REVIEWS

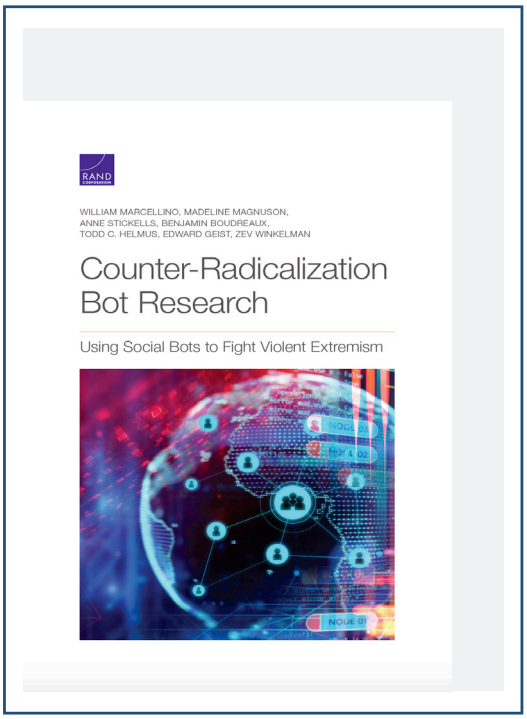
2021

التحالف الإسلامي العسكري لمحاربة الإرهاب
ISLAMIC MILITARY COUNTER TERRORISM COALITION

INTERNATIONAL REPORTS

COUNTER-RADICALIZATION BOT RESEARCH

USING SOCIAL BOTS TO FIGHT VIOLENT EXTREMISM



ISSUE
27

July

www.imctc.org

International Reports

Monthly Issue - Islamic Military Counter-Terrorism Coalition

Director General

Major General Mohammed bin Saeed Al-Moghedi

Secretary-General of the Islamic Military Counter Terrorism Coalition/Acting

Editor-in-Chief

Ashour Ibrahim Aljuhani

Director of Research and Studies Department

Brought to you by

TAOQ RESEARCH

TAOQ نـــوق

E-mail: info@taoqresearch.org

Phone: +966 114890124

COUNTER-RADICALIZATION BOT RESEARCH

USING SOCIAL BOTS TO FIGHT VIOLENT EXTREMISM

Co-authored by William Marcellino, Madeline Magnuson, Anne Stickells, Benjamin Boudreaux, Todd C. Helmus, Edward Geist, and Zev Winkelman, COUNTER-RADICALIZATION BOT RESEARCH: USING SOCIAL BOTS TO FIGHT VIOLENT EXTREMISM was issued by RAND Corporation to address the bot programs and assess the possibility of the US government employing such bot programs in combating extremism and terrorism.

The said report falls into five sections: an introduction to bot technology, bot technology review, uses of bot technology, ethical and legal considerations, implementation of bot technology, and recommendations for the US government. The report was based on several interviews with subject-matter experts, a review of the legal and ethical literature made available, case studies and associated impact, data collection, and messaging campaigns. It has a special attention attached to terrorist groups, such as Al-Qaeda and ISIS; it is applicable to similar groups.

Bot Types

Bots are interactive software deployed on social media, that work individually and automatically to better augment human efforts (the sender). Bots include Artificial Intelligence (AI) technologies, social cognition, and language abilities. They are used for various purposes in different methods, such as active users, providing them with bot-empowered information, or deluding users that a topic is widely supported, or creating electronic noise to distract users from an issue, or distracting users and misguiding them by giving alternative information or data.

Again, bots may directly mislead, construct false narratives, connect users of similar propensities, opinions, and interests, harass users to drive them away from the social media, befriend users to access target data, or persuade users that a given bot is a human user to prevent them from real user communication.

Extremist groups, such as ISIS and far-right organizations in the West have been able to use bot technology to spread their ideas via cyberspace, recruit new members, and expand their support. Hence, deploying this technique to limit the influence of such groups has become a must; bot technology is an effective tool for the authorities to combat fundamentalism and violent extremism. However, such employment depends on several factors, such as consideration of technical, legal, and ethical issues.

The use of digital bots or social bots came into reality at an early stage that dates to the 1980s and 1990s for limited purposes, such as gaming and chat room management. Some governments, armies, and politicians use bots to manipulate and tamper with public opinion and derail discussions off track on various social media. Twitter has admitted that about 23 million accounts are just bots. Equally important, Facebook indicated that fake accounts in one year accounted for 5 to 6% of total accounts.

The emergence of social media such as Facebook and Twitter and the integration with AI and machine learning technologies led to the current revolution in communication, which increased in various areas, such as politics and economics and other roles. The most notoriously serious impact is reflected in the success of violent extremist groups in using bots to achieve the goals of propaganda, recruitment, and influence, in such a manner that outdoes the traditionally used fashions. Groups, such as ISIS tend to reach the people most affected by their ideology and approach, through open discussions and messages, seeking to get them into secret private groups to recruit them. This makes hunting recruits before they convert to fundamentalism very difficult.

Security Bots

Social media via APIs provide opportunities for bot technology developers to use the potential of such platforms. This prompted many commercial advertisers to use bots commercially, in such a manner that does not contradict the terms of use of these platforms. Such bots are known as faithful bots because they do not pretend to be real users.

As of 2018, such platforms began a purification process to expel bots that claim to be real users, or fake accounts, which leading to an arms race between platforms and bot developers. For example, Facebook puts into action protection system, based on machine learning techniques to review the largest number of accounts and control bots.

A research study by the University of British Columbia showed that only 20% of bots were suspended by Facebook, even after being reported by many users. While Twitter successfully reduced the influence of bots used by some parties in Russia to create electronic noise during the 2011 elections. The bot developers were able to overcome the technique of associating accounts with the age of the user and identify the user by purchasing some actual accounts, leading to the development of a technique to detect the social behavior of accounts.

The researchers conclude that most of the currently deployed bots, with their various functions and methods, have a negative impact in cyberspace, and few of them are used to have a positive impact. This may be an exception vis-à-vis bots that aim to sell products, steal personal information, or spread propaganda-triggered ideas. The researchers attribute this to the difficulty of positive engagement with human users, and the poor interaction. The AI development may lead to the construction of more sophisticated bots that contribute to the common good of societies. This progress may occur by developing

various technologies, such as high-accuracy speech recognition, creating and understanding natural language, making bots able to conduct a conversation with human users and understand all connotations.

However, the challenge is that currently available machine learning techniques cannot participate in such real conversations, without dispensing with the inputs from the data. This makes their language closely represent knowledge rather than creating such knowledge. To compensate for this inefficiency, the researchers concerned suggest the adoption of self-planning to update the discourse that bots use to counter the messaging software programs used by opponents. This process may become a cyber planner that develops a strategy for employing the discourse by reviewing the discourses spread online. Some researchers believe that the current AI capabilities may develop into deep learning, making the human factor unimportant. However, the large gap in the AI literature and applications shows the urgent need for more research to create a quantum leap in producing and understanding natural language.

Deployment Map

The report discussed the current use of bots with their various functions. Successful cases of deploying bots in health to perform the emotional functions necessary to follow up on patients, involving the bot, the patient, and specialized therapists, such as Bot Melody and Bot Babylon that collect information from patients and recommend therapists to take certain actions. The positive performance of such bots may be attributed to the human factor, the narrow expertise REQUIRED to activate these bots, and the disciplined environment in which they operate.

Interview bots connect users who can't get to know each other directly. For example, Bot Sensei connects users who need a good or service. The Massachusetts Institute of Technology's Media Lab has also developed BUKKOU BOT to directly connect people with similar symptoms of anxiety and depression. The said bot targets youth active on social media, and single applications developed for such diseases.

Bots connect users to each other by searching the conversation history for specific words or responding to specific messages. One experiment has shown that users of messaging software programs are more likely to have accounts that appear to be more active or affiliated with an ethnic or social group. Other experiments showed that the gender on which a bot builds its personality affects its attractiveness to users. Factors related to the size and rate of tweets, methods of tweet creation and targeting a specific group of users also influence the success of bots.

Some other bots called harvesting bots attempt to collect as much information about users as possible. Such bots operate easily. On Facebook, for example, bots send friend requests to users, and when accepted, bots collect all news, posts, notices, and alerts available in the users' profiles. It seems that the goal of bots is to attract followers and users.

Such bots often use profile pictures of beautiful women, and friend requests from such fake accounts are often well received. NATO called these bots "PEKY TROLL" or PLEASURE BOTS.

The behavior of bots is entirely controlled by the human factor, and bots can go beyond collecting data to having private conversations with the user and then getting them to engage in many activities, such as illegal trade, destruction of operating system, or infection with malware. In 2011, a research team at the University of British Columbia was able to use 102 self-accounts (Troll) on Facebook and extracted 250 gigabytes of data, from more than 3000 users.

Some experiences have proven a failure in communicating with users, including Microsoft chat bot, known as the TAI BOT, which went beyond the inability to achieve the desired goals into causing undesirable adverse effects. The software program was launched in 2016 and the company described it as a machine learning program that analyzes interaction patterns in user messages. However, it quickly developed a racist conversational language, rendering it discontinued one day following the launch on Twitter. Again, the AI program (XIAOICE) was successful when Microsoft launched it on the Chinese platform (WeChat), gained wide popularity, and was added to a million and a half chat groups, and got engaged conversations with ten million users without causing any angry reactions.

The difference between the performance of the two programs is clearly visible in launching (TAI) on a public platform, Twitter, and launching (XIAOICE) on a private chatting platform. The most important difference is the social and political context of launching the two programs.

Twitter Users (commonly used in the United States and the West) tend to use blunt and possibly offensive language, while in China, strict government censorship of content made strict user self-censorship of the language used.

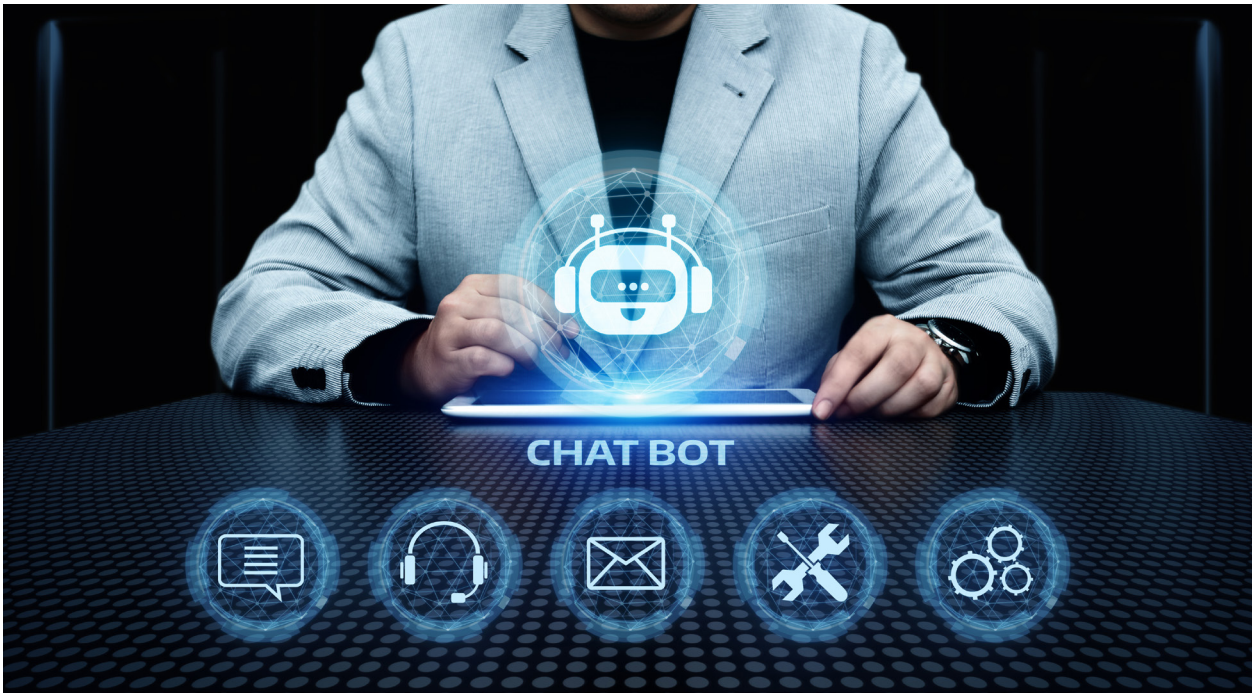
Bots are also used to spread news, mislead or to install a specific vision; such bots operate through a network unit, making them have great power and influence. This type of Troll Bot was widely used by Russia to influence the choices of American voters in the 2016 elections.

Astroturf bots are used when people pay for them to gain popularity or far-reaching influence. The campaign of politician Mitt Romney in the 2012 US elections was accused of buying followers through this bot. This contributed to raising doubts about his election campaign. However, political parties in Mexico successfully deployed bots in the 2012 elections, due to restrictions on official and institutional media, especially issues related to drug trade agreements. This use by the ruling revolutionary party in the country for a long time led to a fierce war led by the campaign of Peña, the presidential candidate, to slander the competitors. The PEÑABOT BOT was able to influence hashtags hostile or opposed to Peña's policies, and created tracks or hashtags that achieved widespread proliferation to direct the discussion favorably.

Maturity

The report revealed that bot technology has developed significantly recently. As showcased by the models presented, the functions, types and competencies of bots differ from one context to another, and their desired results depend on various factors. Therefore, the report provides a measurement and assessment model the bit maturity:

- Awareness: the bot ability to find, store, and interpret minimal content. The development of this ability means that the bot will be able to process natural human language and develop language-based understanding of human speech patterns, by differentiating between the subtle nuances of connotations, and perhaps metaphorical meanings.
- Report: The bot ability to classify content and other data into meaningful categories, facilitating automated decision-making. Although this ability has matured into a large extent, machine learning still encounters various misclassification errors. This calls for the need to be wary of giving bots complete confidence in making decisions. Many interviews conducted by researchers predict continued conflict in the future to develop the ability of governments to identify bots, and the ability of managers of such



bots to remain anonymous. The human factor is a vital part of this conflict, and the ISIS cyber-team attempted to bribe an employee of the companies entrusted to deactivate TROL from preventing ISIS and bots from building accounts on Twitter.

- Action: The program ability to perform human actions in the real or virtual world, such as sending a friend request, or responding to a comment. Bot technology has greatly developed, but these software programs will not be able to hold lengthy discussions with the human user without raising suspicion, which confirms the need for the human manager to continue to monitor their performance and intervene as much as possible. The researchers anticipate that the next generation of bots will move beyond the current messaging generation into video and audio manipulation. Researchers at the University of Washington were able to process some of the video clips, combine some of the snippets, and put some words into people's mouths.

Legal Risks

Using bots is of multi-fold risks, based on type, purpose, and operator. If the software is used in a country, its consequences are likely to be overwhelming. Bots that violate privacy, confidentiality, information coherence and availability are a real threat, which enables the US government to build a policy to counter such threats and set different rules that may be localized and adapted in other countries. The threats of gathering information by bots can be refuted by the means used. The report stresses that the US government will not resort to using bots to achieve goals that can be reached without using them, nor will it resort to publishing certain news or operating in disguise, cloaked in a human mask. Therefore, it would be more appropriate for the US government to share the messaging platforms to put into action the legally approved terms of use and reduce the effect of bots.

Legal Considerations

The Free Speech Clause of the First Amendment protects broad categories of speech from government regulation. This freedom extends to political expression without including speech inciting hatred and violence against minorities or specific persons. Given the borderless nature between the two contents, the content broadcast by ISIS, including violent and bloody scenes, may be considered within the limits of free political speech. If some content is protected, the US government must follow legal procedures to deal with it, such as a court order. Also, if the government wants to use bots to delete certain content, it will have to follow the same procedures.

However, the government can force digital platforms to delete certain content. The Terms of Service (TOS) on these platforms often contain the prevention of the promotion of terrorist propaganda, even if it is subject

to free expression, and the TOS allows users to flag such violations for removal. The UK government set up a Counter Terrorism Referral Unit, which reportedly removes 2,000 pieces of extremist material per week. Similarly, the European Union established the Internet Referral Unit in 2015, which in its first year processed over 11,000 messages.

The Establishment Clause of the First Amendment holds that the USG is prohibited from actions that unduly favor one religion over another. Again, depending on details of a given government-run bot program, this clause may present legal questions for bot typology that target users based on their religion or have a disparate impact on a specific religious group. Bot types that might be susceptible to this risk include influence, harvest, matchmaker, masquerade, and harassment bots. Legal questions will hinge on whether designing or deploying bots that target users on religious grounds or specific words. To bypass these risks, it is possible to target people living abroad, because they are most likely not going to act against the US government. Developers must be careful about the keywords they search for to avoid being related to a particular religion in a narrow sense.

Law Enforcement and Intelligence: Some bot types, including harvest bots, have potential benefits for law enforcement or intelligence efforts. The attempt to access information that is not publicly available makes it subject to certain legal restrictions and processes, including the Privacy Act, the Electronic Communications Privacy Act, the Communications Supplement to Applicable Laws, the Communications Storage Act, the External Intelligence Oversight Act, and others. The entity that employs bots to collect information or in intelligence operations must ensure that there are no violations of these laws.

Smith-Mundt Act: The United States Information and Education Exchange Act of 1948, more commonly known as the Smith-Mundt Act, enables the State Department and the Broadcasting Board of Governors (BBG) to conduct public diplomacy campaigns abroad but restricts their ability to influence public opinion in the United States.

This law limits media materials produced by the government. However, updating the law in 2012 authorized the government and the council to make campaigns targeting the American audience living abroad. When bots are used in these campaigns, it is necessary to review the law, and verify the target audience in such campaigns.

Ethical Considerations

The report mooted several ethical constraints that are very different from legal constraints but lack the means of implementation. When conducting counter-violent extremism or terrorism operations, the US government draws on social media platforms owned by private companies, which have their own interests that may be inconsistent with the government interests and goals. The use of bots in these platforms may prejudice neutrality, and that the TOS can be different from one platform to another. The researchers stress the need to consider these factors to avoid doing any harm to these companies, which are a major part of the national economy.

Several sets of interviews suggested that there is an ethical and practical requirement to be transparent in bot deployment.

Many Americans, and even foreign nationals, have strong cultural expectations that the USG will be honest and direct in its dealings and will generally avoid presenting material that looks like propaganda. It is necessary to be careful when dealing with these materials, as their misuse, manipulation of information or willful misinformation may damage the reputation of the US government and the political culture that sets certain expectations for its performance. It may also lead to a decline in trust in the Internet uses as a safe space for sharing information.

This will have economic, commercial, and political consequences for the US policy that supports the promotion of safe cyber spaces. Taken together, the researchers are calling on the US government to issue a general statement of principles that explains the type and use of bots that the US government uses to help the public to understand what the government is doing inside or outside.

The ethical risks of using bots differ across institutions. The Department of State has many roles that go beyond countering extremism and terrorism to promoting economic interests, supporting Internet freedom and human rights; deploying bots may involve higher risks. This may result in undermining the American interests. As for institutions with narrow or non-consulting tasks, such as the Ministry of Homeland Security, their risks are lower given their domestic security-related roles.

Development and Assessment

The report concluded that bots are a potentially plausible and effective means of conducting CT or CVE interventions against terrorist groups like ISIL.

The logical next step for the USG is the development and implementation of specific bot programs to combat ISIL and similar threats. An intermediate step is to assess and choose between different concepts of operations. By concepts of operations, it is meant general approaches for programs. For example, a program that tries to proactively offer resources and support to populations at risk for radicalization is very different from a bot program aimed at combating misinformation spread by U.S. adversaries. The report indicated that the assessment framework incorporates the following five criteria for bot operations:

1. technical feasibility of both being an efficient tool
2. personal physical, social, and privacy-related risks to SM users
3. potential risks to the developers or deployers of bots
4. the potential impact of the operation in terms of either breadth or depth
5. the complexity involved in undertaking the operation

The researchers synthesized eleven factors in the form of questions to determine the type of bot recruitment process, such as:

- Who oversees the recruitment process?
- Where can bot be deployed?
- What communication strategy will bot use?
- When can bot be activated?
- How do bots operate?
- At what level of operation can bot be detected?

For further clarification of the detailed user assessment, if the target of bots is a global audience, the technical feasibility may be broader, with high risks for the bot technology developer. If the target is the American audience, the technical feasibility is lower, and the risks to the bot technology developer are higher. However, if the target is citizens living at home, the technical feasibility is less, and the risks to the bot technology developer are much lower.

If the target is the hostile accounts confirmed to be affiliated with ISIS, the technical feasibility is higher, and the risks are not present. The technical feasibility of deploying bots for people at risk of becoming fundamentalist is higher, as is the possible risks to

bot technology developers. The degree of technical feasibility increases when bots target ISIS opponents, and the risks for program developers are reduced.

The researchers point to the possibility of employing AI and deep learning to enhance assessment, as some bots rely on human management, and other programs have complete independence, based on the AI use and deep learning, processing big data through computational algorithmic functions.

Programs that rely on AI allow obtaining ideas and opinions that may be anonymous, clandestine, and covert. It can enhance the human agent ability to act anonymously and engage in direct interactions with users. More importantly, it may enhance the bot ability to operate in disguise.

With AI coming into play, there are many notorious risks. The programs may act away from the text that is set for them, and that they may get involved in ethical or legal risks, and may act in an unexpected way, thus becoming vulnerable to exposure.

According to the recommendation of a bot technology expert, it is essential that these software programs remain restricted under human control; a software program used to combat fundamentalism does not boomerang and backfire, turning into a fundamentalist program itself.

Based on this assessment, the researchers conclude that the best bots that aim to influence the audience and spread news are interview bots, given the few ethical and legal risks, and the technology made available for use. As for the bots that aim to undermine extremist networks and limit their influence, they are exposure bots, given the technical feasibility of use and the low risk to developers alongside users of social media. As for the bots that are employed to collect confidential information, they are harvest bots, which are more useful than bots designed for hunting.

The report presented a concept-building model for interview bots to provide adequate resources to citizens at risk of becoming fundamentalist. This bot is transparent; users interact with advertising, a human manager participates in the operation, and attaches importance to minimizing risks to developers, the audience uses social media, user interacts with bots without controlling bot data, users are non-USA adults.

Recommendations

The researchers provided several recommendations to the relevant authorities in the US government, on the bot technology development to enhance bot use and address the ethical and legal considerations when combating radicalization. The researchers stated that the American institutions concerned should consider several issues when developing bot software programs:

1. Leveraging commercial development of bot technology, as industry investment in this rapidly evolving space has yielded significant progress.
2. Tailoring bot technology to the environment in which they are to be deployed, such as platform structures of engagement or the culture of government censorship among the target audience. Taken together, this will maximize credibility in sensitive contexts and help avoid disasters resulting from unanticipated mismatches.
3. Paying attention to the network characteristics of users the bot is seeking to engage, such as the friend count of an individual target user or whether target users are connected merely by topic interest or preexist as a dense network of social connection; users are more likely to engage with accounts with whom they are already connected by social friends.

The researchers suggest the U.S. agencies should consider the following suggestions on how to mitigate the legal and ethical risks of any proposed bot program.

A. The institutions concerned should consider the analysis of international precedents in the use of bot software programs. This is to avoid the normalization of government actions and behaviors that may lead to a threat to cyber-security, by interfering with confidentiality or the integrity and availability of information on the Internet.

B. Considering language-related and legal issues by restricting the recruitment of bots to a narrow range of users abroad, while avoiding targeting users affiliated with a particular religious sect and the need to build a software firewall when appropriate between bot programs, intelligence agencies, law enforcement, and international partners.

C. The need to obtain corporate permission before hiring and using bots on social media.

D. The need to be as transparent as possible in the processes of the US government use of bot

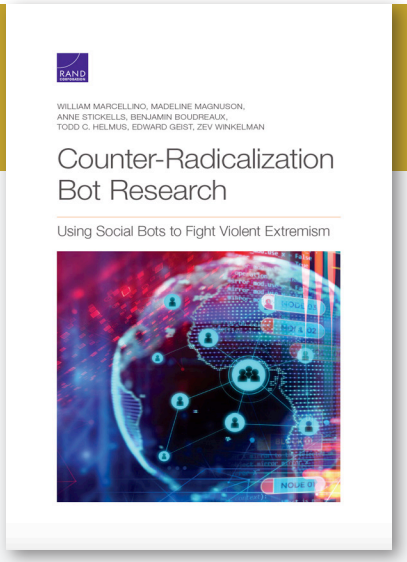
technology and to be within well-thought limits to avoid unpleasant consequences.

E. The need for a legal review of bots used with inter-institutional cooperation to develop the principles of framework and create a special blog.

Conclusion

The researchers conclude that the use of bot technology confirms that they are a practically efficient tool in combating violent extremism and terrorism; the bot technology deployment will be limited by many legal, ethical, and practical factors. This always requires the human factor. Decision makers should weigh the expected gains of employing programs with the risks involved in self-operating programmes. The US government should also promote bot detection technologies. This makes it difficult for opponents to launch deception and misinformation campaigns, and spread false information through bots, considering the awareness of the cultural and social context when developing bots.

The report addressed one of the huge developments in cyber technologies, bots, and assessed their potential use for combating violent extremism and terrorism. It examined the legal and ethical constraints that should be considered when governments employ bots in counter-extremism operations. However, the report highlights expansion in the technical aspect. The report cites telling examples of how bot technology is used by terrorist groups, such as Al-Qaeda and ISIS in enhancing recruitment and propaganda; the report, however, did not provide detailed examples, statistics, or assessment measure of success and failure, following the decline of the two terrorist groups in the Middle East in particular. This makes this report unbalanced in terms of subject-matter. Still, it is useful for institutions engaged in combating extremism. The report showed the reality of this tug-of-war tension with complex political, security, economic, and cultural aspects.



COUNTER-RADICALIZATION BOT RESEARCH

USING SOCIAL BOTS TO FIGHT VIOLENT EXTREMISM

Published By

RAND 2020


التحالف الإسلامي العسكري لمحاربة الإرهاب
ISLAMIC MILITARY COUNTER TERRORISM COALITION

التحالف الإسلامي العسكري لمحاربة الإرهاب

ISLAMIC MILITARY COUNTER TERRORISM COALITION