INPUTS FOR THEMATIC REPORT ON NEW INFORMATION TECHNOLOGIES, RACIAL EQUALITY, AND NON-DISCRIMINATION

15 November 2019

INTRODUCTION

The International Movement Against All Forms of Discrimination and Racism (IMADR) is an international non-profit, non-governmental human rights organisation devoted to eliminating discrimination and racism, forging international solidarity among discriminated groups and advancing the international human rights system. IMADR is grateful to the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, Ms. E. Tendayi Achiume, for providing stakeholders the opportunity to contribute to her next thematic report on new information technologies, racial equality, and non-discrimination.

With the rapid growth of social media, disinformation, and hate speech have become one of the major human rights concerns in the today's world. Social media have been increasingly questioned for its role in spreading disinformation, incitements to violence, distrust in media and democratic institutions, which were once celebrated as powerful tools for freedom of democracy.[1] For example, one study on the 2019 EU Parliamentary elections revealed that populist topics such as anti-immigration and Islamophobic contents, some of which were linked to Euroscepticism, political parties or leaders in the region, tended to lead to the most successful junk news stories in social media.[2] The study analysed that a number of those junk news stories associated Muslims and immigrants with reporting on terrorism or crimes such as sexual violence and honour killings.[3] Such junk news can fuel hostile prejudice against people holding the Islamic faith or those perceived as Muslims, migrants, as well as those with intersectional identities.

Many online instances of hate speech have been observed including incitements to discrimination, hatred and violence based on race, colour, descent, or national or ethnic origin, often combined with other characteristics such as religion or belief, gender, sexual orientation or gender identity. Indigenous peoples, minorities, migrants including asylum-seekers and refugees, are particularly targeted in racist expressions. While direct human engagement is in no doubt the major cause for manifestations of racism, this submission addresses the role of bots[4] in racist hate speech in the online space.

BOTS AND DISINFORMATION

Social media platforms are efficient at engaging a vast number of people, while at the same time allowing

[1] Samantha Bradshaw and Philip N. Howard. (2019). *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation*. Page 21

[2] Nahema Marchal, Bence Kollanyi, Lisa-Maria Neudert and Philip N. Howard. (2019). *Junk News During the EU Parliamentary Elections: Lessons from a Seven-Language Study of Twitter and Facebook*. Page 2

[3] Ibid

[4] This paper uses the following definition of bots – *"social media accounts that automate interaction with other users"*. See Ben Nimmo. (2019). *Measuring Traffic Manipulation on Twitter*. Page 7

personalised interactions with targeted individuals.[5] In particular, Twitter is a dominant social media tool which provides a space for politicians, lobbyists and activists in different parts of the world to promote communications and campaigns for their political purposes.[6] Yet, Twitter is proven to be vulnerable to traffic manipulation[7] due to its high level of data transparency and low level of user transparency.[8] Developers can request the access to Twitter's Application Programming Interface (API) which allows both professional and non-professional programmers to design apps to automate tweets, likes and follows.[9] Its low level of transparency for users makes it possible to create hundreds to thousands of individual accounts.[10] Hence, a significant number of bots can be easily created. Its vulnerability for traffic manipulation by small organised groups and bots has been well documented.[11] Such manipulation includes allowing targeted hashtags and phrases to appear in the "trending" lists and rapidly multiplying twitter traffic by a small number of users.[12] Those bot activities can mislead users about what dominant conversations are in social media.

The role of bots in manipulating conversations in social media is recognised by an increasing number of studies. For example, one research found that out of 70 countries, bots were used in 50 countries for social media manipulation campaigns in 2019.[13] Another research concluded that there is a significant tendency in the use of bots to widely disseminate false information.[14] Those accounts quickly promote viral posts in early stages and target influential social media users.[15] Bots take advantage of people's vulnerability to such manipulation, encouraging them to spread false information.[16] In other words, bots are often identified as powerful promoters of false and biased information.[17] In addition to promoting targeted hashtags into the "trending" lists, bots are also used to silence or intimidate other users.[18] For instance, it was reported that bots were used to overwhelm social media posts critical of the Government of Indonesia's policy over West Papua during the recent unrest.[19]

---

[5] Samantha Bradshaw and Philip N. Howard. (2018). *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation.* Page 4

[6] Ben Nimmo. (2019). *Measuring Traffic Manipulation on Twitter,* page 4

[7] Ibid. Page 6. Ben Nimmo defines 'Twitter traffic manipulation' as *"an attempt by a small group of users to generate a large flow of Twitter traffic, disproportionate to the number of users involved."*

[8] Ibid. Page 5

[9] Ibid

[10] Ibid

[11] Ibid. Page 4

[12] Ibid

[13] Samantha Bradshaw and Philip N. Howard. (2019). *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation.* Page 11

[14] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Alessandro Flammini, and Filippo Menczer. (2017). *The spread of fake news by social bots*. Page 1

[15] Ibid

[16] Ibid

[17] Ibid

[18] Ben Nimmo. (2019). *Measuring Traffic Manipulation on Twitter*. Page 7

[19] BBC News. (11 October 2019). *Papua unrest: Social media bots 'skewing the narrative'.* Last accessed on 11 November 2019: https://www.bbc.com/news/world-asia-49983667

BOTS AND RACIST HATE SPEECH

False and biased information spread by bots include messages that exploit racism, racial discrimination, xenophobia and related intolerance, which have been documented especially during election campaigns. One study found that a botnet campaign was responsible for promoting the term *"Minami Chosen"*, a term for South Korea with a discriminatory connotation often used by right-wing extremists, to one of the most common terms in Twitter at the time of the 2014 general election in Japan.[20] The word appeared in 12,389 tweets which were almost identical, and only 0.3% of them were retweets.[21] Those tweets' duplication ratio was extremely high (97%) which were distributed by 271 accounts.[22] In another instance, a Twitter bot kept posting information about Buraku neighbourhoods in Osaka in any one minute, despite multiple attempts to delete the posts.[23] The bot was clearly created with ill-intention to expose Buraku areas in order to provoke discrimination based on descent against Buraku people.

During the 2018 U.S. midterm elections, messages attacking the Jewish community were spread through the internet by automation.[24] Anti-Defamation League (ADL) found that 30%-40% of accounts which used degrading words against Jewish people were bots.[25] Prior to the Swedish general election in September 2018, 6% (1,429) of Twitter accounts discussing the national politics were identified as bots.[26] The same study found that the bots were tweeting about topics related to immigration and Islam including sharia and jihad more frequently than genuine accounts.[27] The Swedish Defence Research Agency recently reported that traditional authoritarian and nationalistic sentiments were employed by the majority of bots which were removed or suspended by Twitter.[28] In Poland, one report identified that a majority of Twitter bots it analysed were right-wing accounts, many of which promoted provocative and anti-immigration messages.[29]

Although some of the abovementioned bot activities may not amount to incitement to racial discrimination, hatred or violence, they suggest that bots are used to help spreading biased information against racialised groups such as minorities and migrants. Those bots promote discriminatory discourses to create a false impression that such narratives are mainstreamed in the online space. Thus, they nurture a hostile online environment against racialised communities which would fuel racist hate speech.

---

[20] Schäfer, F., Evert, S. and P. Heinrich. (2017). *Japan's 2014 General Election: Political Bots, Right-Wing Internet Activism, and Prime Minister Shinzō Abe's Hidden Nationalist Agenda.* Page 302

[21] Ibid

[22] Ibid

[23] Yasushi Kawaguchi. (2018). *Reality of Buraku Discrimination in the Internet Society [published in Japanese]. Page 30*

[24] Anti-Defamation League. (October 2018). *Computational Propaganda, Jewish-Americans and the 2018 Midterms: The Amplification of Anti-Semitic Harassment Online.* Page 4. Last accessed on 11 November 2019: https://www.adl.org/resources/reports/computational-propaganda-jewish-americans-and-the-2018-midterms-the-amplification

[25] Ibid. Page 11

[26] Johan Fernquist, Lisa Kaati and Ralph Schroeder. (2018). *Political Bots and the Swedish General Election.* Page 127

[27] Ibid. Page 128. TABLES V and VI

[28] Freja Hedman, Fabian Sivnert, Bence Kollanyi, Vidya Narayanan, Lisa-Maria Neudert, and Philip N. Howard. (2018). *News and Political Information Consumption in Sweden: Mapping the 2018 Swedish General Election on Twitter.* Page 2

[29] Robert Gorwa. (2017). *Computational Propaganda in Poland: False Amplifiers and the Digital Public Sphere.* Page 24

At the same time, there are initiatives to use bots to counter racist expressions in social media platforms. Each bot takes a different approach such as regularly posting contents on racial equality and counter-hate speech, sharing and liking posts to promote human rights values. Currently, those counter bots are not developed to have the capacity to engage with conversations in a genuine manner without human moderators. However, researchers have been looking into the possibility of creating bots which automatically intervene in hateful online conversations in an effective way.[30]

Furthermore, there have been challenges in regulating malicious bots while allowing other bots to counter racist hate speech. One notable example is a Twitter bot called "Impostor Buster" which was created to expose neo-Nazi accounts impersonating as members of minority groups. The account was repeatedly suspended by Twitter due to a large volume of complaints from the group it was targeting.[31] This example seems to be only the tip of iceberg of similar instances. It is becoming more evident that social media tools like bots are often not benefitting communities which they meant to empower.[32]

CONCLUSION

Social media platforms face challenges in tackling disinformation and racist hate speech. There are a growing number of evidences that bots play a considerable role in spreading biased and false information about racialised communities such as minorities and migrants. Some bots may directly incite racial discrimination, hatred and violence, while many others contribute to creating a toxic online environment by spreading racist narratives. They are also used to harass and supress other users. At the same time, bots are employed to counter racial hatred and promote equality and human rights. Nevertheless, balancing between allowing bots to empower marginalised communities and protecting them from malicious bot activities remains a challenge. There have been limited studies, especially from the human rights angle, on the role of bots in racist hate speech in the internet. It is critical for all stakeholders working for the protection and promotion of human rights to address this new challenge for racial equality and non-discrimination.

[30] Anna Bethke. (26 September 2019). *NLP Techniques to Intervene in Online Hate Speech*. Last accessed on 11 November 2019: https://www.intel.ai/nlp-techniques-to-intervene-in-online-hate-speech/#gs.c3odu8

[31] Yair Rosenberg. (27 December 2017). *Confessions of a Digital Nazi Hunter* in the New York Times. Last accessed on 11 November 2019: https://www.nytimes.com/2017/12/27/opinion/digital-nazi-hunter-trump.html

[32] Robert Gorwa and Douglas Guilbeault. (2018). *Unpacking the Social Media Bot: A Typology to Guide Research and Policy.* Page 15-16