



ISBN: 978-9934-564-46-8

Authors: Jakob Willemo

Project manager: Sebastian Bay

Editors: Sebastian Bay, Monika Hanley, Rueban Manokara, Baris Kirdemir

Design: Kārlis Ulmanis

Riga, August 2019

NATO STRATCOM COE

11b Kalciema Iela

Riga LV1048, Latvia

[www.stratcomcoe.org](http://www.stratcomcoe.org)

Facebook/[stratcomcoe](https://www.facebook.com/stratcomcoe)

Twitter: [@stratcomcoe](https://twitter.com/stratcomcoe)

This publication does not represent the opinions or policies of NATO or NATO StratCom COE.

© All rights reserved by the NATO StratCom COE. Reports may not be copied, reproduced, distributed or publicly displayed without reference to the NATO StratCom COE. The views expressed here are solely those of the author in his private capacity and do not in any way represent the views of NATO StratCom COE. NATO StratCom COE does not take responsibility for the views of authors expressed in their articles.



# TRENDS AND DEVELOPMENTS IN THE MALICIOUS USE OF SOCIAL MEDIA

The malicious use of social media is a widespread phenomenon, targeting individuals, public opinion, and in some cases even the functioning of the state.

In recent years, social media platforms have been abused by foreign governments, private companies, and individuals to influence the outcomes of democratic elections and to undermine public trust in the societies in which we live.<sup>1</sup> Today, social media platforms are manipulated by malign actors in order to pursue their political and military goals. In other words, social media platforms have developed into an effective tool for waging information warfare.<sup>2</sup> Although information warfare is nothing new, social media platforms offer a cheaper, more efficient, and less demanding stage for influencing larger numbers of people than ever before.<sup>3</sup>

While the social media platforms are conduits facilitating the free passage of information, the companies that own them are active participants wielding significant influence over what takes place in the social media space and, increasingly, over how we communicate, interact, and socialise in the 21<sup>st</sup> century.

Social media interactions are not determined solely by users; they are also shaped by the terms, policies, and algorithms adopted by various social media platforms. These structures, functions, and rules (and the loopholes in between) are constantly being studied and exploited by malicious actors working to influence their target audiences.

This study outlines the salient developments in the malicious use of social media. The first chapters examine seven developments in the malicious use of social media by examining case studies supported by interviews with experts in the respective fields. These chapters also comment on the future trajectory of developments and recommends policy changes. The final chapter provides conclusions and takeaways from this research.

Trends and developments in social media manipulation:

- The current state of play is a cat-and-mouse game between malicious actors, governments and the new media industry. As social media companies and other actors take action to counter abuse, malicious actors adapt to the



new environment. This has led to, among other things, an increase in the sophistication of cyborgs and trolls as simple automated accounts are being taken down.

- Impersonation is commonly used both for the spread of disinformation and for social engineering attacks with different degrees of sophistication, sometimes attempting to create real-life events through online activity. Continued technological development in the field of artificial intelligence

and frighteningly realistic 'deepfake' video and audio techniques may allow impersonation attacks to become even more credible in the future.

- The methods and platforms used to disseminate disinformation are changing. The increased use of encrypted platforms, such as WhatsApp or closed Facebook groups, makes it increasingly difficult to identify ongoing information operations. Furthermore, malicious actors are more effective than before in covering their own tracks.

## 1. IMPERSONATION

Let us consider *Impersonation* as the first example of the malicious use of social media. Impersonation can be defined as 'pretending to be another person for the purpose of entertainment or fraud'.<sup>4</sup> Facebook and Twitter prohibit the use of accounts impersonating both real and non-existent people, organisations, etc., to gain anonymous access to their services, while other social media platforms, such as YouTube, are less clear in their stance.

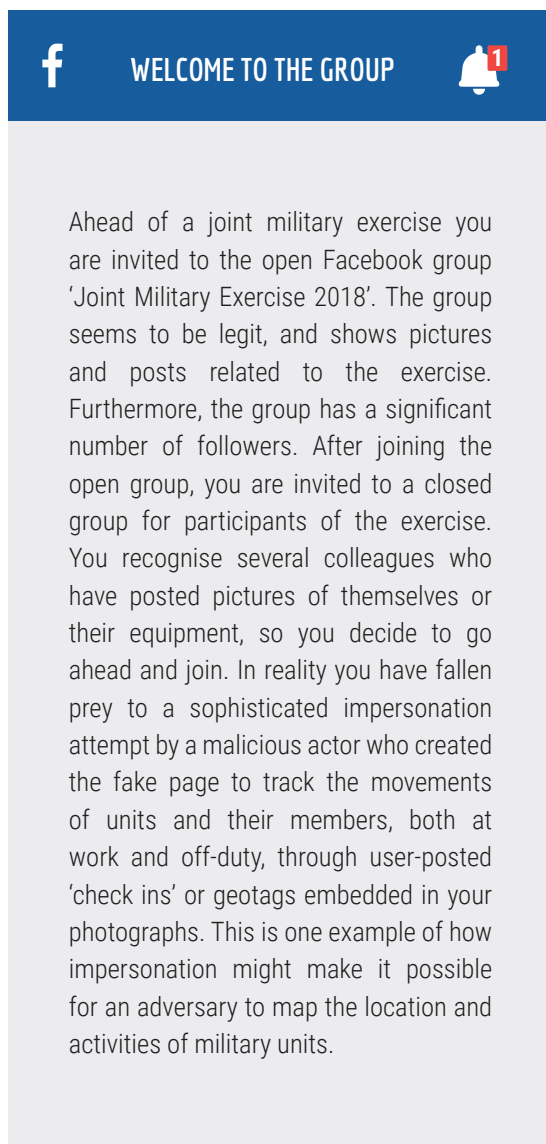
In the United States, impersonation is considered not only a breach of a platform's

user agreement, but may also be treated as a criminal offence.<sup>5</sup> Impersonation is often favoured by white collar thieves engaged in financial fraud, however it is not only individuals that are impersonated, as newspapers and websites are also mimicked. By making small changes to the name of an existing website or domain name, fake news sites are able to emulate respectable news sources and exploit public trust to promote disinformation.

The same methods deployed by white collar thieves could also be used by a malicious



foreign actor pretending to be a newly hired army officer at the regiment where one works, which could have severe implications for operational security. For example, consider the following scenario:



Recent technological advances pose a further, perhaps even more significant challenge. The development of 'deepfakes'

is one such particular worry. 'Deepfakes' can be as the 'digital manipulation of sound, images, or video to impersonate someone or make it appear that a person did something—and to do so in a manner that is increasingly realistic, to the point that the unaided observer cannot detect the fake'. Although this technology is still in its infancy, it has the potential to significantly impact the information environment, leveraging our cognitive biases, and blur the line between truth and falsehood in ways we cannot yet fully comprehend.<sup>6</sup>

### When perception becomes reality

The Russian interference in the 2016 US election revealed how a foreign government was able to interfere in democratic elections<sup>7</sup>. The US Department of Justice Grand Jury indictment of offenses related to the 2016 US election demonstrates how Russian propagandists used political advertisements in combination with addressing ordinary citizens and creating false grassroots movements to support a particular agenda in the virtual world. This technique is known as 'astroturfing'. Impersonation techniques was also used to influence existing social movements, such as the infamous case of the West Palm Beach rally where an American citizen was paid by a troll factory to dress up as an imprisoned Hillary Clinton<sup>8</sup>. This case reveals how powerful social media can be in the hands of a malign actor and how manipulation of social media can generate effects in the physical world.



## 2. SOCIAL ENGINEERING ATTACKS

The social media space has created new opportunities for malign actors to disrupt and manipulate the information environment, exploiting human cognitive biases to affect the perception of targeted populations.<sup>9</sup> All human beings have cognitive biases. Malign actors manipulate these processes to exert influence.<sup>10</sup>

### EXPERT INSIGHT—DAVE BITTNER FROM THE CYBERWIRE AND JOSEPH CARRIGAN FROM THE JOHNS HOPKINS UNIVERSITY INFORMATION SECURITY INSTITUTE.

*The Cyberwire podcast team, Dave Bittner and Joseph Carrigan, has been exploring developments and challenges of social engineering, phishing schemes, and other cyber-related criminal activity since mid-2018. Together with external experts they explore the art of deception and provide real-life cases in their weekly podcast, Hacking Humans. This section is based on an interview with Dave Bittner and Joseph Carrigan conducted by NATO StratCom CoE.*

#### **Latest developments in social engineering attacks using social media**

There have been no dramatic developments regarding methods for social engineering; the ideas and methods haven't changed much in the last hundred years. What has changed is that social media has greatly broadened the 'attack surface', or the sum of our exploitable vulnerabilities. The massive amounts of information available on social media allow malicious actors to develop highly sophisticated and targeted social engineering attacks and deploy them much

more quickly than ever before. Malicious actors can test and fine-tune their techniques with little or no cost, and with much higher success rates, compared with the time before the ubiquitous use of social media. As infrastructure and security systems improve (e.g. firewalls), the incentive for social engineering attacks changed. Today, the manipulation of social media is the most cost-effective way of acquiring sensitive information. We should not forget that it is often the weakest link in a system that is targeted—even if your own social media privacy settings are strong, malicious actors



” Even if your own social media privacy settings are strong, malicious actors may still be able to gather information on your friends or family members.

may still be able to gather information on your friends or family members.

Current strategies for social engineering attacks generally make use of two different and highly successful approaches—sophisticated and credible attacks on the one hand, and simple gambits on the other. Posts or applications that invite you to publish your birthday, the name of your first pet, or the street where you grew up, allow malicious actors to use this information to hack your accounts, as such questions are a common security questions for resetting passwords.

Social engineering attacks have also become an issue for the US armed forces, particularly the so-called romance scams where scammers impersonate military personnel looking for romance. These have proven effective for financial fraud. Similar attacks might be deployed for other purposes, such as eliciting sensitive information, but the exploited human vulnerability remains the same—people like to be helpful.

### **Future challenges**

Continuing developments in the field of artificial intelligence, such as sophisticated chatbots and deepfake videos, pose a fundamental challenge since they make

social engineering attacks seem even more credible. These techniques are not yet fully developed, but are improving almost daily. Although voice-emulators have existed for a long time, they have mostly been used for pranks. However, as this technology is refined, it could be used for highly-sophisticated social engineering; for instance, using an AI model to impersonate a public figure after the model has been taught to sound like the target using public speeches and interviews.

### **What should be done to address the challenges?**

For the armed forces, the best defence against social engineering attacks is proper Operations Security awareness and best practices training. It is essential to enhance training capacity for all personnel and include all relevant actors into these frameworks, including external actors, such as family members.

However, there are a number of issues standing in the way of effective training in this field. First of all, training in a closed or simulated social media environment will not be particularly effective since trainees will be prepared to counter simulated attacks. But training in an open environment is often not possible because the social





media platforms' Terms of Services. The best available solution would, therefore, be to identify and model best practices and provide examples of real-life attacks to raise awareness. Furthermore, it is essential to foster a positive organisational culture, and

to facilitate reporting attacks and suspicious online behaviour.

Finally, social media companies need to step up their game to pro-actively address the vulnerabilities inherent in their platforms.

### 3. RUSSIAN INFLUENCE OPERATIONS

Information influence activities should now be counted among the many techniques hostile actors employ to negatively impact democratic societies together with activities such as espionage, cyber threats, and the deployment of irregular forces. Information influence activities can be conducted as a single activity or as part of a larger information influence operation combining various and multiple activities.

#### **EXPERT INSIGHT—LIUBOV TSYBULSKA, HEAD OF HYBRID WARFARE ANALYTICAL GROUP AT THE UKRAINIAN CRISIS MEDIA CENTRE AND STRATCOM ADVISOR TO THE UKRAINIAN ARMED FORCES**

*The Ukrainian Crisis Media Centre was created in 2014 by several different Ukrainian experts within the fields of international relations, communications, and public relations. The objective of the centre is to provide the international community with accurate information about current events. Their press office publishes daily briefs regarding developments in Ukraine. The centre also offers strategic communications training for officials and journalists and conducts its own in-house research on hybrid warfare. This section is based on an interview conducted by NATO StratCom CoE.*



” Currently, the main method seems to be infiltrating the Ukrainian media to disseminate narratives and disinformation congenial to the Kremlin—especially through television, as this is the most common medium used by the largest number of voters.

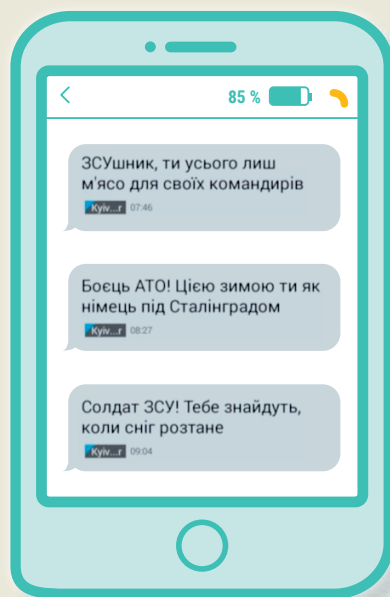
### Latest developments in Russian information operations using social media

Russian social media networks, such as Vkontakte and Odnoklassniki, have been used extensively for Russian information operations. First, significant attempts have been made to recruit Russian agents for rebel groups through these networks. Second, they have been used to profile both armed forces personnel and the Ukrainian electorate. Effective psychological operations were deployed towards the Ukrainian army using the collected data. Soldiers, officers, and their relatives were targeted with personalised

messages aimed at disrupting both the morale and the mobilisation of Ukrainian army units. There is ample evidence of messages sent directly to soldiers and officers, often combining sensitive personal information with threats towards their families. Third, Russian social networks have been used to disseminate disinformation about Ukrainian military leadership to divide and demoralise the armed forces.

Such activities were observed especially during 2014–15, and they proved to be highly effective. In 2017, Ukraine adopted sanctions toward several Russian social

### Examples of messages directed towards Ukrainian servicemen



ЗСУ-er, you are just the meat for your commanders.

ATO warrior! This winter you are just like a German in the Battle of Stalingrad.

ZSU soldier! They will find you when the snow melts.



media networks, thereby limiting public access to them in Ukraine. However, Russian information operations adapted to the new circumstances and deployed new methods. As Facebook became the biggest social media platform in Ukraine, influence activities on Facebook increased. The impersonation of pro-Ukrainian pages and channels and the use of accounts claiming to belong to Ukrainian soldiers have increased on the majority of the social media platforms in the country. Furthermore, as trust in Russian media declined, an increase in attempts to infiltrate Ukrainian media and political leadership has been observed.

### Future challenges

It is likely that Ukraine will be used as a testing ground for information activities influencing in the future as well. Currently, the main method seems to be infiltrating the Ukrainian media to disseminate narratives and disinformation

congenial to the Kremlin—especially through television, as this is the most common medium used by the largest number of voters. It is also probable that Facebook will be used to sow discord among the population and animosity towards the political elite. We have already observed examples of Facebook and YouTube being used as platforms to promote pro-Russian candidates and deliver political advertisements. This is often done by publishing content under different pseudonyms not affiliated with a specific party.

### What should be done to address the challenges?

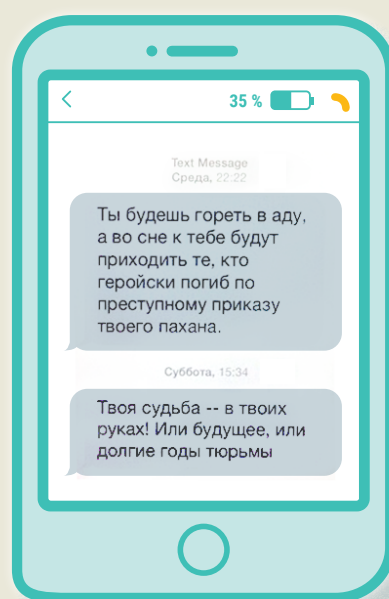
In 2014–15, the Ukrainian Armed Forces took several steps to defend against information operations. One of these was to establish a press office to support the strategic communications of its own personnel. This proved to be an effective solution. Another step was to establish a mobile group of

You will burn in hell, and those who died heroically under the criminal order of your boss will come after you while you sleep.

Your fate is in your hands! Choose a bright future or many years of imprisonment.

*ATO is an abbreviation of 'АнтиТерористична Операція', or Anti-Terrorist Operation.*

*ЗСУ is an abbreviation of 'Збройні Сили України', or the Armed Forces of Ukraine. 'Meat' here means 'meat shield', another idiom for 'cannon fodder'.*



psychologist and other professionals who travelled along the frontline, gathering information about the issues soldiers perceived as problematic and then provided soldiers with the information they lacked; this proved to be an efficient mechanism for creating trust. Effective communication with army personnel and an established presence on social media are vital for bolstering resilience towards information operations. Lower-ranking personnel have

a legitimate need to understand how and why important decisions are taken. If this need is not addressed through genuine trust-building action, discord might be further exacerbated by an adversary through the spread of disinformation. Continued education about strategic communications, raising awareness about current information operations techniques, and best practices training are all vital components of healthy resilience to information operations.

## 4. DISINFORMATION INCITED VIOLENCE





### WHATSAPP MURDERS




The incidents of dissemination and misinformation affecting WhatsApp users in India have been alarming. In June 2018, rumours regarding child kidnappers were circulated on WhatsApp resulting in the killing of several presumed kidnappers in India. One probable catalyst for the lynchings was the spread of a manipulated video of a boy being kidnapped by two men. The video was initially created as a part of a Pakistani child safety campaign, but the end of the video showing the boy's safe return was cut from the version virally circulated on WhatsApp.<sup>18</sup>

An abundance of examples show how social media platforms are being used to incite violence. In 2017, human rights activists in Myanmar reported that social media was being used to disseminate disinformation regarding the Muslim Rohingya population—this was part of an action that eventually forced over 700 000 Rohingya to flee their homes.<sup>19</sup> Today peer-to-peer encryption, widespread smartphone usage, and machine learning are also being used by malicious actors. Governments and social media companies currently struggle to prevent these platforms from being harnessed for malicious activity.

## 5. DISINFORMATION IN ELECTIONS

I'M SORRY I ACCEPTED BRIBES, OR DID I?



Imagine that, the night before Election Day, a well-known political candidate posts a video on Facebook in which she admits to accepting bribes ahead of the election. The video spreads rapidly over social media causing massive anger. The next morning, the video is quickly deleted from the politician's account. She declares that her account was hacked and the video was a fake. The genuine

declaration reaches only half as many people as the fake post did, and by the time the polls close many are still unaware of the deception.

This scenario depicts the dangers we are facing as impersonation, deepfakes, and disinformation continue to develop. These techniques might be applied to any actor, adapted for any situation. Debunking the video might seem to be an easy task, but damage control after such an event is a highly demanding task requiring a deep understanding of the art of countering information activities.

### EXPERT INSIGHT—DONARA BAROJAN, ASSISTANT DIRECTOR FOR RESEARCH AND DEVELOPMENT, ATLANTIC COUNCIL DIGITAL FORENSICS LAB

*The Digital Forensics Lab was created in 2016 as an integrated part of the US-based think tank, the Atlantic Council. Since 2016, they have been one of the most prominent actors in uncovering disinformation around the world. Their purpose is to identify, expose, and explain how disinformation spreads in order to further protect democracy. In 2018, they became one of the first think tanks to partner with Facebook with the explicit purpose of exposing disinformation on the platform in the run-up to and during elections. This section is based on an interview conducted by NATO StratCom CoE.*



” The methods for influence operations on social media are still similar to those identified during the 2016 US election, but are growing increasingly more sophisticated.

### **Latest developments regarding the malicious use of social media during elections**

The use of videos to spread disinformation is on the rise. In some cases the videos are doctored, but more often than not they are simply shown out of context. One such example was observed in the run-up to the Moldovan election in 2018, when an Al Jazeera video in Arabic went viral on Facebook and on the Russian social network Odnoklassniki. The video itself was real, but the subtitles in Romanian did not match what the journalist was saying. The fake subtitles claimed that mayoral candidate Andrei Nastase was proposing to lease the city of Chisinau to the UAE for 50 years, which was not true. The original Arabic narration focused on strained relations between the UAE and Yemen, and then went on to discuss the size of Chisinau and its water pipelines.

Furthermore, there seems to be an increase in the use of ‘impersonation’ accounts and ‘astroturfing’. Actual users are also displaying inorganic behaviour, such as posting tweets at a suspiciously rapid pace.<sup>11</sup>

In other words, not only are bots becoming increasingly similar to people, but people are also becoming more similar to bots. During the Italian election this year, supporters of the now-ruling Lega party behaved like bots, posting in a coordinated rapid pace to amplify their party’s content online.<sup>12</sup> The use of bots has been an overarching theme during election campaigns in many countries, especially in Mexico, where political bots are commonly referred to as [Peñabots](#), named after the head of the PRI party and current Mexican President Enrique Peña Nieto.

The methods for influence operations on social media are still similar to those identified during the 2016 US election, but are growing increasingly more sophisticated. Malicious actors today are better at covering their tracks. Since major social media platforms began actively suppressing hate speech on their networks, there has been an exodus of fringe groups to alternative social media sites like Gab, Voat, and others. This was particularly visible during the Brazilian elections last year, when Twitter’s ban of extreme right-wing accounts resulted in several radical groups transitioning to the more peripheral social media platform Gab.



Similar trends have also been identified in the US and Germany. This makes it more difficult to monitor extremist activities on social media, but also isolates extremists in their own echo chambers, making it harder for them to influence mainstream conversations.

### **Future challenges**

Upcoming developments in the field of artificial intelligence will likely make disinformation operations more dangerous and more influential. For example, bots can now be detected by automated algorithms; automated counter-measures will stop working as they become more sophisticated. Further advances in natural language processing will soon make it possible to develop sophisticated chatbots that seemingly engage users in one-to-one conversation, further personalising disinformation.

### **What should be done to address the challenges?**

Journalists must be more proactive in countering disinformation and ensure that the work of fact-checkers reaches affected audiences. International think tanks and organisations can equip journalists with the skills they need to identify and analyse disinformation, but it is up to journalists to actively counter disinformation by doing what they do best—reporting on these threats and holding malicious actors accountable.

Social media platforms also have an important role to play. They need to improve content moderation and enforce their own rules across all languages and countries of operation. Each social media company should have enough moderators with the necessary language skills to tackle this issue at scale, and to provide support for countries beyond Europe and North America. Finally, social networks must learn from their past mistakes and build resilience to future attacks by coordinated 'red-teaming', foresight exercises, and extensive communication with researchers and organisations specialising in these issues.



## 6. RUSSIAN DISINFORMATION AND THE EUROPEAN UNION

### EXPERT INSIGHT—GILES PORTMAN, HEAD OF EAST STRATCOM TASKFORCE, EUROPEAN EXTERNAL ACTION SERVICE

*The EEAS East StratCom Taskforce was established in 2015 as a response to the increase of Russian disinformation campaigns directed towards Europe and its partners. The objectives of the taskforce are to increase the efficiency of strategic communications, and to promote EU-policies towards the Eastern Neighbourhood, to strengthen the media environment in the EU and in the partner countries, and to enhance the capacity to identify and respond to disinformation. This section is based on an interview conducted by NATO StratCom CoE.*

#### **Latest developments in the malicious use of social media regarding the EU and its partners**

The current Russian information operations strategy seems to be based on ‘trial and error’—by acquiring information through constantly shifting methods and changing target audiences, and then using the insights gained to adapt to new circumstances. In that sense, even unsuccessful disinformation campaigns contribute to the development of more sophisticated approaches. Furthermore, the strategy currently employed is multi-layered, multi-faceted, and sometimes even contradictory.

On the one hand, there are indications that Russian information operations are covering their tracks a greater extent than before. The recently launched RT project, *In the Now*, is an example of the new preferred way of presenting disinformation—through viral and entertaining content. The use of impersonation is also common, often in combination with real-life fake events. Yet, Russian information operations have also become more blatant and more aggressive, for example, during the Skripal-case. Furthermore, current information operations are also closely connected to kinetic activities, such as cyberattacks.





” On the one hand, there are indications that Russian information operations are covering their tracks a greater extent than before. Yet, Russian information operations have also become more blatant and more aggressive.

Until recently, the West has focused on Russian information operations taking place on Twitter and Facebook, but operations on other platforms such as YouTube should not be ignored. Attempts to flood these platforms with disinformation are increasing. This year, we have also seen the weaponisation of peer-to-peer encrypted social media platforms, such as WhatsApp. These changes further complicate the problem of disinformation, as the secret nature of these platforms exacerbates the ‘echo-chamber’ effect, isolating participants from dissenting voices.

### **Future challenges**

As social media platforms take action to combat the malicious use of their services, Russian information operations are constantly adapting to the new circumstances and experimenting with new techniques. A challenge visible on the horizon is the development of highly-credible deepfake videos. As the production of these videos becomes more refined, it will be possible to make any actor say anything—an excellent way of deploying disinformation and further blurring our perception of reality.

### **What should be done to address the challenges?**

A powerful solution would be to reach target

audiences with accurate information. This is currently best done by the traditional media, which is why it is essential to support independent journalism and to provide the journalistic corps with the proper tools and guidelines for identifying malicious activities. The maintenance of a pro-active disinformation-countering environment should be combined with enhancing digital literacy and raising awareness of malicious activities to further augment political pressure.

Social media companies must find a solution to both protect the anonymity users value, while also reducing the vulnerabilities which anonymity creates. Malicious actors can create false accounts and impersonate pages with ease; when these accounts are identified and shut down, new ones pop up within hours and the spread of disinformation continues. To fully tackle these challenges, social media companies must enhance transparency for both users and researchers.

Another important challenge that needs to be addressed is that the Russian population is being fed disinformation about the EU. This audience must be approached cautiously. Rather than debunking disinformation about the EU, we should raise awareness about how the European Union can benefit Russian citizens, such as providing educational opportunities.



## EXPERT INSIGHT—YEHVEN FEDCHENKO, CO-FOUNDER AND CHIEF EDITOR OF STOPFAKE

*StopFake was created in 2014 as a response to Russian disinformation efforts and aggression towards Ukraine. Since 2014, the organisation has been one of the most prominent actors uncovering disinformation in Ukraine and has since debunked more than 2000 cases of disinformation. Furthermore, the organisation has conducted several research projects linked to the development of disinformation in Ukraine and regularly conducts workshops for journalists in identifying disinformation. This section is based on an interview by NATO StratCom CoE.*

### **Latest developments in disinformation on social media in Ukraine**

Russian social media networks pose a significant challenge, as described in the section by Liubov Tsybulska. However, Ukrainian efforts to combat Russian disinformation have been quite successful. Trust in Russian media is low overall in Ukraine, as is the case in Donbass and Luhansk. Furthermore, Russian media penetration in Ukraine is lower than ever. Although Ukrainians can still access Russian sites through using VPNs, they are much less likely to be victims of targeted attacks—if users mask their location, they are also more difficult to target.

However, Russian information operations continually adapt to changes in the online environment and have begun deploying new tactics for influencing Ukrainian society. These tactics consist of attempts

to infiltrate genuine Ukrainian media and television, thereby disseminating manipulated content under the Ukrainian flag. Initially, such Russian-produced media content may contain quality information, but as it gains traction with its target audience, influence operators introduce manipulated information. 2018 saw an increase in disinformation activities using impersonation accounts and false pages to discredit Ukrainian politicians and leadership. Russian information operations have also become increasingly internationalised, linking together disinformation about Ukraine, Syria, and other international events.

### **Future challenges**

It is likely we will see adaptive developments in Russian information operations and the malicious use of Facebook in Ukraine will probably increase. However, Facebook has hired a national representative for



” A worrying development is the Kremlin’s attempt to hijack international discussions regarding disinformation and ‘Fake News’ by taking the lead and setting the agenda

Ukraine as part of its plan to prevent such abuse. As cooperation between Facebook, the Ukrainian government, and relevant NGOs increases, social resilience may also improve.

A worrying development is the Kremlin’s attempt to hijack international discussions regarding disinformation and ‘Fake News’ by taking the lead and setting the agenda. This increases the risk that Russian will strengthen its position in the field.

### **What should be done to address the challenges?**

We can nourish our own credibility by providing educational support for journalists and thus ensuring that the public gets high-quality, truthful information together; this is the key to developing resilience against disinformation. In many ways NGOs can be more effective in combating disinformation than governments; independent organisations are often seen as having a higher level of integrity, which earns them trust.

Furthermore, it is essential that we speak clearly about the cause of the problem—

Russian disinformation activities. We must continue to put pressure on the directors of government-controlled Russian ‘media outlets’ and other notorious disseminators of Russian disinformation. It is especially important to pressure the Kremlin regarding media freedom and to insist that foreign broadcast services be allowed to operate in the country. The keyword here is **reciprocity**—as the Kremlin increases its ‘media’ efforts abroad, we must answer in same way, increasing the presence of credible foreign news outlets in Russia. Finally, if this is not enough, steps taken by Ukraine, such as sanctions on Russian social media companies and television, offer an effective option for diminishing the presence of Russian disinformation, both online and off.



## EXPERT INSIGHT—THE CRITICAL VOICE

*This expert, who chose to remain anonymous, is active within a prominent organisation that has closely followed the development of social media manipulation over the last five years. This section is based on an interview conducted by NATO StratCom CoE.*

### **The latest developments regarding the malicious use of social media**

It is vital that we differentiate between the disinformation campaigns directed towards Western audiences and those aimed at the Russian people. The effects of the Russian disinformation campaigns on Western audiences are somewhat exaggerated. Contrary to common perception, these attempts remain unsophisticated—much of the subversive content produced during 2016 to influence the American election was of poor quality, and it is unlikely that it could actually impact American voters. The same is true of the so-called Troll Farms—these organisations focus is on quantity rather than quality and are probably given more credit than they deserve. However, the picture is quite different where the Russian-speaking audience is concerned. Disinformation disseminated to Russian-speaking audiences is much more sophisticated and leverages the government's deep understanding of the target audience.

### **Future challenges**

The 'deepfake' is one of the most talked-about developments in the field. However,

for the time being, the negative effect of this technology is likely to overestimated, especially since we are already so focused on it. The greater risk is forgetting about the core issue at hand—the dangerous lack of media literacy among the general public and our carelessness regarding truth itself.

### **What should be done to address the challenges?**

*Maybe the worst thing to do when combating the malicious use of social media is to only put a bunch of middle-aged men into a think tank and then let them search freely for the solutions.*

It is important to keep the purpose of the attacker in mind when combating social media manipulation. Counter-measures should be in balance with the desired effect of the information operation. Striking back with irony or sarcasm can be enough—poor attempts don't deserve complex counter-measures. Many of the counter-measures taken today, such as debunking disinformation, tend to exacerbate the negative effects.

It is also important to remember that Russian information operations are most



” Maybe the worst thing to do when combating the malicious use of social media is to only put a bunch of middle-aged men into a think tank and then let them search freely for the solutions.

effective on RUNET where the Russian state-controlled media outlets have a big impact on local audiences. We must focus more effort on countering disinformation activities on these platforms.

Finally, social media companies are implementing measures against the

malicious use of their platforms. Twitter has taken steps that have proven particularly successful, while Facebook seems to be making a big fuss, but has done little to make a change. Overall, platforms tend only to react to immediate challenges, pro-active measures have yet to be developed.



## 7. BOTS, TROLLS, AND IN-BETWEEN THINGS

Bots and trolls have become ubiquitous actors in the social media space, used by malign actors to exploit human biases and vulnerabilities in the social media ecosystem. Revelations about the interference of the Internet Research Agency, a Russian troll factory, in the US elections, was an eye-opener for many because of the scale of coordination behind these activities.<sup>13</sup>

Bots and trolls are used to amplify certain narratives, manipulate the information environment, and make certain views or political movements seem more popular than they are—an effective way of silencing dissidents while creating social acceptance for the promoted narrative.<sup>14</sup> The same methods are also used to destabilise public discourse, undermine cohesion, and fuel chaos, making it more difficult for people to evaluate what is true and what is false. Research has demonstrated that simple automated bots are able to promote specific interests, magnify trending topics, and gain influence on social media, even when not designed to mimic human behaviour.<sup>15</sup>

As social media platforms take action to address these challenges, malign actors are finding new ways of manipulating our information environment. Simple bots can now be easily identified and neutralised,

but more advanced bots and coordinated activities of automated accounts and human-controlled trolls, often referred to as ‘cyborgs’, create more genuine-seeming interactions that can avoid detection mechanisms.<sup>16</sup> We are currently observing a battle between platform moderators and malicious actors racing to find ways to circumvent security measures.

The malicious use of social media through bots and trolls can also negatively affect military activities. For example, social media analysis can enhance situational awareness and help us better understand ‘the big picture’—to get a feel for certain social networks or certain geographical locations. However, if the big picture is being manipulated by malign actors, the military might be tricked into making poor decisions based on disinformation.<sup>17</sup> This is just one of many examples of how the abuse of social media might negatively affect armed forces. The challenges are many and the issue must be taken seriously.



## HASHTAG #HIJACKING

A well-known case of hashtag-hijacking was the Brussels lockdown in 2015, when civilians hijacked the conversation regarding ongoing police operations in the city by publishing pictures of cats.<sup>20</sup> This case portrays both effective collaboration between police and civil society and a worrying tactic for the malicious use of social media. Malicious actors can co-opt the same tactic to disrupt the information environment, as in the case of hijacking the popular hashtag #Syria in 2011, when irrelevant information spread by bots interrupted the necessary information flow on developments in the country.<sup>21</sup>

## EXPERT INSIGHT—ROBOTROLLING, PROJECT MANAGER, DR ROLF FREDHEIM

*RoboTrolling is a research project conducted by the NATO StratCom Centre of Excellence, prepared and authored by Dr Rolf Fredheim. Prior to the start of the RoboTrolling project in 2017, there was no comparative measurement of automated malicious activity on social media. By studying the discussions of NATO on Twitter and VK in the three Baltic states and Poland, the quarterly report identifies current trends and patterns regarding the malicious use of social media through the use of bots, trolls, and cyborgs.*

### **Latest developments in bots, trolls, and cyborgs**

When the RoboTrolling research project began, a high proportion of automated activity was identified on both Russian- and English-language Twitter, but in late December 2017, the automated activity dropped off due to Twitter introducing more effective responses. However, a disparity can be observed between the effectiveness of the responses in the English and Russian language spaces. The counter efforts Twitter introduced proved much more effective

towards automated activities in the English language space, while the Russian language space remains polluted. Furthermore, as automated activity was reduced, more suspicious inauthentic anonymous activity, from what could be described as cyborgs and trolls, took its place.

But this is not all bad—as the spam content produced and shared by simple bots is removed, our efforts can be targeted towards tracking more advanced disinformation campaigns, thereby allowing us to expose and address the real dangers.





As automated activity was reduced, more suspicious inauthentic anonymous activity, from what could be described as cyborgs and trolls, took its place.

### **Future challenges**

It is probable that non-traditional election campaigning techniques using social media will become more dominant. The market for targeted advertisement is booming and flooding our social media flows; if nothing is done, this trend will grow. The scandal surrounding Cambridge Analytica has also persuaded many actors that such techniques can be very successful. Furthermore, as demonstrated by the French and US elections, political manipulation on social media will likely include cybercrime activities, such as hacking email accounts.

Finally, the techniques and methods for political manipulation of social media are dependent on developments in the commercial sector. We need to stay abreast of developments to identify the cutting-edge techniques that could potentially be used for political manipulation on social media.

### **What should be done to address the challenges?**

Social media platforms must stop treating political manipulation as separate from regular spam. Traditional spamming

techniques use a 'plug and play' tool for manipulating social media—this is a simple toolkit requiring little skill. Platforms must deter such manipulation of the social media environment by making it costlier and more difficult.

We are currently in a 'wild west' phase when it comes to social media. There is little control and anyone can create a fake account to manipulate public discourse. Current verification systems are not adequately equipped to handle the multitudes of malicious actors. Social media companies need to establish more creative and advanced verification systems that both safeguard user anonymity and make anonymous malicious activity more expensive and time-consuming.

It is worth repeating that the economic model social media platforms depend on is based on selling advertisements, and only customers can exert the external pressure necessary to force platforms to change their ways. A key step toward creating incentives for change is growing the awareness of the platform's main customers, the advertisers, that the malicious use of social media for financial gain negatively affects their businesses.





## EXPERT INSIGHT—THE LITHUANIAN ELVES

*The Lithuanian Elves consist of a loose network of civil society activists with the common goal—to combat the vast numbers of pro-Kremlin trolls attempting to denigrate Lithuania, NATO, and the European Union. The network of Elves was created in 2014 as a response to Russia’s illegal annexation of Crimea, aggressions towards Ukraine, and the increase of information operations directed towards Lithuania. At the start, the network consisted of about 40 volunteers but has grown rapidly to a resistance force of several thousand volunteers. This section is based on an interview conducted by NATO StratCom CoE.*

### **Latest developments in Internet trolls and disinformation**

The new frontline for Russian information operations in Lithuania is Facebook. Online trolls are increasing their efforts to polarise Lithuanian opinion. The methods they have adopted are very similar to those used by the Russian Information Agency during the 2016 US elections. Rather than pushing certain narratives, the trolls are disrupting public discourse by adopting extremist positions on both sides of the political spectrum thereby attempting to create divisions within Lithuanian society, often by exploiting already existing dividing issues. There have also been cases in which the trolls attempt to use social media to transcend the virtual world and influence real world events, such as demonstrations.

Russian information operations tend to start in neutral groups on Facebook, such

as fan groups for popular movies or famous actors that attract large numbers followers. The posts initially published in such groups are related to the subject of the group, but after some time disinformation is actively inserted between other posts, thereby exposing a large number of people, such as a fan base, to malicious disinformation. These trolls normally organise through other social media networks, mostly VKontakte, and then engage on Facebook. Although Facebook is the main target, other social media networks such as YouTube have also been affected by these malicious activities. For example, the pro-Kremlin trolls report comments by people who oppose their disinformation videos to great effect, resulting in the suppression and removal of e.g. pro-Navalny voices.

The social media platforms have not responded effectively. For example, during a larger NATO exercise in Lithuania in 2018, disinformation claiming that three tanks



## ” The social media platforms have not responded effectively

had sunk in a river during the exercise was disseminated by a ‘news outlet’ on Facebook; the report was completely false. Lithuanian Elves reported the ‘news outlet’ thousands of times to have it removed from the platform, but Facebook claimed the page was not in breach their terms of service, so the protest had no effect. Even so, the work of the Elves has not been in vain. Since they began their efforts, the need to intervene in discussions and debunk disinformation has measurably decreased and ordinary citizens have become more involved in public debate.

### **Future challenges**

It is highly probable that such activities will continue, and that pro-Kremlin trolls will adapt to future obstacles, however, there are positive tendencies as well. Social media companies are slowly taking action to minimise abuse of their platforms and reduce the amount of malicious activity. Civil society is more active than ever before, and tools such as *debunk.lt* are also being developed to provide an effective means of identifying ongoing information operations.

### **What should be done to address the challenges?**

Online armies of pro-Kremlin trolls can only be countered by armies of Elves—in other words,

we must engage civil society in the fight against disinformation. A comprehensive approach with extensive cooperation between civil society, government, and social media companies must be the point of departure for defence against malicious online activity.

However, there is a clear disparity of financial resources available for information warfare. Therefore, it is vital that civil society and NGOs are supported by governmental institutions to provide journalists and civil activists with the training and tools necessary for the fight against disinformation. But we should not forget that the Kremlin’s efforts are not always as successful as we might think—civil society is strong and resilient, and the work of the Elves constitutes a good example of how civil society can be effectively integrated into countering information influence activities.



# FUTURE TRENDS OF DISINFORMATION

The technological advancements brought about by automation, 5G and Artificial Intelligence (AI) bring about new avenues for information manipulation on a larger scale with wider reaching effects.

Traditionally confined to traditional and social media, development in automation technology has enabled the large-scale spread of disinformation even within encrypted peer-to-peer social networks such as WhatsApp. A dangerous trend, especially in emerging economies such as Brazil and Indonesia where these closed networks account for a high traffic of information.<sup>22</sup> In the 2018 Brazilian elections, it was observed that data scraping technology enabled automated accounts to send up to 300,000 messages at a time.<sup>23</sup> Alarmingly, these accounts also allowed users choose a desired target audience by searching for keywords, pages or public groups on Facebook.<sup>24</sup> The spread of disinformation in encrypted peer-to-peer social networks is especially difficult to combat where its end-to-end encryption makes it challenging to deter, detect or monitor the spread of disinformation within these networks.

5G is touted to be the next big shift in technology. Although disinformation can be spread regardless of the network used (i.e 4G or 5G), 5G arguably brings with it the potential to spread disinformation faster

and with greater reach while reducing the effectiveness of existing disinformation monitoring and debunking mechanisms. The ability to have greater-than-fibre speeds anywhere also presents a greater ability to be manipulated anywhere. Inter alia, 5G will enable IoT (Internet of Things), AI, AR (Augmented Reality) and VR (Virtual Reality) to be used at industrial levels in factories, sea ports, airports and national highways. With 5G connectedness, malicious actors could manipulate, impair or even destroy these 5G-connected industries and transport links by disrupting these lines of communication. 5G would also change the way traditional media gathers and delivers information. With a high-speed connected viewership, media agencies could deliver news in AR or VR mediums. In the context of disinformation, this means more believable fake news. The lack of data bandwidth restrictions also means people would have access to much more information in real time, at all times. News agencies, and malicious actors, could deliver a constant stream of information to viewers, together with live video press conferences and human analysis debates. While 5G in itself does not generate disinformation, it could possibly increase the reach, resonance and material effect of disinformation significantly.

Homes are increasingly becoming automated through Artificial Intelligence



and the Internet of Things (IoT). Current estimates project the value of the IoT industry to be in excess of US\$151 billion and an estimated 7 billion IoT devices in homes.<sup>25</sup> That number is projected to grow exponentially to over 21 billion devices in the next six years.<sup>26</sup> Everything from the heating systems to media devices and home security systems could be connected to the internet. This in itself creates a potential avenue for manipulation through information. While cyber security measures could be put in place as a deterrent to manipulation, the relative cost of doing so has led smart home security somewhat lacking.<sup>27</sup> An example of disinformation

enabled by IoT was seen in San Francisco where a family was alerted to an alarm signal stating North Korean ballistic missiles were en route to sites in Los Angeles, Chicago and Ohio via a speaker in their home Nest camera. Convinced the message was real, the family scrambled to evacuate their home, only to learn later that the message was from a hacker rather than a government warning system.<sup>28</sup>

In essence, while the concepts related to the spread of disinformation are not new, these developments in technology are likely to make the spread of disinformation more believable, faster and more accessible.

## CONCLUSIONS

The technological advancements of our time have created new ways to influence public opinion. These tools are now available to anyone, often in ways we do not yet fully comprehend.

During 2018, there was a relative decrease in the use of social media for news consumption around the world. Reuters Institute assessed that this might be a consequence of Facebook changing its algorithm to downgrade news content.<sup>29</sup> Other observers have identified a partial shift from traditional social media, such as Facebook, to peer-to-peer encrypted chat applications, such as WhatsApp, Signal,

and Snapchat. This change has worrying consequences affecting how disinformation is disseminated as the design of these platforms makes it even more difficult for regulators to identify and counter malicious use of social media.<sup>30</sup> As a result, individual users have a greater responsibility to critically evaluate information they consume, and social media companies should take the necessary steps to tackle this growing threat.

Another shift making disinformation more difficult to monitor is that information operations are increasingly moving from open pages to closed groups, seeding



disinformation to 'invisible' groups that later disseminate it to the public. In other words, just as malign actors are now taking advantage of peer-to-peer encrypted applications, they are also leveraging the potential reach of social media in combination with the platform's closed group function.

As things stand, malign actors are able to hide behind anonymous social media accounts, pages, or groups, exploiting a system designed to protect privacy rights. Much of the news we now consume is being promoted without source attribution and without advertising transparency. This provides many opportunities for malign actors to target unsuspecting audiences with disinformation and other forms of information activities without users ever knowing about it.

The malicious use of social media has developed into a flourishing economy and there are too few obstacles standing in the way of this malicious practice. The social media companies base their economic model on advertising. This model is being harnessed by malicious actors who can pay to promote destructive content. The bot-industry has developed into a lucrative market where people make a living of creating more or less advanced bots that boost views, likes, and shares in the social media space, manipulating the information environment on social media.<sup>31</sup> Buying bots is neither demanding nor expensive. A NATO StratCom CoE publication on the

black market of social media manipulation provides an in-depth analysis of this problem.

These are the broader vulnerabilities that enable the abuse of the online information environment through which malign actors can manipulate public opinion, trick people, and undermine trust in society. Other vulnerabilities, such as lack of training and education, and trust in media and governmental actors, are contextual and vary from nation to nation. The malicious use of social media is not merely a question of abuse of the terms and policies of the social media platforms; it is as much a question of abuse of the human mind and the fundamental tenets on which our democratic societies are based.

Social media companies themselves are also inadvertently creating vulnerabilities when they update their platforms. Recent changes to Facebook Graph Search better protect user information but also seriously hamper the ability of external researchers to identify and analyse malicious use of Facebook.<sup>32</sup> The trade-offs between privacy and transparency – between the right to be anonymous and the need for accountability we will need to find better answers to in the months and years to come.

Trends and developments in social media manipulation:

- The current state of play is a cat and mouse game between malicious actors and governments and the new media



industry. As social media companies and other actors take action to counter abuse, malicious actors adapt to the new environment.

- Impersonation is commonly used both for the spread of disinformation and for social engineering attacks with different degrees of sophistication, sometimes attempting to create real-life events through online activity. Continued technological development in the field of artificial intelligence and frighteningly realistic 'deepfake' video techniques may allow impersonation attacks to become even more credible.
- The methods and platforms used to disseminate disinformation are also changing. The increased use of encrypted platforms, such as WhatsApp or closed Facebook groups, makes it increasingly difficult to identify ongoing information operations. Furthermore, malicious actors are more effective than before in covering their own tracks.
- There has been an increase in the use of cyborgs and trolls in response to social media platforms taking action to defend themselves against attacks from simple automated accounts.



# ENDNOTES

- 1 House of Commons, "Disinformation and 'Fake News,'" n.d., 89; Samantha Bradshaw and Philip N Howard, "Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation," n.d., 26.
- 2 Beata Bialy and Sanda Svetoka, "New Trends in Social Media," NATO Strategic Communications Centre of Excellence, 2016.
- 3 Rand Waltzman, *The Weaponization of Information: The Need for Cognitive Security* (RAND Corporation, 2017), <https://doi.org/10.7249/CT473>.
- 4 Oxford Dictionaries, "[Impersonation | Definition of Impersonation in English by Oxford Dictionaries](#)," accessed August 10, 2018.
- 5 "[18 U.S. Code § 912 - Officer or Employee of the United States | US Law | LII / Legal Information Institute](#)," accessed August 10, 2018.
- 6 Robert Chesney and Danielle Citron, "[Deep Fakes: A Looming Crisis for National Security, Democracy and Privacy?](#)," *Lawfare*, February 21, 2018.
- 7 "United States of America v. Internet Research Agency, No. 18 U.S.C. 2, 371, 1349, 1028A.," n.d.
- 8 "[Factbox: Effort to Sway Election Included 'Clinton' in Prison Garb...](#)," Reuters, February 16, 2018.
- 9 Giorgio Bertolin, ed., "Digital Hydra: Security Implications of False Information Online.," NATO Strategic Communications Centre of Excellence, n.d.
- 10 Bertolin, "Digital Hydra: Security Implications of False Information Online.," 7.
- 11 "United States of America v. Internet Research Agency, No. 18 U.S.C. 2, 371, 1349, 1028A."
- 12 For a more detailed view, please see @DFRLab, "[#ElectionWatch: Italy's Self-Made Bots](#)," DFRLab (blog), January 25, 2018.
- 13 "United States of America v. Internet Research Agency, No. 18 U.S.C. 2, 371, 1349, 1028A."
- 14 "Countering Information Influence Activities - A Handbook for Communicator," n.d., 22.
- 15 Luca Maria Aiello et al., "[People Are Strange When You're a Stranger: Impact and Influence of Bots on Social Networks](#)," ArXiv:1407.8134 [Physics], July 30, 2014.
- 16 Bradshaw and Howard, "Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation," 12.
- 17 "NATO Multinational Capability Development Campaign. 'Applied Concept Social Media in Support of Situation Awareness' Version 1.0, 2014," 2014, 24, 29.
- 18 Michael Safi, "[WhatsApp Murders: India Struggles to Combat Crimes Linked to Messaging Service](#)," *The Guardian*, July 3, 2018, sec. World news.
- 19 "[Facebook Now Linked to Violence in the Philippines, Libya, Germany, Myanmar, and India](#)," *Columbia Journalism Review*, accessed September 10, 2018.
- 20 "[Belgians Tweet Cat Pics to Help Police](#)," *BBC News*, accessed August 14, 2018.
- 21 "[Spam Bots Flooding Twitter to Drown Info About #Syria Protests \[Updated\]](#)," *Global Voices Advocacy* (blog), April 18, 2011.
- 22 Fernanda Saboia, "[The Rise of WhatsApp in Brazil Is About More than Just Messaging](#)," *Harvard Business Review*, April 15, 2016.
- 23 Matheus Magenta, Juliana Gagnani, and Felipe Souza, "[WhatsApp 'weaponised' in Brazil Election](#)," October 24, 2018, sec. Technology.
- 24 Ibid.
- 25 Knud Lasse Lueth, "[State of the IoT 2018: Number of IoT devices now at 7B – Market accelerating](#)," August 8, 2018.
- 26 Ibid.
- 27 K. Ghirardello et al., "Cyber security of smart homes: Development of a reference architecture for attack surface analysis," *Living in the Internet of Things: Cybersecurity of the IoT - 2018*, London, 2018, pp. 1-10.
- 28 Meira Gebel, "[A California Woman Says Her Family Experienced 'sheer Terror' after Their Nest Security Camera Was Hacked, Warning Them of a North Korean Missile Attack](#)," *Business Insider*, Accessed July 4, 2019.
- 29 "Digital News Report 2018," 2018, 10.
- 30 "Digital News Report 2018," 10; Bradshaw and Howard, "Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation," 10.
- 31 Michael H. Keller, "[The Flourishing Business of Fake YouTube Views](#)," accessed August 13, 2018.
- 32 "[Changes to Facebook Graph Search Leaves Online Investigators in a Lurch](#)," *TechCrunch* (blog), Accessed July 4, 2019.









Prepared and published by the  
**NATO STRATEGIC COMMUNICATIONS  
CENTRE OF EXCELLENCE**

The NATO Strategic Communications Centre of Excellence (NATO StratCom COE) is a NATO accredited multi-national organisation that conducts research, publishes studies, and provides strategic communications training for government and military personnel. Our mission is to make a positive contribution to Alliance's understanding of strategic communications and to facilitate accurate, appropriate, and timely communication among its members as objectives and roles emerge and evolve in the rapidly changing information environment.

Operating since 2014, we have carried out significant research enhancing NATO nations' situational awareness of the information environment and have contributed to exercises and trainings with subject matter expertise.

[www.stratcomcoe.org](http://www.stratcomcoe.org) | [@stratcomcoe](https://twitter.com/stratcomcoe) | [info@stratcomcoe.org](mailto:info@stratcomcoe.org)