Influential Article Review - The Role of Twitter and Trolls in Health Communication

Corey Terry

Evan Ramos

Marcia Hall

This paper examines social media. We present insights from a highly influential paper. Here are the highlights from this paper. To understand how Twitter bots and trolls ("bots") promote online health content. We compared bots' to average users' rates of vaccine-relevant messages, which we collected online from July 2014 through September 2017. We estimated the likelihood that users were bots, comparing proportions of polarized and anti vaccine tweets across user types. We conducted a content analysis of a Twitter hashtag associated with Russian troll activity. Whereas bots that spread malware and unsolicited content disseminated anti vaccine messages, Russian trolls promoted discord. Accounts masquerading as legitimate users create false equivalency, eroding public consensus on vaccination. For our overseas readers, we then present the insights from this paper in Spanish, French, Portuguese, and German.

SUMMARY

- Results suggest that Twitter bots and trolls have a significant impact on online communications about vaccination.
- Russian trolls and sophisticated Twitter bots post content about vaccination at significantly higher rates than does the average user. Content from these sources gives equal attention to pro- and anti vaccination arguments.
- Unlike troll accounts, content polluters 21 post anti vaccine messages 75% more often than does the average nonbot Twitter user. This suggests that vaccine opponents may disseminate messages using bot networks that are primarily designed for marketing. By contrast, spambots,3,4 which can be easily recognized as nonhuman, are less likely to promote an anti vaccine agenda than are nonbots.
- Several accounts could not be positively identified as either bots or humans because of intermediate or unavailable Botometer scores. These accounts, together constituting 93% of our random sample from the vaccine stream, tweeted content that was both more polarized and more opposed to vaccination than is that of the average nonbot account. Although the provenance of their tweets is unclear, we speculate that these accounts may possess a higher proportion of trolls or cyborgs—accounts nominally controlled by human users that are, on occasion, taken over by bots or otherwise exhibit bot-like or malicious behavior.15 Cyborg accounts are more likely to fall into this middle

range because they only display bot-like behaviors sometimes. This middle range is also likely to contain tweets from more sophisticated bots that are designed to more closely mimic human behaviors.

- Finally, trolls—exhibiting malicious behaviors yet operated by humans—are also likely to fall within this middle range. This suggests that proportionally more antivaccine tweets may be generated by accounts using a somewhat sophisticated semi automated approach to avoid detection.
- Survey data show a general consensus regarding the efficacy of vaccines in the general population.35 Consistent with these results, accounts unlikely to be bots are significantly less likely to promote polarized and anti vaccine content. Nevertheless, bots and trolls are actively involved in the online public health discourse, skewing discussions about vaccination. This is vital knowledge for risk communicators, especially considering that neither members of the public nor algorithmic approaches may be able to easily identify bots, trolls, or cyborgs.
- Malicious online behavior varies by account type. Russian trolls and sophisticated bots promote both pro- and anti vaccination narratives. This behavior is consistent with a strategy of promoting political discord. Bots and trolls frequently retweet or modify content from human users. Thus, well-intentioned posts containing provaccine content may have the unintended effect of «feeding the trolls,» giving the false impression of legitimacy to both sides, especially if this content directly engages with the anti vaccination discourse.

HIGHLY INFLUENTIAL ARTICLE

We used the following article as a basis of our evaluation:

Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., Quinn, S. C., & Dredze, M. (2018). Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. American Journal of Public Health, 108(10), 1378–1384.

This is the link to the publisher's website:
https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2018.304567

INTRODUCTION

Health-related misconceptions, misinformation, and disinformation spread over social media, posing a threat to public health.1 Despite significant potential to enable dissemination of factual information,2 social media are frequently abused to spread harmful health content,3 including unverified and erroneous information about vaccines.1,4 This potentially reduces vaccine uptake rates and increases the risks of global pandemics, especially among the most vulnerable.5 Some of this information is motivated: skeptics use online platforms to advocate vaccine refusal.6 Anti Vaccine advocates have a significant presence in social media,6 with as many as 50% of tweets about vaccination containing anti vaccine beliefs.7

Proliferation of this content has consequences: exposure to negative information about vaccines is associated with increased vaccine hesitancy and delay.8–10 Vaccine-hesitant parents are more likely to turn to the Internet for information and less likely to trust health care providers and public health experts on the subject.9,11 Exposure to the vaccine debate may suggest that there is no scientific consensus, shaking confidence in vaccination.12,13 Additionally, recent resurgences of measles, mumps, and pertussis and increased mortality from vaccine-preventable diseases such as influenza and viral pneumonia14 underscore the importance of combating online misinformation about vaccines.

Much health misinformation may be promulgated by "bots"15—accounts that automate content promotion—and "trolls"16—individuals who misrepresent their identities with the intention of promoting discord. One commonly used online disinformation strategy, amplification,17 seeks to create impressions of false equivalence or consensus through the use of bots and trolls. We seek to understand what role, if any, they play in the promotion of content related to vaccination.

Efforts to document how unauthorized users—including bots and trolls—have influenced online discourse about vaccines have been limited. DARPA's (the US Defense Advanced Research Projects Agency) 2015 Bot Challenge charged researchers with identifying "influence bots" on Twitter in a stream of vaccine-related tweets. The teams effectively identified bot networks designed to spread vaccine misinformation,18 but the public health community largely overlooked the implications of these findings. Rather, public health research has focused on combating online anti vaccine content, with less focus on the actors who produce and promote this content.1,19 Thus, the role of bots' and trolls' online activity pertaining to vaccination remains unclear.

We report the results of a retrospective observational study assessing the impact of bots and trolls on online vaccine discourse on Twitter. Using a set of 1 793 690 tweets collected from July 14, 2014, through September 26, 2017, we quantified the impact of known and suspected Twitter bots and trolls on amplifying polarizing and anti vaccine messages. This analysis is supplemented by a qualitative study of #VaccinateUS—a Twitter hashtag designed to promote discord using vaccination as a political wedge issue. #VaccinateUS tweets were uniquely identified with Russian troll accounts linked to the Internet Research Agency—a company backed by the Russian government specializing in online influence operations.20 Thus, health communications have become "weaponized": public health issues, such as vaccination, are included in attempts to spread misinformation and disinformation by foreign powers. In addition, Twitter bots distribute malware and commercial content (i.e., spam) masquerade as human users to distribute anti vaccine messages. A full 93% of tweets about vaccines are generated by accounts whose provenance can be verified as neither bots nor human users yet who exhibit malicious behaviors. These unidentified accounts preferentially tweet anti vaccine misinformation. We discuss implications for online public health communications.

CONCLUSION

Results suggest that Twitter bots and trolls have a significant impact on online communications about vaccination. The nature of this impact differs by account type.

Russian Trolls

Russian trolls and sophisticated Twitter bots post content about vaccination at significantly higher rates than does the average user. Content from these sources gives equal attention to pro- and anti vaccination arguments. This is consistent with a strategy of promoting discord across a range of controversial topics—a known tactic employed by Russian troll accounts.20,26 Such strategies may undermine the public health: normalizing these debates may lead the public to question long-standing scientific consensus regarding vaccine efficacy.13 Indeed, several anti vaccine arguments claim to represent both sides of the debate—like the tactics used by the trolls identified in this study—while simultaneously communicating a clear gist (i.e., a bottom-line meaning). We recently found that this strategy was effective for propagating news articles through social media in the context of the 2015 Disneyland measles outbreak.32

Commercial and Malware Distributors

Unlike troll accounts, content polluters (i.e., disseminators of malware, unsolicited commercial content, and other disruptive material that typically violates Twitter's terms of service)21 post anti vaccine messages 75% more often than does the average nonbot Twitter user. This suggests that vaccine opponents may disseminate messages using bot networks that are primarily designed for marketing. By contrast, spambots,3,4 which can be easily recognized as nonhuman, are less likely to promote an anti vaccine agenda than are nonbots. Notably, content polluters and traditional spambots are both less likely to discuss vaccine-preventable illnesses than is the average Twitter user, suggesting that when they do tweet vaccine-relevant messages, their specific focus is on vaccines per se, rather than the viruses that require them. Thus, it is unclear to what extent their promotion of vaccine-related content is driven by true anti vaccine sentiment or is used as a tactic designed to drive up click-through rates by propagating motivational content ("clickbait").

Unidentified Accounts

Several accounts could not be positively identified as either bots or humans because of intermediate or unavailable Botometer scores. These accounts, together constituting 93% of our random sample from the vaccine stream, tweeted content that was both more polarized and more opposed to vaccination than is that of the average nonbot account. Although the provenance of their tweets is unclear, we speculate that these accounts may possess a higher proportion of trolls or cyborgs—accounts nominally controlled by human users that are, on occasion, taken over by bots or otherwise exhibit bot-like or malicious behavior.15 Cyborg accounts are more likely to fall into this middle range because they only display bot-like behaviors sometimes. This middle range is also likely to contain tweets from more sophisticated bots that are designed to more closely mimic human behaviors.

Finally, trolls—exhibiting malicious behaviors yet operated by humans—are also likely to fall within this middle range. This suggests that proportionally more antivaccine tweets may be generated by accounts using a somewhat sophisticated semi automated approach to avoid detection. This creates the false impression of grassroots debate regarding vaccine efficacy—a technique known as "astroturfing"17 (as in the #VaccineUS tweets shown in the box on page 1383). There are certainly standard human accounts that also fall within this middle range. Although technological limitations preclude us from drawing definitive conclusions about these account types, the fact that middle-range tweets tend to post proportionately more anti vaccine messages suggests strongly that these anti vaccine messages may be disseminated at higher rates by a combination of malicious actors (bots, trolls, cyborgs, and human users) who are difficult to distinguish from one another.

This interpretation is supported by the fact that users within this intermediate range tended to produce more tweets, and especially anti vaccine tweets, per account, suggesting that anti vaccine activists may preferentially use these channels. In addition, users whose accounts had been deleted posted more polarized messages per user and were also significantly more likely to post anti vaccine messages. Although reasons for account deletion vary, recent movement by Twitter to remove bots,33,34 trolls,20 cyborgs, and other violators of Twitter's terms of service suggests that these violators may be overrepresented among the deleted accounts in our sample. We cannot claim that all, or even most, accounts with unknown bot scores are malicious actors; however, we expect that a higher proportion of malicious actors are present in this subset of the data. By contrast, randomly sampled accounts that were easily identifiable as bots generated more neutral, but not polarized tweets per account. Presumably, accounts that are obviously automated are more frequently used to disseminate content such as news and may not be considered credible sources of grassroots anti vaccine information.

Public Health Implications

Survey data show a general consensus regarding the efficacy of vaccines in the general population.35 Consistent with these results, accounts unlikely to be bots are significantly less likely to promote polarized and anti vaccine content. Nevertheless, bots and trolls are actively involved in the online public health discourse, skewing discussions about vaccination. This is vital knowledge for risk communicators, especially considering that neither members of the public nor algorithmic approaches may be able to easily identify bots, trolls, or cyborgs.

Malicious online behavior varies by account type. Russian trolls and sophisticated bots promote both pro- and anti vaccination narratives. This behavior is consistent with a strategy of promoting political discord. Bots and trolls frequently retweet or modify content from human users. Thus, well-intentioned posts containing provaccine content may have the unintended effect of "feeding the trolls," giving the false impression of legitimacy to both sides, especially if this content directly engages with the anti vaccination discourse. Presuming bot and troll accounts seek to generate roughly equal numbers of tweets for both sides, limiting access to pro vaccine content could potentially also reduce the incentive to post anti vaccine content.

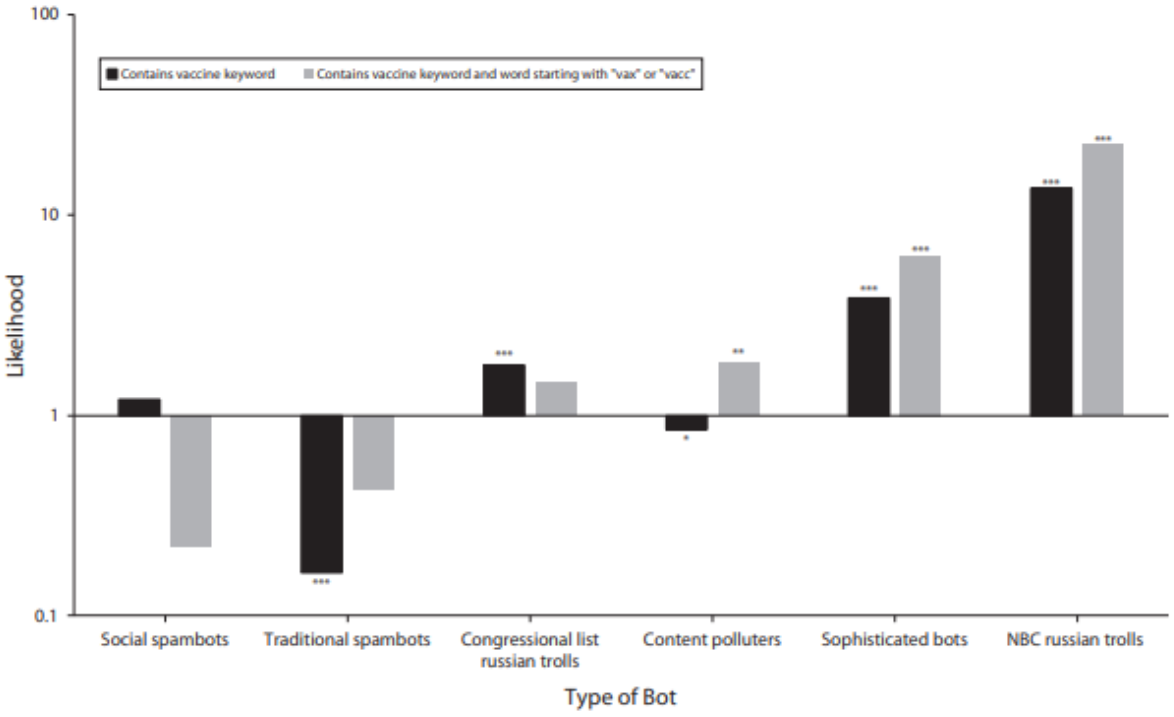
By contrast, accounts that are known to distribute malware and commercial content are more likely to promote anti vaccination messages, suggesting that anti vaccine advocates may use pre existing

infrastructures of bot networks to promote their agenda. These accounts may also use the compelling nature of anti vaccine content as clickbait to drive up advertising revenue and expose users to malware. When faced with such content, public health communications officials may consider emphasizing that the credibility of the source is dubious and that users exposed to such content may be more likely to encounter malware. Anti Vaccine content may increase the risks of infection by both computer and biological viruses.

The highest proportion of anti vaccine content is generated by accounts with unknown or intermediate bot scores. Although we speculate that this set of accounts contains more sophisticated bots, trolls, and cyborgs, their provenance is ultimately unknown. Therefore, beyond attempting to prevent bots from spreading messages over social media, public health practitioners should focus on combating the messages themselves while not feeding the trolls. This is a ripe area for future research, which might include emphasizing that a significant proportion of anti vaccination messages are organized "astroturf" (i.e., not grassroots) and other bottom-line messages that put anti vaccine messages in their proper contexts.

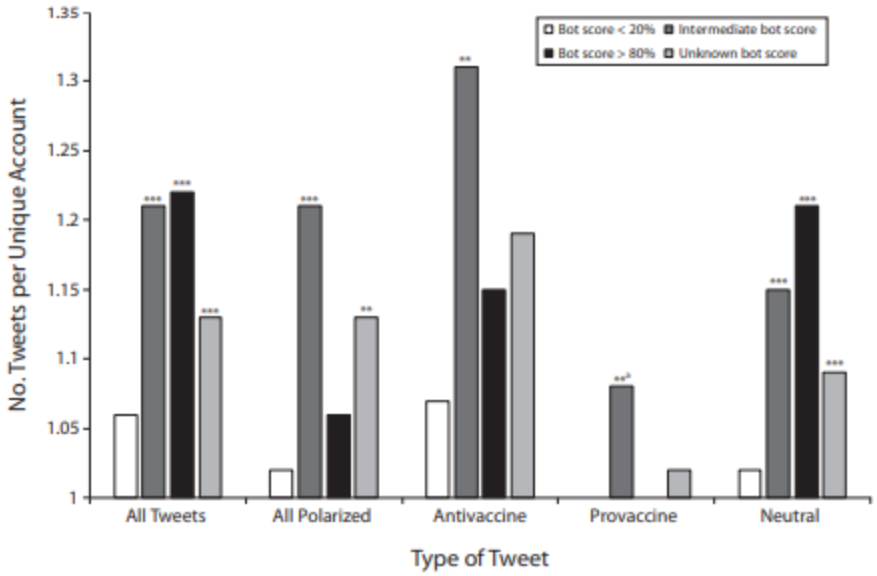
APPENDIX

FIGURE 1
BOTS' LIKELIHOOD OF TWEETING ABOUT VACCINES COMPARED WITH AVERAGE TWITTER USERS: JULY 14, 2014–SEPTEMBER 26, 2017



Note. NBC = National Broadcasting Network. All results remained significant after controlling for multiple comparisons using the Holm–Bonferroni procedure. Raw counts are given in Table B (available as a supplement to the online version of this article at http://www.ajph.org).
*P<.05; **P<.01; ***P<.001.

FIGURE 2
NUMBER OF TWEETS PER UNIQUE ACCOUNT, SEPARATED BY SENTIMENT AND BOT SCORE CATEGORY: JULY 14, 2014–SEPTEMBER 26, 2017



[a]Not significant after controlling for multiple comparisons using the Holm–Bonferroni procedure. Raw counts are given in Table D (available as a supplement to the online version of this article at http://www.ajph.org).

**$P < .01$; ***$P < .001$.

TABLE 1
PROPORTIONS OF POLARIZED AND ANTIVACCINE TWEETS BY USER TYPE: JULY 14, 2014–SEPTEMBER 26, 2017

User Type	Polarized, %	Antivaccine, %
Assorted users, bot score, %		
< 20	31	35
20–80	39***	60***
> 80	26	49*,a
Unknown	37*,a	62***
Known bots and trolls		
NBC Russian trolls[20]	20*,a	47
Content polluters[21]	38	60***
Fake followers[22]	0	NA
Traditional spambots[23,24]	3***	0
Social spambots[23,24]	18**	56*,a
Sophisticated bots[25]	28	44
Congressional list Russian trolls[26]	39	48

Note. NA = not applicable because of insufficient data; NBC = National Broadcasting Network. A statistically significant result indicates that a certain type of account posts polarized or antivaccine tweets at a rate that differs significantly from that of accounts with bot scores < 20% (likely humans). Polarized proportion is the ratio of all nonneutral tweets to all tweets. Antivaccine proportion is the ratio of antivaccine tweets to polarized tweets. Raw counts are shown in Table E (available as a supplement to the online version of this article at http://www.ajph.org).

[a]No longer significant after controlling for multiple comparisons using the Holm–Bonferroni procedure.

*P < .05; **P < .01; ***P < .001.

REFERENCES

Betsch C, Brewer NT, Brocard P, et al. Opportunities and challenges of Web 2.0 for vaccination decisions. Vaccine. 2012;30(25):3727–3733.

Breland JY, Quintiliani LM, Schneider KL, May CN, Pagoto S. Social media as a tool to increase the impact of public health research. Am J Public Health. 2017;107(12):1890–1891.

Broniatowski DA, Hilyard KM, Dredze M. Effective vaccine communication during the Disneyland measles outbreak. Vaccine. 2016;34(28):3225–3228. Crossref, Medline, Google Scholar

Centers for Disease Control and Prevention. Mortality. 2017. Available at: https://www.cdc.gov/nchs/health_policy/mortality.htm. Accessed April 27, 2018.

Chu Z, Gianvecchio S, Wang H, Jajodia S. Detecting automation of Twitter accounts: are you a human, bot, or cyborg? IEEE Trans Depend Secure Comput. 2012;9(6):811–824.

Collins English Dictionary. Troll definition and meaning. Available at: https://www.collinsdictionary.com/dictionary/english/troll. Accessed April 27, 2018.

Cresci S, Di Pietro R, Petrocchi M, Spognardi A, Tesconi M. Fame for sale: efficient detection of fake Twitter followers. Decis Support Syst. 2015;80:56–71.

Cresci S, Di Pietro R, Petrocchi M, Spognardi A. The paradigm-shift of social spambots: evidence, theories, and tools for the arms race. In: Barrett R; International World Wide Web Conferences Steering Committee, eds. Proceedings of the 26th International Conference on World Wide Web Companion.Republic and Canton of Geneva: ACM Press; 2017:963–972.

Cresci S, Pietro RD, Petrocchi M, Spognardi A, Tesconi M. Social fingerprinting: detection of spambot groups through DNA-inspired behavioral modeling. IEEE Trans Dependable and Secure Comput. 2018;15(4):561–576.

Davis CA, Varol O, Ferrara E, Flammini A, Menczer F. BotOrNot: A System to evaluate social bots. Available at: http://dl.acm.org/citation.cfm?doid=2872518.2889302. Accessed July 25, 2018.
Dixon G, Clarke C. The effect of falsely balanced reporting of the autism-vaccine controversy on vaccine safety perceptions and behavioral intentions. Health Educ Res. 2013;28(2):352–359.
Dredze M, Broniatowski DA, Smith MC, Hilyard KM. Understanding vaccine refusal: why we need social media now. Am J Prev Med. 2016;50(4):550–552.
Dubé E, Vivion M, MacDonald NE. Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications. Expert Rev Vaccines. 2015;14(1):99–117.
Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods. 2006;5(1):80–92.
Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. Commun ACM. 2016;59(7):96–104.
Flynn K. Twitter influencers suspect a bot "purge." Available at: https://mashable.com/2018/01/29/twitter-bots-purge-influencers-accounts. Accessed April 27, 2018.
Frommer D. Twitter's list of 2,752 Russian trolls. 2017. Available at: https://www.recode.net/2017/11/2/16598312/russia-twitter-trump-twitter-deactivated-handle-list. Accessed March 11, 2018.
Funk C, Kennedy B, Hefferon M. Vast majority of Americans say benefits of childhood vaccines outweigh risks. 2017. Available at: http://www.pewinternet.org/2017/02/02/vast-majority-of-americans-say-benefits-of-childhood-vaccines-outweigh-risks. Accessed February 14, 2017.
Jolley D, Douglas KM. The effects of anti-vaccine conspiracy theories on vaccination intentions. PLoS One. 2014;9(2):e89177.
Jones AM, Omer SB, Bednarczyk RA, Halsey NA, Moulton LH, Salmon DA. Parents' source of vaccine information and impact on vaccine attitudes, beliefs, and nonmedical exemptions. Adv Prev Med. 2012;2012:932741.
Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the internet. Vaccine. 2010;28(7):1709–1716.
Kata A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm—an overview of tactics and tropes used online by the anti-vaccination movement. Vaccine. 2012;30(25):3778–3789.
Lee K, Eoff BD, Caverlee J. Seven months with the devils: a long-term study of content polluters on Twitter. Available at: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2780. Accessed March 11, 2018.
Luxton DD, June JD, Fairall JM. Social media and suicide: a public health perspective. Am J Public Health. 2012;102(suppl 2):S195–S200.
Madrak S. Wingers melt down as twitter finally purges Russian bots. Available at: https://crooksandliars.com/2018/02/wingers-have-sad-twitter-purges-russian-0. Accessed April 27, 2018.
Popken B. Twitter deleted Russian troll tweets. So we published more than 200,000 of them. Available at: https://www.nbcnews.com/tech/social-media/now-available-more-200-000-deleted-russian-troll-tweets-n844731. Accessed March 11, 2018.
Quinn SC. Probing beyond individual factors to understand influenza and pneumococcal vaccine uptake. Am J Public Health. 2018;108(4):427–429.
Sandelowski M, Barroso J. Classifying the findings in qualitative studies. Qual Health Res. 2003;13(7):905–923.
Smith MJ, Marshall GS. Navigating parental vaccine hesitancy. Pediatr Ann. 2010;39(8):476–482.
Subrahmanian VS, Azaria A, Durst S, et al. The DARPA Twitter bot challenge. Computer. 2016;49(6):38–46.
Tomeny TS, Vargo CJ, El-Toukhy S. Geographic and demographic correlates of autism-related anti-vaccine beliefs on Twitter, 2009–15. Soc Sci Med. 2017;191:168–175.

Varol O, Ferrara E, Davis CA, Menczer F, Flammini A. Online human–bot interactions: detection, estimation, and characterization. Available at: https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587. Accessed March 11, 2018.
Ward JK, Peretti-Watel P, Verger P. Vaccine criticism on the Internet: propositions for future research. Hum Vaccin Immunother. 2016;12(7):1924–1929.
Witteman HO, Zikmund-Fisher BJ. The defining characteristics of Web 2.0 and their potential influence in the online vaccination debate. Vaccine. 2012;30(25):3734–3740.
Wojcik S, Messing S, Smith A, Rainie L, Hitlin P. Bots in the Twittersphere. 2018. Available at: http://www.pewinternet.org/2018/04/09/bots-in-the-twittersphere. Accessed April 26, 2018.

TRANSLATED VERSION: SPANISH

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

VERSION TRADUCIDA: ESPAÑOL

A continuación se muestra una traducción aproximada de las ideas presentadas anteriormente. Esto se hizo para dar una comprensión general de las ideas presentadas en el documento. Por favor, disculpe cualquier error gramatical y no responsabilite a los autores originales de estos errores.

INTRODUCCIÓN

Conceptos erróneos relacionados con la salud, desinformación y desinformación difundidas por las redes sociales, lo que representa una amenaza para la salud pública.1 A pesar de un potencial significativo para permitir la difusión de información fáctica,2 se abusa con frecuencia de las redes sociales para difundir contenido sanitario dañino,3 incluida la información no verificada y errónea sobre las vacunas.1,4 Esto reduce potencialmente las tasas de captación de vacunas y aumenta los riesgos de pandemias mundiales, especialmente entre los más vulnerables.5 Parte de esta información es motivada. : los escépticos utilizan plataformas en línea para abogar por la negativa a las vacunas.6 Los defensores de la antivacuna tienen una presencia significativa en las redes sociales6, con hasta el 50% de los tweets sobre la vacunación que contienen creencias antivacunas.7

La proliferación de este contenido tiene consecuencias: la exposición a información negativa sobre las vacunas se asocia con un aumento de la vacilación de las vacunas y el retraso.8-10 Los padres que dudan de las vacunas son más propensos a recurrir a Internet para obtener información y menos propensos a confiar en los proveedores de atención médica y expertos en salud pública sobre el tema.9,11 La exposición al debate sobre la vacuna puede sugerir que no hay consenso científico , sacudiendo la confianza en la vacunación.12,13 Además, los recientes resurgimientos del sarampión, las paperas y la tos ferina y el aumento de la mortalidad por enfermedades prevenibles vacunables como la gripe y la neumonía viral14 subrayan la importancia de combatir la desinformación en línea sobre las vacunas.

Mucha desinformación sanitaria puede ser promulgada por "bots"15 (cuentas que automatizan la promoción de contenido) y "trolls"16— individuos que tergiversan sus identidades con la intención de promover la discordia. Una estrategia de desinformación en línea comúnmente utilizada, amplificación,17 busca crear impresiones de falsa equivalencia o consenso a través del uso de bots y trolls. Tratamos de entender qué papel, si los hay, desempeñan en la promoción de contenidos relacionados con la vacunación.

Los esfuerzos para documentar cómo los usuarios no autorizados, incluidos bots y trolls, han influido en el discurso en línea sobre las vacunas han sido limitados. EL Desafío de proyectos avanzados de defensa de LOS Estados Unidos (Agencia de Proyectos de Investigación Avanzada de Defensa de los Estados Unidos) 2015 encargó a los investigadores la identificación de "bots de influencia" en Twitter en una corriente de tweets relacionados con vacunas. Los equipos identificaron efectivamente las redes de bots

diseñadas para difundir la desinformación de las vacunas18, pero la comunidad de salud pública pasó por alto en gran medida las implicaciones de estos hallazgos. Más bien, la investigación en salud pública se ha centrado en la lucha contra el contenido antivacuna en línea, con menos atención a los actores que producen y promueven este contenido.1,19 Por lo tanto, el papel de la actividad en línea de los robots y trolls en relación con la vacunación sigue sin estar claro.

Informamos de los resultados de un estudio observacional retrospectivo que evalúa el impacto de los robots y trolls en el discurso de la vacuna en línea en Twitter. Usando un conjunto de 1 793 690 tweets recopilados desde el 14 de julio de 2014 hasta el 26 de septiembre de 2017, cuantificamos el impacto de los robots y trolls conocidos y sospechosos de Twitter en la amplificación de los mensajes polarizadores y antivacunas. Este análisis se complementa con un estudio cualitativo de #vaccinateus, un hashtag de Twitter diseñado para promover la discordia utilizando la vacunación como un tema de cuña política. #vaccinateus tweets se identificaron de manera única con las cuentas de trolls rusos vinculadas a la Agencia de Investigación de Internet, una empresa respaldada por el gobierno ruso especializada en operaciones de influencia en línea.20 Por lo tanto, las comunicaciones sanitarias se han "armado": los problemas de salud pública, como la vacunación, se incluyen en los intentos de difundir la desinformación y la desinformación por parte de potencias extranjeras. Además, los bots de Twitter que distribuyen malware y contenido comercial (es decir, spam) se hacen pasar por usuarios humanos para distribuir mensajes antivaccina. Un 93% de los tweets sobre vacunas son generados por cuentas cuya procedencia puede ser verificada como ni bots ni usuarios humanos que exhiben comportamientos maliciosos. Estas cuentas no identificadas tuitean preferentemente la desinformación antivaccina. Discutimos las implicaciones para las comunicaciones de salud pública en línea.

CONCLUSIÓN

Los resultados sugieren que los robots y trolls de Twitter tienen un impacto significativo en las comunicaciones en línea sobre la vacunación. La naturaleza de este impacto varía según el tipo de cuenta.

Trolls rusos

Los trolls rusos y los sofisticados robots de Twitter publican contenido sobre la vacunación a tasas significativamente más altas que el usuario medio. El contenido de estas fuentes presta la misma atención a los argumentos pro y antivacunación. Esto es coherente con una estrategia de promoción de la discordia en una serie de temas controvertidos: una táctica conocida empleada por los relatos de trolls rusos.20,26 Tales estrategias pueden socavar la salud pública: normalizar estos debates puede llevar al público a cuestionar el consenso científico de larga data sobre la eficacia de la vacuna.13 De hecho, varios argumentos antivacunas afirman representar ambos lados del debate, como las tácticas utilizadas por los trolls identificados en este estudio, al mismo tiempo que comunican un claro gist (es decir. , un significado de la línea inferior). Recientemente descubrimos que esta estrategia era eficaz para propagar artículos de noticias a través de las redes sociales en el contexto del brote de sarampión de Disneylandia de 2015.32

Distribuidores comerciales y de malware

A diferencia de las cuentas de troll, los contaminadores de contenido (es decir, los diseminadores de malware, contenido comercial no solicitado y otro material disruptivo que normalmente infringe los términos de servicio de Twitter)21 mensajes post antivaccina 75% más a menudo que el usuario promedio de Twitter no robot. Esto sugiere que los oponentes de la vacuna pueden difundir mensajes utilizando redes de bots que están diseñados principalmente para el marketing. Por el contrario, los robots de spam,3,4 que pueden ser fácilmente reconocidos como no humanos, son menos propensos a promover una agenda antivacuna que los no robots. En particular, los contaminadores del contenido y los robots de spam tradicionales son menos propensos a discutir enfermedades prevenibles estoy tratando de enfermedades prevenibles estoy en el caso de lo que es el usuario medio de Twitter, lo que sugiere que cuando tuitean mensajes relevantes para las vacunas, su enfoque específico se centra en las vacunas per se, en lugar de los virus que las requieren. Por lo tanto, no está claro en qué medida su promoción del contenido relacionado

con la vacuna está impulsada por un verdadero sentimiento antivacuna o se utiliza como una táctica diseñada para aumentar las tasas de clics mediante la propagación de contenido motivacional ("clickbait").

Cuentas no identificadas

Varias cuentas no se pudieron identificar positivamente como bots o humanos debido a puntuaciones intermedias o no disponibles de Botometer. Estas cuentas, que juntos constituyen el 93% de nuestra muestra aleatoria de la corriente de la vacuna, tuitearon contenido que era más polarizado y más opuesto a la vacunación que el de la cuenta promedio no bot. Aunque la procedencia de sus tweets no está clara, especulamos que estas cuentas pueden poseer una mayor proporción de trolls o cyborgs, cuentas controladas nominalmente por usuarios humanos que, en ocasiones, son tomadas por bots o de otra manera exhiben un comportamiento similar a un bot o malintencionado.15 Las cuentas cyborg son más propensas a caer en este rango medio porque solo muestran comportamientos similares a bots a veces. Este rango medio también es probable que contenga tweets de bots más sofisticados que están diseñados para imitar más de cerca los comportamientos humanos.

Por último, los trolls, que exhiben comportamientos maliciosos pero operados por humanos, también son propensos a estar dentro de este rango medio. Esto sugiere que proporcionalmente más tweets antivaccina pueden ser generados por cuentas utilizando un enfoque semiautomatizado algo sofisticado para evitar la detección. Esto crea la falsa impresión del debate de base sobre la eficacia de la vacuna, una técnica conocida como "astroturfing"17 (como en el #vaccineus tweets que se muestran en el recuadro de la página 1383). Ciertamente hay cuentas humanas estándar que también caen dentro de este rango medio. Aunque las limitaciones tecnológicas nos impiden sacar conclusiones definitivas sobre estos tipos de cuentas, el hecho de que los tweets de rango medio tiendan proporcionalmente más mensajes antivacunas sugiere fuertemente que estos mensajes antivacunas pueden ser difundidos a tasas más altas por una combinación de actores maliciosos (bots, trolls, cyborgs y usuarios humanos) que son difíciles de distinguir entre sí.

Esta interpretación está respaldada por el hecho de que los usuarios dentro de este rango intermedio tendían a producir más tweets, y especialmente tweets antivaccina, por cuenta, lo que sugiere que los activistas antivaccina pueden utilizar preferentemente estos canales. Además, los usuarios cuyas cuentas se habían eliminado publicaron mensajes más polarizados por usuario y también eran significativamente más propensos a publicar mensajes antivaccina. Aunque las razones para la eliminación de la cuenta varían, el movimiento reciente de Twitter para eliminar bots, 33,34 trolls, 20 cyborgs y otros infractores de los términos de servicio de Twitter sugiere que estos infractores pueden estar sobrerrepresentados entre las cuentas eliminadas de nuestra muestra. No podemos afirmar que todas las cuentas, o incluso la mayoría, con puntuaciones de bots desconocidas son actores maliciosos; sin embargo, esperamos que una mayor proporción de actores malintencionados estén presentes en este subconjunto de los datos. Por el contrario, las cuentas muestreadas aleatoriamente que eran fácilmente identificables como bots generaban tweets más neutros, pero no polarizados por cuenta. Presumiblemente, las cuentas que obviamente están automatizadas se utilizan con más frecuencia para difundir contenidos como noticias y pueden no considerarse fuentes creíbles de información antivacuna de base.

Implicaciones para la salud pública

Los datos de la encuesta muestran un consenso general sobre la eficacia de las vacunas en la población general.35 En consonancia con estos resultados, es poco probable que las cuentas sean bots tienen menos probabilidades de promover el contenido polarizado y antivacuna. Sin embargo, los bots y trolls participan activamente en el discurso de salud pública en línea, sesgando discusiones sobre la vacunación. Este es un conocimiento vital para los comunicadores de riesgos, especialmente teniendo en cuenta que ni los miembros del público ni los enfoques algorítmicos pueden ser capaces de identificar fácilmente bots, trolls o cyborgs.

El comportamiento malintencionado en línea varía según el tipo de cuenta. Trolls rusos y robots sofisticados promueven narrativas pro y antivacunación. Este comportamiento es consistente con una estrategia de promoción de la discordia política. Los bots y trolls frecuentemente retuitea o modifican el contenido de los usuarios humanos. Por lo tanto, los mensajes bien intencionados que contienen contenido de provacuna pueden tener el efecto no intencional de "alimentar a los trolls", dando la falsa impresión de legitimidad a ambas partes, especialmente si este contenido se involucra directamente con el discurso de la antivacunación. La presunción de cuentas de bots y trolls busca generar aproximadamente el mismo número de tweets para ambos lados, lo que limita el acceso al contenido de provaccina podría potencialmente también reducir el incentivo para publicar contenido antivaccina.

Por el contrario, las cuentas que se sabe que distribuyen malware y contenido comercial son más propensas a promover mensajes antivacunación, lo que sugiere que los defensores de la antivacuna pueden utilizar infraestructuras preexistentes de redes de bots para promover su agenda. Estas cuentas también pueden utilizar la naturaleza convincente del contenido antivacuna como clickbait para aumentar los ingresos publicitarios y exponer a los usuarios a malware. Cuando se enfrentan a este tipo de contenido, los funcionarios de comunicaciones de salud pública pueden considerar hacer hincapié en que la credibilidad de la fuente es dudosa y que los usuarios expuestos a dicho contenido pueden ser más propensos a encontrar malware. El contenido de antivacunas puede aumentar los riesgos de infección por parte de virus informáticos y biológicos.

La mayor proporción de contenido antivacuna se genera en cuentas con puntuaciones de bot desconocidas o intermedias. Aunque especulamos que este conjunto de cuentas contiene robots, trolls y cyborgs más sofisticados, su procedencia es en última instancia desconocida. Por lo tanto, más allá de intentar evitar que los bots difundan mensajes a través de las redes sociales, los profesionales de la salud pública deben centrarse en combatir los mensajes ellos mismos mientras no alimentan a los trolls. Este es un área madura para futuras investigaciones, que podría incluir enfatizar que una proporción significativa de los mensajes de vacunación están organizados "astroturf" (es decir, no de base) y otros mensajes de resultados que ponen mensajes antivaccina en sus contextos adecuados.

TRANSLATED VERSION: FRENCH

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

VERSION TRADUITE: FRANÇAIS

Voici une traduction approximative des idées présentées ci-dessus. Cela a été fait pour donner une compréhension générale des idées présentées dans le document. Veuillez excuser toutes les erreurs grammaticales et ne pas tenir les auteurs originaux responsables de ces erreurs.

INTRODUCTION

Les idées fausses, la désinformation et la désinformation liées à la santé se propagent sur les médias sociaux, ce qui constitue une menace pour la santé publique.1 Malgré un potentiel important pour permettre la diffusion d'informations factuelles2, les médias sociaux sont fréquemment utilisés pour diffuser des contenus nocifs pour la santé3, y compris des informations non vérifiées et erronées sur les vaccins.1,4 Cela pourrait réduire les taux d'utilisation des vaccins et accroître les risques de pandémies mondiales5. : Les sceptiques utilisent les plateformes en ligne pour préconiser le refus du vaccin.6 Les défenseurs des antivaccins sont présents dans les médias sociaux6, avec jusqu'à 50 % des tweets sur la vaccination contenant des croyances antivaccin.7

La prolifération de ce contenu a des conséquences : l'exposition à des informations négatives sur les vaccins est associée à une augmentation de l'hésitation et du retard des vaccins.8–10 Les parents hésitants

à se tourner vers Internet pour obtenir de l'information et moins susceptibles de faire confiance aux fournisseurs de soins de santé et aux experts en santé publique à ce sujet.9,1 l'exposition au débat sur les vaccins peut suggérer qu'il n'y a pas de consensus scientifique. , secouant la confiance dans la vaccination.12,13 En outre, les récentes résurgences de la rougeole, des oreillons et de la coqueluche et l'augmentation de la mortalité due aux maladies évitables par la vaccination telles que la grippe et la pneumonie virale14 soulignent l'importance de lutter contre la désinformation en ligne sur les vaccins.

Une grande partie de la désinformation sur la santé peut être promulguée par des « bots »15 — des comptes qui automatisent la promotion du contenu — et des « trolls » 16 — des individus qui dénaturent leur identité dans l'intention de promouvoir la discorde. Une stratégie de désinformation en ligne couramment utilisée, l'amplification,17 cherche à créer des impressions de fausse équivalence ou de consensus par l'utilisation de bots et de trolls. Nous cherchons à comprendre quel rôle, le cas échéant, ils jouent dans la promotion du contenu lié à la vaccination.

Les efforts visant à documenter comment les utilisateurs non autorisés, y compris les bots et les trolls, ont influencé le discours en ligne sur les vaccins ont été limités. Darpa (l'us Defense Advanced Research Projects Agency) 2015 Bot Challenge a chargé les chercheurs d'identifier les « bots d'influence » sur Twitter dans un flux de tweets liés au vaccin. Les équipes ont effectivement identifié des réseaux de robots conçus pour diffuser la désinformation sur les vaccins18, mais la communauté de la santé publique a largement négligé les implications de ces résultats. Au contraire, la recherche en santé publique s'est concentrée sur la lutte contre le contenu antivaccin en ligne, en mettant moins l'accent sur les acteurs qui produisent et font la promotion de ce contenu.1,19 Ainsi, le rôle de l'activité en ligne des bots et des trolls concernant la vaccination reste flou.

Nous rapportons les résultats d'une étude d'observation rétrospective évaluant l'impact des bots et des trolls sur le discours des vaccins en ligne sur Twitter. À l'aide d'un ensemble de 1 793 690 tweets collectés du 14 juillet 2014 au 26 septembre 2017, nous avons quantifié l'impact des bots et trolls Twitter connus et soupçonnés sur l'amplification des messages polarisants et antivaccinations. Cette analyse est complétée par une étude qualitative de #vaccinateus, un hashtag Twitter conçu pour promouvoir la discorde en utilisant la vaccination comme un problème de coin politique. #vaccinateus tweets ont été identifiés de manière unique avec des comptes de trolls russes liés à l'Internet Research Agency, une société soutenue par le gouvernement russe spécialisée dans les opérations d'influence en ligne.20 Ainsi, les communications de santé sont devenues « armées » : les questions de santé publique, telles que la vaccination, sont incluses dans les tentatives de diffusion de désinformation et de désinformation par des puissances étrangères. En outre, les bots Twitter distribuant des logiciels malveillants et du contenu commercial (c.-à-d. Du spam) se font passer pour des utilisateurs humains pour distribuer des messages antivaccin. Un total de 93% des tweets sur les vaccins sont générés par des comptes dont la provenance ne peut être vérifiée que ni les bots ni les utilisateurs humains encore qui présentent des comportements malveillants. Ces comptes non identifiés tweetent de préférence la désinformation antivaccin. Nous discutons des répercussions sur les communications en ligne en matière de santé publique.

CONCLUSION

Les résultats suggèrent que les bots Twitter et les trolls ont un impact significatif sur les communications en ligne sur la vaccination. La nature de cet impact diffère selon le type de compte.

Trolls russes

Les trolls russes et les bots Twitter sophistiqués publient du contenu sur la vaccination à des taux nettement plus élevés que l'utilisateur moyen. Le contenu de ces sources accorde une attention égale aux arguments pro et antivaccination. Ceci est compatible avec une stratégie de promotion de la discorde sur une série de sujets controversés — une tactique connue employée par les comptes trolls russes.20,26 De telles stratégies peuvent saper la santé publique : la normalisation de ces débats peut amener le public à remettre en question le consensus scientifique de longue date concernant l'efficacité du vaccin.13 En effet, plusieurs arguments antivaccins prétendent représenter les deux côtés du débat — comme les tactiques

utilisées par les trolls identifiés dans cette étude, tout en communiquant simultanément un virus clair (c.-à-d. , un sens de base). Nous avons récemment constaté que cette stratégie était efficace pour propager des articles de presse à travers les médias sociaux dans le contexte de l'épidémie de rougeole à Disneyland en 201532.

Distributeurs commerciaux et malveillants

Contrairement aux comptes trolls, les pollueurs de contenu (c.-à-d. Les diffuseurs de logiciels malveillants, le contenu commercial non sollicité et d'autres documents perturbateurs qui violent généralement les conditions de service de Twitter)21 publient des messages antivaccins 75 % plus souvent que l'utilisateur moyen de Twitter nonbot. Cela suggère que les opposants au vaccin peuvent diffuser des messages à l'aide de réseaux de bots qui sont principalement conçus pour la commercialisation. En revanche, les spambots,3,4 qui peuvent être facilement reconnus comme non humains, sont moins susceptibles de promouvoir un programme antivaccin que ne sont pasbots. Notamment, les pollueurs de contenu et les spambots traditionnels sont tous deux moins susceptibles de discuter de maladies évitables par la vaccination que l'utilisateur moyen de Twitter, ce qui suggère que lorsqu'ils tweetent des messages pertinents pour le vaccin, leur attention spécifique est sur les vaccins en soi, plutôt que sur les virus qui en ont besoin. Ainsi, il n'est pas clair dans quelle mesure leur promotion de contenu vaccinal est motivée par un véritable sentiment antivaccin ou est utilisée comme une tactique conçue pour augmenter les taux de clics en propageant du contenu motivationnel (« clickbait »).

Comptes non identifiés

Plusieurs comptes n'ont pas pu être identifiés positivement comme des bots ou des humains en raison des scores intermédiaires ou indisponibles de Botomètre. Ces comptes, qui représentent ensemble 93 % de notre échantillon aléatoire provenant du flux vaccinal, ont tweeté des contenus à la fois plus polarisés et plus opposés à la vaccination que celui du compte moyen de nonbot. Bien que la provenance de leurs tweets ne soit pas claire, nous spéculons que ces comptes peuvent posséder une proportion plus élevée de trolls ou de cyborgs — des comptes nominalement contrôlés par des utilisateurs humains qui, à l'occasion, sont repris par des bots ou présentent autrement des comportements bot-like ou malveillants.15 les comptes Cyborg sont plus susceptibles de tomber dans cette gamme moyenne parce qu'ils n'affichent que des comportements comme des bots parfois. Cette gamme moyenne est également susceptible de contenir des tweets de bots plus sophistiqués qui sont conçus pour imiter plus étroitement les comportements humains.

Enfin, les trolls, qui présentent des comportements malveillants encore opérés par des humains, sont également susceptibles de se situer dans cette fourchette moyenne. Cela suggère que proportionnellement plus de tweets antivaccins peuvent être générés par des comptes en utilisant une approche semi-automatique quelque peu sophistiquée pour éviter la détection. Cela crée la fausse impression d'un débat populaire sur l'efficacité du vaccin, une technique connue sous le nom d'« astroturfing »17 (comme dans les tweets #vaccineus indiqués dans la case à la page 1383). Il y a certainement des comptes humains standard qui relèvent également de cette fourchette moyenne. Bien que les limitations technologiques nous empêchent de tirer des conclusions définitives sur ces types de comptes, le fait que les tweets de milieu de gamme ont tendance à publier proportionnellement plus de messages antivaccin suggère fortement que ces messages antivaccin peuvent être diffusés à des taux plus élevés par une combinaison d'acteurs malveillants (bots, trolls, cyborgs, et les utilisateurs humains) qui sont difficiles à distinguer les uns des autres.

Cette interprétation est étayée par le fait que les utilisateurs de cette gamme intermédiaire avaient tendance à produire plus de tweets, et en particulier des tweets antivaccinaires, par compte, suggérant que les militants antivaccins peuvent utiliser de préférence ces canaux. En outre, les utilisateurs dont les comptes avaient été supprimés ont publié des messages plus polarisés par utilisateur et étaient également beaucoup plus susceptibles de publier des messages antivaccin. Bien que les raisons de la suppression de compte varient, le mouvement récent par Twitter pour supprimer les bots,33,34 trolls,20 cyborgs, et d'autres contrevenants des conditions de service de Twitter suggère que ces contrevenants peuvent être surreprésentés parmi les comptes supprimés dans notre échantillon. Nous ne pouvons pas prétendre que tous, ou même la plupart, des comptes avec des scores bot inconnus sont des acteurs malveillants;

cependant, nous nous attendons à ce qu'une plus grande proportion d'acteurs malveillants soient présents dans ce sous-ensemble des données. En revanche, les comptes échantillonnés au hasard qui étaient facilement identifiables comme des bots ont généré des tweets plus neutres, mais pas polarisés par compte. On peut supposer que les comptes qui sont manifestement automatisés sont plus fréquemment utilisés pour diffuser du contenu comme les nouvelles et peuvent ne pas être considérés comme des sources crédibles d'information antivaccin de base.

Incidences sur la santé publique

Les données de l'enquête montrent un consensus général quant à l'efficacité des vaccins dans la population générale.35 Conformément à ces résultats, les comptes peu susceptibles d'être des bots sont beaucoup moins susceptibles de promouvoir la teneur polarisée et antivaccine. Néanmoins, les bots et les trolls sont activement impliqués dans le discours de santé publique en ligne, biaisant les discussions sur la vaccination. Il s'agit d'une connaissance essentielle pour les communicateurs de risque, d'autant plus que ni les membres du public ni les approches algorithmiques ne peuvent être en mesure d'identifier facilement les bots, les trolls ou les cyborgs.

Le comportement malveillant en ligne varie selon le type de compte. Les trolls russes et les bots sophistiqués font la promotion des récits pro et antivaccination. Ce comportement est compatible avec une stratégie de promotion de la discorde politique. Les bots et les trolls retweetent ou modifient fréquemment le contenu des utilisateurs humains. Ainsi, les messages bien intentionnés contenant du contenu provaccine peuvent avoir l'effet involontaire de « nourrir les trolls », donnant la fausse impression de légitimité aux deux parties, surtout si ce contenu s'engage directement dans le discours antivaccination. En supposant que les comptes de bot et de troll cherchent à générer un nombre à peu près égal de tweets pour les deux parties, limiter l'accès au contenu provaccine pourrait également réduire l'incitation à afficher du contenu antivaccin.

En revanche, les comptes qui sont connus pour distribuer des logiciels malveillants et du contenu commercial sont plus susceptibles de promouvoir les messages antivaccination, ce qui suggère que les défenseurs antivaccins peuvent utiliser des infrastructures préexistantes de réseaux de bots pour promouvoir leur ordre du jour. Ces comptes peuvent également utiliser la nature convaincante du contenu antivaccin comme clickbait pour augmenter les recettes publicitaires et exposer les utilisateurs à des logiciels malveillants. Face à un tel contenu, les responsables des communications de santé publique peuvent envisager de souligner que la crédibilité de la source est douteuse et que les utilisateurs exposés à un tel contenu peuvent être plus susceptibles de rencontrer des logiciels malveillants. La teneur en antivaccin peut augmenter les risques d'infection par les virus informatiques et biologiques.

La plus forte proportion de contenu antivaccin est générée par des comptes avec des scores de bot inconnus ou intermédiaires. Bien que nous spéculions que cet ensemble de comptes contient des bots plus sophistiqués, trolls, et cyborgs, leur provenance est finalement inconnue. Par conséquent, au-delà de tenter d'empêcher les bots de diffuser des messages sur les médias sociaux, les praticiens de la santé publique devraient se concentrer sur la lutte contre les messages eux-mêmes tout en ne nourrissant pas les trolls. Il s'agit d'un domaine mûr pour la recherche future, qui pourrait inclure de souligner qu'une proportion importante de messages antivaccination sont organisés « astroturf » (c.-à-d. Pas de base) et d'autres messages de base qui mettent les messages antivaccins dans leur contexte approprié.

TRANSLATED VERSION: GERMAN

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

ÜBERSETZTE VERSION: DEUTSCH

Hier ist eine ungefähre Übersetzung der oben vorgestellten Ideen. Dies wurde getan, um ein allgemeines Verständnis der in dem Dokument vorgestellten Ideen zu vermitteln. Bitte entschuldigen Sie alle grammatikalischen Fehler und machen Sie die ursprünglichen Autoren nicht für diese Fehler verantwortlich.

EINLEITUNG

Gesundheitsbezogene Missverständnisse, Fehlinformationen und Desinformation, die über soziale Medien verbreitet werden und eine Bedrohung für die öffentliche Gesundheit darstellen.1 Trotz erheblicher Möglichkeiten, die Verbreitung von Fakten zu ermöglichen,2 werden soziale Medien häufig missbraucht, um schädliche Gesundheitsinhalte zu verbreiten3, einschließlich nicht überprüfter und fehlerhafter Informationen über Impfstoffe.1,4 Dies verringert potenziell die Aufnahmeraten von Impfstoffen und erhöht das Risiko globaler Pandemien, insbesondere unter den schwächsten.5 Einige dieser Informationen sind motiviert.1 : Skeptiker nutzen Online-Plattformen, um die Verweigerung von Impfstoffen zu befürworten.6 Antiimpfsbefürworter haben eine signifikante Präsenz in sozialen Medien,6 mit bis zu 50 % der Tweets über Impfungen, die Impfgegner enthalten.7

Die Verbreitung dieses Inhalts hat Folgen: Die Exposition gegenüber negativen Informationen über Impfstoffe ist mit einer erhöhten Impfverzögerung und Verzögerung verbunden.8–10 Impfzögernde Eltern wenden sich eher an das Internet, um Informationen zu erhalten, und weniger wahrscheinlich, gesundheitsdienstleistern und Gesundheitsexperten zu diesem Thema zu vertrauen.9,11 Die Exposition gegenüber der Impfstoffdebatte könnte darauf hindeuten, dass es keinen wissenschaftlichen Konsens gibt, dass es keinen wissenschaftlichen Konsens gibt. 12,13 Darüber hinaus unterstreichen die jüngsten Wiederaufstände von Masern, Mumps und Keuchhusten sowie die erhöhte Sterblichkeit durch durch Impfungen vermeidbare Krankheiten wie Grippe und virale Lungenentzündung14 die Bedeutung der Bekämpfung von Online-Fehlinformationen über Impfstoffe.

Viele Gesundheitsfehlinformationen können von "Bots"15 – Konten, die die Verbreitung von Inhalten automatisieren – und "Trolls"16 – Personen, die ihre Identität falsch darstellen, mit der Absicht, Zwietracht zu fördern, verbreitet werden. Eine häufig verwendete Online-Desinformationsstrategie, Amplifikation,17 versucht, Eindrücke von falscher Äquivalenz oder Konsens durch den Einsatz von Bots und Trollen zu schaffen. Wir versuchen zu verstehen, welche Rolle sie, wenn überhaupt, bei der Förderung von Impfinhalten spielen.

Die Bemühungen, zu dokumentieren, wie nicht autorisierte Benutzer – einschließlich Bots und Trolle – den Online-Diskurs über Impfstoffe beeinflusst haben, waren begrenzt. Darpa es (the US Defense Advanced Research Projects Agency) 2015 Bot Challenge beschuldigte Forscher, "Einflussbots" auf Twitter in einem Strom von Impf-Tweets zu identifizieren. Die Teams identifizierten effektiv Bot-Netzwerke, die darauf abzielten, Fehlinformationen über Impfstoffe zu verbreiten18, aber die öffentliche Gesundheit übersah die Auswirkungen dieser Ergebnisse weitgehend. Vielmehr konzentrierte sich die Öffentliche Gesundheitsforschung auf die Bekämpfung von Online-Impfinhalten, wobei weniger auf die Akteure ausgerichtet ist, die diese Inhalte produzieren und fördern.1,19 Daher bleibt die Rolle der Online-Aktivitäten von Bots und Trollen im Zusammenhang mit Impfungen unklar.

Wir berichten über die Ergebnisse einer retrospektiven Beobachtungsstudie, die die Auswirkungen von Bots und Trollen auf den Online-Impfdiskurs auf Twitter bewertet. Anhand eines Satzes von 1 793 690 Tweets, die vom 14. Juli 2014 bis zum 26. September 2017 gesammelt wurden, quantifizierten wir die Auswirkungen bekannter und mutmaßlicher Twitter-Bots und -Trolle auf die Verstärkung polarisierender und impfender Nachrichten. Ergänzt wird diese Analyse durch eine qualitative Studie über #vaccinateus – einen Twitter-Hashtag, der Zwietracht fördern soll, indem Impfungen als politisches Keilproblem verwendet werden. #vaccinateus Tweets wurden eindeutig mit russischen Troll-Konten identifiziert, die mit der Internet Research Agency verbunden sind – einem Unternehmen, das von der russischen Regierung unterstützt wird, das sich auf Online-Einflussoperationen spezialisiert hat.20 So sind Gesundheitskommunikationen "bewaffnet": Fragen der öffentlichen Gesundheit, wie Impfungen, werden in Versuche einbezogen, Fehlinformationen und Desinformation durch ausländische Mächte zu verbreiten.

Darüber hinaus twitter Bots Verteilen von Malware und kommerziellen Inhalten (d. H. Spam) Maskerade als menschliche Nutzer, um Anti-Impfstoff-Nachrichten zu verteilen. Ganze 93% der Tweets über Impfstoffe werden von Accounts generiert, deren Herkunft als weder Bots noch als menschliche Nutzer überprüft werden kann, die noch böswilliges Verhalten zeigen. Diese nicht identifizierten Konten twittern bevorzugt Anti-Impfstoff-Fehlinformationen. Wir diskutieren die Auswirkungen auf die Online-Kommunikation im Bereich der öffentlichen Gesundheit.

SCHLUSSFOLGERUNG

Die Ergebnisse deuten darauf hin, dass Twitter-Bots und Trolle einen erheblichen Einfluss auf die Online-Kommunikation über Impfungen haben. Die Art dieser Auswirkungen unterscheidet sich je nach Kontotyp.

Russische Trolle

Russische Trolle und ausgeklügelte Twitter-Bots posten Inhalte über Impfungen mit deutlich höheren Raten als der durchschnittliche Nutzer. Inhalte aus diesen Quellen geben den Argumenten der Pro- und Antiimpfung die gleiche Aufmerksamkeit. Dies steht im Einklang mit einer Strategie zur Förderung von Zwietracht in einer Reihe kontroverser Themen – einer bekannten Taktik, die von russischen Trollkonten angewandt wird.20,26 Solche Strategien können die öffentliche Gesundheit untergraben: Die Normalisierung dieser Debatten kann die Öffentlichkeit dazu veranlassen, den langjährigen wissenschaftlichen Konsens über die Wirksamkeit von Impfstoffen in Frage zu stellen.13 Tatsächlich behaupten mehrere Impfargumente, beide Seiten der Debatte zu repräsentieren – wie die Taktik, die von den in dieser Studie identifizierten Trollen verwendet wird – und gleichzeitig einen klaren wissenschaftlichen Konsens in Bezug auf die Wirksamkeit von Impfstoffen zu kommunizieren.13 Tatsächlich behaupten mehrere Impfbekämpfungsargumente, beide Seiten der Debatte zu repräsentieren – wie die Taktik, die von den in dieser Studie identifizierten Trollen verwendet wird – und gleichzeitig einen klaren Wissenschaftlichen Konsens zu kommunizieren (d.h.). , eine untere Bedeutung). Kürzlich haben wir festgestellt, dass diese Strategie wirksam war, um Nachrichtenartikel über soziale Medien im Zusammenhang mit dem Ausbruch der Disneyland-Masern 2015 zu verbreiten.32

Kommerzielle und Malware Distributoren

Im Gegensatz zu Troll-Accounts, Content-Verschmutzer (d. H. Verbreiter von Malware, unerwünschte kommerzielle Inhalte, und andere störende Material, das in der Regel Twitters Nutzungsbedingungen verletzt)21 poste Anti-Impfstoff-Nachrichten 75% häufiger als der durchschnittliche Nicht-Bot Twitter-Nutzer. Dies deutet darauf hin, dass Impfgegner Nachrichten über Bot-Netzwerke verbreiten können, die in erster Linie für Marketing konzipiert sind. Im Gegensatz dazu sind Spambots,3,4, die leicht als nicht menschlich erkannt werden können, weniger wahrscheinlich, eine Anti-Impfstoff-Agenda zu fördern, als Nichtbots. Bemerkenswert ist, dass Inhaltsverschmutzer und herkömmliche Spambots weniger wahrscheinlich über durch Impfung vermeidbare Krankheiten diskutieren als der durchschnittliche Twitter-Nutzer, was darauf hindeutet, dass, wenn sie impfrelevante Nachrichten twittern, ihr spezifischer Fokus auf Impfstoffen an sich liegt, und nicht auf den Viren, die sie benötigen. Es ist daher unklar, inwieweit ihre Förderung von impfenbezogenen Inhalten von einer echten Impfgegner-Stimmung getrieben wird oder als Taktik verwendet wird, um die Klickraten durch die Verbreitung von Motivationsinhalten zu erhöhen ("Clickbait").

Nicht identifizierte Konten

Mehrere Konten konnten aufgrund von zwischengeschalteten oder nicht verfügbaren Botometer-Scores weder als Bots noch als Menschen positiv identifiziert werden. Diese Konten, die zusammen 93% unserer Zufallsstichprobe aus dem Impfstoffstrom ausmachen, twitterten Inhalte, die sowohl polarisiert er als auch mehr gegen Impfungen waren als die des durchschnittlichen Nonbot-Kontos. Obwohl die Herkunft ihrer Tweets unklar ist, spekulieren wir, dass diese Konten einen höheren Anteil an Trollen oder Cyborgs

besitzen können – Konten, die nominell von menschlichen Benutzern kontrolliert werden, die gelegentlich von Bots übernommen werden oder anderweitig botartiges oder bösartiges Verhalten aufweisen.15 Cyborg-Konten fallen eher in diesen mittleren Bereich, weil sie nur manchmal botähnliche Verhaltensweisen zeigen. Dieser mittlere Bereich enthält wahrscheinlich auch Tweets von anspruchsvolleren Bots, die entwickelt wurden, um menschliches Verhalten genauer nachzuahmen.

Schließlich werden Trolle– die böswillige Verhaltensweisen aufweisen, die noch von Menschen betrieben werden – wahrscheinlich auch in diesen mittleren Bereich fallen. Dies deutet darauf hin, dass proportional mehr Impf-Tweets von Accounts generiert werden können, die einen etwas ausgeklügelten halbautomatischen Ansatz verwenden, um eine Erkennung zu vermeiden. Dies erzeugt den falschen Eindruck einer Basisdebatte über die Wirksamkeit von Impfstoffen – eine Technik, die als "Astroturfing"17 bekannt ist (wie in den #vaccineus Tweets, die im Kasten auf Seite 1383 gezeigt werden). Es gibt sicherlich Standard-Human-Accounts, die auch innerhalb dieses mittleren Bereichs fallen. Obwohl technologische Beschränkungen uns daran hindern, endgültige Schlussfolgerungen über diese Kontotypen zu ziehen, deutet die Tatsache, dass Tweets aus dem mittleren Bereich dazu neigen, proportional mehr Impfbotschaften zu posten, stark darauf hin, dass diese Impfschutzbotschaften mit höheren Raten durch eine Kombination von böswilligen Akteuren (Bots, Trolle, Cyborgs und menschliche Nutzer) verbreitet werden können, die schwer voneinander zu unterscheiden sind.

Diese Interpretation wird durch die Tatsache unterstützt, dass Nutzer innerhalb dieses mittleren Bereichs dazu neigten, mehr Tweets und insbesondere Impf-Tweets pro Konto zu produzieren, was darauf hindeutet, dass Impfgegner diese Kanäle bevorzugt nutzen könnten. Darüber hinaus veröffentlichten Benutzer, deren Konten gelöscht worden waren, mehr polarisierte Nachrichten pro Benutzer und waren auch deutlich wahrscheinlicher, Impfnachrichten zu posten. Obwohl die Gründe für die Löschung von Konten variieren, deutet die jüngste Bewegung von Twitter zur Entfernung von Bots, 33,34 Trollen, 20 Cyborgs und anderen Verletzern der Twitter-Nutzungsbedingungen darauf hin, dass diese Verletzer unter den gelöschten Accounts in unserer Stichprobe überrepräsentiert sein könnten. Wir können nicht behaupten, dass alle oder sogar die meisten Konten mit unbekannten Bot-Scores böswillige Akteure sind; Wir erwarten jedoch, dass ein höherer Anteil böswilliger Akteure in dieser Teilmenge der Daten vorhanden ist. Im Gegensatz dazu generierten zufällig abgetastete Accounts, die leicht als Bots identifizierbar waren, neutralere, aber nicht polarisierte Tweets pro Konto. Vermutlich werden Konten, die offensichtlich automatisiert sind, häufiger zur Verbreitung von Inhalten wie Nachrichten verwendet und werden möglicherweise nicht als glaubwürdige Quellen von Informationen zur Bekämpfung von Impfstoffen an der Basis angesehen.

Auswirkungen auf die öffentliche Gesundheit

Umfragedaten zeigen einen allgemeinen Konsens über die Wirksamkeit von Impfstoffen in der allgemeinen Bevölkerung.35 In Übereinstimmung mit diesen Ergebnissen sind Konten, die wahrscheinlich keine Bots sind, deutlich weniger wahrscheinlich, polarisierte und impfende Inhalte zu fördern. Nichtsdestotrotz sind Bots und Trolle aktiv am Online-Diskurs über die öffentliche Gesundheit beteiligt und verzerren die Diskussionen über Impfungen. Dies ist ein wichtiges Wissen für Risikokommunikatoren, insbesondere wenn man bedenkt, dass weder Mitglieder der Öffentlichkeit noch algorithmische Ansätze in der Lage sein können, Bots, Trolle oder Cyborgs leicht zu identifizieren.

Bösartiges Onlineverhalten variiert je nach Kontotyp. Russische Trolle und ausgeklügelte Bots fördern sowohl Pro- als auch Antiimpfungsnarrative. Dieses Verhalten steht im Einklang mit einer Strategie zur Förderung politischer Zwietracht. Bots und Trolle retweeten oder ändern häufig Inhalte von menschlichen Nutzern. So können gut gemeinte Beiträge, die Pro-Impfstoff-Gehalt enthalten, den unbeabsichtigten Effekt haben, "die Trolle zu füttern", was beiden Seiten den falschen Eindruck von Legitimität vermittelt, insbesondere wenn dieser Inhalt direkt mit dem Anti-Impfdiskurs zugreift. Die Annahme von Bot- und Troll-Konten soll ungefähr gleich viele Tweets für beide Seiten generieren, was die Beschränkung des Zugriffs auf Pro-Impfstoff-Inhalte potenziell auch den Anreiz verringern könnte, Impfinhalte zu veröffentlichen.

Im Gegensatz dazu sind Konten, die dafür bekannt sind, Malware und kommerzielle Inhalte zu verbreiten, eher dazu geeignet, Impfbotschaften zu fördern, was darauf hindeutet, dass Impfgegner bereits vorhandene Infrastrukturen von Bot-Netzwerken nutzen können, um ihre Agenda zu fördern. Diese Konten können auch die zwingende Natur von Impfschutz-Inhalten als Clickbait verwenden, um Werbeeinnahmen zu steigern und Benutzer Malware auszusetzen. Wenn sie mit solchen Inhalten konfrontiert werden, können Beamte der öffentlichen Gesundheit in Betracht ziehen, zu betonen, dass die Glaubwürdigkeit der Quelle zweifelhaft ist und dass Benutzer, die solchen Inhalten ausgesetzt sind, eher auf Malware stoßen. Impfschutzkannien können das Infektionsrisiko sowohl durch Computer- als auch durch biologische Viren erhöhen.

Der höchste Anteil an Impfinhalten wird durch Konten mit unbekannten oder mittleren Bot-Scores generiert. Obwohl wir spekulieren, dass diese Reihe von Konten anspruchsvollere Bots, Trolle und Cyborgs enthält, ist ihre Herkunft letztlich unbekannt. Daher sollten sich die Praktiker der öffentlichen Gesundheit nicht nur versuchen, Bots daran zu hindern, Botschaften über soziale Medien zu verbreiten, sondern sich auf die Bekämpfung der Botschaften selbst konzentrieren, ohne die Trolle zu füttern. Dies ist ein reifer Bereich für zukünftige Forschung, wozu auch die Betonung gehören könnte, dass ein erheblicher Teil der Anti-Impfbotschaften "Astroturf" (d. H. Nicht von der Basis) und andere Untermundbotschaften sind, die Impfbotschaften in ihren richtigen Kontext stellen.

TRANSLATED VERSION: PORTUGUESE

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

VERSÃO TRADUZIDA: PORTUGUÊS

Aqui está uma tradução aproximada das ideias acima apresentadas. Isto foi feito para dar uma compreensão geral das ideias apresentadas no documento. Por favor, desculpe todos os erros gramaticais e não responsacule os autores originais responsáveis por estes erros.

INTRODUÇÃO

Equívocos relacionados com a saúde, desinformação e desinformação difundidas nas redes sociais, representando uma ameaça para a saúde pública.1 Apesar do potencial significativo para permitir a divulgação de informação factual, 2 redes sociais são frequentemente abusadas para espalhar conteúdo nocivo para a saúde,3 incluindo informação não verificada e errónea sobre vacinas.1,4 Isto potencialmente reduz as taxas de absorção de vacinas e aumenta os riscos de pandemias globais, especialmente entre os mais vulneráveis.5 Algumas desta informação são motivadas. : os céticos usam plataformas online para defender a recusa da vacina.6 Os defensores da antivaccine têm uma presença significativa nas redes sociais,6 com até 50% dos tweets sobre vacinação contendo crenças antivaccine.7

A proliferação deste conteúdo tem consequências: a exposição a informações negativas sobre vacinas está associada ao aumento da hesitação da vacina e ao atraso.8-10 Os pais hesitantes em vacinas são mais propensos a recorrer à Internet para obter informações e menos propensos a confiar em prestadores de cuidados de saúde e especialistas em saúde pública sobre o assunto.9,11 A exposição ao debate sobre a vacina pode sugerir que não existe consenso científico. , sacudindo a confiança na vacinação.12,13 Adicionalmente, os recentes ressurgimentos de sarampo, papeira e pertussis e o aumento da mortalidade por doenças evitáveis pela vacinação, como a gripe e a pneumonia viral14, sublinham a importância do combate à desinformação online sobre as vacinas.

Muita desinformação sobre saúde pode ser promulgada por "bots" 15 — contas que automatizam a promoção de conteúdos — e "trolls" 16 — indivíduos que deturpam as suas identidades com a intenção de promover a discórdia. Uma estratégia de desinformação online comumente usada, a amplificação,17

procura criar impressões de falsa equivalência ou consenso através do uso de bots e trolls. Procuramos perceber qual o papel que desempenham na promoção de conteúdos relacionados com a vacinação.

Os esforços para documentar como os utilizadores não autorizados , incluindo bots e trolls, influenciaram o discurso online sobre as vacinas têm sido limitados. A DARPA's (agência norte-americana de projetos de investigação avançada de investigação) 2015 Bot Challenge acusou os investigadores de identificarem "bots de influência" no Twitter num fluxo de tweets relacionados com a vacina. As equipas identificaram efetivamente redes de bots destinadas a espalhar desinformação sobre vacinas, 18, mas a comunidade de saúde pública ignorou largamente as implicações destas descobertas. Pelo contrário, a investigação em saúde pública tem-se centrado no combate ao conteúdo de antivaccine online, com menos foco nos atores que produzem e promovem este conteúdo.1,19 Assim, o papel da atividade online de bots e trolls relativo à vacinação permanece incerto.

Relatamos os resultados de um estudo observacional retrospetivo que avalia o impacto dos bots e trolls no discurso da vacina online no Twitter. Utilizando um conjunto de 1 793 690 tweets recolhidos entre 14 de julho de 2014 e 26 de setembro de 2017, quantificámos o impacto de conhecidos e suspeitos bots e trolls do Twitter na ampliação de mensagens polarizadas e antivaccine. Esta análise é complementada por um estudo qualitativo de #vaccinateus - uma hashtag do Twitter projetada para promover a discórdia usando a vacinação como uma questão de cunha política. #vaccinateus tweets foram identificados exclusivamente com contas troll russas ligadas à Internet Research Agency — uma empresa apoiada pelo governo russo especializada em operações de influência online.20 Assim, as comunicações de saúde tornaram-se "armadas": as questões de saúde pública, como a vacinação, estão incluídas em tentativas de espalhar desinformação e desinformação por potências estrangeiras. Além disso, os bots do Twitter que distribuem malware e conteúdo comercial (ou seja, spam) mascaram-se como utilizadores humanos para distribuir mensagens de antivaccine. 93% dos tweets sobre vacinas são gerados por contas cuja proveniência não pode ser verificada como nem bots nem utilizadores humanos que exibam comportamentos maliciosos. Estas contas não identificadas preferencialmente tweet antivaccine desinformação. Discutimos implicações para as comunicações de saúde pública online.

CONCLUSÃO

Os resultados sugerem que os bots e trolls do Twitter têm um impacto significativo nas comunicações online sobre a vacinação. A natureza deste impacto difere por tipo de conta.

Trolls russos

Trolls russos e bots sofisticados do Twitter publicam conteúdo sobre vacinação a taxas significativamente mais altas do que o utilizador médio. O conteúdo destas fontes dá igual atenção aos argumentos pró e antivaccination. Isto é consistente com uma estratégia de promoção da discórdia em vários tópicos controversos — uma conhecida tática utilizada pelas contas troll russas.20,26 Tais estratégias podem minar a saúde pública: normalizar estes debates pode levar o público a questionar um consenso científico de longa data sobre a eficácia da vacina.13 Na verdade, vários argumentos antivaccine afirmam representar ambos os lados do debate — como as táticas usadas pelos trolls identificados neste estudo — ao mesmo tempo que comunicam simultaneamente uma clara essência (ou seja. , um significado de fundo). Recentemente, descobrimos que esta estratégia foi eficaz para a propagação de artigos noticiosos através das redes sociais no contexto do surto de sarampo da Disneyland 2015.32

Distribuidores comerciais e malware

Ao contrário das contas troll, os poluidores de conteúdo (isto é, os difusores de malware, conteúdos comerciais não solicitados e outros materiais disruptivos que normalmente violam os termos de serviço do Twitter)21 publicam mensagens de antivaccine 75% mais frequentemente do que o utilizador médio não-bot do Twitter. Isto sugere que os oponentes da vacina podem divulgar mensagens usando redes bot que são projetadas principalmente para marketing. Em contraste, os spambots,3,4 que podem ser facilmente reconhecidos como não humanos, são menos propensos a promover uma agenda de antivaccine do que os

não-bots. Notavelmente, os poluidores de conteúdo e os spambots tradicionais são ambos menos propensos a discutir doenças evitáveis pela vacinação do que o utilizador médio do Twitter, sugerindo que quando fazem mensagens de tweet relevantes para a vacina, o seu foco específico é nas vacinas em si, em vez dos vírus que as exigem. Assim, não é claro até que ponto a sua promoção de conteúdos relacionados com a vacina é impulsionada por um verdadeiro sentimento de ativaccine ou é usada como uma tática concebida para aumentar as taxas de clique através da propagação de conteúdo motivacional ("clickbait").

Contas não identificadas

Várias contas não puderam ser identificadas positivamente como bots ou humanos devido a pontuações botómetros intermédias ou indisponíveis. Estas contas, que juntos constituem 93% da nossa amostra aleatória do fluxo da vacina, twittaram conteúdo que era tanto mais polarizado e mais oposto à vacinação do que a média da conta nonbot. Embora a proveniência dos seus tweets não seja clara, especulamos que estas contas podem possuir uma maior proporção de trolls ou cyborgs — contas nominalmente controladas por utilizadores humanos que são, ocasionalmente, assumidas por bots ou de outra forma exibem comportamentos semelhantes a bots ou maliciosos.15 as contas cyborg são mais propensas a cair nesta gama média porque só exibem comportamentos semelhantes a bots às vezes. Esta gama média também é suscetível de conter tweets de bots mais sofisticados que são projetados para imitar mais de perto comportamentos humanos.

Finalmente, os trolls - exibindo comportamentos maliciosos mas operados por humanos - também são suscetíveis de se enquadrarem nesta gama média. Isto sugere que proporcionalmente mais tweets de antivaccine podem ser gerados por contas usando uma abordagem semi-autóma um pouco sofisticada para evitar a deteção. Isto cria a falsa impressão do debate popular sobre a eficácia da vacina - uma técnica conhecida como "astroturfing" 17 (como nos tweets #vaccineus mostrados na caixa na página 1383). Há certamente contas humanas padrão que também se enquadram neste intervalo médio. Embora as limitações tecnológicas nos impeçam de tirar conclusões definitivas sobre estes tipos de conta, o facto de os tweets de médio alcance tendem a publicar mensagens proporcionalmente mais antivaccine sugere fortemente que estas mensagens de antivaccine podem ser divulgadas a taxas mais elevadas através de uma combinação de atores maliciosos (bots, trolls, ciborgues e utilizadores humanos) que são difíceis de distinguir uns dos outros.

Esta interpretação é apoiada pelo facto de os utilizadores dentro desta gama intermédia tendem a produzir mais tweets, e especialmente tweets de antivaccine, por conta, sugerindo que os ativistas da antivaccine podem utilizar preferencialmente estes canais. Além disso, os utilizadores cujas contas tinham sido eliminadas publicavam mensagens mais polarizadas por utilizador e também eram significativamente mais propensos a publicar mensagens de antivaccine. Embora as razões para a eliminação de contas variem, o recente movimento do Twitter para remover bots,33,34 trolls,20 cyborgs, e outros violadores dos termos de serviço do Twitter sugerem que estes violadores podem estar sobre-representados entre as contas eliminadas na nossa amostra. Não podemos afirmar que todas, ou mesmo a maioria, contas com pontuações de bots desconhecidas são atores mal-intencionados; no entanto, esperamos que uma maior proporção de atores maliciosos estejam presentes neste subconjunto dos dados. Em contraste, contas de amostras aleatórias que eram facilmente identificáveis como bots geraram tweets mais neutros, mas não polarizados por conta. Presumivelmente, contas que são obviamente automatizadas são mais frequentemente usadas para divulgar conteúdos como notícias e podem não ser consideradas fontes credíveis de informação de antivaccine popular.

Implicações para a saúde pública

Os dados do inquérito mostram um consenso geral sobre a eficácia das vacinas na população em geral.35 Consistente com estes resultados, as contas improváveis de serem bots são significativamente menos propensos a promover o teor polarizado e a anacrâneo. No entanto, bots e trolls estão ativamente envolvidos no discurso de saúde pública online, distorindo as discussões sobre a vacinação. Este é um conhecimento vital para os comunicadores de risco, especialmente tendo em conta que nem membros do público nem abordagens algorítmicas podem ser capazes de identificar facilmente bots, trolls ou ciborgues.

O comportamento online malicioso varia por tipo de conta. Trolls russos e bots sofisticados promovem narrativas pró e antivaccination. Este comportamento é consistente com uma estratégia de promoção da discórdia política. Bots e trolls frequentemente retweetam ou modificam conteúdo de utilizadores humanos. Assim, publicações bem intencionadas que contenham conteúdo comprovado podem ter o efeito não intencional de "alimentar os trolls", dando a falsa impressão de legitimidade a ambas as partes, especialmente se este conteúdo se envolver diretamente com o discurso de antivaccination. As contas de bot e troll presumindo-se procuram gerar um número aproximadamente igual de tweets para ambos os lados, limitando o acesso ao conteúdo provaccine poderia potencialmente também reduzir o incentivo para publicar conteúdo antivaccine.

Em contrapartida, as contas conhecidas por distribuir malware e conteúdo comercial são mais propensas a promover mensagens de antivaccination, sugerindo que os defensores da antivaccine podem usar infraestruturas pré-existentes de redes de bots para promover a sua agenda. Estas contas também podem usar a natureza convincente do conteúdo antivaccine como clickbait para aumentar as receitas publicitárias e expor os utilizadores a malware. Quando confrontados com este conteúdo, os funcionários das comunicações de saúde pública podem considerar sublinhar que a credibilidade da fonte é duvidosa e que os utilizadores expostos a esses conteúdos podem ser mais propensos a encontrar malware. O teor de antivaccine pode aumentar os riscos de infeção por vírus informáticos e biológicos.

A maior proporção de conteúdo de antivaccine é gerada por contas com pontuações de bots desconhecidas ou intermédias. Embora especulemos que este conjunto de contas contenha bots, trolls e ciborgues mais sofisticados, a sua proveniência é, em última análise, desconhecida. Por isso, para além de tentarem evitar que os bots espalhem mensagens através das redes sociais, os profissionais de saúde pública devem concentrar-se em combater as mensagens em si, sem alimentar os trolls. Esta é uma área madura para pesquisas futuras, que pode incluir sublinhar que uma parte significativa das mensagens de antivaccination são organizadas "astroturf" (isto é, não populares) e outras mensagens de fundo que colocam mensagens de antivaccine nos seus contextos adequados.