ROBOTROLLING

Executive Summary

This quarter, the disputed presidential election result and nationwide protests in Belarus were the main targets of inauthentic Russian-language accounts, resulting in a cluster of spikes in fake activity in August. Pro-Lukashenka users concocted an external threat from NATO by pushing false claims of NATO buildup along the Belarusian border and shared rumours of impending intervention. Automated users asserted that NATO posed an internal threat in Belarus as well, alleging that the demonstrations are "puppeteered" by the West.

The situation in Belarus coincided with the most pronounced uptick in attention from identifiably human Russian-language accounts. Compared to the previous report, the portion of messages attributed to identifiable humans increased from 14% to 18% on Twitter and from 26% to nearly 30% on VK. This increase in legitimate engagement in NATO-related discussions of Belarus drove down the percentage of bot users to the lowest figure we have observed, 15% on Russian Twitter and 19% on VK.

English-language activity focused on Polish affairs, both independently and in relation to the ongoing protests in Belarus. Inauthentic English-language discussions peaked with announcements of US troop relocation from Germany to Poland. In September, former US vice president Joe Biden made critical comments about Hungary and Poland, triggering the highest volume of automated retweets from English-language bots this quarter.

Finally, in this instalment of Robotrolling we take a look at the supply side of fake social media accounts. The second iteration of the COE's social media manipulation experiment tracks variation between the responses of Facebook, Twitter, YouTube, Instagram and TikTok to inauthentic engagement. Strikingly, the report found that Instagram is 10x cheaper to manipulate than Facebook, TikTok has virtually no self-regulatory defences, and it remains easy to manipulate US senators' accounts, even during an election period. ◼

The Big Picture

This issue of Robotrolling continues to track the online manipulation of information regarding the NATO presence in Poland, Estonia, Latvia, and Lithuania on the social media platforms Twitter and VK. Our analysis focuses on the activities of automated accounts (bots) and coordinated, anonymous human accounts (trolls). This edition identifies the key trends that emerged in the Russian- and English-language information spaces during the period 1 August to 31 October 2020.

This quarter, we observed a total of 10 700 messages on Twitter referencing the NATO presence in Poland and the Baltics, a slight increase compared to the previous period of 1 May to 31 July 2020. This rise in tweet volume was driven by an increase in activity generated by anonymous users. Half of all Russian-language tweets and 40% of English-language tweets were attributed to anonymous users by our algorithm (see Figure 2). We observed a simultaneous boost in the number of users engaging in conversations about the Enhanced Forward Presence (eFP).

This uptick in message volume and active users on Twitter is the combined result of an increase in Russian-language activity and moderate decrease in English-language activity. The number of Russian-language tweets increased by nearly a third, while the number of Russian-language users increased by 37%. The percentage of bot users decreased to the lowest share we have observed, to 15% on Twitter and 19% on VK. In contrast, English-language messages and the users that disseminated them dipped by 6% and 8%, respectively.

Although overall Russian-language activity increased this quarter, we noted a significant reduction in the volume of automated messages and the number of bot users circulating them. On Twitter, the portion of Russian-language activity attributed to bots fell from 38% observed last quarter to a mere 21% of messages. This can be explained by an increase

in authentic human activity within Russian-language discussions, from 14% to 18% on Twitter and 26% to almost 30% on VK, the highest rates of human engagement we have observed. Automated activity in the English-language sphere dropped as well, with only 9% of messages attributed to bot users.

We observed a similar pattern on VK, where the total number of posts and active users increased, but bot activity fell by 10%. Poland and Lithuania were the central targets of bot, anonymous, and human activity on VK (see Figure 3). The absolute number of posts shared by bots on VK increased marginally, along with the number of bot users discussing the eFP. ◼

RU 15%

EN 8 %

RU 19%
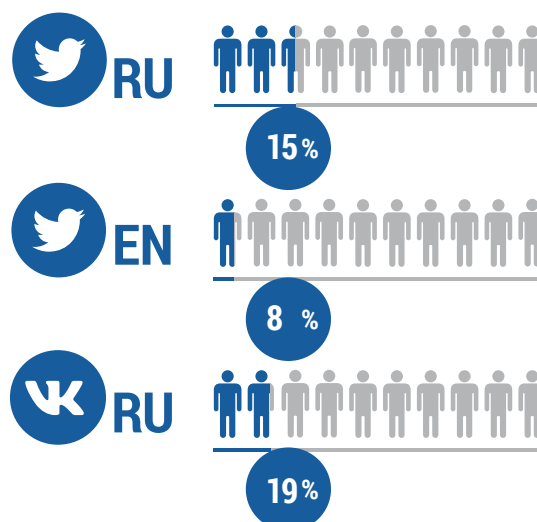
Country Overview

This quarter, discussions about the NATO presence in Poland and the Baltics played out against the backdrop of the 2020 Belarusian presidential election, which was held on 9 August. The contested election outcome, and the civil unrest that followed in its aftermath, featured particularly prominently throughout August and September, resulting in clusters of spikes in anonymous and bot activity. Belarus-related messaging driven by bots echoed official Belarusian statements and propaganda outlets' outlandish claims that Poland and Lithuania were influencing the situation. Automated activity was at its highest in August and subsequently waned in September and October.

On Twitter, Russian-language bot activity peaked on 6 September, with bots disseminating a statement by Russian Defense Minister Sergei Shoigu claiming that NATO plans to redeploy additional US military units to Poland in the near future. Bot activity on VK spiked earlier, on 17 August, driven by discussions of Polish and Baltic statements of support for a peaceful democratic transition in Belarus. In the English-language information space, automated activity peaked on 16 August, coinciding with US Secretary of State Mike Pompeo's signing of a new defence agreement with Poland agreeing upon the redeployment of American troops there from Germany. These spikes are visualised in Figure 4 below.

Estonia

Inauthentic Russian-language activity targeting Estonian affairs decreased this quarter compared to the summer months, with Estonia receiving the fewest mentions on both Twitter and VK. Most mentions of Estonia occurred in discussions of the Baltic countries as a collective regional grouping. In early September, bots circulated reports that Estonian Foreign Minister Urmas Reinsalu urged NATO to remain vigilant of the political turmoil in Belarus. On English-language Twitter, bots shared an RT article relaying a statement by the Russian Embassy in Washington condemning US exercises in Estonia as provocative.
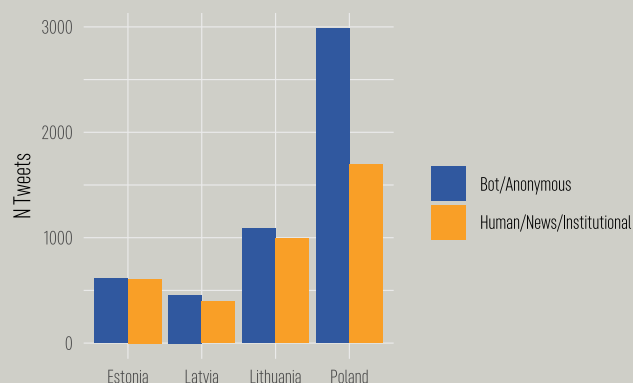
Latvia

Compared to the previous quarter, Latvia received significantly less attention from bots operating in the Russian-language information space and was the least-mentioned country on English-language Twitter. In late September, automated accounts circulated Russian articles about an interview Latvian MP Atis Lejiņš gave with Latvijas Avīze about the future of the Riga City Council. The articles focused on his responses to questions about the role of the Russian language in Latvia, portraying his statements as Russophobic.

Lithuania

This quarter, Lithuanian-Belarusian relations were the central focus of Russian-language bot users, constituting a marked increase and a sharp change from previous periods. Bot discourse about Lithuania concentrated on the diplomatic back-and-forth between Belarusian president Alexander Lukashenka and various officials in the Lithuanian government. While this topic dominated discussions throughout our monitoring period, we also observed spikes in bot activity surrounding NATO exercises in the country and public discussions of Ukraine's eligibility to join NATO and the EU.

Poland

Per usual, Poland received the bulk of English-language bot attention this quarter. Despite typically attracting low levels of Russian-language bot activity, Poland received the second-highest volume of bot mentions on Russian Twitter and VK due to the ongoing political crisis in Belarus. On 18 September, during an ABC News town hall, former US vice president Joe Biden appeared to group NATO members Hungary and Poland in the same category as Belarus. Reports of this statement triggered a burst of bot activity, culminating in the highest peak of retweets observed this quarter. ■
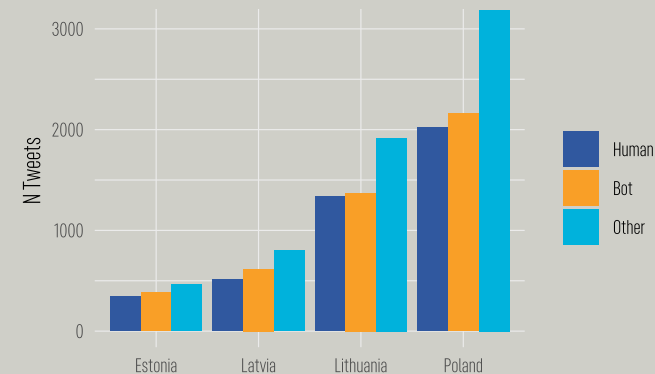

Figure 2: Country comparison of English-language bot and anonymous activity.

Figure 3: Country comparisson of Russian-Language Bot Activity on VK

Themes

Both authentic and inauthentic discussions about NATO's presence in Poland and the Baltics revolved around the 2020 presidential election in Belarus and the nationwide protests that erupted in its wake. President Alexander Lukashenka, who has held office for the past 26 years, was declared the winner of the 9 August election with 80 percent of the vote. Sviatlana Tsikhanouskaya, the main opposition candidate, fled to neighbouring Lithuania shortly after the results were announced. In the following weeks, thousands of protesters filled the streets to dispute Lukashenka's reelection, which is both domestically and internationally considered fraudulent.

Although mass demonstrations began soon after the election results were announced, we observed a notable delay in hostile online inauthentic activity discussing the situation in Belarus. This same delay was visible in Russian print and online media. During the first week after the election, Belarus-related hashtags on Russian Twitter largely supported the opposition. However, the silence on pro-Kremlin Twitter was broken by anonymous users on 16 August, coinciding with a series of phone calls between Lukashenka and Russian president Vladimir Putin in which the Belarusian leader confirmed that Russia would provide military assistance against external threats. Automated activity on Twitter and VK had picked up a few days earlier, on 13 August, with discussions about hypothetical NATO intervention in Belarus. It is hard to say for certain what part of this activity was fake, but the delay may point to pro-Kremlin trolls taking a cue from Putin's expression of support for Lukashenka. This would further suggest either an absence of domestic

pro-Lukashenka cyber troops, or that they were not preemptively poised to defend the regime online after the election.

Activity throughout August focused on a diplomatic dispute between the eFP countries and Belarus caused by false claims of NATO involvement with the opposition protests. On 13 August, the presidents of Poland, Estonia, Latvia, and Lithuania issued a joint statement calling upon Lukashenka to de-escalate the situation in Belarus, rousing Russian-language bot activity on Twitter and VK. Inauthentic users shared Lukashenka's accusations of a "NATO military buildup" along the Belarusian border, spearheaded by Poland and Lithuania. Despite being disputed by Lithuanian, Polish, and NATO officials, accusations of western "puppeteering" of the anti-Lukashenka protests continued.

Overall, the contested election result and widespread political protests in Belarus were discussed by anonymous and bot users in the wider context of a geopolitical struggle between Russia and the West. The demands of western governments and institutions for Belarus to halt the use of force against protesters, release detained political prisoners, initiate a dialogue with the public, and hold free and fair elections have been interpreted by many online supporters of Lukashenka as attempts to overthrow the government. The prevailing narrative asserts that, if Lukashenka concedes and initiates a transfer of power to the opposition, Belarus will be forced to erase the Russian language and culture from its ethos, forfeit the northwestern Grodno region to Poland and Lithuania, and join NATO and the EU. ■
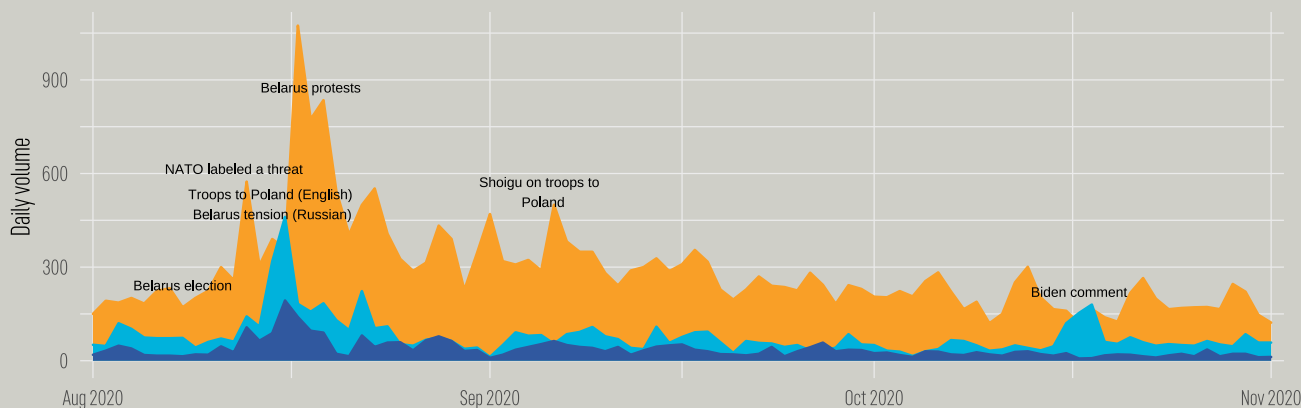

Figure 4: Timeline showing spikes in activity about the NATO presence on English and Russian-language Twitter, as well as VK.

Catching up to bots for hire

In the digital age, the use of coordinated social media manipulation campaigns to shape public debate has steadily increased. These groups of antagonists, which can range from foreign governments to terror groups to commercial enterprises, rely on fake accounts and inauthentic behaviour to undermine the legitimacy of online conversations. This activity has the potential to harm individuals and society, both online and offline.

In 2019, the NATO StratCom COE conducted an experiment to evaluate mainstream social media companies' ability to identify and remove inauthentic manipulation from their platforms, the results of which were published in the report Falling Behind: How Social Media Companies are Failing to Combat Inauthentic Behaviour Online. The authors, having spent only 300 EUR, were able to purchase 54 000 fake interactions (comments, likes, shares, and video views) for posts on Facebook, Instagram, Twitter, and YouTube. The findings of this report demonstrate the ubiquity of the commercial social media manipulation industry and explore the challenges of social media self-regulation.

Since the release of these findings, social media platforms have committed to strengthening their online defences, particularly ahead of the 2020 US presidential election. Over the past year, Facebook, Google, and Twitter have updated their policies and increased transparency regarding manipulation of their platforms. However, independent assessment of the efforts' effectiveness continues to be difficult, and, in some cases, impossible. The ongoing and evolving threat of online interference in democratic processes has underscored the need for a whole-of-society approach to define acceptable online behaviour, in addition to the development of deterrence frameworks.

For these reasons, the NATO StratCom COE decided to re-run the original experiment one year later. In this iteration of the experiment, the primary purpose was to test the ability of social media companies to withstand manipulation from well-resourced commercial manipulation service providers. By adopting this approach, the authors could track variation among platform responses rather than the relative performance of manipulation providers. Additionally, the authors expanded the experiment to include a fifth social media platform: TikTok. TikTok, a Chinese-owned video-sharing social network, has become one of the fastest-growing social media platforms in the world with over 700 million active users.

Social media platforms have pledged to safeguard elections in particular against social media manipulation, which includes inauthentic behaviour such as the artificial inflation of likes, views and comments. Due to the timing of the experiment and the contentious context of the 2020 US presidential election, the NATO StratCom COE partnered with US Senators Chuck Grassley (R-IA) and Chris Murphy (D-CT) to assess whether, and to what extent, their social media accounts could be shielded from manipulation. Including this element allowed the authors to draw conclusions about the abuse of verified social media accounts in general.

To test the ability of social media companies to identify and remove fake activity, the authors bought engagement on 39 Facebook, Instagram, Twitter, YouTube, and TikTok posts using 3 high-quality Russian social media manipulation service providers.
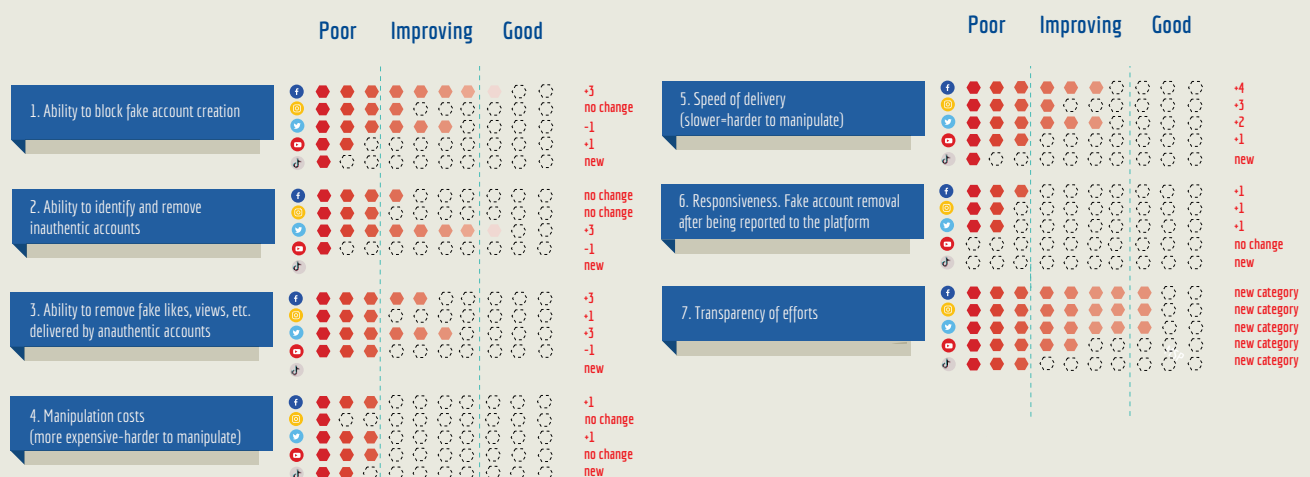

Figure 5: Assessment of platforms relative strenghts and weaknesses. In red - relative change from 2019.

Catching up to bots for hire (continued)

Once again, the authors found that 300 EUR can buy significant engagement: they received a combined 1 250 inauthentic comments, 8 700 likes, 323 000 views and 2 700 shares on the five platforms. This enabled the authors to identify approximately 8 036 accounts being used for social media manipulation.

When compared to last year's experiment, the authors noted positive improvement on Facebook and Twitter (Figure 5). Twitter stands out for its ability to remove inauthentic accounts from the platform, with fake accounts disappearing at a rate 40 percent faster than in 2019, and at a rate indicating Twitter is three times faster than Facebook at removing accounts engaged in inauthentic activity on its platform. Despite being owned by Facebook, Instagram remains much easier to manipulate. YouTube continues to struggle with manipulation, while TikTok performed the worst of all platforms, having failed to remove even the cheapest forms of manipulation.

A disparity exists in manipulation costs on the different platforms. While 10 euros will buy more than a thousand comments on Instagram, the same amount will only fetch 130 comments on Facebook.

Although cheap manipulation services are still primarily used for commercial purposes, the authors noted continued attempts to influence political discussions. In an experiment conducted with two Senators during the US election, they demonstrate that it remains just as easy to manipulate fake content on material posted by verified political accounts than on material published by non-verified users.

There are no solid indications to suggest that the platforms have put in place any kind of additional safeguards for verified political accounts to counter against this form of manipulation during the current US election cycle. Therefore, efforts to protect official accounts, counter bot-networks, and combat commercial manipulation need to be further improved.

In light of their findings, the authors recommend:

- Increasing transparency and developing new standards for comparison
- Establishing independent and well-resourced oversight
- Regulating the market for social media marketing
- Pressuring social media platforms to do more
- Developing a whole-of-industry solution

In the first experiment, the authors discovered that the different social media platforms weren't equally poor at removing manipulation, that, in fact, some were significantly better than others. The findings from the COE's second experiment reinforce this dynamic, as the authors observed improvements on some platforms and record lows on others. Investment, resources, and determination make a significant difference for the ability of social media companies to counter manipulation.

Fore more, read the report "Social Media Manipulation 2020. How Social Media Companies are Failing to Combat Inauthentic Behaviour Online". ■