

## SOCIAL MEDIA POLICY AGAINST RUSSIAN PROPAGANDA IN UKRAINE

Yuskiv Kh. V.

**Problem statement and its relation to important scientific or practical tasks.** Before 2014, pro-Russian accounts have been spreading false information regarding Russia's role in the Donetsk area, fueling uncertainty and destabilization and aiding Russia's takeover. Russia's "hybrid warfare" strategy included this as a key component.

As early as 2014, General Philip Breedlove, NATO's Supreme Allied Commander Europe, described the Russian communication strategy in Ukraine as "the most amazing information warfare blitzkrieg we have ever seen in the history of information warfare"<sup>1</sup>. Destructive communications of the Russian Federation in the information space of Ukraine, caused by the illegal aggression of the Russian Federation and the annexation of Crimea, led to violations of the information space of the state, the use of hybrid information campaigns, the spread of hostility, and the transformation of information policy in the context of countering modern challenges and threats. The processes that take place in society affect language, which is also a weapon in hybrid warfare. Ukraine is constantly the object of information aggression, propaganda (even non-military but negative socio-economic and socio-political) and pressure in all spheres of public relations from Russia. Since the Russian attack began on February 24, the world has rallied in large numbers in support of Ukraine, and the vast majority of the international community has also condemned Russian aggression. In fact, this war is commonly regarded as the most black-and-white conflict in contemporary history.

**Analysis of recent researches and publications, which have initiated problem solution, the author relies on.** International legal aspects of propaganda of war and information warfare have been analyzed by domestic and foreign researchers, in particular Ya. Malik, O. Frolova, S. Lyulko, M. Zinchenko, M. Kearney, G. Michael, D. Skrynka. Among the domestic scientists studying this direction, it is worth mentioning D. Dubov, M. Ozhevan, S. Parkhomenko.

**Identification of previously unresolved matters of the generic problem the article deals with.** At the same time, there is a need to clarify the issue of countering the Kremlin's propaganda at the level of global IT giants.

**Goal statement** – an analysis of changes in the policy of social networks after the military invasion of the Russian Federation on February 24, 2022, in particular the opposition of IT companies to Russian propaganda against Ukraine.

**Presentation of research material with full justification of findings.** With regard to Ukraine and Russian propaganda, it should be noted that this is not only about propaganda, but also about such a phenomenon as the "war of sense/meanings". A wide range of possible information communications methods are used to spread such meanings; in particular, one of the main elements is simulacra – a copy of something that does not really exist<sup>2</sup>. Propaganda in the context of this study should be understood as the whole set of information messages from Russia, Russian media, Russian opinion leaders, Russian diplomatic missions in the world, etc., which have a negative context for Ukraine. The information war is perhaps the main instrument of Russian aggression against Ukraine, at this stage in this war, the main weapon is propaganda, in particular the use of "hate speech" and misinformation on social networks<sup>3</sup>.

W. Curtis, a political scientist at the University of Portland (USA), emphasizes that "hate speech" is about spreading harmful stereotypes and is aimed at inciting hatred and violence between certain groups and communities<sup>4</sup>. As T. Pechonchik notes, the definitions of "hate speech" are based on establishing the fact of incitement to hatred, humiliation, or discrimination on certain grounds in statements, with a list of these signs<sup>5</sup>.

In our study, we use the definition of "hate speech" proposed by the Committee of Ministers of the Council of Europe: "the term «hate speech» shall be understood as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant

<sup>1</sup> Pomerantsev P. Russia and the menace of unreality. URL: <https://www.theatlantic.com/international/archive/2014/09/russia-putin-revolutionizing-information-warfare/379880/>.

<sup>2</sup> Зозуля О. Фейк як інструмент інформаційної війни. URL: <https://yur-gazeta.com/publications/practice/inshe/feyk-yak-instrument-informaciynoi-viyni.html>.

<sup>3</sup> Ісакова Т. Мова ворожнечі як проблема українського інформаційного простору. Стратегічні пріоритети. 2016. № 4 (41). С. 90–97. URL: <http://ippi.org.ua/sites/default/files/isakova.pdf>.

<sup>4</sup> Curtis W. M. Hate speech. URL: <https://www.britannica.com/topic/hate-speech>.

<sup>5</sup> Печончик Т. Чому мова ворожнечі з'являється у ЗМІ. URL: <http://www.cje.org.ua/ua/blog/chomu-mova-vorozhnechi-zyavlyayetsya-u-zmi>.

origin”<sup>6</sup>. Content that has an extremely negative assessment of the addressee’s actions and even threats does not belong to “hate speech” if it concerns a specific person and does not identify him with a specific social group<sup>7</sup>.

Researchers say new media today has “become the world’s biggest hate speech mouthpiece”<sup>8</sup>. Social networks’ administration is aware of its position in modern information and social processes, thus they are becoming increasingly involved in the fight to ensure that their platforms are a “clean and healthy space”.

Companies such as Meta, Twitter, and Google have all released statements, that have one thing in common: emphasizing the necessity of countering disinformation. Companies have frequently reported suspending accounts or erasing botnet networks in recent years, notably those linked to Russia.

**Meta.** Facebook launched a one-click option for Ukrainians to block their accounts to outsiders almost soon after the full-scale Russian invasion of Ukraine began<sup>9</sup>. Russia’s war against Ukraine has forced social media platforms to take unprecedented steps.

Despite the established standards of content filtering, Meta surprised the world on March 10, 2022 with an uncommon decision for such a company: temporarily allow Facebook and Instagram users in some countries urging violence, namely to call for the deaths of Russian President Vladimir Putin, Belarusian dictator Alexander Lukashenko, for violence against the Russian military in the context of war against Ukraine. A statement from a Meta spokesman said – “As a result of the Russian invasion of Ukraine, we have temporarily made allowances for forms of political expression that would normally violate our rules like violent speech such as ‘death to the Russian invaders’. We still won’t allow credible calls for violence against Russian civilians”. Calls for the death of leaders will be allowed if they do not contain other goals or do not have specific instructions such as the place or method of death<sup>10</sup>. The decision was made a few days after Russian troops dropped powerful aerial bombs on a children’s hospital in Mariupol.

“Hate speech” is not prohibited, according to the moderators’ letter, whether it (a) targets Russian soldiers other than prisoners of war, or (b) attacks Russians in the context of Russia’s invasion of Ukraine (for example,

the content mentions an invasion, self-defence, etc.). Only in the context of protecting Ukraine from Russian invasion is it also temporarily permitted to glorify the far-right Azov Battalion, which was previously banned<sup>11</sup> (the social media platform Facebook has classified Azov as a dangerous organization<sup>12</sup>).

It can be assumed that this was done to avoid blocking messages from Ukrainians and citizens of other nations who were critical of the Russian and Belarusian governments as a result of the Russian war against Ukraine. Despite the fact that Meta swiftly reversed its decision, banning the ability to call for Putin and Lukashenko’s deaths on its platforms on March 14<sup>13</sup>, – Russia labeled the American corporation as a “extremist group” and banned Facebook and Instagram on its territory<sup>14</sup>.

Since February 24, Meta has made 12 changes to its content policy. This has resulted in internal confusion, especially among content moderators who are searching posts depicting blood, hostile statements, and incitement to violence. Sometimes the platform changed the rules every day. According to The New York Times, this was similar to the obsession with moderating content related to Russia’s war against Ukraine<sup>15</sup>. Facebook has often been criticized for allowing hostile statements and calls for violence on its platform, and has been described as “a place where calls for violence thrive”. According to a study by the Associated Press, Facebook has been used to fuel tensions and offline violence in various conflicts such as Myanmar, Syria, and Ethiopia<sup>16</sup>.

In order to counter and eradicate “hate speech”, Facebook has introduced its own algorithms for combating and defending. For example, the network uses a reactive model (a response to a message or complaint from a user); emphasizes responsible journalism and moderation, deletes information that violates international and national standards (but does not delete

<sup>6</sup> Ісакова Т. Мова ворожнечі як проблема...

<sup>7</sup> Загальнополітична рекомендація Європейської комісії проти расизму та нетерпимості (ЄКРН) № 15: протидія мові ворожнечі. URL: <https://rm.coe.int/ecri-general-policy-recommendation-no-15-on-hatespeech-ukrainian-tran/1680a11674>.

<sup>8</sup> Ісакова Т. Мова ворожнечі як проблема...

<sup>9</sup> Facebook дозволив українцям закривати акаунти від незнайомих. URL: <https://ms.detector.media/sotsmerezhi/post/29033/2022-02-25-facebook-dozvoliv-ukraintsyam-zakryvaty-akaunty-vid-neznayomtsiv/>.

<sup>10</sup> Facebook allows war posts urging violence against Russian invaders. URL: <https://www.reuters.com/world/europe/exclusive-facebook-instagram-temporarily-allow-calls-violence-against-russians-2022-03-10/>.

<sup>11</sup> Facebook та Instagram не цензуруватимуть заклики до смерті російським окупантам. URL: <https://hmarochos.kiev.ua/2022/03/11/facebook-ta-instagram-ne-cenzuruvatymut-zaklyky-do-smerti-rosijskym-okupantam/>.

<sup>12</sup> «Азов» належить до переліку небезпечних організацій, визначеного соціальною мережею Facebook. URL: <https://ms.detector.media/sotsmerezhi/post/28294/2021-10-13-15-ukrainskykh-organizatsiy-ie-u-zaboronnomu-spysku-facebook-zdebilshogo-ultrapravi/>.

<sup>13</sup> Meta narrows guidance to prohibit calls for death of a head of state. URL: <https://www.reuters.com/technology/meta-narrows-guidance-restrict-calls-death-head-state-2022-03-14/>.

<sup>14</sup> Russia bans Instagram and Facebook as court declares Meta an 'extremist organisation'. URL: <https://www.euronews.com/next/2022/03/21/ukraine-war-facebook-temporarily-allows-posts-calling-for-violence-against-russians-or-put>.

<sup>15</sup> How War in Ukraine Roiled Facebook and Instagram. URL: <https://www.nytimes.com/2022/03/30/technology/ukraine-russia-facebook-instagram.html>.

<sup>16</sup> How Meta refused to act on anti-Russia propaganda. URL: [https://www.youtube.com/watch?v=6biHFiTdiXU&ab\\_channel=WION](https://www.youtube.com/watch?v=6biHFiTdiXU&ab_channel=WION).

profiles of users who systematically violate the rules of use of the social network)<sup>17</sup>.

As for Instagram, during the war, users started writing about things that are marked online as “hate speech”. How does the ban mechanism (one of the accepted methods of monitoring the actions of users on the Internet) of content or an entire profile on the social networking sites Instagram and Facebook works: Instagram’s hate speech is a direct attack on people (not concepts or institutions) on the basis of so-called “protected characteristics”. Direct attacks include: harsh or derogatory remarks, harmful stereotypes, statements of inferiority, expressions of contempt, disgust or rejection, abusive expressions or calls for isolation or segregation; the use of harmful stereotypes, including degrading comparisons that have been used in the past to attack certain groups, intimidate or exclude them, often associated with real-world violence.

Instagram’s “protected characteristics” are: race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease; the age at which it is mentioned together with another protected feature; refugees, migrants, immigrants and asylum seekers; sphere of activity, when it is mentioned together with another characteristic with special protection; sometimes, based on the nuances of local culture, these may be certain words or phrases as code symbols for groups of people with special protection<sup>18</sup>.

Any of the above actions is the basis for the ban. That is, if the photo/video in the post and the stories, or the text of the post and the stories can be interpreted as cruel derogatory statements and (or) images on some of the “protected characteristics”, this post can be blocked, ie deleted, and profile can be pessimized on news feed, or even be blocked<sup>19</sup>.

Who makes the decision? There are two ways:

- automatically Instagram algorithm based on powerful AI<sup>20</sup>;

Automatically, AI views tons of content based on labels, text, the location, and hashtags used. If photos are recognized by artificial intelligence as “hate speech”, and a large number of similar and identical images occur at the same time, as was with the content under the hashtags #buchas and #Buchamassacre<sup>21</sup>, the system

can automatically hide the issuance of such content. On April 4, amid the horrors in Bucha, users began to complain en masse about Instagram for allegedly blocking truthful information. However, this happened automatically, and the Meta team quickly unlocked hashtags and access to information.

There is a reverse story when Meta deliberately blocks content, such as Russian propaganda media and opinion leaders. There is a blockade of content in the region – at the beginning of the war, before Roskomnadzor recognized Meta products as extremist organizations, Instagram blocked the issuance of content by Ukrainians to Russian users and vice versa.

manually, based on reports from users.

Here we are dealing with user complaints (reports). Each user has the right to complain about the post, story, comment, live broadcast and profile of anyone. There are several reasons for the report: spam, nudity or sexual activity, hate speech or symbols, violence or dangerous organisations, bullying or harassment, selling illegal or regulated goods, intellectual property violations, suicide or self-injury, eating disorders, scams or fraud, false information and category “other”<sup>22</sup>. Each category has a number of subcategories. Thus, there are many reasons to complain about content.

For example, at the moment when users start complaining about a particular post or profile en masse, manual content moderation is no longer automatic, but manual. Specialists from the support service (or the relevant department on Instagram) check the content for a complaint and, if everything is correct, ban a post or even a profile. This happened to a number of profiles of millionaire bloggers in both Ukraine and Russia. Even more, they were not just blocked from their profiles, their pages were deleted. This means that these profiles cannot be restored.

Often, such actions are coordinated and take place as a flash mob, which, incidentally, is a valuable and effective tool on the information front. It should be noted that such actions can be used against Ukraine too. Therefore, it is important to ensure that your content does not fall under the potential mark of hate speech and does not violate the current rules of conduct online.

Since 2020, Meta has been marking the Facebook pages and Instagram accounts of Russian state media. After the start of a new phase of the war in Ukraine, Meta continued to label Russian media content as well as lower the priority of its issuance and make it more difficult to find these resources on these social networks around the world. Posts containing references to Russian-controlled media and media outlets are subject to the same restrictions.

Simultaneously, at the end of February 2022, the state media were banned from advertising or monetizing posts on the company’s platforms around the world.

<sup>17</sup> Європейська практика протидії «мови ворожнечі». *Ізбірком*. 31 липня 2019. URL: <https://izbirkom.org.ua/publications/medialiteracy/2019/ievropeiska-praktika-protidiyi-movi-vorozhnechi/>.

<sup>18</sup> Мова ворожнечі. Детальна інформація про правила. URL: <https://transparency.fb.com/uk-ua/policies/community-standards/hate-speech/>.

<sup>19</sup> An update on our work to tackle abuse on Instagram. URL: <https://about.instagram.com/blog/announcements/an-update-on-our-work-to-tackle-abuse-on-instagram>.

<sup>20</sup> Shedding more light on how Instagram works. URL: <https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works>.

<sup>21</sup> Facebook owner Meta briefly blocks hashtags tied to Bucha killings. URL: <https://www.reuters.com/technology/facebook-owner-meta-briefly-blocks-hashtags-tied-bucha-killings-2022-04-05/>.

<sup>22</sup> How do I report a post or profile on Instagram? URL: <https://help.instagram.com/192435014247952>.

This was announced by the head of the cyber security department of Facebook Nathaniel Gleicher. Meta does not specify the list of banned Russian media outlets. At the same time, the company makes it clear that the list will be updated with new media outlets that are controlled by the Russian government<sup>23</sup>.

Banning or deleting Russian state media's accounts is done according to the requests of the governments of different countries, not on the company's own initiative. In particular, at the request of the Ukrainian government restricted access (but did not delete the account) to the Facebook page of blogger Andriy Shariy, to the Instagram accounts of Margarita Simonyan, head of the Russian RT propaganda channel, and Anton Krasovsky, director of the Russian-language RT service (Instagram accounts are currently up and running, but with restrictions), Russian propagandists, TV presenters and others. At the same time, the EU and the UK, limited access to Russia Today and Sputnik in response to government requests<sup>24</sup>.

The corporation conducts a separate segment for work in the field of combating propaganda and misinformation. It:

- works with independent specialists to combat the spread of false information on the platforms;
- has increased the number of independent fact-checkers in the region who operate in Russian and Ukrainian, and is seeking to secure more money for Ukrainian information verification partners;
- removes information that violates the platform's standards and collaborates with regional fact-checkers to debunk claims. The fake information is marked and the importance of its release is reduced. Alternatively, place the account on the "shadow banning" list;
- when users try to spread war-related photographs, they get warnings in addition to fact-check marks. Meta validation systems cover a period of more than a year, giving users extra information concerning out-of-context or false photos;
- Messenger, Instagram and WhatsApp limit message forwarding and label those that haven't originated with the sender. WhatsApp, by the way, continues to work in Russia<sup>25</sup>, although his parent company is recognized in Russia as "extremist";
- those who have previously distributed or tried to share unverified content are notified, allowing users to chose whether or not to continue doing so;
- additional penalties will be levied against Facebook pages, groups, accounts, and domains that re-distribute false information. For example, this pages, groups,

<sup>23</sup> We are now prohibiting Russian state media... Nathaniel Gleicher, *Twitter*. URL: <https://twitter.com/ngleicher/status/1497417241947607043>.

<sup>24</sup> Meta, Twitter і Google проти Росії. Як соцмережі протистоять пропаганді Кремля. URL: <https://ms.detector.media/sotsmereszhi/post/29241/2022-03-27-meta-twitter-i-google-proti-rosii-yak-sotsmereszhi-protystoyat-propagandi-kremlya/>.

<sup>25</sup> Russia bans Instagram and Facebook...

accounts are removed from the recommendations and their content is shown with lower priority;

- users who visit Facebook or Instagram accounts that have persistently published false information, will see notification about it<sup>26</sup>.

Meta also announced the formation of an association of experts and native speakers on February 25, 2022 — special operations center — to respond as soon as possible and eliminate "hate speech", to combat disinformation amid a military special operation in Donbas. Even if false reports are published in state media, the center's experts will detect and recognize these as "fakes".

Earlier, on February 24, Facebook imposed restrictions on the materials of the Russian TV channel "Zvezda", RIA Novosti agency, Lenta.ru, Gazeta.ru through the dissemination of false information. Media reports will be shown below in the news feed for 90 days.

Many technology companies have suspended activities in Russia. Russian Foreign Ministry spokeswoman Mariia Zakharova said on April 15, 2022 that Meta and other American digital corporations, are actively involved in the information war waged by the United States against Russia. She also mentioned the words of the company's vice president Nick Clegg, who said that the IT giant would step up efforts to counter Russian propaganda. According to Zakharova, propaganda means any opinion different from the Western mainstream<sup>27</sup>.

A report on cyber threats for the first quarter of 2022 was recently released by Meta<sup>28</sup>. On February 26-27, the Russian hacker group Ghostwriter attacked the Facebook accounts of dozens of Ukrainian military, according to the document. In early 2022, the authors of the report noticed an upsurge in misinformation on the social network, including content concerning Russia's war in Ukraine. False reports of an invasion of Ukraine are being spread by Kremlin-linked groups, who accuse Kiev of all wrongdoing.

Meta's efforts to battle disinformation and disseminate the war in Ukraine have been criticized for not doing enough to stop the spread of misinformation. The Meta report states this fact. A variety of disinformation campaigns relating to Russia's invasion of Ukraine are mentioned in the report, including dozens of fake accounts that spread anti-Ukrainian rhetoric. The paper also details the operations of a cyber network that, prior to the Russian invasion, submitted hundreds of false complaints against Ukrainian Facebook users in an attempt to force them off the platform<sup>29</sup>.

<sup>26</sup> Meta, Twitter і Google проти Росії...

<sup>27</sup> Захарова снова обвинила Facebook в информационной войне. URL: <https://news.ru/society/zaharova-zayavila-chto-facebook-otkryto-uchastvuet-v-informacionnoj-vojne/>.

<sup>28</sup> Adversarial Threat Report. URL: [https://about.fb.com/wp-content/uploads/2022/04/Meta-Quarterly-Adversarial-Threat-Report\\_Q1-2022.pdf](https://about.fb.com/wp-content/uploads/2022/04/Meta-Quarterly-Adversarial-Threat-Report_Q1-2022.pdf).

<sup>29</sup> Безпорадна мережа. Чому Meta не справляється з дезінформацією Кремля у Facebook та Instagram. URL: <https://www.epravda.com.ua/publications/2022/04/25/686192/>.

Meta stated that additional policies would be implemented in the next weeks and months. They will enable to detect groups on social media which intend to use the platforms for their own goals in advance. This, however, will not be simple. Communities that spread Russian propaganda and disinformation are adapting as well.

**YouTube.** YouTube, like other social media platforms, takes steps to combat misinformation on its platform and regulates the spread of content related to Russia's war in Ukraine. YouTube operates on a "proactive" approach, allowing administrators to remove (or restrict) access to content before it is distributed; uses artificial intelligence to recognize and detect "hate speech". The policies of Facebook and YouTube allow any network user to report hate speech: on Facebook it is an option to "report a problem" (including offensive content, "hate speech" (unfortunately, does not contain a Ukrainian version). YouTube includes a "Hate Speech Policy" section where users may learn more about how to report inappropriate content<sup>30</sup>.

Some of the company's decisions were announced publicly. On March 11, video hosting sites disabled monetization<sup>31</sup> for Russian residents. Furthermore, foreign bloggers with a Russian audience will be unable to profit from the Russian market. This could have a significant impact on the content of bloggers, particularly in Ukraine. Russian bloggers and media companies that have profited from integrated advertising on YouTube could lose up to \$100 million per month as a result of this decision alone<sup>32</sup>. On April 9, 2022, the platform blocked the YouTube channel of the Russian State Duma<sup>33</sup>. Because of numerous violations, a single case of severe abuse, or when they are dedicated to content that violates the rules, channels may be permanently closed, according to the service policy.

Unlike Facebook and Instagram, YouTube continues to work in Russia. Valeriy Fadeyev, the head of the Human Rights Council, admitted in an interview that YouTube is the most important propaganda platform for Russians to promote its own interpretation of the war, their own truth<sup>34</sup>. Fadeev's position, in our opinion, is the most likely reason why YouTube works in Russia and will most likely continue to work. The blocking of the largest video hosting should be considered in the case of the

service's principled stand and the subsequent blocking of some propaganda fakes about the war or entire video channels of the official Russian media or government.

YouTube Premium, Music Premium, sponsorship, super chat, super stickers and merch subscriptions will not be available to all viewers in Russia. For breaking anti-violence regulations, the video provider also declared an immediate global blockade of Russia's state media. Previously, the video platform blocked channels across Europe, including Russia Today and Sputnik. Russia's state media have called unjustified censorship of restrictions imposed on them by distributors, including app stores and other social media. YouTube has refused to reveal which and how many channels have been shut globally, or whether they will be restored at any point in the future<sup>35</sup>. In particular, YouTube will remove ads and content about Russia's war in Ukraine that violates video service policies. Google, for example, announced<sup>36</sup> the removal of hundreds of channels and thousands of videos relating to the Ukraine conflict on February 25.

YouTube, like other social platforms, has blocked Russian propaganda channels RT and Sputnik across Europe, and in Ukraine, Channel One, Russia 24 and Russia 1, TASS, RIA Novosti, RBC and Zvezda TV Channel. Video hosting companies, in addition to censoring YouTube media channels, also restrict the accounts of individual pro-Kremlin propagandists, which we believe is more effective than blocking individual videos. Google AdSense has also stopped monetizing content that denies the war in Ukraine, according to the corporation<sup>37</sup>.

**Twitter.** Meta's and Twitter's efforts are similar. The company also has a team that keeps an eye on the situation in Ukraine and detects perceived risks. The team was formed few weeks ago, according to Ronan Costello, public policy director for Europe, Turkey, and Israel.

In 2017, when the US began to record and talk openly on behalf of the Russian government about the impact on the US presidential election, Twitter initiated a campaign against Russian media propagandists. That year, the social media platform decided to stop promoting the content of all Russia Today and Sputnik accounts<sup>38</sup>. The decision was based on Twitter's own research and the findings of US intelligence.

<sup>30</sup> Європейська практика протидії...

<sup>31</sup> YouTube channel monetisation policies. URL: <https://support.google.com/youtube/answer/1311392?hl=en-GB>.

<sup>32</sup> Російські ютуб-канали втрачатимуть \$100 мільйонів щомісяця через блокування монетизації. URL: <https://ms.detector.media/sotsmerezhi/post/29181/2022-03-15-rosiyski-yutub-kanaly-vtrachatymut-100-milyoniv-shchomisyatsya-cherez-blokvannya-monetyzatsii/>.

<sup>33</sup> Google заблокував YouTube-канал «Дума ТВ». Государственная Дума. Телеграм-канал. URL: [https://t.me/duma\\_gov\\_ru/188](https://t.me/duma_gov_ru/188).

<sup>34</sup> В России рассказали, почему до сих пор не заблокировали YouTube. URL: [https://tech.24tv.ua/ru/rossii-rasskazali-pochemu-do-sih-por-ne-zablokirovali-youtube\\_n1951413](https://tech.24tv.ua/ru/rossii-rasskazali-pochemu-do-sih-por-ne-zablokirovali-youtube_n1951413).

<sup>35</sup> YouTube blocks Russian state-funded media channels globally. URL: <https://www.reuters.com/business/media-telecom/youtube-blocks-russian-state-funded-media-channels-globally-2022-03-11/>.

<sup>36</sup> Повідомлення від Google, 25 лютого 2022 р. URL: <https://ukraine.googleblog.com/2022/02/google.html?m=1>.

<sup>37</sup> AdSense & Google Ads breaking up with Russian media (by Ad Revenue). URL: <https://www.monetizemore.com/blog/google-adsense-disabled-russian-publishers-ukraine-invasion>.

<sup>38</sup> Sinéad McSweeney Our ongoing approach to the war in Ukraine. URL: [https://blog.twitter.com/en\\_us/topics/company/2022/our-ongoing-approach-to-the-war-in-ukraine](https://blog.twitter.com/en_us/topics/company/2022/our-ongoing-approach-to-the-war-in-ukraine).

In August 2020, the company took the next step, identifying and limiting the coverage of accounts controlled by the Russian government, as well as twenty other countries. In 2021, the social network has expanded the list of countries and accounts that it restricts and designates. It now has around 100 media accounts identified as being linked to Russian government. Sinéad McSweeney, Twitter's Vice President for Public Policy and Communications in Europe, Africa, and the Middle East, tweeted that the social media platform has begun tagging Belarusian state media accounts as well<sup>39</sup>. Twitter does not say who was included on this list; all they say is that it is being reviewed and updated, with newly created Russian accounts being included.

The labels were placed to accounts and tweets posting links to official media in Belarus after covering their role in the Ukraine war, according to Yoel Roth, head of the company's professional ethics section. He also mentioned that Twitter launched a label for Russian state media on February 28, 2022<sup>40</sup>. It's worth noting that almost 45,000 tweets per day containing connections to Russian official media were recorded by Twitter in just four days, but as of March 11, the number of shows has dropped<sup>41</sup>.

Twitter is implementing a feature that allows users to re-read messages before posting them, with the goal of preventing "hate speech" and misinformation from spreading. This function was tested on Android users in June 2020, and the results revealed that "40 percent of readers opened articles more often than before" after it was introduced<sup>42</sup>.

Twitter, on the other hand, abides by EU sanctions and restricts access to some information in EU member states. Similar work is done by the company outside of the European Union. The early results of the administration's efforts have already been reported: as of March 16, more than 50,000 tweets containing misinformation concerning Russia's war against Ukraine had been removed or labelled as false<sup>43</sup>. Over 75,000 profiles have been deleted due to violations of the platform's spam and manipulation policies. The deleting of accounts has noteworthy consequences: since the large cleanup of bot accounts on Twitter, the number of pro-Russian politicians' subscribers has decreased considerably.

<sup>39</sup> We've begun labelling state-affiliated media accounts belonging to Belarus. *Sinéad McSweeney, Twitter*. URL: <https://twitter.com/smcs/status/1502242278500454407>.

<sup>40</sup> We're adding labels... *Yoel Roth, Twitter*. URL: <https://twitter.com/yoyoel/status/1502117185124065294>.

<sup>41</sup> Since the invasion... *Yoel Roth, Twitter*. URL: <https://twitter.com/yoyoel/status/1498343851819368452>.

<sup>42</sup> Твітер буде пропонувати користувачам читати статті перед тим, як ретвітнути. *MediaSapien*. 25 вересня 2020. URL: <https://ms.detector.media/sotsmerezhi/post/25579/2020-09-25-twitter-budeproponuvati-koristuvacham-chitati-statti-pered-tim-yak-retvitnuti/>.

<sup>43</sup> Sinéad McSweeney Our ongoing approach...

This has attracted the attention of Ukrainian government officials on several occasions<sup>44</sup>.

In addition, posts from the Russian embassy in the United Kingdom were removed from Twitter and Facebook, saying that photographs of a hospital explosion in Mariupol, Ukraine, were staged. One of the deleted tweets quoted unreasonable remarks by Russian Foreign Minister Sergei Lavrov, claiming that the hospital was "not functional" and that the Ukrainian military forces and "radicals" were using it<sup>45</sup>. Similar posts were immediately deleted from other embassies' accounts. However, official Russian government accounts on social media sites around the world are not being deleted. According to Kevin McAllister, policy manager, even if social platform disagrees with the content that users share, it does not delete accounts; nevertheless, social platform does take action when users violate its rules.

The artificiality of the distribution of postings from pro-Kremlin accounts is clearly demonstrated in these situations. The labeling of Russian-controlled media accounts has delivered comparably interesting results: a 30% decline in tweets two weeks after the invasion began, as a result of this extended policy (as of March 11)<sup>46</sup>. However, following Russia's full-scale invasion of Ukraine, the social network registered more than 45,000 tweets each day from ordinary users with links to Russian state media.

This demonstrates that the vast majority of Russian state media content is shared on Twitter via personal accounts rather than via labeled official accounts of these media. As a result, the company recently changed its policy and began marking tweets that include links to state-owned Russian media. Such tweets also have less reach: they will not appear in the top search results and will not be recommended<sup>47</sup>.

Twitter stopped advertising in Ukraine and Russia for a similar reason. Political advertising (since 2019); monetization of misleading or false content linked to the Russian-Ukrainian war; monetization of search queries related to the Russian-Ukrainian war; and promoting content published by Russian-associated media are also prohibited.

Twitter also recorded a huge increase in the number of forged and manipulated content, for instance footage from video games masked as actual real video; footage from other conflicts or military operations masked as footage from Ukraine. The social network either labels or deletes such content.

<sup>44</sup> Meta, Twitter і Google проти Росії. Як соцмережі протистоять пропаганді Кремля. URL: <https://ms.detector.media/sotsmerezhi/post/29241/2022-03-27-meta-twitter-i-google-proti-rosii-yak-sotsmerezhi-protystoyat-propagandi-kremlya/>.

<sup>45</sup> Twitter and Facebook remove Russian embassy posts claiming Mariupol hospital bombing was faked. URL: <https://www.independent.co.uk/news/uk/home-news/russia-ukraine-hospital-bomb-twitter-b2033340.html>.

<sup>46</sup> Last week, we launched labels... *Yoel Roth, Twitter*. URL: <https://twitter.com/yoyoel/status/1502117187821064197>.

<sup>47</sup> Sinéad McSweeney Our ongoing approach...

Twitter is also taking additional steps to protect the service:

- adds context to crisis-related content, in particular through the Moments and Events feature;
- scans tweets for signs of manipulation or other untrustworthy behavior, and removes or limits the spread of tweets that misrepresent the picture of events;
- checks vulnerable accounts, such as those of journalists, activists, government officials and bodies, in order to prevent any hacking efforts;
- created extra recommendations on how to comply with digital security and protect your account in English, Ukrainian, and Russian;
- for users in Ukraine and Russia, some recommended tweets from people they don't follow have been disabled so that they don't receive offensive content<sup>48</sup>.

On April 26, Twitter accepted Elon Musk's offer to buy the \$ 44 billion Twitter account. Elon Musk announced a short list of his goals shortly after buying Twitter. Aside from freedom of speech, there is a desire to improve Twitter by introducing new features, as well as the adoption of open algorithms, the defeat of spam bots, and universal authentication<sup>49</sup>. However, misinformation, propaganda, and extremist ideas continue to abound on the site, particularly in regards to Russia's invasion of Ukraine. A softer content moderation policy can lead to even more fakes, propaganda and more toxic user behavior.

Due to such actions of social media during the first ten days of the war, Moscow banned Facebook and Twitter, shut down most of the remaining independent media, and enacted new harsh laws that threaten long-term imprisonment for anyone who dares to question the Kremlin's position on the war in Ukraine<sup>50</sup>.

## Conclusions

For audiences throughout the world attempting to understand the reasons for and make logical sense of

<sup>48</sup> Meta, Twitter і Google проти Росії...

<sup>49</sup> Elon Musk Yesss! Free speech... URL: <https://twitter.com/elonmusk/status/1518677066325053441>.

<sup>50</sup> Anders Åslund Why Vladimir Putin is losing the information war to Ukraine. URL: <https://www.atlanticcouncil.org/blogs/ukrainealert/why-vladimir-putin-is-losing-the-information-war-to-ukraine/>.

Russia's conflict against Ukraine, social media has become a main source of information. At the same time, the Russian and Ukrainian governments are using it to set the agenda for broader media reportage. On Twitter, Facebook, Instagram, and YouTube, official Russian government accounts have been revealed to be promoting pro-Russia disinformation. Meanwhile, the Ukrainian government has gone to social media to ask for help from its followers.

Information warfare is no longer an independent strategy arm, but rather a complementary component of military tactics. Because of the rise of social media, it is now simpler than ever to understand how governments use mass communication as a weapon.

The efforts of Meta, Twitter, and Google intended to counteract misinformation have had a considerable influence on the spreading of Russian propaganda and fake news among Ukrainians. Russia, on the other hand, disseminates its propaganda not just through state-controlled media and pro-Kremlin bloggers, but also through diplomatic missions, government organizations, and other websites.

At the same time, despite the corporations' willingness to engage in dialogue, there remain several ongoing issues – Russian disinformation and its consequences cannot be removed from the social network. The massive network of disinformation channels is a challenge for social networks, which they can not always cope with; so fakes spread sooner than platforms can not just block them, but at least recognize this information as a fake.

By banning Russian official media in Western Europe and adding labels to identify Russian government accounts on social media, companies have taken extraordinary steps to combat disinformation regarding the war in Ukraine. However, more than a dozen YouTube, Twitter, and Instagram accounts linked to persons and corporations on the sanctions list promote many of the same narratives that Sputnik and RT do so freely. While social media platforms have been at the center of information warfare, they have not been aggressive enough in removing not only content but also user profiles, because they find value in providing access to information.

## Information about the author:

**Yuskiv Khrystyna Victorivna,**

Assistant at the Department of International Information  
Lviv Polytechnic National University  
12, Stepan Bandera str., Lviv, 79013, Ukraine

## Інформація про автора:

**Юськів Христина Вікторівна,**

асистент кафедри міжнародної інформації  
Національний університет «Львівська політехніка»  
12, вул. Степана Бандери, м. Львів, 79013, Україна