

# Bad Bots: an Ethical Perspective on Influencing Discourse with Machines

2nd edition\*

Marin Beijerbacht

June 2022

*\*Edit of first version from May 2019 co-authored by Jiska Warmels*

Automation is everywhere in the 21st century, from production processes to interactions. For example, when you want to contact the customer service of a company it is likely you will have to talk to a bot first. The term bot is short for “robot” (1). However, the bots mentioned here are not physical robots, but automated programs that often run over the internet. Bots can run automatically or execute certain commands when receiving specific input. With technology developing at a tremendous speed, the use of bots, both by users and organisations, is expanding greatly in today’s society(2).

A specific type of bot is a social bot, although no agreed upon definition exists, it can generally be described as “computer-generated programs that post, tweet, or message of their own accord” (3). Some other definitions also specifying that the bot simulates how a human user would act.

Bots in the media are often portrayed with a negative undertone (4; 5). Therefore, it is not strange that people feel anxious towards bots. This feeling of anxiousness can also be referred to as ‘distrust’ in many cases. However, it is not always justified that people feel that way. Many social bots are also used for useful causes, like preventing loneliness or taking the role of personal assistant and some are used for harmless fun applications, like tweeting the weather or recommending books to read(6; 7). Yet the negative undertone when reporting on social bots is there for a reason, there are many cases where bots are used for wrong purposes.

Nowadays the acceptance of social bots is extremely relevant, as many companies are shifting from human workers to digital workers. These digital workers are, in other words, social bots. Therefore it is important to understand why a heap of the current use of social bots is often morally wrong and how bots can also be used for morally good purposes. In this paper the malicious applications of social bots will be discussed. First we will lay out the problematic uses of social bots, what they do and how they do it. Then the morality of the deployment of such bots is discussed from the philosophical viewpoints of consequentialism, utilitarianism and Kantian deontology. As mentioned before not all bots are bad, so to give a more complete oversight of social bot use a few of the good applications are discussed in the next section. The paper concludes with a discussion on what is and can be done about social bots online.

## **Misuse of Social Bots**

Social bots are active on social media platforms to insert themselves into discourse. One platform where many social bots notably reside is Twitter. In 2016 it has been estimated that at the time between 9% and 15% of active Twitter accounts were bots (8), though the current numbers are hard to estimate. When talking about the malicious application of social bots people generally refer to the influencing of public opinion and people their behaviour. Though the goals of the discourse manipulation can vary. One such goal could be popularising a brand or person by artificially boosting interaction with their content or defaming them by posting negative or hate comments(9). Another common goal is influencing the financial market. Which is done by spreading (fake) news about a company or brand or making it seem as if a stock is high in demand(10). Another goal, the one with the most media attention, is social bots influencing and manipulating politics(11; 12). Here an actor like a (foreign) government, organisation or a political party themselves can

popularise or defame a certain political party or personage. Another goal can be influencing peoples opinions on politically charged matters such as vaccination, 5G or immigration to steer public opinion(13). In all of these applications fake news and misinformation is often employed. This has become a serious problem which can even be called a threat to democracy with the way (political) conversations are hijacked by bots(8).

There are a few ways these bots insert themselves into conversations and influence the users behaviour on the platform. First of all there is smoke screening, this is a technique that uses context-related hashtags on Twitter to distract users from the actual point of the discussion. Related to this is misdirecting, where the context-related hashtags are used without referring to the debate-topic whatsoever and point the attention to irrelevant information. Lastly there is astroturfing, which is a deceptive practice where bots all present the same opinion, marketing or campaign in the guise of comments from members of the public (8; 14).

According to the book *Designing sociable robots* it is stated that an ideal social bot should be capable of communicating in such a way that users can understand it socially, relate to the bot and empathize with it (15). For social bots with a specific use case, like being a personal assistant, it can be important to incorporate these qualities. However, when applying these qualities to bots for manipulation purposes it can become problematic. On a social networks these bots frequently pose as real humans, because of which people often do not recognize whether they are dealing with a bot or a human being.

### **Ethics of Influencing Discourse**

Having discussed some of the goals of bad social bots we will now have a look at the moral and ethical implications. Taking on some of the views from normative ethics, it can be argued that the uses discussed are morally wrong. Take for example consequentialism, here the consequences of an action are the basis for moral judgement. The consequences of defamation of a person or organisation can be minimal like less business for a while, however it can also be very bad ruining the life of a person or a business. When artificially popularising a person/brand at first glance that does not seem too bad. However, the consequences of this still has bad sides. It can cause a brand that does not deliver what it promises to get popularised and competing honest brands do not get attention. Even when a good brand uses these tactics it is not fair as this creates an inequality with competing brands. Similarly, the consequences for influencing political discourse has bad consequences. As mentioned before it can be argued to be a threat for democracy. People are being influenced to have another opinion than they might actually have or act differently as they are exposed to wrong or partial information. Also functional rational debate is stopped when bots start using astroturfing or smoke screening techniques which halts or polarises the discussion, damaging the principles on which a democracy is built(8).

In practice the ones using these bad bots are organisations, small groups of people and sometimes only a single person, all with a certain agenda(16). When it comes to utilitarianism the usage of bots to influence the public opinion for example. This would not benefit the majority of the population and would therefore not be viewed as morally sound. With just one organisation or person who has control over the bots, there will be no benefitting majority. The reasons as discussed for consequentialism also make bad bots morally bad in the eyes of utilitarianism given the consequences of bad social bots (17).

The philosophy of deontology would also disapprove of such use of social bots. The action of influencing other individuals for one's self-interest is a selfish act. As Kantian deontology states: "Act only according to that maxim whereby you can, at the same time, will that it should become a universal law" (18). This implies that an action is morally good or bad based on a set of universal rules, like do not lie or cheat. When taking this stance into consideration, the use of bots for one's self-interest is a morally bad action. Influencing discourse does not bring the best consequences, does not benefit the majority of the population and is not an action that one would want to be a universal law.

Under the ethical views discussed it can be argued that the use of social bots for discourse influencing is morally not sound. This poses the question of what should be done about social bots online to which we will return. First it also needs to be considered that not all social bots are bad. Many of them are even very useful, so what about those?

## But What About Good Social Bots?

To refresh, social bots are bots that can post, tweet or message of their own accord and often imitate humans in doing so. Often the bad social bots are getting publicity, however luckily not all are bad. Social bots can be very helpful and are implemented in such ways. For example you can think of weather bots or news bots which report on matters of their domain in users' news feeds(8). There are also recommender bots that can help recommend new shows, books or recipes. Automatic responder bots are also implemented online and can be helpful to inform users of status updates or answer simple questions on products and services(19). Social bots can also be used in helping people feel less lonely and provide a listening ear so to speak(20). These are just a few examples of the good uses of social bots, though there are many more and ever increasing as technologies advance.

When looking at the ethical implications of these good uses, it can generally be said that they are good and beneficial to online users. Take, for example, recommender bots. These can help users find new and interesting media or goods to consume, which can generally be classified as a good outcome of which most people will benefit. The goals in these examples of good uses are very different from the bad social bots. Here the goal is to help anyone who interacts with the bot and improve their situation in some way. Adding to this is the fact that in these applications it is almost always communicated to the user in some form that they are interacting with a bot(21).

So the bad behaviour of some of the social bots online should not impact our use of the good and beneficial ones. However, even for the good social bots some bounds need to be set as they can still produce bad outcomes when not careful. For example news summary bots can still accidentally report fake news, chatbots for helping people can store too much personal information which can leak or any of them can have inadvertent bias(8; 22). So something needs to be done to both tackle the bad bots and regulate the good ones.

## So What To Do With Social Bots?

Social Media companies like Twitter itself try to make an effort of detecting and removing bad bots(23). There are also tools users themselves can use to check if an account is a human or bot, like Botometer<sup>1</sup>. However, detecting and removing or reporting accounts does not completely solve the problem. Combatting bad social bots is a bit of a cat and mouse game. There are ways to recognise bot accounts using things such as time signatures of posting or amount of friend interactions (24). However, with technologies in artificial intelligence improving it is becoming harder to spot accounts using the traditional methods, as the bad bots are becoming indistinguishable from humans. With improving technologies bot detection methods also improve, but not quite fast enough(25).

Another important aspect to help minimize issues with bad social bots is educating users on the matter(26). This helps them spot bots or use tools to do so and it helps spread awareness of how to interact online. Laws governing the digital domain can also help in managing bad social bots. For example, help incentivise companies to spend time on removing bad actors from their platforms and spend time on updating their methods to do so. It is important to take care when drafting such laws, as too strict a law can suppress (political) discourse or be seen as an attempt at censorship(26).

For good bots user education and regulations are also needed, to guide implementation and use. With regulation good development practices can be encouraged. Things such as examining bias or providing transparency on identity or actions of the bot could be enforced. Though it remains a challenge on how to implement such laws, more work on this is required(22). User education on how to interact with bots in turn can help in users recognising when a good bot is safe, help prevent them confusing a bad bot for a good one. Additionally, as mentioned before, more companies are shifting to digital workers. Hence, it is favourable users know how to safely and efficiently interact with these social bots(26).

There are social bots active online that have bad intentions. In seeking profit in some form for the ones employing them, they have bad consequences for others. One can even go as far and argue these bots form a threat to democracy. Though banning all bots would not do, as there are many

---

<sup>1</sup><https://botometer.osome.iu.edu/>

good social bots. With user education and regulations, it is the hope the social bot landscape can be shaped such that the democracy stays intact and the highest amount of benefit can be gained. The work on this subject is still ongoing and should be encouraged.

### Acknowledgements

This piece was first written as an assignment for philosophy of Artificial Intelligence together with Jiska Warmels in 2019. This current edited version was commissioned for the WishWillWay Foundation<sup>2</sup> by Jake Blok. It is also part of the 'Log on to X' talks of log on to the web<sup>3</sup>.

## References

- [1] P. Christensson, "Bot defenition," february 2014. <https://techterms.com>.
- [2] GVR, "Chatbot market analysis by end user, by application/business model, by type, by product landscape, by vertical, by region (north america, europe, apac, mea), and segment forecasts, 2018 - 2025," tech. rep., Grand View Research, august 2017.
- [3] M. Forelle, P. N. Howard, A. Monroy-Hernández, and S. Savage, "Political bots and the manipulation of public opinion in venezuela," *Available at SSRN 2635800*, 2015.
- [4] F. J. McEvoy, "Social bots are ruining the internet for the rest of us," december 2017. <https://venturebeat.com/2017/12/05/social-bots-are-ruining-the-internet-for-the-rest-of-us/>.
- [5] G. Stocking and N. Sumida, "Social media bots draw public's attention and concern," tech. rep., october 2018. <https://www.journalism.org/2018/10/15/social-media-bots-draw-publics-attention-and-concern/>.
- [6] S. MacPherson, "I built a personal chat bot — why i love it, and you'll love yours too," 2016. <https://chatbotsmagazine.com/i-built-a-personal-chat-bot-why-i-love-it-and-youll-love-yours-too-6c2d019f3725>.
- [7] A. Gordon, "This bot will send you free recommendations of books by black authors," 2020. <https://edition.cnn.com/2020/06/28/us/black-authors-bot-book-recommendations-trnd/index.html>.
- [8] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Commun. ACM*, vol. 59, pp. 96–104, June 2016.
- [9] R. M. Santini, D. Salles, C. E. Estrella, C. E. Barros, and D. Orofino, "Bots as online impersonators: automated manipulators and their different roles on social media," *The International Review of Information Ethics*, vol. 30, no. 1, 2021.
- [10] R. Fan, O. Talavera, and V. Tran, "Social media bots and stock markets," *European Financial Management*, vol. 26, no. 3, pp. 753–777, 2020.
- [11] J. Susskind, "Chatbots are a danger to democracy," 2018. <https://www.nytimes.com/2018/12/04/opinion/chatbots-ai-democracy-free-speech.html>.
- [12] A. Bessi and E. Ferrara, "Social bots distort the 2016 us presidential election online discussion," *First monday*, vol. 21, no. 11-7, 2016.
- [13] C. Cheng, Y. Luo, and C. Yu, "Dynamic mechanism of social bots interfering with public opinion in network," *Physica A: Statistical Mechanics and its Applications*, vol. 551, p. 124163, 2020.

---

<sup>2</sup><https://wishwillway.org>

<sup>3</sup><https://logontotheweb.com>

- [14] ASMC, “How powerful are social bots,” tech. rep., Academic Society for Management & Communication, 2018.
- [15] C. L. Breazeal, *Designing Sociable Robots*. MIT Press, 2002.
- [16] S. C. Woolley, “Automating power: Social bot interference in global politics,” *First Monday*, 2016.
- [17] A. Quinton, *Utilitarian ethics*. Springer, 1973.
- [18] I. Kant and J. Ellington, *Grounding for the Metaphysics of Morals (Third Edition): with On a Supposed Right to Lie because of Philanthropic Concerns*. HPC Classics Series, Hackett Publishing Company, 1993.
- [19] S. Stieglitz, F. Brachten, B. Ross, and A.-K. Jung, “Do social bots dream of electric sheep? a categorisation of social media bot accounts,” *arXiv preprint arXiv:1710.04044*, 2017.
- [20] M. Wezel, E. A. Croes, M. L. Antheunis, *et al.*, ““i’m here for you”: Can social chatbots truly support their users? a literature review,” in *International Workshop on Chatbot Research and Design*, pp. 96–113, Springer, 2020.
- [21] Twitter, “The secret world of good bots,” 2021. <https://blog.twitter.com/common-thread/en/topics/stories/2021/the-secret-world-of-good-bots>.
- [22] M. L. Jones, “Silencing bad bots: Global, legal and political questions for mean machine communication,” *Communication Law and Policy*, vol. 23, no. 2, pp. 159–195, 2018.
- [23] Y. Roth and N. Pickles, “Bot or not? the facts about platform manipulation on twitter,” 2020. [https://blog.twitter.com/en\\_us/topics/company/2020/bot-or-not](https://blog.twitter.com/en_us/topics/company/2020/bot-or-not).
- [24] K.-C. Yang, O. Varol, C. A. Davis, E. Ferrara, A. Flammini, and F. Menczer, “Arming the public with artificial intelligence to counter social bots,” *Human Behavior and Emerging Technologies*, vol. 1, no. 1, pp. 48–61, 2019.
- [25] S. Cresci, “A decade of social bot detection,” *Communications of the ACM*, vol. 63, no. 10, pp. 72–83, 2020.
- [26] D. Schmuck and C. von Sikorski, “Perceived threats from social bots: The media’s role in supporting literacy,” *Computers in Human Behavior*, vol. 113, p. 106507, 2020.