ROBOTROLLING
2019. ISSUE 1

ROBOTROLLING

Executive Summary

This report presents top-level findings from the first research project to systematically track and measure the scale of inauthentic activity on the Russian social network VK.

On VK, a vocal core consisting of loyal news media, pro-Kremlin groups, and bots and trolls dominates the conversation about NATO. The volume of material from this core group is such, that overall genuine users account for only of 14% of the total number of messages about NATO in the Baltic States and Poland.

The spread of demonstrably fake content can offer a starting point for measuring how social media manipulation impacts genuine conversations. In the case of one story about a fictitious Finnish blogger, our algorithm estimates that at least 80% of users who shared the fake story were authentic.

This quarter, messages appeared in more than 2 000 different group pages on VK. Setting aside messages from group pages, 37% of VK posts came from 'bot' accounts—software that mimics human behaviour online. This level of activity is comparable to what we have seen on Russian-language Twitter. Unlike on Twitter, where the vast majority of human-controlled accounts are operated anonymously, on VK most accounts are likely to be authentic.

Western social media companies have belatedly taken an active role in reducing the reach of the Kremlin's social media manipulation efforts. However, it remains hard for researchers to evaluate the effectiveness of these measures on platforms such as Facebook and Instagram. In this context, VK offers a cautionary view of a network with minimal privacy, regulation, and moderation. ◼

The Big Picture

Robotrolling analyses the manipulation of information regarding the NATO presence in the Baltic States and Poland on the social media platforms Twitter and VK. It focuses on two main types of manipulation: automated activity from robotic accounts, and messaging from fake human accounts. Fake human accounts may be operated by patriotically minded individuals or groups, or generated for profit by so-called troll factories.
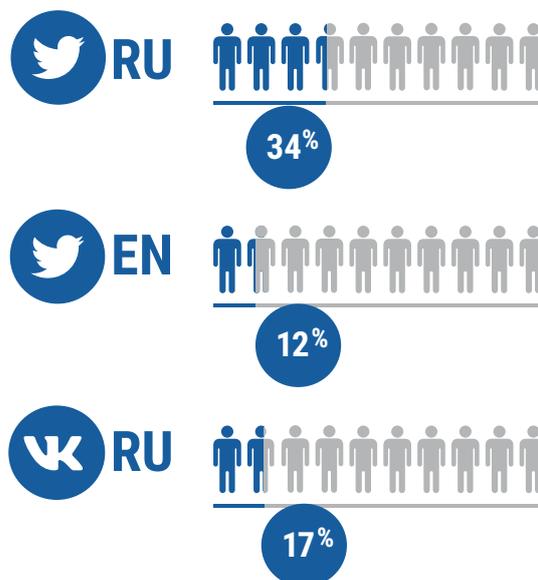
This quarter the NATO presence was mentioned in 10 100 posts from 4 000 accounts. Of these messages, 6 000 were posted by accounts of groups or pages, whereas the remaining 4 100 were posted by individual accounts. This issue focuses on the messages from individual accounts, presenting the first estimates about the levels of robotic and troll activity on the platform.

By comparison, the total number of Twitter posts is 5 520, of which 30% were in Russian. The number of active users was 3 740. The number of posts and active users is on par with the previous quarter.

The level of activity on VK was six times higher than on Russian-language Twitter. Much of this activity is inauthentic. We estimate that only 14% of the public messaging on the platform about NATO in the Baltics and Poland comes from recognisably human accounts.

Setting aside posts from groups and pages, VK bot levels are on par with those on Russian-language Twitter. In the period analysed, bots created 38% of messages about NATO in the Baltic States and Poland on VK. On Twitter, bots created 37% of all Russian-language messaging. English-language Twitter stands in stark contrast, with bots creating 13% of all messages.

On VK, bots are a hyperactive minority, accounting for 38% of the messages but only 17% of users. On Russian-language Twitter, bots post a comparable 39% of messages but make up 34% of active users. On average, VK bots are more focused on a single political agenda. ◼

Country Overview

Commentary about the NATO presence in the Baltics and Poland peaked in November, at the time of Latvia's Centenary and the Kerch Strait incident. Following a lull in troll and other activity over the holiday period (lasting until just after Orthodox Christmas on January 7) the conversation turned to the Alliance's willingness to go to war with Russia over the Baltic States.

On VK, human users posted similar levels of messages about all four countries, but bots and group pages paid disproportionate attention to Lithuania and Poland. On Twitter, both Russian- and English-language bots targeted Poland, as did anonymous English-language accounts.

The English-language conversation was dominated by reactions to the January 2019 US House of Representatives legislation aimed at preventing any US move to withdraw from NATO. In late January comments by Tucker Carlson and other conservative US commentators gave fuel to the Russian narrative that NATO would not defend the Baltics.

Estonia

This quarter, robotic activity about Estonia was comparatively low until late January. In the US, the conservative commentator Tucker Carlson questioned whether the US would wish to protect, say, 'the territorial integrity of Estonia'. Bots and anonymous accounts on VK and Twitter alike pushed multiple sources citing a Sputnik piece about Latvia, Lithuania, Estonia, and Poland being expendable 'second-rate states' that NATO would neither wish to nor be able to defend.

Latvia

The Latvian Centenary on 18 November drew a split reaction on social media. The English-language space was dominated by formal greetings from international figures. The Russian-language conversation peaked the next day, with critical reactions from anonymous Twitter users. On January 17, choice quotes from an interview with Harmony politician and military expert Raimonds Rublovskis about the weakness of NATO bases were widely shared.

Lithuania

In early December, Lithuania introduced sanctions in response to Russian aggression against Ukraine in the Kerch Strait. This decision was widely attacked on VK, in particular by bots and by pro-Kremlin/ pro-Donbass groups. The number of bot-generated posts on VK about Lithuania was twice the number of posts about Estonia and three times greater than for Latvia. In military news, the Polish air force took over the lead for NATO's air policing mission in the Baltics. The handover from Belgium, which took place in Šiauliai on January 3, caught the attention of bots on Twitter, but was largely absent from VK.

Poland

In early November, the Polish-led multinational exercise ANAKONDA attracted high levels of attention, especially on VK where it was branded 'anti-Russian'. On 30 November 2018 the Russian Deputy Minister of Defence, General Alexander Fomin, characterised the exercises as a smokescreen for a NATO troop build-up on Russia's border. Maria Zakharova spoke of NATO's efforts to 'militarise the region'. The volume of English-language Twitter posts about NATO in Poland remains high. Anonymous human-operated accounts posted 50% of all English-language messages, compared to 39% for the Baltic States. ■
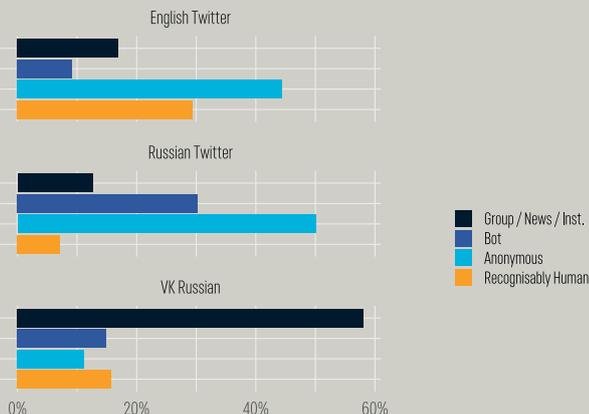

Figure 2: Comparison of account types by platform. News and institutional accounts on Twitter are grouped.
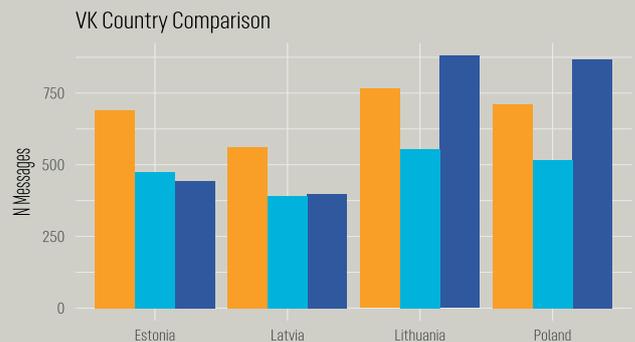

Figure 3: VK Country comparison for activity by different account types. Group pages are omitted.

Themes

This quarter, both bots and group pages on VK reflected a Russian media agenda roundly condemning Lithuania's decision to introduce sanctions. Users classified as humans by our algorithm paid less attention to these and other stories posted in response to NATO activities in the Baltics and Poland. Instead, derivative and recycled content generated sustained interactions with real people using the platform.

The Kerch Strait incident provoked anti-NATO and anti-Ukrainian commentary in nationalist circles. Two pieces in particular went viral, primarily shared by accounts belonging to individuals. The first was the latest in a series of parodies of Anastasia Dmitruk's poem 'Never ever can we be brothers', composed in response to the occupation of Crimea in 2014. The parody appeared online on 28 November, three days after the Kerch Strait incident. It was promoted widely on various websites and on VK groups supporting the imaginary state of Novorossiya, but also by recognisably human VK users.

The second viral post was a message about a fictitious Finnish blogger Veikko Korhonen from Oulu who was supposedly fortunate enough to have a well-educated grandmother who told him the truth about Finno-Russian history. Fed-up with Western anti-Russian sentiment, the blogger set things straight in a Facebook post on his feed, where he advanced a tendentious reading of history that half of Europe and Asia owe Russia eternal gratitude for giving them their independence.

There is no blogger from Oulo called Veikko Korhonen, and the picture attached to the profile is of Atte Korhola, a professor of climate science from Helsinki University. The post was originally published on VK, not Facebook in July 2015 to a group supportive of the self-proclaimed Donetsk People's Republic. Since this time, the post has repeatedly entered the news cycle.

The spread of content about the post illustrates VK's fake content eco system. The Veikko Korhonen account is primarily a bot that automatically posts news stories and social media content from various pro-Kremlin and nationalist Russian news outlets and groups. As many as 230 distinct group pages shared the story in 2018. News reports about Korhonen have been published by established Russian outlets such as Argumenty i Fakty, as well as at least 33 unique web domains with little or no reputation for credible news. VK groups and fake media outlets maintain a feedback loop that ensures constant visibility. The story first went viral in January 2018. During 2018, material about Korhonen was shared on 287 separate days.

Although everything about the Korhonen story is fake, genuine human VK users were more likely to share it than any other content type. Overall our algorithm estimates that only 14% of the public messaging about NATO in the Baltics and Poland on VK this quarter came from recognisably human accounts. In contrast, 67% of messages about Korhonen were published by human users. ■
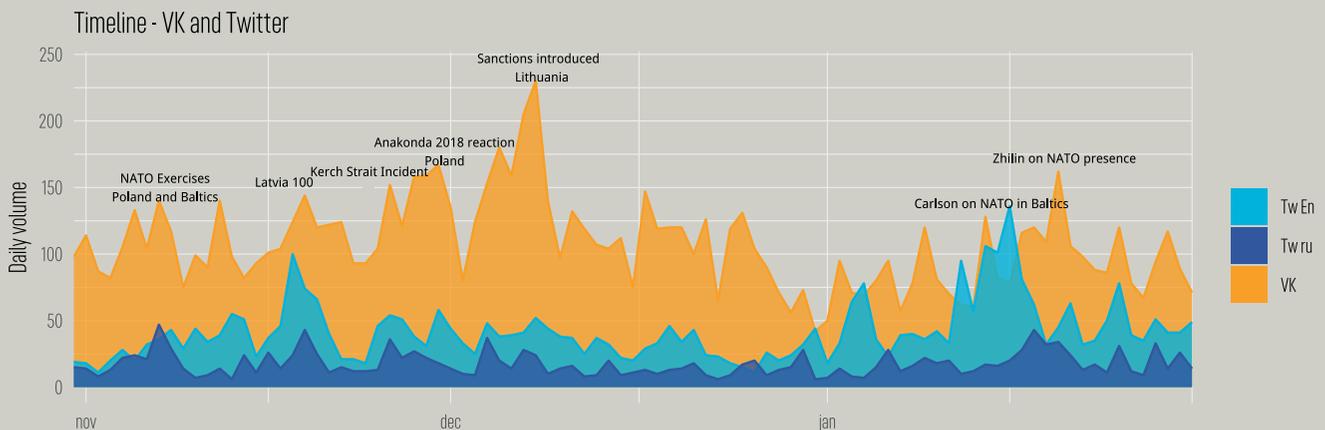

Figure 4: Timeline contrasting VK, English-language Twitter posts, and Russian-language Twitter posts for November 2018 – January 2019.

Robo-topics

It bears repeating that fake social media accounts have a real impact. Today advertisers, journalists, politicians, and researchers look to social media to test the appeal of ideas, policies, and products. Researchers measure metrics to quantify the mood of the population. And algorithms automatically promote trending content to real users, meaning what people see by default can be manipulated.

Actors seeking to game social metrics or the public conversation may use fake accounts to do so. Such fake citizens may use geographical markers—place or residence, birth, school or university—to create a false impression of the opinions and preferences of a particular demographic.

To test whether manipulation on VK is concentrated in particular geographic areas, we combined our algorithm's prediction with self-declared user data for 125 000 accounts which mentioned NATO in the period January 2018—January 2019. Together, these accounts listed 5 200 different places of residence. We identified the geographical coordinates of each location and calculated the proportion of genuine users from a given area to discover the location of suspicious accounts.

For example, 13 200 VK users gave Moscow as their location. Our algorithm identified 60% of these users as human, 32% as anonymous, an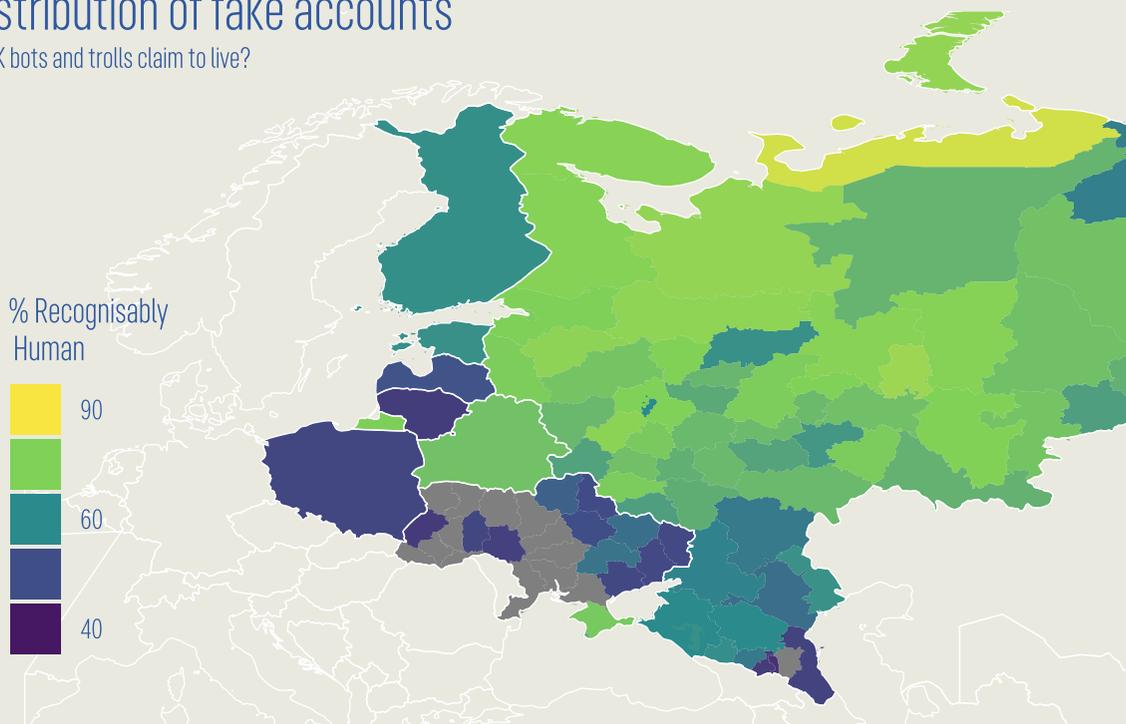d 8% as bots. Errors, possibly large errors, are inevitable: some people choose to maintain anonymous accounts, and some accounts will be misclassified. However, the errors will be similar for all areas, so the numbers form a good baseline for comparison.

Consider all 8 300 VK users in the sample self-reporting as located in Ukraine. Our algorithm estimates 47% of these users are human, 45% are anonymous, and 8% are bots. The difference to Moscow, where recognisable humans outnumbered anonymous accounts 2 to 1, is striking. Perhaps we should expect such statistics from a conflict territory where warring parties are likely to be manipulating the information space and individuals may prefer to remain anonymous given the circumstances.

Figure 5 shows the concentration of people (lighter) and fake accounts (darker) by self-reported location information. It is based on the location provided by VK users who mentioned NATO in the period 1 January 2018 to 31 January 2019. Greyed-out areas have very low user numbers. The figure shows that Ukraine is not the only place where the proportion of suspicious accounts is high. Across all states on Russia's periphery the proportion of probably fake accounts is much higher than in Russia itself. This holds true for the Baltic States and Poland, but also for Finland. Another area with high levels of suspicious activity is the troubled Caucasus region. ■

Geodistribution of fake accounts
Where do VK bots and trolls claim to live?

% Recognisably Human

90

60

40

In Depth: Fakes on VK

VK is the top social network in Russia, and it is popular with Russian-speaking populations throughout the former Soviet territories. It was also the top social network in Ukraine until the Poroshenko administration's ban in 2017.

VK has been mired in controversy. In 2014, its founder, Pavel Durov, was forced to hand over control of the network to oligarchs Igor Sechin and Alisher Usmanov, both of whom have intimate ties to the Kremlin and Russia's Federal Security Service (FSB).

Despite heightened awareness of the use of bots and trolls for political manipulation, only a handful of academics and think-tanks analyse VK data. None attempt to quantify the extent to which bots and other fake accounts are present on the platform. This oversight is surprising, not least because the platform has a generous API providing abundant data access.

VK has taken a hands-off approach to moderation. The platform removes pornography and terrorist content, but has a more relaxed approach to hate speech, and it has consequently gained a reputation as a safe space for political extremists. Additionally, a combination of push and pull factors have created a heavily politicised social environment. Suspicion that the FSB uses VK to snoop on its citizens stifles opposition, while financial incentives make it a lucrative space for interests that overlap with the Kremlin's.

On VK we divide accounts into three types: recognisably human, anonymous, and bot. Various types of accounts are included within the anonymous category. The first is the so-called burner account—a disposable persona with a largely empty profile and few friends. Such accounts are often disproportionately active, especially within group conversations. Another common type is the stolen, hacked, or copied account, containing genuine but outdated content that end sudden-ly and is replaced by impersonal posts. Troll accounts usually produce a high-volume of highly political content, have no verifiable personal information, and do not appear to be embedded in a social graph (have no genuine friends). Our algorithm also identifies automated accounts that post material without human input based on patterns in posting behaviour.

A vocal core of VK users consisting of loyal news media, pro-Kremlin groups, and high-volume bots and trolls dominates the conversation around the NATO presence in the Baltics and Poland. This configuration should not be simply dismissed as controlled and bankrolled by the Kremlin. It is created and maintained by a heterogeneous hodgepodge of patriotic youth, veterans and other interest groups, political parties, media outlets of all stripes, social media and political entrepreneurs, true believers, armchair warriors, and—yes—troll factory employees.

The vocal core operates through the interplay of groups, unconventional online sources, and hybrid accounts—part bot, part troll. They are active across all major social media platforms, but VK seems to be their main area of focus. For instance, many groups and accounts maintain Facebook and Twitter pages that automatically cross-post messages from VK to fill their feeds.

Bots certainly exist on VK. They operate at high frequencies to promote the visibility of certain voices and to game metrics. Compared to Twitter, however, they play a less dominant role in the conversation, which on VK largely takes place within groups. This quarter, messages about the NATO presence appeared on more than 2 000 different group pages, and the posts made up 60% of the total messaging. Whereas it is comparatively rare to find authentic Russian-language voices on Twitter, the bulk of accounts are plausibly authentic, even within the subset commenting on the politically-charged subject of NATO's presence in the Baltics and Poland. ■