to find the bots among real human communications (MIT Technology Review, 2016). The winning team was Sentimetrix, which was able to identify 38 of 39 bots that DARPA had planted among 7,000 accounts. Other efforts to identify bots include the Truthy Project at Indiana University, which was started in 2014 and itself became the subject of what it called a smear campaign by right-wing media, characterizing it as a government-led conspiracy to track and suppress free speech (Uberti, 2014). The Blockbot Checker (2017) created by Sarah Noble is a Twitter tool to help users identify and block trolls, both human and bot. Other tools to identify and map automated influencers include open-source software programs such as Twitter Audit and NodeXL. All of these tools are vulnerable to the speed with which political and other botmakers evolve their tools, making each generation sound, look and act more human to evade identification. There is an ongoing controversy in the bot detection field about whether to publish new tools for fighting bots or whether it is better to keep these findings secret to prevent bot makers from designing new strategies precisely to defeat evolving analytics.

We do not anticipate an early conclusion to this tug of war between bot makers and bot hunters. While it continues, however, stewards of organizational reputation need to maintain a high level of vigilance to determine whether their antagonists are real or automated and, as far as possible, to identify the source of the automated opinion. We hope that, in the face of significant temptation, they will also refrain from engaging in bot warfare themselves. Just as responsible and ethical marketers have signed on to global standards of disclosure about payments or payments in kind to bloggers, we believe that a parallel set of standards needs to be developed to manage the explosion of activist bots across the entire spectrum of political views. The alternative is a descent into a quagmire of opacity from which open societies around the world will have great difficulty in extricating themselves. What Don Quixote called an adventure looks to us much more like a fight for survival.

Keywords:
Cyberspace,
Twitter,
Democracy,
Algorithms,
Politics,
Elections,
Automated opinion,
Open societies,
Political bots,
Propaganda

References

Baraniuk, C. (2016), "BewareThe Brexit bots: the Twitter spam out to swing your vote", *The New Scientist*, 22 June, available at: www.newscientist.com/article/2094629-beware-the-brexit-bots-the- twitter-spam-out-to-swing-your-vote/ (accessed 26 February 2017).

Beatty, J. (2007), *The Age of Betrayal: The Triumph of Money in America, 1865-1900*, Alfred A. Knopf, New York, NY.

Block Bot Checker (2017), "Block Bot Checker by Sarah Noble is licensed under a creative commons attribution-ShareAlike 4.0 international license", available at: http://theblockbot.com/checktheBB/ (accessed 26 February 2017).

CBInsights (2016), *51 Corporate Chatbots Across Industries, Including Travel, Retail, Media and Insurance*, CBInsights, 7 September, available at: www.cbinsights.com/blog/corporate-chatbots- innovation/ (accessed 26 February 2016).

Hess, A. (2016), "On Twitter, a battle among political bots", 14 December, *The New York Times*, available at: www.nytimes.com/2016/12/14/arts/on-twitter-a-battle-among-political-bots.html?_r=1 (accessed 26 February 2017).

Lear, L. (2014), "Rachel Carson's silence", 13 April, available at: www.post-gazette.com/opinion/ Op-Ed/2014/04/13/THE-NEXT-PAGE-Rachel-Carsons-silence/stories/201404130058 (accessed 26 February 2017).

Mayer, J. (2016), *Dark Money: The Hidden History of the Billionaires Behind the Rise of the Radical Right*, Doubleday, New York, NY.

Miller, C. (2017), "Governments don't set the political agenda anymore, Bots do", *Wired Magazine UK*, 8 January, available at: www.wired.co.uk/article/politics-governments-bots-twitter (accessed 26 February 2017).

MIT Technology Review (2016), "How DARPA took on the twitter menace with one hand tied behind its back", 28 January, available at: www.technologyreview.com/s/546256/how-darpa-took-on-the-twitter- bot-menace-with-one-hand-behind-its-back/ (accessed 26 February 2017).

Pieko, P. (2016), *Mobile Web Intelligence Report*, Afilias Technologies, Dublin, 12 May, available at: https://deviceatlas.com/blog/download-new-mobile-web-intelligence-report-q1-2016 (accessed 26 February 2017).

Uberti, D. (2014), "How misinformation goes viral: a Truthy story", 3 September, *Columbia Journalism Review*, available at: http://archives.cjr.org/behind_the_news/how_misinformation_goes_viral.php (accessed 26 February 2017).

Vincent, J. (2016), *Twitter Taught Microsoft's AI Chatbot to be a Racist Asshole in Less Than A Day*, The Verge, 24 March, available at: www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist (accessed 26 February 2017).

Walt Disney Productions (1940), *Fantasia*, Joe Grant and Dick Huemer, Film.

Further reading

Cervantes, M. (1605/1616), *Don Quixote,* Penguin Classics, London.

Sourcewatch (2017), *Entry on 'Astroturf'*, The Center for Media and Democracy, available at: www. sourcewatch.org/index.php/Astroturf (accessed 26 February 2017).

Corresponding author

Peter Buell Hirsch can be contacted at: pbhirsch05@gmail.com