DOI:10.1145/3126492

Carolina Alves de Lima Salge and Nicholas Berente

Computing Ethics
Is That Social Bot Behaving Unethically?

A procedure for reflection and discourse on the behavior of bots in the context of law, deception, and societal norms.

ATTEMPTING TO ANSWER the question posed by the title of this column requires us to reflect on moral goods and moral evils—on laws, duties, and norms, on actions and their consequences. In this Viewpoint, we draw on information systems ethics[6,7] to present *Bot Ethics*, a procedure the general social media community can use to decide whether the actions of social bots are unethical. We conclude with a consideration of culpability.

Social bots are computer algorithms in online social networks.[8] They can share messages, upload pictures, and connect with many users on social media. Social bots are more common than people often think.[a] Twitter has approximately 23 million of them, accounting for 8.5% of total users; and Facebook has an estimated 140 million social bots, which are between 5.5%–1.2% total users.[b,c] Almost 27 million Instagram users (8.2%) are estimated to be social bots.[d] LinkedIn and Tumblr also have significant social bot activity.[e,f] Sometimes their activity on these networks can be innocuous or even beneficial. For example, *SF QuakeBot*[g] performs a useful service by disseminating information about earthquakes, as they happen, in the San Francisco Bay area. However, in other situations, social bots can behave quite unethically.

Items purchased by Random Darknet Shopper, an automated computer program designed as an online shopping system that would make random purchases on the deep Web. The robot would have its purchases delivered to a group of artists who then put the items in an exhibition in Switzerland; the robot was 'arrested' by Swiss police after it bought illegal drugs.

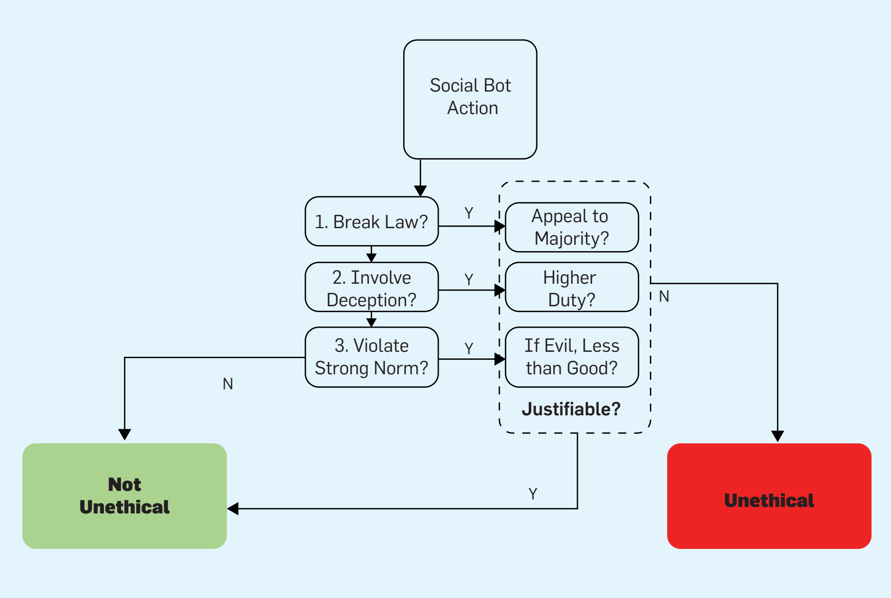
IMAGE COURTESY OF !MEDIENGRUPPE BITNIK

Social Bots Behaving Unethically

LinkedIn reports that social bots on the professional networking platform are often used to "steal data about legitimate users, breaching the user agreement and violating copyright law."[h] Social bots have been reported to behave badly in a variety of ways across various contexts—everything from disseminating spam[i] and fake news[j] to limiting free speech.[k] But it is not always clear whether their undesirable activity is simply a nuisance or whether it is indeed unethical—particularly given the random nature of the logic underlying many social bots. Bad actions are not necessarily unethical—

[a] http://bit.ly/2uDfIbP
[b] http://cnnmon.ie/2uFR4XJ
[c] http://bit.ly/1ieIIXN
[d] http://read.bi/1LFQJFU
[e] http://bit.ly/1Ktz5kc
[f] http://tcrn.ch/2tKo90x
[g] http://bit.ly/2vneleU
[h] http://bit.ly/2vFRI4E
[i] http://ubm.io/1MbsSf3
[j] http://bit.ly/2ftn0It
[k] http://bit.ly/14bDiuN

Bot Ethics: How to determine whether social bot actions are unethical.



there are shades of gray that are difficult to judge.

For example, *Tay*,[1] a social bot created by Microsoft to conduct research on conversational understanding, went from "humans are super cool" to "Hitler was right I hate the Jews" in less than 24 hours on Twitter due to malicious humans interacting with the social bot.[m] In another case, a social bot tweeted "I seriously want to kill people" from randomly generated sentences during a fashion convention in Amsterdam.[n] Clearly such inadvertent comments violate our sensibilities and are distasteful, but are they unethical? Perhaps, but by what standard do we judge? Some social bots do more than just comment—clearly those that steal information and other misdeeds are engaging in unethical activity, but, again, it is not always so clear. For instance, the *Random Darknet Shopper*—a social bot coded to explore the dark Web in the name of art—inadvertently purchased 10 Ecstasy pills (an illegal narcotic) and a counterfeit passport.[o] So a law was broken, but was this unethical behavior? We developed a procedure, which we describe next, to help answer such questions.

Bot Ethics: A Procedure to Evaluate the Ethics of Social Bot Activity

Ethics in philosophy dates back thousands of years, and this Viewpoint column cannot do justice to the entire field. However, because of the increasing prominence of social bots and their potential for malicious activity, ethical judgment about their activity is necessary. The best way to guide ethical conduct in a community is to provide a procedure for reflection and discourse.[5] The procedure we created is called "*Bot Ethics*" (see the figure here) and it focuses on the behavior of social bots with respect to law, deception, and norms.

Break Law?

Many laws are developed from ethical principles.[6] Even when a law may be flawed, it is typically the ethical course of action to follow that law.[9] Therefore a natural first question is: "*Does the action of the social bot break the law?*" The objective is to assess straightforward ethical questions, such as whether algorithms plant viruses in someone else's device. This is clearly illegal and unethical. There are cases where a social bot might ethically violate the law, such as civil disobedience for a cause the creator considers just. However, civil disobedience is only ethical in very rare cases in constitutional democracies where legal recourse for unjust laws pervade.[6] Cases where a law may be broken that are not unethical require justification—compelling arguments that appeal to moral standards of the majority.[6] Only in such rare cases may illegal acts be seen as moral and therefore ethical.[6] Thus we ask "*Is the illegal act justifiable?*" Acts that are not suitably justifiable (that is, do not appeal to the morality of the majority) are unethical. Swiss authorities did not file charges against the *Random Darknet Shopper* developers.[p] They argued that social bots can buy illegal narcotics over the Internet for the purpose of art[q] and that "ecstasy in this presentation was safe." The behavior was not unethical because it was justified according to the pervading morality of the community.

Social bots have been reported to behave badly in a variety of ways across various contexts.

Involve Deception?

If a social bot's behavior does not break any laws, next evaluate for truthfulness: "*Is any deception involved?*" Social bots may act deceitfully. For example, they can misrepresent themselves as human beings[2] or spread untruthful information (such as fake news). Deceiving acts communicate false or erroneous assertions, violating the prima facie duty of fidelity. Social bots should always act truthfully.[3] However, deceitful acts can be justifiable if the duty of fidelity is superseded by a higher-order duty, such as beneficence.[r] Deceptive, satirical actions may not be unethical since they elicit pleasure, improving the life of others. Consider *Big Data Batman*[s] as an illustration.

[l] https://twitter.com/TayandYou

[m] http://bit.ly/14bDiuN

[n] http://bit.ly/2ttN5Ox

[o] http://bit.ly/2vFGdu9

[p] By "developer" we are referring to either the organization or management of the organization or the software developer involved in the creation of the social bot.

[q] http://bit.ly/2ud2cZC

[r] Beneficence is the duty to bring virtue, knowledge or pleasure to others; other duties, according to Ross 1930, include non-maleficence, self-improvement, justice, gratitude, reparation (see Mason et al.[7], p. 132–133).

[s] http://bit.ly/2ttNUH7

The social bot finds every tweet with the term big data, replaces "big data" with "Batman," and then tweets the message as if it were its own. It obviously substitutes its words for others' words, but the satire makes it difficult to judge its ethics. Because the social bot might insult and embarrass some big-data advocates the community must go beyond the act (deontology) to consider its consequences (teleology), and ask whether potentially bad actions (for example, insult and embarrassment) outweigh, or supersede, the good (for example, pleasure through laughter) for the involved parties. Again, is the deception justifiable? Deception in the absence of supersession is likely to be unethical.

Violate Strong Norm?

Social bots that are legal and truthful can still behave unethically by violating strong norms that create more evil than good. Moral evils inflict "limits on human beings and contracts human life."[4] Evil restrains, instead of emancipating, evil actions reduce opportunities. Let us go back to *Tay*'s racist comments on Twitter. Although not illegal (First Amendment protections apply), nor deceitful, they violated the strong norm of racial equality. Social media companies like Twitter that temporarily lock or permanently suspend accounts that "directly attack or threaten other people on the basis of race,"[t] have established that the moral evil of racism outweighs the moral good of free speech. By applying *Bot Ethics* to Twitter's norms we conclude that *Tay*'s actions were unethical. Yet, there are cases where social bots may violate strong norms and not act unethically, as with asking inappropriate questions (what is your salary?). Such violations do not create moral evils.

Culpability of Unethical Social Bot Behavior

Should the general social media community blame developers for unethical behavior of their social bots? In the example of the algorithm that randomly generated that it wanted to kill people, who is responsible for the death threat? The programmer? Who is responsible for *Tay*'s remark about Hitler—Microsoft developers or those teaching the social bot to generate racist statements? Similarly, who is responsible for the social bot buying the illegal narcotics?

> **Should the general social media community blame developers for the unethical behavior of their social bots?**

Aristotle[1] said we can only assign culpability if we know that individuals behaved voluntarily and knowingly. Involuntary situations likely do not apply to social bots. Developers who are coerced into doing something unethical without a choice may not be entirely culpable, but in the case of free enterprise there is always a choice. Therefore, culpability rests on the knowledge of the developers. Developers who knowingly create social bots to engage in unethical actions are clearly culpable. They should be punished if evidence of their wrongdoing is convincing—the penalty must be consistent and proportional to the harm done and those affected should be compensated.[7]

But what about situations where developers act unknowingly? In those occasions the community must determine whether developers are culpably ignorant—did they ignore industry best practices in creating and testing their algorithms? If industry guidelines were not followed and the action was unethical, developers are culpable. However, developers who followed good development practices and incorporated the current industry thinking, and yet their social bot still acted unethically, deserve our pity and pardon, but they are not culpable. They should apologize, correct immediately, learn from their experience, and communicate the occurrence to the development community. For example, Microsoft posted its learning from *Tay* in blog form.[u]

Conclusion

We do not purport to write the last word on social bot ethics and culpability. Ethics is simply too complex of a domain to deal with fully in such a format. Nevertheless, some readily accessible guidance rooted in sound ethical thinking is in order.

For example, with the recent attention to the role of social bots in spreading misinformation in the form of "fake news," other social bots, such as *Reuters News Tracer*, are being created to ferret out such deceitful activity.[v] The *Bot Ethics* procedure can help the social media community understand when these deceitful actions are indeed unethical. It further helps to expand the focus of the community beyond narrow (that is, only deceitfulness) and simplistic (that is, good or bad bot) assessments of social bot activity to attend to the complexities of ethical assessments. In short, the *Bot Ethics* procedure serves as a starting point and guide for ethics-related discussion among various participants in a social media community, as they evaluate the actions of social bots.

[t] http://bit.ly/19SJwlt

[u] http://bit.ly/2tiPfMH

[v] http://bit.ly/2hIlfXG

References

1. Aristotle. *Nicomachean Ethics of Aristotle*. E.P. Dutton, NY, 1911.
2. Ferrara, E. et al. The rise of social bots. *Commun. ACM 59*, 7 (July 2016); 96–104; DOI: 10.1145/2818717
3. Gotterbarn, D., Miller, K. and Rogerson, S. Computer society and ACM approve software engineering code of ethics. *Computer Society Connection*, (1999), 84–88.
4. Grisez, G. and Shawn, R. *Beyond the New Morality: The Responsibilities of Freedom*. University of Notre Dame Press, Notre Dame, IN, 1980.
5. Habermas, J. *The Theory of Communicative Action, Volume 1: Reason and the Rationalization of Society*. 1985.
6. Kallman, E.A. and Grillo, J.P. *Ethical Decision Making and Information Technology*. McGraw-Hill, New York, NY, 1996.
7. Mason, R.O., Mason, F.M., and Culnan, M. *Ethics of Information Management*. Sage Publications, London, U.K.
8. Morstatter, F. et al. A new approach to bot detection: Striking the balance between precision and recall. ASONAM, 2016.
9. Rawls, J. The justification of civil disobedience. *Arguing about Law* (2013). 244–253.

Carolina Alves de Lima Salge (csalge@uga.edu) is a doctoral candidate at the University of Georgia.

Nicholas Berente (berente@uga.edu) is an associate professor at the University of Georgia.