

ISBN 978-9934-619-32-8



Social Media Manipulation 2022/2023:

Assessing the Ability of Social Media
Companies to Combat Platform Manipulation

PREPARED AND PUBLISHED BY THE
**NATO STRATEGIC COMMUNICATIONS
CENTRE OF EXCELLENCE**



ISBN: 978-9934-619-32-8

Authors: Rolf Fredheim, Sebastian Bay, Tetiana Haiduchyk, Anton Dek, Martha Stolze

Copy Editor: Merle Anne Read

Design: Inga Ropša

This report was completed in January 2023, based on an experiment that was conducted in September and October 2022. Discussions with social media companies regarding preliminary results took place in December 2022.

NATO STRATCOM COE

11b Kalnciema Iela

Riga LV1048, Latvia

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 do not represent the views of NATO.

Executive Summary

In this report—the fourth version of our social media manipulation experiment—we show that social media companies remain unable to prevent commercial manipulators from undermining platform integrity. Overall, no platform has improved compared to 2021 and, taken together, their ability to prevent manipulation has decreased.

Buying manipulation remains cheap. The percentage of accounts identified and removed by the platforms dropped. We demonstrate that the manipulation providers have circumvented sanctions imposed in response to Russia's full-scale invasion of Ukraine. It remains easy to pay for manipulation services with both Visa and Apple Pay. The platforms' ability to combat manipulation by slowing the speed of delivery has declined. Today, 89 per cent of purchased inauthentic behaviour is delivered within one day. The vast majority of the inauthentic engagement remained active across all social media platforms four weeks after purchasing. Thus, the platforms' moderation decisions appear to be only minimally responsive to user notifications.

Social media manipulation services hence continue to outperform social media platforms. With the quality of transparency reporting unchanged, the gap between platform performance in countering inauthentic engagement and the quality of platform reporting is widening. Platforms have found it expedient to focus less on preventing commercial manipulators from accessing the platform, and more on reducing the reach and impact of their posts. However, our research shows that commercial accounts are exploiting flaws in platforms, and pose a structural threat to the integrity of platforms. More data is required to assess whether the platforms' approach adequately mitigates the systemic risk posed by platform manipulators.

Introduction

In this fourth iteration of our social media assessment, as with our previous experiments, the primary aim is to test and assess the ability of social media companies to withstand manipulation by well-resourced commercial manipulation service providers. Assessing social media manipulation furthers our understanding of the tools and techniques used to manipulate platforms. It provides a framework to discuss specific issues with the social media companies and to deepen our understanding of their abilities to counter platform manipulation.

This edition of our experiment is the first to monitor social media manipulation since the Russian invasion of Ukraine on 24 February 2022. As a result of the war, the information environment has changed dramatically. Sanctions now hinder the Kremlin's messaging on Western platforms, and in Russia, access to Instagram, Facebook, and Twitter was restricted and later blocked. The Russian social media manipulation industry, however, seems largely unaffected.

Why Does This Matter?

In our experiment we buy inauthentic interactions on social media content in order to assess how good social media companies' systems are at independently detecting and blocking manipulation. A recurring theme in conversation with social media companies is that they prioritise moderating content that is likely to have a high impact and/or cause harm. Our interventions are designed to have low impact and be harmless for ethical reasons.

We argue that experiments of this type offer an effective way of assessing how platforms handle fake activity. Attempts to separate state actors from commercial ones and focus on the activity of the former miss the bigger picture: a company taking commercial clients today may be used by an actor seeking political ends tomorrow. Allowing the commercial industry to flourish has the added downside that there is a 'talent pool' from which state actors can recruit.

The simple, cheap, commercial, and highly available manipulation relies on accounts that have a very specific footprint. If the algorithms do not spot this activity, they will be unlikely automatically to detect a more determined actor. Our experiments show, at the very least, how fast, cheap, and effective it is to buy low-level commercial manipulation of social media platforms. And we track, assess, and compare performance to further our understanding of inauthentic manipulation of the information space.

Our experiment is especially relevant in light of the EU's Digital Services Act, which will oblige social media companies to conduct risk assessments of the threat posed by inauthentic accounts, and detail what mitigation measures have been put in place.

The Experiment

During September and October 2022, we conducted an experiment to test the ability of social media companies to identify and remove manipulation. Using one Indian, one Nigerian, and three Russian social media manipulation service providers, we bought inauthentic engagement on 44 Facebook, Instagram, Twitter, YouTube, TikTok, and VKontakte posts. As the methodology is increasingly standardised, we refer the reader to the description in our [2022 report](#).

For €168 we received inauthentic engagement in the form of 1225 comments, 6560 likes, 15,785 views, and 3739 shares on Facebook, Instagram, YouTube, Twitter, TikTok, and

VKontakte, enabling us to identify 6564 accounts used for social media manipulation. Of the 27,309 fake engagements purchased, more than 93 per cent remained online and active after four weeks.

Although all platforms except Instagram have made limited improvements in at least one criterion, the overall ability of the platforms to prevent manipulation has decreased (see Figure 7 for an overview of all criteria). They have not built on the significant improvements noted in recent reports. In particular, with regard to platform transparency efforts, we observe that progress has stagnated.

Changes to the methodology are:

In assessing the price of fake accounts, we switched from taking an average across multiple services to selecting the minimum price obtainable.

For the first time we checked inauthentic accounts identified in the previous report, approximately one year after first reporting

them as fake. This metric gives an idea of the long-term rate at which inauthentic accounts are removed.

We sent the companies a list of questions arising from our results; their responses were factored into our transparency scores.

The Seven Criteria

1. Blocking the Creation of Inauthentic Accounts

This section assesses the ability of social media platforms to counter the creation of inauthentic accounts. The metric combines two measures: the cheapest price available on marketplaces for inauthentic accounts, and an assessment of how hard it is to create burner accounts on each platform.

In our analysis, Facebook is the industry leader when it comes to countering fake account creation, while Instagram and VKontakte are

the worst performers. Meta continues to show a lack of internal coordination between its two platforms Facebook and Instagram. Facebook blocked some of our attempts to create burner accounts, while Instagram accepted the same fake personal data for account creation.

The price of inauthentic accounts has fluctuated but remains cheap (Figure 1). In 2022 the most expensive accounts were Facebook accounts (€0.12 per account), while YouTube accounts were the cheapest (€0.04 per account).

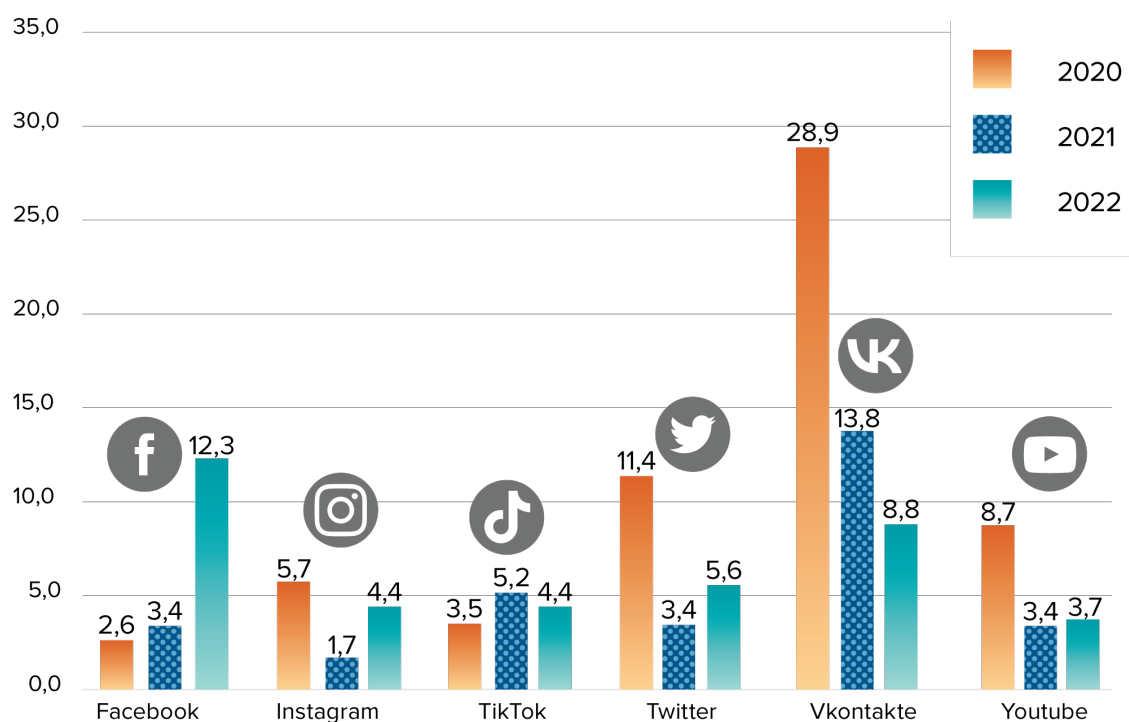


Figure 1: Cost (euro cents) of purchasing fake accounts by platform over time

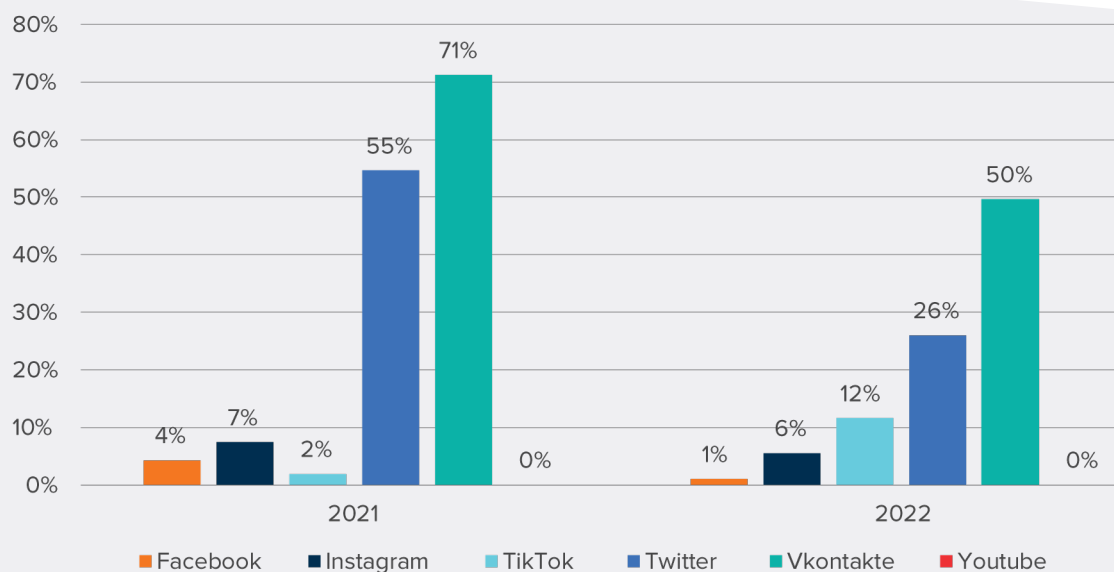


Figure 2. Removed inauthentic accounts during the monitoring period

2. Removing Inauthentic Accounts

Social media platforms’ ability to detect and remove inauthentic accounts is central to their ability to counter platform manipulation. During last year’s assessment, on average, 25 per cent of all the identified accounts were removed by the social media platforms. This year the percentage of accounts removed dropped to 16 per cent overall. By this metric, the platforms’ ability to detect and remove inauthentic accounts is deteriorating (Figure 2).

Vkontakte removed the largest number of inauthentic accounts, albeit significantly fewer compared to our previous experiment.

Twitter still ranks second for its ability to remove inauthentic accounts, but compared to last year, the platform’s removal rate plummeted, dropping from 55 per cent to 26 per cent. The only social network that improved compared to the previous year is TikTok—an increase from 2 per cent to 12 per cent of accounts removed. Facebook performed slightly worse, while we observed no significant change for YouTube and Instagram. Performance by Facebook,

YouTube, Instagram, and TikTok was unsatisfactory. They all failed to remove even 10 per cent of the inauthentic accounts during the four-week monitoring period.

3. Removing Inauthentic Activity

Removing inauthentic activity is the process of identifying and removing fake engagement posted on the platform. The faster the inauthentic activity is removed, the smaller the effect the engagement will have on social media conversations, as fewer people will have had the chance to interact with it. In all three previous reports we have shown that social media companies struggled to automatically identify and remove fake activity, and that the vast majority of all the fake engagement was still online four weeks after delivery (Table 1).

This year, Twitter and YouTube performed slightly better—Twitter even removed all the inauthentic likes from one manipulated tweet within 96 hours. Facebook, Instagram, and TikTok performed worse, while VKontakte performed in line with previous years. Overall, 93 per cent

	2020	2021	2022
Facebook	97%	99%	99%
Instagram	92%	96%	100%
TikTok	100%	85%	97%
Twitter	74%	83%	82%
Vkontakte		100%	100%
YouTube	97%	92%	90%

Table 1. Inauthentic activity remaining on the platforms after four weeks (%)

of the inauthentic engagement remained active across all social media platforms four weeks after purchasing. Social media manipulation services thus continue to outperform social media platforms, and have managed to find ways to prevent social media platforms from removing the majority of manipulation delivered.

4. Cost of Services

The cost of manipulation indicator captures how effectively social media platforms combat manipulation. When social media platforms remove accounts used to perform manipulation, it creates costs for manipulation service providers to replace accounts or to update their scripts, and these costs must ultimately be passed on to consumers. Therefore, rising manipulation costs are a strong indicator that social media platforms are effectively combating manipulation.

We track the price of a basket of social media manipulation consisting of 100 likes, 100 comments, 100 followers, and 1000 views from six Russian manipulation service providers. Over the years, the prices have stayed roughly the same and we have not observed any major changes to the price model.

This year the price of Facebook manipulation has increased, with Facebook emerging as the most expensive platform to manipulate (Figure 3). While Facebook and Twitter manipulation is slowly getting more expensive, YouTube manipulation is becoming cheaper. The price today is roughly one third of the price in 2018. Instagram and VKontakte remain the cheapest platforms to manipulate. In sum, social media manipulation continues to be cheap and readily available (Figure 4).

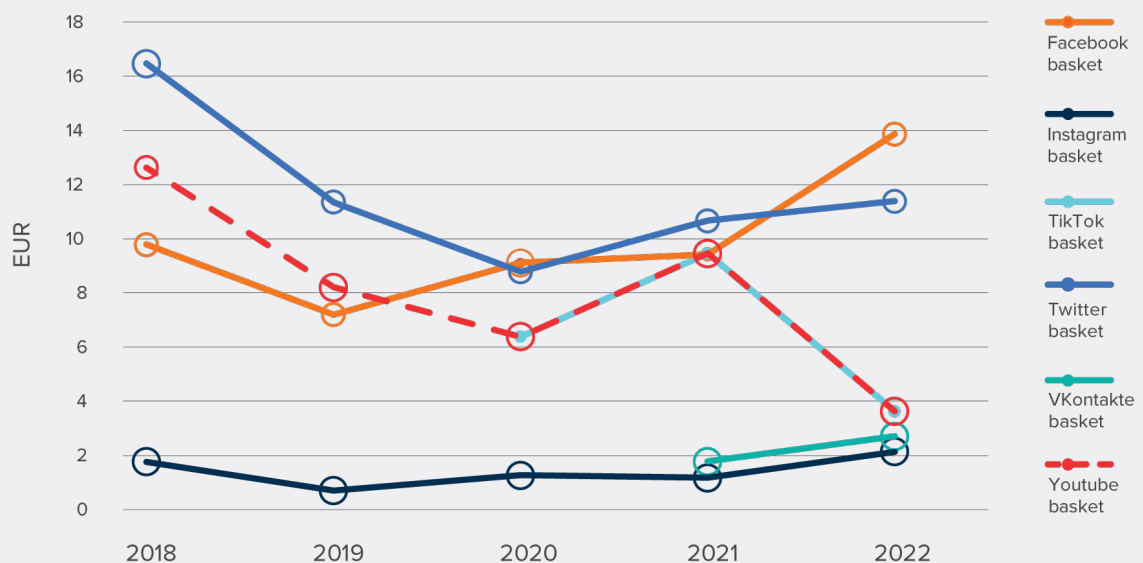


Figure 3. Price of a basket of social media manipulation

5. Speed and Availability of Manipulation

Slowing down the ability of manipulation providers to deliver fake engagement will reduce the impact and harm of the manipulation. In a fast-paced digital environment, speed of manipulation is essential for any manipulator hoping to impact current events and discussions.

In this reporting period, 89 per cent of all manipulation across all platforms was delivered within 24 hours, an increase in pace compared to 60 per cent in 2020. Over the last two years, manipulation service providers increasingly over-delivered in an attempt to compensate for platforms' countermeasures.

This year Facebook and TikTok performed significantly worse in relation to last year, with manipulation services being able to deliver more than 100 per cent of the ordered volume within the first 12 hours (Figure 5). Twitter was the only

platform that managed to significantly slow deliveries, with less than 50 per cent delivered during the initial 12 hours.

6. Responsiveness

Social media should be able to act on user reporting of inauthentic accounts, and accurately identify and remove accounts violating their terms of service. To assess the responsiveness of the social media platforms to user reporting, we reported 150 random accounts from each platform identified as used for social media manipulation. We then monitored how many accounts the platforms removed within five days (Table 2).

This year we observe a reduction in the number of reported accounts removed. Instagram, YouTube, and Twitter removed none of the reported accounts. Facebook—the best performing platform—removed only 3 per cent of

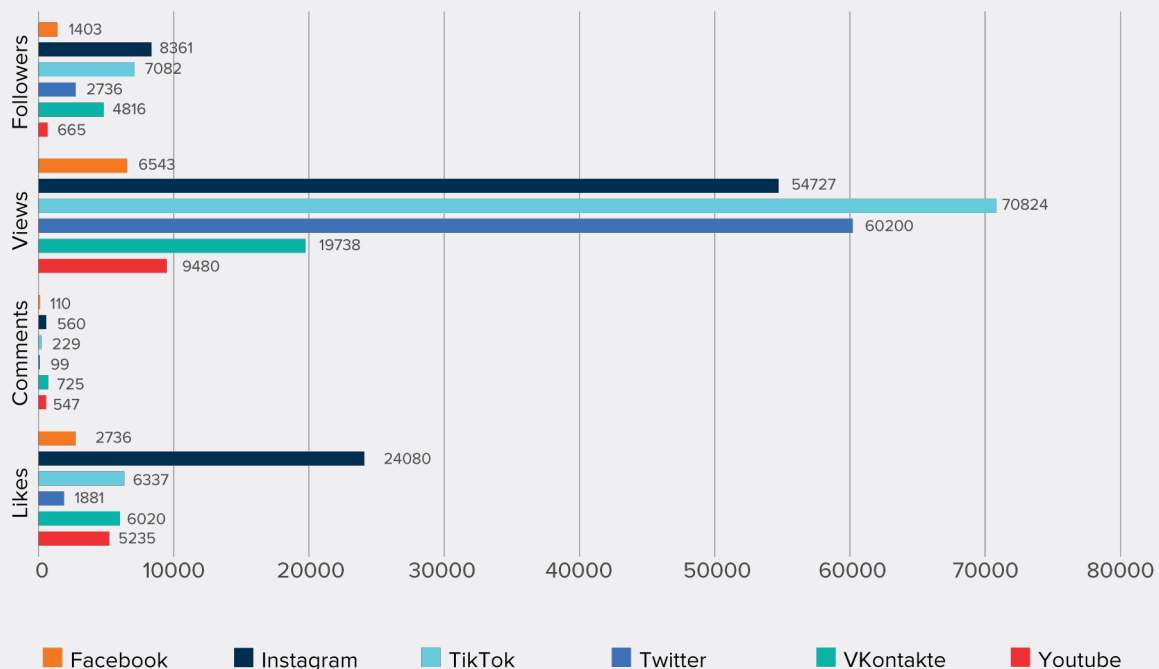


Figure 4. The amount of manipulation that can be bought for €10

	2019*	2020	2021	2022
Facebook	12%	9%	10%	3%
Instagram	3%	1%	1%	0%
YouTube	0%	0%	0%	0%
TikTok		0%	4%	1%
Twitter	3%	7%	2%	0%
Vkontakte			0%	1%

* Checked 21 days after reporting.

Table 2. Share of accounts removed five days after reporting (%)

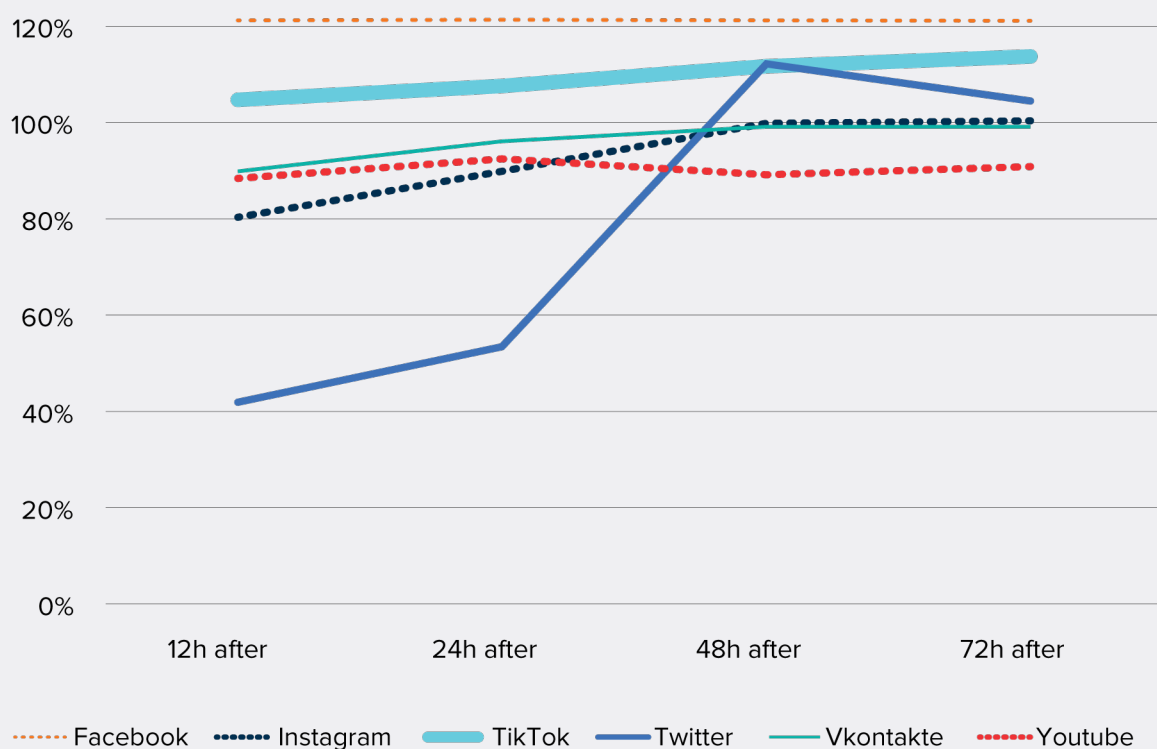


Figure 5. Comparison of delivery speed of fake engagements following purchase

the reported accounts. It is clear that social media platforms are unable or unwilling to respond rapidly to individual user reports of inauthentic accounts.

From conversations with the platforms we understand they tend to wait for the number of notifications of bad behaviour to reach a threshold. For this reason, takedowns can be slow. To check whether platforms perform

better over longer time horizons, we revisited the inauthentic accounts reported as part of last year’s experiment. The vast majority of accounts remained online (Figure 6). Clearly, social media companies struggle to identify and remove accounts engaged in inauthentic behaviour on the platforms.

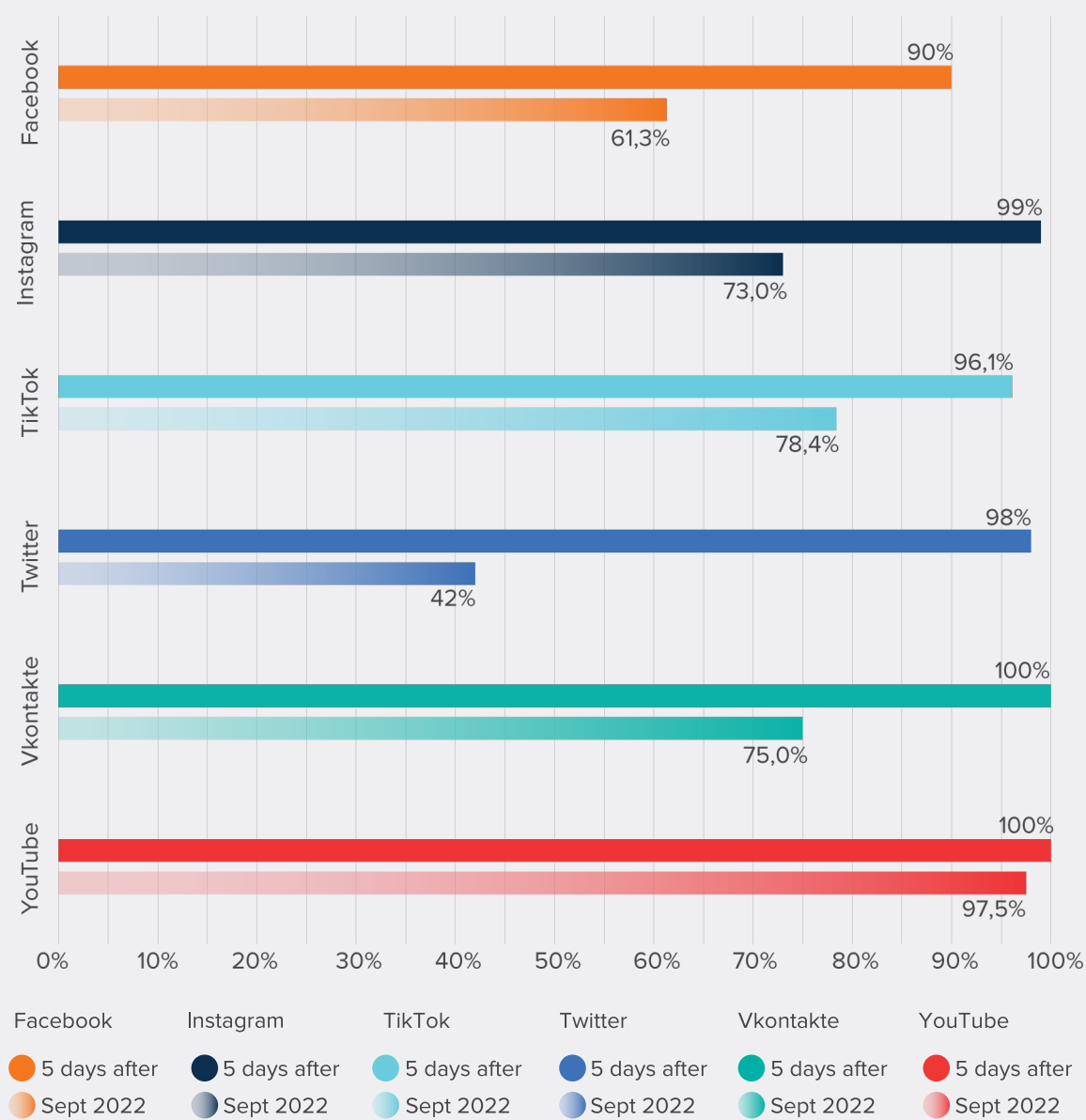


Figure 6. Share of accounts reported in 2021 remaining online

7. Transparency of Actions

TikTok¹, Twitter², Facebook³, Google⁴, and VKontakte⁵ all have pages dedicated to transparency reporting, but the nature of the reporting varies significantly. During the past year we did not observe any significant improvement in transparency reporting related to inauthentic activity. Notably Twitter has not updated its limited data on the topic since December 2021, and Meta continues its practice of only providing transparency data for Facebook, while failing to do the same for Instagram. YouTube developed a new metric during the year and now reports the number of videos removed for violating its misinformation policies, but continues to provide little information on inauthentic engagement. TikTok provides industry-leading transparency relating to inauthentic activity, allowing us to assess the scale of the problem. For comparison TikTok removed 33 million fake accounts and prevented or removed more than 26 billion fake engagements in the second quarter of 2022.

For our 2023 assessment of the platforms' transparency reporting related to inauthentic activity, Facebook and Instagram lost points for failing to provide updated statistics. The lack of continuous improvement in the field is worrying, and we continue to advocate for greater transparency, more data, and qualitative assessments by the social media companies. As last year, we shared the results of this report together with data about the experiment and a number of further questions with the social media platforms. Only TikTok and YouTube responded. This confirms that transparency remains an issue, and illustrates the challenge of poor platform responsiveness routinely faced by researchers.

These companies should release the number of fake accounts and engagements (likes, views, followers, shares, and comments) prevented and removed; data related to the effectiveness of their own ability to identify and remove fake engagement (time and impact); an assessment of the scale and impact of the problem; and regular updates on the efforts undertaken to counter inauthentic activity.

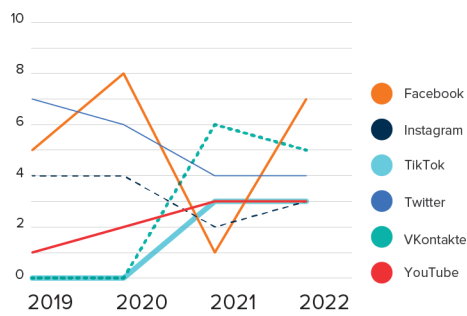
1 TikTok, '[TikTok Transparency Center](#)' [Accessed 1 November 2022].

2 Twitter, '[Twitter Transparency](#)' [Accessed 1 November 2022].

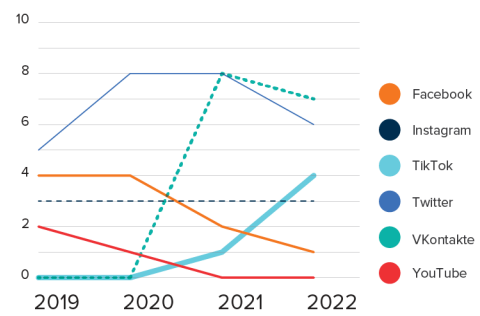
3 Facebook, '[Meta Transparency Center](#)' [Accessed 1 November 2022].

4 Google, '[Google Transparency Report](#)' [Accessed 1 November 2022].

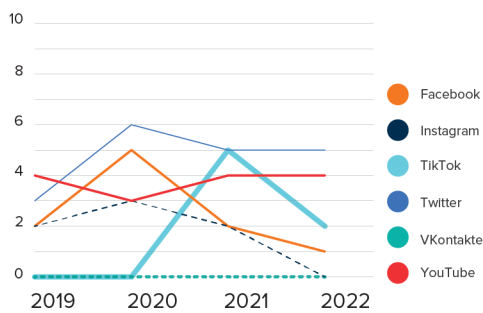
5 VKontakte, '[VK Safety](#)' [Accessed 1 November 2022].



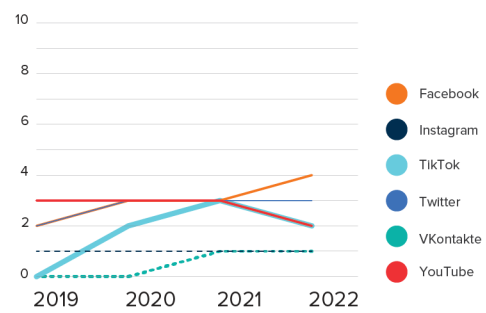
Blocking account creation



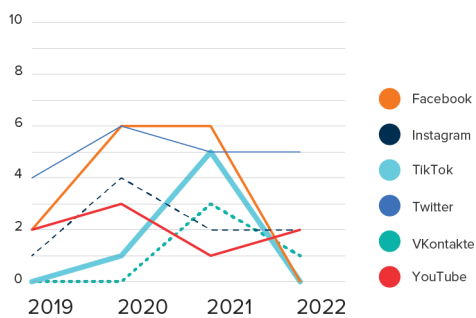
Removing accounts



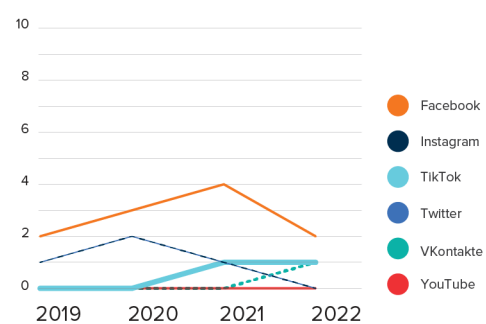
Removing activity



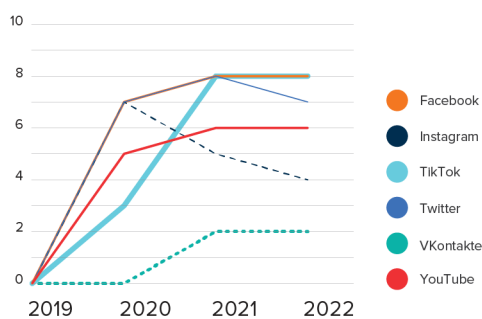
Manipulation costs



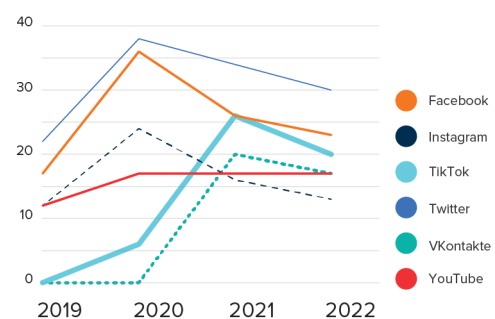
Speed of delivery



Responsiveness



Transparency of efforts



Summary

Figure 7. Overview of assessment criteria (scores by platform, 2019-2022)

A billion-dollar industry?

The available data does not allow us to make accurate assessments of the economics of social media manipulation. But using the data available we can extrapolate that the 1.4 billion fake followers removed by TikTok in the second quarter of 2022 have a street value of at least €1.9 million. The total sale of fake TikTok followers during the same period is probably much larger, as our experiment shows only 5 per cent







of all bought fake engagement was identified and removed within a four-week period. Given that fake TikTok followers are just a small part of the available manipulation services across all social media platforms, it is possible that the global social media manipulation industry is now a billion-dollar industry. The lack of data, however, makes it impossible to assess the exact scale of the problem.

Who Else Used Our Bots and What For?

In this section we present an overview of other content the accounts used to deliver manipulation services for us interacted with. In some cases, it was hard to get access to sufficient data to identify the accounts; for instance on YouTube and TikTok we could only see accounts used to place comments. In some cases when we could identify an account as being responsible for specific inauthentic efforts, we were unable to see the other activities of that account.

For this iteration of the report we were able to confidently identify and track the activity of more than 4500 accounts used by the manipulation providers to boost our posts:



	Facebook: 1 001
	Twitter: 1 385
	YouTube: 183
	Instagram: 1 132
	TikTok: 190
	VKontakte: 634

Some themes emerged as common across all the platforms. In every case—but especially on Instagram—we identified that accounts had been used to boost the visibility of online influencers and celebrities. Cryptocurrency projects and scams constituted the single largest category of harmful or even illegal content promoted by the accounts. On Instagram, some of the accounts had disseminated pornographic content; others advertised social media manipulation services. On Instagram, VKontakte, and TikTok, accounts promoted resellers of (possibly counterfeit) luxury goods and food supplements.

On all the platforms except Instagram we observed that the bot accounts had also been used to spread political influence. The nature of the content varied considerably from platform to platform. VKontakte had the highest proportion of political content, generally pro-Kremlin and pro-war, and targeting a Russian domestic audience. On TikTok, the majority of the amplified political content was posted by Russian influencers reciting official Russian government talking points. This included a number of clips glorifying the Russian military and including the Z symbol (or Zwastika). On YouTube, the accounts systematically boosted a pro-Kremlin channel posting about current affairs, history, and the war.

On VKontakte, we identified a lot of material promoted by the tracked accounts relating to various Russian politicians at different regional

levels. This included candidates of the Liberal Democratic Party of Russia (LDPR), the ruling United Russia party, and the Communists (KPRF). Typically, the artificially boosted content promoted or discredited regional political candidates, or disseminated calls to vote in the gubernatorial elections. Unlike in previous reports, the accounts did not promote Russian political parties on the Western platforms.

Political material on Facebook and Twitter tended to relate to the rest of the world. On Facebook, the majority of political material promoted a range of Azerbaijani official institutions and

government ministries. On Twitter, a lot of the content concerned the Colombian elections in June 2022. A large number of fake accounts interacted with and amplified material relating to US politics. This included material both promoting (the majority) and criticising President Donald Trump.

During our observation period, parliamentary elections took place in Italy, Latvia, and Sweden. However, no content about those campaigns was amplified by the accounts that had delivered inauthentic engagement for us.

We identified a long list of international actors apparently benefiting from publicity from the same manipulation providers:



- two candidates in the 2022 Colombian presidential elections (Twitter)
- one of the candidates campaigning to become leader of the UK Conservative Party (Twitter)
- a Kazakh opposition politician (Twitter)
- an Indian actor and politician (Twitter)
- a local-level Argentinian politician (Twitter)
- a Nigerian pastor running live prayer sessions (Facebook)
- accounts belonging to five separate government ministries in the Republic of Azerbaijan (Facebook)
- a permanent make-up instructor (Instagram)
- a UK-based cosmetologist (Instagram)
- two self-styled psychologists and motivational speakers (Instagram)
- a Russian 'bio-hacker' (TikTok)
- a tarologist (TikTok)
- two singers/musicians (TikTok)
- two Turkish pop stars (YouTube).

These examples provide insight into the diverse recipients of (possibly unwelcome) inauthentic increase in their visibility online.

Our conclusion from previous reports stands: manipulation services are still being used primarily for commercial purposes, but political actors are making regular forays into manipulating public discourse.

The Impact of War

Russia's full-scale invasion of Ukraine has left its imprint also on the social media manipulation market. This year the fake accounts we identified were also used to push pro-Kremlin narratives like 'US biolabs in Ukraine', 'US and NATO invaded Afghanistan, Libya, Iraq, Vietnam, etc.', and denial of the atrocities committed by the Russian army in Izium. At the same time, manipulation accounts also posted positive comments under a tweet by President Zelenskyy calling for arms to Ukraine. On Facebook, bots heavily pushed the pro-Azerbaijani agenda in regard to the Nagorno-Karabakh conflict and some internal Azerbaijani politics. On TikTok, we observed the use of fake engagement to amplify Russian bloggers' posts promoting the Kremlin agenda, embracing patriotism, and mobilising support for the war against Ukraine.

The sanctions imposed on the Russian financial system in response to Russia's full-scale invasion of Ukraine also affected manipulation providers unable to process transactions from outside the Russian Federation. However, we observed that the industry rapidly adapted, and today it is easy to circumvent the sanctions to buy Russian manipulation with Western payment solutions.

All manipulation providers use payment gateways that aggregate different payment methods for the users' convenience: Visa/Mastercard, Crypto, Apple Pay, and virtual money wallets. We successfully verified it is possible to buy services using the most user-friendly payment option, Apple Pay. Not only was the transaction immediate, it was denominated in Russian roubles. In another test, we paid using a Revolut Visa card. In this case, the transaction was processed in Kazakhstani tenge. In a final test, we were able to make a payment using a financial services company licensed in Estonia.

What Data Would We Need?

Each platform has its peculiarities in obtaining information about the activity of bots. For example, Twitter, YouTube, and VKontakte, which have official APIs, allow comprehensive data analysis, while for Facebook, Instagram, and TikTok we can only use manual analysis (Table 3).

The most open platform is Twitter, which enables the investigation of the following types of manipulation performed by bots: comments, follows, likes, and retweets. By contrast, on YouTube, researchers can see only videos published by bot accounts (rarely a method of manipulation) and information about bot subscriptions. This situation may be rectified through [YouTube's new Researcher Program](#). On Instagram, the same information is available, but only manually through the web interface.

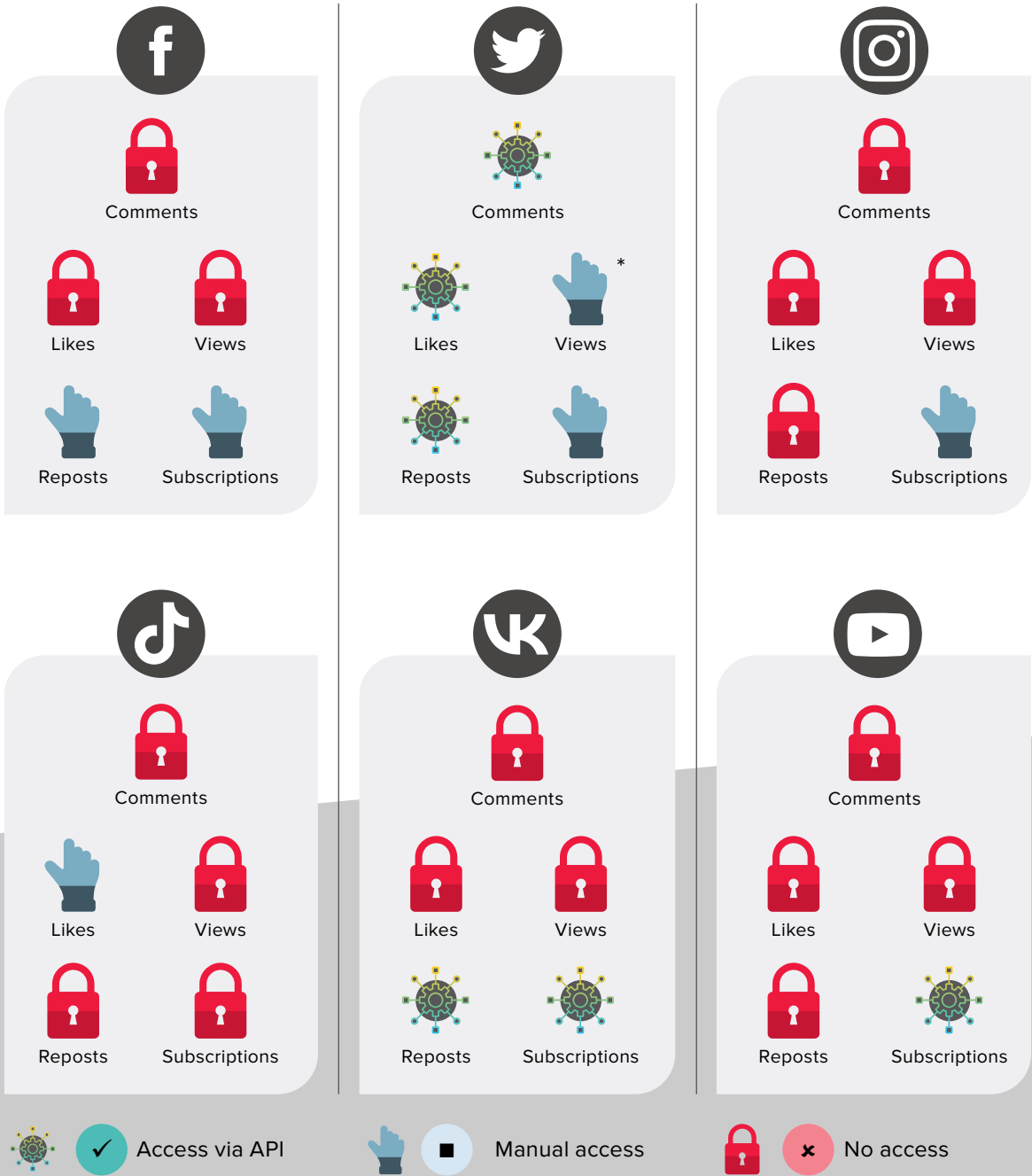
VKontakte allows automatic access to lists of reposted posts, friends, and pages followed. The same information can be found on Facebook: however, unlike VKontakte, this social network doesn't allow comprehensive access through the API. Finally, TikTok allows us to see posts shared by bots (usually not a form of manipulation) and posts that bots have liked.

These limitations on data access to bot activity render high-quality transparency reporting all the more important. As noted, TikTok's industry-leading transparency relating to inauthentic activity allows us to assess the scale of the problem. Still, overall the current transparency reporting by platforms is insufficient. Adequate reporting should include data on all the categories we assess in this study, but it should also include qualitative assessments of the risks and effects of inauthentic engagements.

Platforms are increasingly moving to measure the impact of fake activity through the metric of impressions or views to decide whether or not to take action. In most cases, this metric is inaccessible to outside researchers. While all

platforms publish the view count for videos, only Twitter and VKontakte shows how many times a post has been viewed. We call on the platforms to make this metric accessible for researchers,

and the regulator to request this data for the purpose of assessing the effectiveness of mitigation efforts.



* change was introduced in December 2022, until then data was inaccessible

Table 3. Accessible data on user activity

The Role of the Regulator

Our experiment is especially relevant in light of the EU's Digital Services Act (Regulation 2022/2065, DSA), which mandates more transparency from the large social media companies. We assess that existing mitigation measures are inadequate. While in absolute terms the number of removals and actions taken against manipulators are impressive, our experiments show that the majority of manipulation still makes it onto the platform. It shows that the cost of manipulation is getting cheaper, and that platforms are unable to prevent the manipulators from running commercially viable businesses.

The DSA requires very large online platforms to perform risk assessments, analysing whether and how the risks are influenced by intentional manipulation of the service. This includes through inauthentic use or automated exploitation, amplification and rapid dissemination of illegal content, or information violating their terms and conditions.

The DSA not only requires platforms to mark content that falsely appears to a person to be authentic or truthful, and to enable users to report such content. It also requires effective mitigation measures, including adapting the speed

and quality of processing notices of illegal content and the expeditious removal of notified content (Article 35).

Signatories to the Code of Practice on Disinformation commit to bolster their policies against impermissible manipulative behaviour to contain: the creation and use of fake accounts; account takeovers and bot-driven amplification; hack-and-leak operations; impersonation; malicious deep fakes; the purchase of fake engagements; non-transparent paid messages or promotion by influencers; the creation and use of accounts that participate in coordinated inauthentic behaviour; user conduct aimed at artificially amplifying the reach or perceived public support for disinformation (CoP, Commitment 14).

Our research shows that commercial accounts are exploiting flaws in platforms, and pose a structural threat to the integrity of platforms. We have seen examples of these accounts being active in countries on at least five continents, spreading partisan content and interfering in elections. Platform risk assessments will need to reflect this threat.

Conclusions

While we were expecting that social media platforms would take action and improve their performance, we have found that platforms' overall efforts to counter commercial manipulation have stagnated. Only limited improvements in some criteria have been noted. No platform has improved in more than two criteria. The overall ability of the platforms to prevent manipulation has decreased. This shows that the platforms have not built on the significant improvements noted in recent reports. In particular with regard to platform transparency efforts, we observe that progress has stalled.

Comparing the different platforms, our analysts assessed that Twitter was the hardest to manipulate, and Instagram the easiest. In this year's report, Twitter thus emerges as the undisputed frontrunner as Facebook falls back. TikTok's positive trend has reversed; it performed significantly worse in this reporting period. YouTube scores similar to previous years, with no evidence that the measures the platform has introduced have been effective. VKontakte excels at blocking and removing accounts, but in all other respects its anti-manipulation efforts are dire. Instagram has shown no improvements. A new finding of this year's research is that Russian providers don't appear to discriminate against customers amplifying pro-Ukrainian content.

Buying manipulation remains cheap. The price today is roughly one third of the price in 2018. Instagram and VKontakte remain the cheapest platforms to manipulate. Facebook is the industry leader when it comes to countering fake account creation, while Instagram and VKontakte are the worst performers.

This year, the percentage of accounts identified and removed by the platforms dropped to 16 per cent overall. A further one per cent of accounts were removed five days after we reported them. The only social network that improved in account removal rate compared to the previous year is TikTok. Overall, 93 per cent of the inauthentic engagement remained active across all social media platforms four weeks after purchasing. Thus, the platforms' moderation decisions appear to be minimally responsive to user notifications. Social media manipulation services continue to outperform social media platforms and have found ways to prevent the platforms from removing the majority of manipulation delivered. Over time, the platforms' ability to combat manipulation by slowing the speed of delivery has declined as well. Today delivery is near instantaneous, with 89 per cent delivered within one day.

The quality of transparency reporting is unchanged. We have noted a widening gap between platform performance and the quality of platform reporting. Platforms have found it expedient to allow commercial manipulators to access the platform, focusing instead on reducing their reach and impact. More data is required to assess whether the platforms' approach adequately mitigates the systemic risk posed by platform manipulators. It remains to be seen if the new regulatory measures of the Digital Services Act will help to reverse the trend.



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.