



The Ethics of Automated Behavioral Microtargeting

Dennis G Wilson (IRIT, University of Toulouse; dennis.wilson@irit.fr)

DOI: [10.1145/3137574.3139451](https://doi.org/10.1145/3137574.3139451)

One day AM woke up and knew who he was, and he linked himself, and he began feeding all the killing data, until everyone was dead, except for the five of us, and AM brought us down here.

I was the only one still sane and whole. Really! AM had not tampered with my mind. Not at all.

I Have No Mouth and I Must Scream Ellison (1967)

AI is undeniably powerful in its modern form. It has surpassed human performance in board games, trivia game shows, and even recognizing other humans' handwriting. Soon, it will be evident how much better at driving it is, and then maybe at all forms of navigation. In its wake, we are left to ponder the ethical and social implications of the tools we have created. For a technology that began development, arguably, over 60 years ago, we are woefully unprepared when it comes to ethical, social, and regulatory understanding of AI, and less so concerning precedent.

How should the work a robot does, especially in the case of direct human job replacement, be taxed? How should AI contribute fiscally to society? Are they complicit in a social contract, and are there basic rights that extend to non-human intelligences? Who should be held accountable for the actions of an AI, such as the operator of a self-driving car? How can AI's power be equally distributed across society, to ensure that all benefit and that it isn't used to disadvantage select groups? These and more are now capturing the attention of great thinkers from prestigious universities (Stone et al., 2016) to the White House (Executive Office of the President & Technology Council, 2016).

However, when tasked with finding the most pressing issues related to AI, we must look to the present and to the impact AI has already made. The self driving car fatality count stands at one, which is unfortunate but far from pressing. High frequency trading, some-

times scripted responses to financial cues, sometimes real AI decisions about stocks made in fractions of seconds, has earned the ire of governments worldwide for creating volatile markets with flash crashes and deceptive upswings. This threat is in the process of mitigation, the world now more wary of algorithmic decisions.

However, as politics have recently turned the world upside, with a major force for globalization losing a key member, and a US President who lost the popular vote by almost 3 million, it seems pertinent to investigate the role of AI in politics. Issues abound in this field as they do in others, although separating the symptoms of AI from those of malevolent actors and competing political factions is a daunting task.

AI, by its definition, enables us. It is a tool. The dystopian concerns of a hate-filled machine manipulating and torturing humans are nowhere near our reality. However, as a powerful and novel tool, the ways in which it enables us must be considered. By examining the use of AI in political campaigning, it is evident that AI can realize undesired potential. Specifically, AI can be used to manipulate and suppress human ideas. It can facilitate the formation of ideological barriers that serve to divide people. It can enable the concerted efforts of few to disrupt the marketplace of ideas. These are the most pressing issues related to AI technologies, and we must identify and address them fully.

The Personal Web

Personalized content recommendation has long been a hallmark of AI success. The Netflix Prize, a competition for predicting user ratings of films, was started in 2006 as an effort to increase the quality of film recommendations. Based solely on user ratings of previously watched films, the competitors devised algorithms to accurately predict what rating a user would give a new film, a metric then usable by Netflix to determine recommendation priority of this new film to the user. Re-

searchers from IBM, AT&T Labs, visionary professor Geoffrey Hinton's lab, and many others competed in this prestigious event, improving upon and showcasing the power of what was then mostly called machine learning.

More than a decade later, personalized recommendation is an increasingly normal part of the web. Advertisers have long since understood the benefits of personalizing their message and targeting individuals based on intelligent personality analysis. Google and Facebook have been leaders in this market, with advertising revenues in the billions. The advertisement software platforms from Google, AdSense and AdWords, accounted for 89% of the company's revenue in 2014. Both companies wield intelligent personality metrics to build their advertising renown. Both are now heavily investing in AI.

Google's AdSense uses the term *matched content* to describe showing advertisements to specifically profiled individuals. By their claims, matched content recommendations increase the number of pages viewed by 9% and the time spent on site by 10%. (Google, 2017a) Similarly, Facebook has *Custom Audiences* that advertisers can create for their campaign based on selected demographics. Interestingly, Facebook also allows advertisers to select target users, such as existing followers of the product's Page or previous visitors to their site, to build a *Lookalike Audience*. In their words, "A Lookalike Audience is a way to reach new people who are likely to be interested in your business because they're similar to people who already are." (Facebook, 2017b) Similarity is commonly used in AI problem formulation as it simplifies multiple problems, whether a user will be interested in specific products, to a single one.

As AI capabilities increase, the ability of these platforms to deliver very specifically personalized content increases. The capabilities of AI in media were discussed in a report from Stanford's One Hundred Year Study on AI (AI100). The positives of entertainment that is more interactive, personalized, and engaging were considered, as was the potential for media conglomerates to act as Big Brothers, controlling the ideas and online experiences to which specific individuals are exposed. "Media power-

houses," they note, "will be able to micro-analyze and micro-serve content to increasingly specialized segments of the population down to the individual." (Stone et al., 2016) These media powerhouses will be able to control, on a large scale and yet with a high level of specificity, the exposure to different products, media, and ideas.

The control of idea exposure is not the only ethical issue exacerbated by the increasingly capable personality analysis performed regularly online. The same data that determines a user's product interest can reveal private details and identifying factors. As early as 2011, there was research showing AI capable of determining the political alignment of individuals based on their Twitter data. (Conover, Goncalves, Ratkiewicz, Flammini, & Menczer, 2011). Even the seemingly innocuous Netflix Prize was dogged for years after its termination by lawsuits claiming that the users' data had violated their privacy, with researchers able to identify a number of users from the Netflix Prize datasets by cross-referencing user data from the Internet Movie Database.

In 2013, research demonstrated that it was possible to recover a large amount of personal information from Facebook Likes. In 88% of the tested cases, an AI model correctly discriminated between homosexual and heterosexual men, between African Americans and Caucasian Americans in 95% of cases, and between Democrat and Republican in 85% of cases. Age, intelligence, gender, and the personality metrics openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism (OCEAN) were also estimated from Like data with a high degree of certainty. (Kosinski, Stillwell, & Graepel, 2013) This model is now available online for public use.¹

While the ethical issue of privacy invasion in this analysis is apparent, so is the marketing potential. More recent work from the same researcher uses the information gleaned from Likes to draw users to different posts, as advertisers would. The early findings show that marketing using individual personality analysis, deemed personality targeting or behavioral microtargeting, can attract up to 63%

¹<https://appliedmagicsauce.com/demo.html>

more clicks, a clear profit for advertisers.

In the development of AI, it is often necessary to define a numeric goal for optimization. In the aforementioned Netflix prize, it was the accuracy of user movie rating predictions. Often in advertising, it is a proxy of the interest generated by the ad, using metrics such as clicks through to an advertiser's website. YouTube measures the amount of time a user watches a video, and how much that contributes to a session of watching multiple videos, to determine a video's popularity sorting. Specific videos are suggested to users based on their ability to capture the attention of the user, elongating their session and increasing their exposure to more advertisements.

Yvonne Hofstetter, a lawyer, AI expert, and the Managing Director of Teramark Technologies GmbH, writes

Even Google Search is a control strategy. When typing a keyword, a user reveals his intentions. The Google search engine, in turn, will not just present a list with best hits, but a link list that embodies the highest (financial) value rather for the company than for the user. Doing it that way, i.e. listing corporate offerings at the very top of the search results, Google controls the users next clicks. (Helbing, Frey, Gigerenzer, & Hafen, 2017)

The intent of the actor performing behavioral microtargeting comes through in this numeric goal, for which an AI can be optimized. While the ethical and regulatory issues surrounding the use of this technology for commercial purposes, such as advertising, are potentially troubling, the danger of this technology is apparent when other goals are considered. In 2016, the data firm Cambridge Analytica used behavioral microtargeting in two major US political campaigns, that of Senator Ted Cruz in the Republican primary, and of Donald Trump. These campaigns, and how they used data-driven AI to profile and persuade the public, are useful as a case study on the current issues surrounding behavioral microtargeting. The efficacy of microtargeting in the mentioned political campaigns is not the focus of this work; while Senator Cruz lost the primary despite aid from AI, and many factors contributed to President Trump's election, it

is the use of this technology that raises concerns, not its potential influence on outcome. The CEO of Cambridge Analytica, Alexander Nix, argues in favor of behavioral microtargeting:

Your behavior is driven by your personality and actually the more you can understand about peoples personality as psychological drivers, the more you can actually start to really tap in to why and how they make their decisions, Nix explained to Bloombergs Sasha Issenberg. We call this behavioral microtargeting and this is really our secret sauce, if you like. This is what were bringing to America. (Anderson & Horvath, 2017)

Cambridge Analytica informed the campaigns of individuals who matched specific psychological profiles for canvassing. It analyzed communities to determine talking points and campaign strategies for visiting candidates. "We can use hundreds or thousands of individual data points on our target audiences to understand exactly which messages are going to appeal to which audiences," Nix claimed in a lecture on the Cruz campaign. (Nix, 2016)

Developing political strategies based on citizen information such as demographics is neither novel nor ethically questionable. However, the use of AI has enabled profiling to a degree that violates citizen privacy. It is founded on a basis of data that some would argue belongs first to the citizens and only to political campaigns with explicit consent. Most importantly, though, when this form of analysis is used to deliver personalized political content, the diversity of opinions citizens are exposed to becomes artificially limited.

Facebook is not only a vehicle for advertisement. Of the 67% of American adults who use Facebook, two thirds of them, being 44% of the adult population, cite it as a part of their news sources. Facebook is not the only social media site that functions as a news source for Americans, but it is by far the largest in terms of reach. The majority of users, 64%, who get news from a social networking site rely solely on that site for their news. Most commonly, that solitary news source is Facebook. (Gottfried & Shearer, 2016)

Mark Zuckerberg is aware of the implications

of this reliance on Facebook. In a recent letter to the community of Facebook, he detailed a plan to report and remove terrorist propaganda from the site. The plan involves using AI to flag suspect content for administrative review, a system that currently generates one-third of all reports to Facebook's content review team. (Zuckerberg, 2016)

However, the site is already widely used to push political agendas. Facebook played a pivotal role in fundraising for the Trump campaign and was a main focus of their advertisement. User feedback from political ads, such as clicks or shares, informed the usage of over forty thousand different ad variants the campaign used. Many have argued that this barrage of tightly focused advertisement lead to the creation of virtual echo chambers, spaces where a limited set of ideas were constantly reinforced. (Anderson & Horvath, 2017)

The social issues at hand were captured well in an article of Scientific American:

In order for manipulation to stay unnoticed, it takes a so-called resonance effect: suggestions that are sufficiently customized to each individual. In this way, local trends are gradually reinforced by repetition, leading all the way to the "filter bubble" or "echo chamber effect": in the end, all you might get is your own opinions reflected back at you. This causes social polarization, resulting in the formation of separate groups that no longer understand each other and find themselves increasingly at conflict with one another. In this way, personalized information can unintentionally destroy social cohesion. (Helbing et al., 2017)

While geographic boundaries or social class have in the past limited the landscape of ideas available to individuals, it is surprising that this issue has resurfaced in the Age of Information. As this divide was enabled by novel technologies, among them the AI used in behavioral microtargeting, it is fitting that we evaluate the appropriate use of these technologies and propose methods to maximize their societal benefit. This will be discussed further in section 3. First, however, we will highlight the technologies used to create these echo chambers and to push specific messages.

Automated Interaction

In a lecture describing the company's approach during the Cruz campaign, Nix used the example of a private beach owner showing an intentionally misleading sign warning of shark sightings as an example of behavioral communication, the new technique that trumps older techniques of informational communication, like a sign that states that the beach is private property. (Nix, 2016) While Cambridge Analytica itself did not appear to support any intentional misleading during the campaign, it became a focal issue in a campaign based on behavioral communication.

Large-scale manipulation of public opinion and understanding is a growing ethical issue related to AI. While much of the existing threat is due simply to automation, bots that have no independent intelligence, the potential for damage is already visible. AI is poised to replace existing bots and worsen this issue if allowed. To illustrate this potential, further examples of political manipulation are shown.

Standing less popular nationally than Facebook, Twitter is used by 16% of US adults. Of those, 56% use the site as a news source. Twitter is an attractive platform for automated users, or bots, as there are accessible application programming interfaces (APIs) in multiple programming languages, and programs for tasks such as tweet repetition and automatic liking.

By simply selecting random popular words and parroting other users' tweets, one researcher's Twitter bot was able to reach influence scores close to celebrities and higher than many human users. (Messias, Schmidt, Oliveira, & Benevenuto, 2013) This bot was intended to deceive human users in to believing it was also human, and it appears to have succeeded. The difficulty of separating a bot, even a simple scripted one, from a human user on Twitter is so difficult that modern AI has been utilized to perform the task. BotOrNot² uses random decision forests, an AI classification technique, to determine if a Twitter user is a bot or not. (Davis, Varol, Ferrara, Flammini, & Menczer, 2016)

With the difficulty of discerning humans from bots on the platform, and the ease with which

²<https://botometer.iuni.iu.edu/>

bots can be created and updated, the stage is set for technical users to exert influence far beyond what their single human account could have. Bots have been shown to participate and potentially manipulate Venezuelan politics on Twitter, with nearly 10 percent of all politician retweets coming from bot-related platforms. The most active bots in this study were those used by Venezuelas radical opposition. (Forelle, Howard, Monroy-Hernandez, & Savage, 2015)

Political bots were also highly active during the 2016 US election, perhaps unprecedentedly so. Highly automated pro-Trump activity increased until the final results, outnumbering pro-Clinton bot activity 5:1. (Kollanyi, Howard, & Woolley, 2016) One group, using BotOrNot, found that roughly 400,000 bots engaged in political discussion about the Presidential election, responsible for roughly 3.8 million tweets, about one-fifth of the entire conversation. (Bessi & Ferrara, 2016)

The AI100 report details one of the ethical issues of this trend:

AI technologies are already being used by political actors in gerrymandering and targeted robocalls designed to suppress votes, and on social media platforms in the form of bots. They can enable coordinated protest as well as the ability to predict protests, and promote greater transparency in politics by more accurately pinpointing who said what, when. Thus, administrative and regulatory laws regarding AI can be designed to promote greater democratic participation or, if ill-conceived, to reduce it. (Stone et al., 2016)

However, there is a specific danger in the combination of behavioral microtargeting and the use of bots: users can be targeted by other seemingly human users for coercion and idea suppression. In a psychology study, anonymous users were more likely to make riskier gambles if they knew other users had chosen to do so, even if the other users were anonymous strangers. (Chung, Christopoulos, King-Casas, Ball, & Chiu, 2015) The reward mechanism of targeted users can be manipulated by artificially inflating retweets or likes of their posts, which will then inform their future behavior, and artificially raise their standing in a

social network with other humans. Humans use social information to modify their behavior and make decisions, and when that social information is easily manipulated, human decision can also be manipulated. (Bhanji & Delgado, 2014)

Networks can be created with a high density of bots, or to connect individuals who have similar personality traits seen by a campaign as exploitable. Users already tend to aggregate around common interests in a phenomena known as homophily, but this can be enhanced with automated users that link previously unknown users together via follows and retweets. Echo chambers can be created with a mix of bots and human users, unknowingly selected together. Beyond limiting their exposure to ideas, this type of organization has been shown to facilitate rumor spreading (Aiello et al., 2012). Polarization is another factor in misinformation spreading (Anagnostopoulos et al., 2014), meaning a campaign with knowledge of polarized individuals, based on behavioral analysis, could facilitate rumor spreading by linking these individuals with automated accounts that reinforce desired rumors.

This is not a new phenomena. The technology behind these bots is far from sophisticated, and more technical AI has been used to study it. Truthy, an earlier project of the same team that created BotOrNot, used SVM and AdaBoost to determine how factual a trending idea was. (Ratkiewicz et al., n.d.) In the course of this study, they noted the alarming ease with which false information could be encouraged to spread widely on Twitter.

Even while presenting honest content from human users, the combination of automation with behavioral microtargeting has troubling consequences, and it is not restricted to Twitter. The platform's automation accessibility facilitates it, but these tactics are possible on other platforms as well. While Facebook strictly verifies the identities of its users, posts can be automated to maximally convey their message. The phrasing and presentation of a post, regardless of its content, has been shown to affect its potential for spreading. (Alhabash et al., 2013)

Facebook's automated rules allow advertising campaigns to create rules that modify

their advertisement based on assigned conditions. (Facebook, 2017a) While this can be as simple as stopping an ad if it isn't performing well, Cambridge Analytica appears to have done much more complex automated advertisement administration. Based on the ads selected by users, content was added to their feed in posts personalized for them, determined by their behavior profile. Automatically selecting from the thousands of ad variants available, these rules targeted specific individuals and seem to have created the same echo chambers as described in Twitter. (Grassegger & Krogerus, 2017) Even without bots, these tightly networked groups are still restricted from exposure to a diversity of opinions and are susceptible to the spread of false information. (Anagnostopoulos et al., 2014)

Twitter bots and Facebook ad manipulation are not using state of the art AI and natural language processing, for the most part. Some bots aren't even fully artificial. Users like Daniel Sobieski have automated programs that tweet more than 1,000 times a day using schedulers that work through a queue of their previously written tweets. (*As a conservative Twitter user sleeps, his account is hard at work*, 2017) While scripts are far from AI, their use informs a discussion on human machine interaction that is vital as AI capabilities increase. Microsoft's disastrous attempt at a teenage Twitter chatbot, Tay, must be given credit for creating seemingly human responses, albeit tainted by the preferences of users that hijacked the experiment. As bots on social media gain increased social capability, and as artificially generated content further resembles human generated content, our interactions on social media must be well informed.

The ethical issue at hand is therefore the large scale manipulation of human ideas, opinions, and agency using AI. The same technologies have created anew the social issue of ideologically isolated communities, now manufactured artificially to reduce opinion diversity and facilitate misinformation. The first technology behind these issues is the powerful personality analysis now possible due to greater data availability and more accurate AI. The second is automation on social media platforms, which, for now, is mostly rudimentary scripting and does not resemble intelligent decision

making.

For this reason, there is currently a human barrier between the two technologies. Personalities are analyzed using AI, and then a human actor uses the information to decide and design automated strategies on social media. Cambridge Analytica is an exception to this, as their posts seem to be selected from a large pool based on input from the analytical technology, but this selection is also rudimentary compared to state of the art generative AI.

When this gap between the technologies is closed by AI, and fully autonomous processes go from personality profiling to specialized content delivery and generation, we must have well established guidelines for the ethics of such systems.

Propositions

To address these concerns, proposed directions for the government, industry, and public organizations and academia are examined in the next three sections. These issues can not be resolved by any one sector alone. Rather, there must be a coordinated effort of those that work with AI in all three sectors. While there are many other strides that could be taken to address the issues related to AI, the initiatives proposed below are those best suited to combat the pressing issues raised in this article.

Government

The current drive of AI is data. The companies that own the most data have been making the greatest strides in AI, and this data is largely generated by their users. The European Union has been a powerful force in countering corporate data ownership by declaring citizen's rights over their data. Governments must continue to enforce and expand this type of law. The right of a person to all of the data associated with their identity, and the agency of each person to control that data, must be respected.

Second to that is the funding of AI initiatives. While there is a surplus of funding for the development of AI, it mostly fits the individual desires of the company using that AI. Quality AI research that doesn't appear to have

a corporate application should be supported by the government. Furthermore, research into the study of AI itself and how it affects society must be done without corporate influence. The Obama Administration was very supportive of increasing AI research funding. ([Executive Office of the President & Technology Council, 2016](#))

Lastly, the government must apply strict advertisement laws to new forms of marketing as AI continues to change marketing. By requiring that advertisements are clearly marked as such, the issue of unaware manipulation becomes much less concerning.

Industry

Industries can not be expected to sacrifice potential profits by not utilizing the powerful user analysis enabled by AI. Nor can they be expected to invest their resources in endeavors that do not benefit them in return. However, in the event of government reforms of data policy, it would be in the interest of companies to develop tools that allow users to personally perform the type of analysis being done with their data currently. For example, on Google News, specific news items are suggested based on personality. Users can disable these articles and they can modify their interests in different pre-selected categories, but they can only influence the personality metrics Google has built around them by indicating their interest for or against new articles. ([Google, 2017b](#)) This does not afford the user understanding nor control of their data.

Furthermore, industries that allow automated users to interact with human users need to make dedicated efforts to allow their human users to distinguish between actions of bots and of humans. While tools such as BotOrNot are good research efforts, they should not be necessary. An example of good policy is found in Slack, a messaging app that allows for bots and software service integration. Bot users are clearly marked, even though they come from a variety of automation tools. The distinction is important; human reaction in video games has been markedly different when players are aware that the opponent is a bot as opposed to a human. ([Smith & Delgado, 2015](#))

Lastly, media sites, including social network

sites, should be cautious about applying in-house AI to combat what they see as negative trends in their content or user interaction. The numerical optimization used in AI should be considered carefully, as overemphasis on a particular metric or disregard of another can have drastic consequences once that optimized AI is put to use. Zuckerberg's community initiative, while seemingly well-intended, comes off as naive in its understanding of both AI and democracy, as AI must enforce local Community Standards by potentially limiting content while simultaneously fostering open democratic policies. ([Zuckerberg, 2016](#))

Organizations and Academia

Organizations and academia have the greatest potential to shape the discourse around AI. A first major step in that is recognizing, as the AI100 initiative has, that what we call AI is a moving target, and is often one placed just out of reach. ([Stone et al., 2016](#)) Various forms of AI have existed and been in use for half a century, but there is a great hesitation to call something that would have been considered AI 10 years ago the same now. Furthermore, technologies that were or are outside the technical term's strict definition should be considered when discussing the impact of AI, such as the Twitter automation scripts discussed in this article.

Academics can also fill in the research gaps that industry will not, and organizations can support this effort. The BotOrNot and Truthy applications are examples of useful tools outside the corporate interest of the platform they interact with. By making these tools independently and available to the public, society is able to better understand the tool it is wielding.

Lastly, but importantly, organizations and academia must remain independent and unbiased in their evaluation of industrial and governmental use of AI. The dangers of good AI in the wrong hands have already been demonstrated, and they can come from a number of sources. Independent organizations must support academic research into the fair and appropriate use of AI in all sectors.

Conclusion

Power is in tearing human minds to pieces and putting them together again in new shapes of your own choosing.
1984 Orwell (1949)

This is not the dystopia Ellison nor Orwell imagined. Efforts are being made to check the use of private data. Social media is re-evaluating its newfound place as the news provider to many. People are learning to think critically about what they see online before believing it. Some in the private sector, like Google, have offered their research capabilities in the form of open source code and publications. All of these are marked progress against the concerns of AI manipulating and suppressing human ideas, slicing up the marketplace of ideas into small despots.

Still, there is a ways to go and basic attitudes must change. The tech mantra of “Move fast and break things” must give way to cautious, considered approaches when AI is concerned. Whether the fault of technology, and specifically AI, or not, things have broken enough already.

References

- Aiello, L. M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., & Menczer, F. (2012). Friendship prediction and homophily in social media. *ACM Transactions on the Web*, 6(2), 1–33.
- Alhabash, S., McAlister, A. R., Hagerstrom, A., Quilliam, E. T., Rifon, N. J., & Richards, J. I. (2013). Between Likes and Shares: Effects of Emotional Appeal and Virality on the Persuasiveness of Anticyberbullying Messages on Facebook. *Cyberpsychology, Behavior, and Social Networking*(3), 130201060931000.
- Anagnostopoulos, A., Bessi, A., Caldarelli, G., Del Vicario, M., Petroni, F., Scala, A., ... Quattrociocchi, W. (2014). Viral Misinformation: The Role of Homophily and Polarization. *arXiv:1411.2893*, 1–12. <http://arxiv.org/abs/1411.2893>.
- Anderson, B., & Horvath, B. (2017). *The rise of the weaponized ai propaganda machine*. <https://scout.ai/story/the-rise-of-the-weaponized-ai-propaganda-machine>.
- As a conservative twitter user sleeps, his account is hard at work. (2017). https://www.washingtonpost.com/business/economy/as-a-conservative-twitter-user-sleeps-his-account-is-hard-at-work/2017/02/05/18d5a532-df31-11e6-918c-99ede3c8cafa_story.html?utm_term=.0938a67alce5.
- Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 u.s. presidential election online discussion. *First Monday*, 21(11). <http://journals.uic.edu/ojs/index.php/fm/article/view/7090>.
- Bhanji, J. P., & Delgado, M. R. (2014). The social brain and reward: Social information processing in the human striatum. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(1), 61–73.
- Chung, D., Christopoulos, G. I., King-Casas, B., Ball, S. B., & Chiu, P. H. (2015). Social signals of safety and risk confer utility and have asymmetric effects on observers' choices. *Nature neuroscience*(6), 912–6.
- Conover, M. D., Goncalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. (2011, oct). Predicting the Political Alignment of Twitter Users. In *2011 ieee third int'l conference on privacy, security, risk and trust and 2011 ieee third int'l conference on social computing* (pp. 192–199). IEEE. <http://ieeexplore.ieee.org/document/6113114/>.
- Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). BotOrNot: A System to Evaluate Social Bots. *arXiv*, : 1602.009, 2. <http://arxiv.org/abs/1602.00975>.
- Ellison, H. (1967). *I have no mouth and i must scream*. Galaxy Publishing Corp.
- Executive Office of the President, N. S., & Technology Council, C. o. T. (2016). *Preparing for the future of artificial intelligence*. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

- Facebook. (2017a). *Automated rules*. <https://www.facebook.com/business/help/247173332297374/>.
- Facebook. (2017b). *Lookalike audiences*. <https://www.facebook.com/business/help/231114077092092>.
- Forelle, M. C., Howard, P. N., Monroy-Hernandez, A., & Savage, S. (2015). Political Bots and the Manipulation of Public Opinion in Venezuela. *SSRN Electronic Journal*, 1–8.
- Google. (2017a). *Matched content*. https://support.google.com/adsense/answer/6111336?hl=en&ref_topic=6111161.
- Google. (2017b). *Personalize your news settings*. https://support.google.com/news/answer/1146405?hl=en&ref_topic=2428815.
- Gottfried, J., & Shearer, E. (2016). *News use across social media platforms 2016*. <http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/>.
- Grassegger, H., & Krogerus, M. (2017). *The data that turned the world upside down*. <https://motherboard.vice.com/en-us/article/how-our-likes-helped-trump-win>.
- Helbing, D., Frey, B. S., Gigerenzer, G., & Hafen, E. (2017). *Will democracy survive big data and artificial intelligence?* <https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/>.
- Kollanyi, B., Howard, P. N., & Woolley, S. C. (2016). *Bots and Automation over Twitter during the First U.S Presidential Debate*.
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), 5802–5.
- Messias, J., Schmidt, L., Oliveira, R., & Benvenuto, F. (2013). You followed my bot! transforming robots into influential users in twitter. *First Monday*, 18(7).
- Nix, A. (2016). *The power of big data and psychographics*. <https://www.youtube.com/watch?v=n8Dd5aVXLcC>.
- Orwell, G. (1949). 1984. Secker and Warburg.
- Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F. (n.d.). Detecting and Tracking the Spread of Astroturf Memes in Microblog Streams. *Proceedings of the 20th International Conference Companion on World Wide Web*, 249–252.
- Smith, D. V., & Delgado, M. R. (2015). Social nudges: utility conferred from others. *Nature Neuroscience*(6), 791–792.
- Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., ... others (2016). Artificial intelligence and life in 2030. *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*.
- Zuckerberg, M. (2016). *Building global community*. <https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10103508221158471/>.



Dennis G Wilson is a PhD candidate at the Institut de Recherche en Informatique de Toulouse studying artificial neural models, after studying evolutionary computation and gene regulatory networks at the Anyscale Learning For All group at CSAIL, MIT. Their professional and personal information can be found at <https://d9w.xyz/>