

OVERTURNED

2022-008-FB-UA

# Russian poem

The Oversight Board has overturned Meta’s original decision to remove a Facebook post comparing the Russian army in Ukraine to Nazis and quoting a poem that calls for the killing of fascists

## Policies and topics

-  Art / Writing / Poetry,  
War and conflict
-  Hate speech

## Region and countries

-  Europe
-  Latvia, Ukraine,  
Russia

## Platform

-  Facebook

## Attachments

[Latvian translation](#)

[Ukrainian translation](#)

[Russian poem public comments](#)

This decision is also available in Russian (via the “language” tab accessed through the menu at the top of this screen), in Latvian ([here](#)) and in Ukrainian ([here](#)).

Lai iepazītos ar lēmumu latviski, spied [šeit](#).

Щоб прочитати це рішення українською мовою, натисніть [тут](#).

## Case summary

The Oversight Board has overturned Meta’s original decision to remove a Facebook post comparing the Russian army in Ukraine to Nazis and quoting a poem that calls for the killing of fascists. It has also overturned Meta’s finding that an image of what appears to be a dead body in the same post violated the Violent and Graphic Content policy. Meta had applied a warning screen to the image on the grounds that it violated the policy. This case raises some important issues about content moderation in conflict situations.

### About the case

In April 2022, a Facebook user in Latvia posted an image of what appears to be a dead body, face down, in a street. No wounds are visible. Meta confirmed to the Board that the person was shot in Bucha, Ukraine.

The Russian text accompanying the image argues that the alleged atrocities Soviet soldiers committed in Germany in World War Two were excused on the basis that they avenged the crimes Nazi soldiers had committed in the USSR. It draws a connection between the Nazi army and the Russian army in Ukraine, saying the Russian army “became fascist.”

The post cites alleged atrocities committed by the Russian army in Ukraine and says that “after Bucha, Ukrainians will also want to repeat... and will be able to repeat.” It ends by quoting the poem “Kill him!” by Soviet poet Konstantin Simonov, including the lines: “kill the fascist... Kill him! Kill him! Kill!”

The post was reported by another Facebook user and removed by Meta for violating its Hate Speech Community Standard. After the Board selected the case, Meta found it had wrongly removed the post and restored it. Three weeks later, it applied a warning screen to the image under its Violent and Graphic Content policy.

### Key findings

The Board finds that removing the post, and later applying the warning screen, do not align with Facebook’s Community Standards, Meta’s values, or its human rights responsibilities.

The Board finds that, rather than making general accusations that “Russian soldiers are Nazis,” the post argues that they acted like Nazis in a particular time and place, and draws historical parallels. The post also targets Russian soldiers because of their role as combatants, not because of their nationality. In this context, neither Meta’s human rights responsibilities nor its Hate Speech Community Standard protect soldiers from claims of egregious wrongdoing or prevent provocative comparisons between their actions and past events.

The Board emphasizes the importance of context in assessing whether content is urging violence. In this case, the Board finds that the quotes from the poem “Kill him!” are an artistic and cultural reference employed as a rhetorical device. When read in the context of the whole post, the Board finds that the quotes are being used to describe, rather than encourage, a state of mind. They warn of cycles of violence and the potential for history to repeat itself in Ukraine.

Meta’s internal guidance for moderators clarifies that the company interprets its Violence and Incitement Community Standard to allow such “neutral reference[s] to a potential outcome” and “advisory warning[s].” However, this is not explained in the public Community Standards. Likewise, the Violent and Graphic Content policy prohibits images depicting a violent death. Internal guidance for moderators describes how Meta determines whether a death appears violent, but this is not included in the public policy.

In this case, a majority of the Board finds that the image in the post does not include clear indicators of violence which, according to Meta’s internal guidance for moderators, would justify the use of a warning screen.

Overall, the Board finds that this post is unlikely to exacerbate violence. However, it notes that there are additional complexities in evaluating violent speech in international conflict situations where international law allows combatants to be targeted. The Russian invasion of Ukraine is internationally recognized as unlawful. The Board urges Meta to revise its policies to take into consideration the circumstances of unlawful military intervention.

## The Oversight Board's decision

The Oversight Board overturns Meta's original decision to remove the post and its subsequent determination that the image in the post violated the Violent and Graphic Content policy, as a result of which Meta applied a warning screen.

The Board recommends that Meta:

- Amends the public Violence and Incitement Community Standard to clarify, based on Meta's interpretation of the policy, that it permits content that makes “neutral reference to a potential outcome.”
- Include an explanation of how it determines whether an image depicts “the violent death of a person,” in the public Violent and Graphic Content Community Standard.
- Assess the feasibility of introducing tools that allow adult users to decide whether to see graphic content at all and, if so, whether to see it with or without a warning screen.

\* Case summaries provide an overview of the case and do not have precedential value.

## Full case decision

### 1. Decision summary

The Oversight Board overturns Meta's original decision to remove a Facebook post addressing the conflict in Ukraine. The content, which is in the Russian language and was posted in Latvia, comprises a photographic image of a street view with a person lying – likely deceased – on the ground, accompanied by text. The text includes quotations from a well-known poem by the Soviet poet Konstantin Simonov calling for resistance against the German invaders during World War II, and it implies that Russian invaders are playing a similar role in Ukraine to that which German soldiers played in the USSR. After the Board selected this post for review, Meta changed its position and restored the content to the platform. The content raises important definitional questions under Meta's Hate Speech and Violence and Incitement policies. A few weeks after deciding to restore the post, Meta affixed a warning screen to the photo. A majority of the Board finds that the photographic image does not violate the Violent and Graphic Content policy, as the image lacks clear visual indicators of violence as described in Meta's internal guidelines to content moderators, which would

of violence, as described in Meta's internal guidelines to content moderators, which would justify the use of the warning screen.

## 2. Case description and background

In April 2022, a Facebook user in Latvia posted a photo and text in Russian to their News Feed. The post was viewed approximately 20,000 times, shared approximately 100 times, and received almost 600 reactions and over 100 comments.

The photo shows a street view with a person lying, likely deceased, on the ground, next to a fallen bicycle. No wounds are visible. The text begins, "they wanted to repeat and repeated." The post comments on alleged crimes committed by Soviet soldiers in Germany during the Second World War. It says such crimes were excused on the basis that soldiers were avenging the horrors that the Nazis had inflicted on the USSR. It then draws a connection between the Second World War and the invasion of Ukraine, arguing that the Russian army "became fascist." The post states that the Russian army in Ukraine "rape[s] girls, wound[s] their fathers, torture[s] and kill[s] peaceful people." It concludes that "after Bucha, Ukrainians will also want to repeat... and will be able to repeat" such actions. At the end of the post, the user shares excerpts of the poem "Kill him!" by Soviet poet Konstantin Simonov, including the lines: "kill the fascist so he will lie on the ground's backbone, not you"; "kill at least one of them as soon as you can"; "Kill him! Kill him! Kill!"

The same day the content was posted, another user reported it as "violent and graphic content." Based on a human reviewer decision, Meta removed the content for violating its Hate Speech Community Standard. Hours later, the user who posted the content appealed and a second reviewer assessed the content as violating the same policy.

The user appealed to the Oversight Board. As a result of the Board selecting the appeal for review on May 31, 2022, Meta determined that its previous decision to remove the content was in error and restored it. On June 23, 2022, 23 days after the content was restored, Meta applied a warning screen to the photograph in the post under the Violent and Graphic Content Community Standard, on the basis that it shows the violent death of a person. The warning screen reads "sensitive content – this photo may show violent or graphic content," and gives users two options: "learn more" and "see photo."

The following factual background is relevant to the Board's decision and is based on research commissioned by the Board:

- The Russian invasion of Ukraine was recognized as unlawful by the United Nations General Assembly on March 2, 2022 ( [A/RES/ES-11/1](#)).
- Michelle Bachelet, UN High Commissioner for Human Rights, expressed concerns over “serious and disturbing questions about possible war crimes, grave breaches of international humanitarian law and serious violations of international human rights law” in Ukraine.
- In post-Soviet cultures, the term “fascist” often refers to German Nazis, especially in the context of World War II. However, the definition of the term “fascist” is no longer directly associated with the German Nazis. In Russia and in Ukraine, the term became blurred due to political discussions where the term “fascist” has been used to stigmatize opponents across different parts of the political spectrum.
- “They wanted to repeat and repeated,” the first sentence of the user’s post, is a reference to “[w]e can repeat,” a popular slogan in Russian, which is used by some to refer to the Soviet victory against the Nazis during World War II. The user further alludes to this slogan when stating that, “after Bucha, Ukrainians will also want to repeat... and will be able to repeat” the Russian army’s actions.

### **3. Oversight Board authority and scope**

The Board has authority to review Meta’s decision following an appeal from the user whose content was removed (Charter Article 2, Section 1; Bylaws Article 3, Section 1). The Board may uphold or overturn Meta’s decision (Charter Article 3, Section 5), and this decision is binding on the company (Charter Article 4). Meta must also assess the feasibility of applying its decision in respect of identical content with parallel context (Charter Article 4). The Board’s decisions may include policy advisory statements with non-binding recommendations that Meta must respond to (Charter Article 3, Section 4; Article 4).

When the Board selects cases like this one, where Meta has agreed that it made an error, the Board reviews the original decision to help increase understanding of why errors occur, and to make observations or recommendations that may contribute to reducing errors and to enhancing fair and transparent procedures.

### **4. Sources of authority**

The Oversight Board considered the following authorities and standards:

## *I. Oversight Board decisions:*

The most relevant previous decisions of the Oversight Board include:

- “Tigray Communication Affairs Bureau” case ( [2022-006-FB-MR](#)). In this case, the Board upheld Meta’s decision to take down the content after concluding that, given the user’s profile (a regional government ministry), the language adopted (an explicit call to kill soldiers who do not surrender) and reach of the page (about 260,000 followers), there was a high-risk the post could have led to further violence.
- “Sudan graphic video” case ( [2022-002-FB-MR](#)). In this case, the Board recommended Meta: (i) amend the Violent and Graphic Content Community Standard to allow videos of people or dead bodies when shared for the purpose of raising awareness of or documenting human rights abuses; and (ii) undertake a policy development process that develops criteria to identify videos of people or dead bodies when shared for the purpose of raising awareness of or documenting human rights abuses.

## *II. Meta’s content policies:*

Under the [Hate Speech Community Standard](#), Meta does not permit “violent” or “dehumanizing” speech that is directed at people or groups on the basis of their protected characteristics. The policy states that “dehumanizing speech” includes “comparisons, generalizations, or unqualified behavioral statements to or about... violent and sexual criminals.” The policy explicitly does not apply to qualified behavioral statements. Groups described as “having carried out violent crimes or sexual offenses” are not protected from attacks under the Hate Speech policy.

Under the [Violence and Incitement Community Standard](#), Meta does not allow “threats that could lead to death (and other forms of high-severity violence)” where “threat” is defined as, among other things, “calls for high-severity violence” and “statements advocating for high-severity violence.” Meta’s internal guidelines for content reviewers clarify that the company interprets this policy to allow content containing statements with “neutral reference to a potential outcome of an action or an advisory warning.”

Under the [Violent and Graphic Content Community Standard](#), “imagery that shows the

violent death of a person or people by accident or murder” is covered with a warning screen.

### *III. Meta’s values:*

Meta has described Facebook’s values of “Voice,” “Dignity,” and “Safety,” among others, in the introduction to the Community Standards. This decision will refer to those values if and as relevant to the decision.

### *IV. International human rights standards:*

The UN Guiding Principles on Business and Human Rights ( [UNGPs](#)), endorsed by the UN Human Rights Council in 2011, establish a voluntary framework for the human rights responsibilities of private businesses. In 2021, Meta [announced](#) its [Corporate Human Rights Policy](#), where it reaffirmed its commitment to respecting human rights in accordance with the UNGPs. The Board’s analysis of Meta’s human rights responsibilities in this case was informed by the following human rights standards:

- [The rights of freedom of opinion and expression](#): Article 19, International Covenant on Civil and Political Rights ( [ICCPR](#)), [General Comment No. 34](#), Human Rights Committee, 2011; UN Special Rapporteur on freedom of opinion and expression, reports: [A/HRC/38/35](#) (2018), [A/74/486](#) (2019) and [A/HRC/44/49/Add.2](#) (2020); UN High Commissioner for Human Rights, report: [A/HRC/22/17/Add.4](#) (2013).
- [Equality and non-discrimination](#): Article 2, para. 1 (ICCPR).
- [Right to physical security](#): Article 9 (ICCPR).
- [Right to life](#): Article 6 (ICCPR).

## **5. User submissions**

In their appeal to the Board, the user states that the photo they shared is the “most innocuous” of the pictures documenting the “crimes of the Russian army in the city of Bucha,” “where dozens of dead civilians lie on the streets.” The user says that their post does not call for violence and is about “past history and the present.” They say the poem was originally dedicated to the “struggle of Soviet soldiers against the Nazis,” and that they posted it to show how “the Russian army became an analogue of the fascist army.” As part of their appeal, they state they are a journalist and believe it is important for people to understand what is happening, especially in wartime.

## 6. Meta's submissions

In the rationale Meta provided to the Board, the company analyzed the content of this case in light of three different policies, starting with Hate Speech. Meta focused on why the company reversed its original decision, rather than explaining how it had come to its original decision. According to Meta, claiming that Russian soldiers committed crimes in the context of the Russia-Ukraine conflict does not constitute an attack under the Hate Speech policy because “qualified behavioral statements” are allowed on the platform. Meta also explained that fascism is a political ideology, and merely linking the Russian army to a certain political ideology does not constitute an attack because “the Russian army is an institution and therefore not a protected characteristic group or subset covered by the Hate Speech policy (as compared to Russian soldiers, who are people).” Finally, Meta indicated that the different excerpts of the poem “Kill him!” quoted in the text of the post (e.g., “kill a fascist,” “kill at least one of them,” “kill him!”) refer to “Nazis” in the context of World War II, and Nazis are not a protected group.

Meta also analyzed the post in light of its Violence and Incitement policy. In this regard, Meta explained that stating that “Ukrainians will also want to repeat... and will be able to repeat” after the events in Bucha does not advocate violence. Meta claimed that this is a “neutral reference to a potential outcome,” which the company interprets as permitted under the Violence and Incitement policy. Meta also stated that quoting Simonov’s poem was a way of raising awareness of the potential for history to repeat itself in Ukraine. Finally, the company explained that advocating violence against individuals covered in the Dangerous Individuals and Organizations policy, such as the Nazis (referred to as “fascists” in Simonov’s poem), is allowed under the Violence and Incitement Community Standard.

Meta then explained that a warning screen and appropriate age restrictions were applied to the post under its Violent and Graphic Content policy because the image included in the content shows the violent death of a person. Meta confirmed that the image depicts an individual who was shot in Bucha, Ukraine.

In response to the Board’s questions, Meta provided further explanation of initiatives it has developed in the context of the conflict in Ukraine. Meta confirmed, however, that none of these initiatives were relevant to the initial removal of the content in this case or the decision to restore it with a warning screen. The company added that it has taken several steps

consistent with the I IN Guiding Principles on Business and Human Rights to ensure due

consistent with the UN Guiding Principles on Business and Human Rights to ensure due diligence in times of conflict. These steps included engagement with Ukrainian and Russian civil society, and Russian independent media, to seek feedback on the impact of the measures Meta adopted at the onset of the conflict.

The Board asked Meta 11 questions, and Meta responded to them all fully.

## **7. Public comments**

The Oversight Board received eight public comments related to this case. Three of the comments were submitted from Europe, four from the United States and Canada and one from Latin America and the Caribbean. The submissions covered the following themes: international armed conflict; the importance of context in content moderation; the role of journalists in conflict situations; the documentation of war crimes; and artistic expression.

To read public comments submitted for this case, please click [here](#).

## **8. Oversight Board analysis**

The Board initially examined whether the content was permissible under Meta's content policies, interpreted where necessary in light of the values of the platform, and then assessed whether the treatment of the content comports with the company's human rights responsibilities.

This case was selected by the Board because the removal of this post raised important concerns about artistic expression and cultural references repurposed in new contexts that potentially risk inciting violence in conflict situations. Online spaces for expression are particularly important for people impacted by war and social media companies must pay particular attention to protecting their rights. This case demonstrates how a lack of contextual analysis, which is common in content moderation at scale, may prevent users from expressing opinions regarding conflict, and drawing provocative historical parallels.

### **8.1 Compliance with Meta's content policies**

The Board finds that the content in this case does not violate the Hate Speech Community Standard, or the Violence and Incitement Standard. A majority of the Board finds that the content does not violate the Violent and Graphic Content Community Standard.

## I. Hate Speech

Meta's Hate Speech policy prohibits attacks against people based on protected characteristics, including nationality. Profession receives "some protections" when referenced along with a protected characteristic. The company's internal guidelines for moderators further clarify that "quasi-protected subsets," including groups defined by a protected characteristic plus a profession (e.g., Russian soldiers), are generally entitled to protection against Hate Speech Tier 1 attacks. Such attacks include, among others, "violent speech" and "dehumanizing speech... in the form of comparisons, generalizations, or unqualified behavioral statements" "to or about... violent and sexual criminals." Meta's Hate Speech policy does not offer protection to "groups described as having carried out violent crimes or sexual offences."

Generic claims that Russian soldiers have a propensity to commit crimes could, depending on content and context, violate Meta's Hate Speech policy. Such claims could fall within the prohibition against "dehumanizing speech" under the policy, in the form of "unqualified behavioral statements." Meta's policy distinguishes between: (i) the attribution of bad character or undesirable traits to a group on account of its ethnicity, national origin or other protected characteristics (this is what Meta means by "generalizations"); (ii) the criticism of members of a group without context (this is what Meta means by "unqualified behavioral statements"); and (iii) the criticism of members of a group for their past behavior (this is what Meta means by "qualified behavioral statements"). In this case, the claims about invading Russian soldiers are made in the context of their actions in the Ukraine conflict, not in general.

The question here is whether the user's comparison of Russian soldiers to World War II-era German fascists, and their assertion that they raped women, killed their fathers, and killed and tortured innocent persons, violated Meta's Hate Speech policy. Both Meta and the Board conclude that the post does not constitute violent or dehumanizing speech under the Hate Speech policy, though for different reasons. This was the only part of the Hate Speech policy that has been identified as relevant to this post.

Meta argues that the post is not violating because the accusatory statements are directed at the Russian *army*, which is an institution, and not at Russian *soldiers*, who are people. The Board finds that this distinction does not hold in this case, since the user refers to "army" and "soldiers" interchangeably.

Nonetheless, the Board finds that the user's accusation that Russian soldiers committed crimes comparable to the Nazis' in the context of Russia's invasion of Ukraine is permitted. This is the case because ascribing specific actions (e.g., "they began to really take revenge – rape girls, cut their fathers, torture and kill peaceful people of peaceful outskirts Kyiv") and comparing Russian soldiers' actions in Ukraine with other armies known to have committed war crimes (e.g., "the Russian army, after 70 years, completely repeated itself in Germany and the German [army] in Ukraine") are "qualified" statements, related to behavior observed during a specific conflict.

The Board therefore finds that, comparing Russian soldiers' actions in a specific context to the crimes of the Nazis is permitted under the Community Standards, regardless of whether a generic comparison to the Nazis is or is not permissible. According to Meta's internal guidelines, material that may otherwise constitute hate speech does not violate the policy if it is targeted against groups "described as having carried out violent crimes or sexual offenses." More broadly, it does not violate the hate speech policy to report instances of violations of human rights in a particular context, even if those people are identified by reference to their national origin.

The Board further finds that the different excerpts of the poem "Kill him!" quoted in the content (e.g., "kill a fascist," "kill at least one of them," "kill him!") should not be considered "violent speech" because, when read together with the rest of the post, the Board understands that the user is calling attention to the cycle of violence, rather than urging violence.

Finally, the Board concludes that Russian soldiers are targeted in the post because of their role as combatants, not their nationality. The claims are not attacks directed at a group "on the basis" of their protected characteristics. It follows that the content is not hate speech, because no protected characteristic is engaged.

## *II. Violence and Incitement*

Under the Violence and Incitement policy, Meta removes "calls for high-severity violence," "statements advocating for high-severity violence," and "aspirational or conditional statements to commit high-severity violence," among other types of expression. The company's internal guidelines for moderators further clarify that Meta interprets this Community Standard to allow statements with "neutral reference to a potential outcome of an

action or an advisory warning.” Additionally, the internal guidelines explain “content that condemns or raises awareness of violent threats” is also allowed under the Violence and Incitement policy. This applies to content that “clearly seeks to inform and educate others about a specific topic or issue; or content that speaks to one’s experience of being a target of a threat or violence,” including academic and media reports.

Meta explained to the Board that it decided to restore this content because the post did not advocate violence. Meta characterizes the user’s statement that “Ukrainians will also want to repeat ... and will be able to repeat” after the events in Bucha as a “neutral reference to a potential outcome,” which the company interprets as permitted under the Violence and Incitement policy. Meta also states that quoting Simonov’s poem was a way of raising awareness of the potential for history to repeat itself in Ukraine. Finally, pointing to the fact that Simonov’s poem is directed against German fascists, it notes that advocating violence against individuals covered in the Dangerous Individuals and Organizations policy, such as the Nazis, is allowed under the Violence and Incitement Community Standard.

The Board is partially persuaded by this reasoning. It agrees that the sentence “Ukrainians will also want to repeat... and will be able to repeat” neither calls for nor advocates violence. Read literally, this portion of the post merely states that Ukrainians might well respond as violently to the Russian army’s actions as the Soviets did to the Nazis’. In other words, it is a “neutral reference to a potential outcome,” permitted as per Meta’s interpretation of the Violence and Incitement policy, clarified in the internal guidelines provided to content moderators.

The Board also finds that the excerpts with violent language of the poem “Kill him!” cited in the section above, may be read as describing, not encouraging, a state of mind. When read together with the entire post, including the photographic image, the excerpts are part of a broader message warning of the potential for history to repeat itself in Ukraine. They are an artistic and cultural reference employed as a rhetorical device by the user to convey their message. Therefore, the Board concludes that this part of the content is also permitted by Meta’s internal guidelines.

The Board concludes, however, that Meta is being unrealistic when it analyzes the post as if it were merely a call to violence against Nazis. The user makes clear that they regard Russian soldiers in Ukraine today as akin to Germans in Russia during World War II. To the extent that the post, with its quotations from Simonov’s poem, could be considered to refer to soldiers

now committing atrocities against civilians, there is a risk that readers will read this as a call to violence against Russian soldiers today. The Board nonetheless agrees with Meta's conclusion that the post does not violate the Violence and Incitement Standard because its primary meaning, in context, is a warning against a cycle of violence.

### *III. Violent and Graphic Content*

Under its Violent and Graphic Content policy, Meta adds warning labels to content "so that people are aware of the graphic or violent nature before they click to see it." That is the case for "[i]magery that shows the violent death of a person or people by accident or murder." In its internal guidelines for content moderators, Meta describes as "indicators of a violent death" graphic imagery of the "aftermath of violent death where the victim appears dead or visibly incapacitated, and there are additional visual indicators of violence," such as "blood or wounds on the body, blood surrounding the victim, bloated or discolored body, or bodies excavated from debris." The internal guidelines further explain that bodies without "any visible indicator of violent death" or without "at least one indicator of violence" should not be considered as a depiction of a "violent death."

The photo included in the content shows a view of a street with a person lying still on the ground. No wounds are visible. Meta was able to confirm that the person was shot in Bucha, Ukraine. The Board notes that content moderators working at scale would not necessarily have access to this type of information. A majority of the Board finds that the Violent and Graphic Content policy was not violated because the photographic image lacks clear visual indicators of violence, as described in Meta's internal guidelines for content moderators. Therefore, the majority concludes that a warning screen should not have been applied.

Considering the context of armed conflict in Ukraine and the debris depicted in the image, a minority of the Board finds that the content does violate the Violent and Graphic Content policy.

### *IV. Enforcement action*

The content was reported by a user for Violent and Graphic Content but was taken down under the Hate Speech policy. It was restored only after being brought to Meta's attention by the Board. In response to questions from the Board, Meta explained that the case content was not escalated to policy or subject matter experts for an additional review. "Escalated for

review" means that, instead of the decision being revisited by content moderators conducting at-scale review, it is sent to an internal team at Meta that is responsible for the relevant policy or subject area.

## 8.2 Compliance with Meta's values

The Board finds that removing the content, and placing a warning screen over the image included in it, are not consistent with Meta's values.

The Board is concerned about the situation of Russian civilians, and possible effects of violent speech targeting Russians in general. However, in this case the Board finds the content does not pose a real risk to the "Dignity" and "Safety" of those people that would justify displacing "Voice," especially in a context where Meta should make sure users impacted by war are able to discuss its implications.

## 8.3 Compliance with Meta's human rights responsibilities

The Board finds that Meta's initial decision to remove the content and Meta's decision to apply a warning screen to the content were both inconsistent with Meta's human rights responsibilities as a business. Meta has committed itself to respect human rights under the UN Guiding Principles on Business and Human Rights ( [UNGPs](#)). [Facebook's Corporate Human Rights Policy](#) states that this includes the International Covenant on Civil and Political Rights ( [ICCPR](#)).

### *Freedom of expression (Article 19 ICCPR)*

The scope of the right to freedom of expression is broad. Article 19, para. 2, of the ICCPR gives heightened protection to expression, including artistic expression, on political issues, and commentary on public affairs, as well as to discussions of human rights and of historical claims ( [General Comment No. 34](#), paras. 11 and 49). Even expression which may be regarded as "deeply offensive" is entitled to protection (General Comment No. 34, para. 11). The content under analysis by the Board in this case contains strong language. However, it amounts to political discourse and draws attention to human rights abuses in a war context.

The content in this case included quotes from a well-known war poem, which the user employed as a provocative cultural reference to educate and warn their audience of the potential consequences of Russian soldiers' actions in Ukraine. The UN Special Rapporteur

on freedom of expression has highlighted that artistic expression includes “the fictional and nonfictional stories that educate or divert or provoke” ([A/HRC/44/49/Add.2](#), para. 5).

ICCPR Article 19 requires that where restrictions on expression are imposed by a state, they must meet the requirements of legality, legitimate aim, and necessity and proportionality (Article 19, para. 3, ICCPR). The UN Special Rapporteur on freedom of expression has encouraged social media companies to be guided by these principles when moderating online expression ([A/HRC/38/35](#), paras. 45 and 70).

### *1. Legality (clarity and accessibility of the rules)*

The principle of legality requires rules used by states to limit expression to be clear and accessible (General Comment 34, para. 25). The Human Rights Committee has further noted that rules “may not confer unfettered discretion for the restriction of freedom of expression on those charged with [their] execution” (General Comment 34, para. 25). Individuals must have enough information to determine if and how their expression may be limited, so that they can adjust their behavior accordingly. Applied to Meta’s content rules for Facebook, users should be able to understand what is allowed and what is prohibited.

Two sections from Meta’s internal guidelines on how to enforce the Violence and Incitement Community Standard are particularly relevant to the conclusion reached by both Meta and the Board that the content should stay on Facebook. First, Meta interprets this policy to allow messages that warn of the possibility of violence by third parties, if they are statements with “neutral reference to a potential outcome of an action or an advisory warning.” Second, otherwise violating content is allowed if it “condemns or raises awareness of violent threats.” The Board notes, however, that these internal guidelines are not included in the public-facing language of the Violence and Incitement Community Standard. This might cause users to believe that content such as the post in this case is violating, when it is not. Meta should integrate these sections into the public-facing language of the Violence and Incitement policy, so that it becomes sufficiently clear to users. The Board is aware that publicizing more detail on Meta’s content policies might enable users with malicious intent to circumvent the Community Standards more easily. The Board considers, however, that the need for clarity and specificity prevails over the concern that some users might attempt to “game the system.” Not knowing that “neutral references to a potential outcome,” “advisory warnings,” “condemning” or “raising awareness” of violent threats are permitted might cause users to avoid initiating or engaging in public interest discussions on Meta’s platforms.

The Board is also concerned that in this case, Meta's decision to affix a warning screen is inconsistent with its internal guidelines to content moderators. Additionally, Meta's interpretation of the Violent and Graphic Content policy may not be clear to users. Meta should seek to clarify, in the public-facing language of the policy, how the company interprets the policy, and how it determines whether an image "shows the violent death of a person or people by accident or murder," in the context of conflict, as per Meta's internal guidelines to content moderators.

Additionally, Meta informed the Board that no message was sent to the user who originally reported the content to inform them that the post was later restored by the company. This raises legality concerns, as the lack of relevant information for users may interfere with "the individual's ability to challenge content actions or follow up on content-related complaints" ([A/HCR/38/35](#), at para. 58). The Board notes that notifying reporters of the enforcement action taken against the content they reported, and the relevant Community Standard enforced, would help users to better understand and follow Meta's rules.

## *II. Legitimate aim*

Any restriction on freedom of expression should also pursue a "legitimate aim." The Board has previously recognized that the Hate Speech Community Standard pursues the legitimate aim of protecting the rights of others (General Comment No. 34, para. 28), including the rights to equality and non-discrimination based on ethnicity and national origin (Article 2, para. 1, ICCPR). Protecting Russians targeted by hate is, therefore, a legitimate aim. However, the Board finds that protecting "soldiers" from claims of wrongdoing is not a legitimate aim, when they are being targeted because of their role as combatants during a war, not because of their nationality or another protected characteristic; criticism of institutions such as the army should not be prohibited (General Comment No. 34, para. 38).

The Violence and Incitement Standard, properly framed and applied, pursues the legitimate aim of protecting the rights of others. In the context of this case, this policy seeks to prevent the escalation of violence which could lead to harm to the physical security (Article 9, ICCPR) and life (Article 6, ICCPR) of people in the areas impacted by the Russian-Ukraine conflict.

The Board notes that there are additional complexities involved in evaluating violent speech in the context of armed resistance to an invasion. The Russian invasion of Ukraine is

internationally recognized as unlawful (A/RES/ES-11/1), and the use of force as self-defense against such acts of aggression is permitted (Article 51, UN Charter). In a context of international armed conflict, international humanitarian law on the conduct of parties to hostilities allows active combatants to be lawfully targeted in the course of armed conflict. This is not the case with people no longer taking active part in the hostilities, including prisoners of war (Article 3, [Geneva Convention relative to the Treatment of Prisoners of War](#)). When violence is itself lawful under international law, speech urging such violence presents different considerations that must be examined separately. Although the Board has found the content in this case to be non-violating, the Board urges Meta to revise its policies to take into consideration the circumstances of unlawful military intervention.

With reference to the company's decision to affix a warning screen to the photograph, Meta notes that the Violent and Graphic content policy aims to promote an environment that is conducive to diverse participation by limiting "content that glorifies violence or celebrates the suffering or humiliation of others." The Board agrees that this aim is legitimate in the context of Meta's goal to promote an inclusive platform.

### *III. Necessity and proportionality*

The principle of necessity and proportionality provides that any restrictions on freedom of expression "must be appropriate to achieve their protective function; they must be the least intrusive instrument amongst those which might achieve their protective function; [and] they must be proportionate to the interest to be protected" (General Comment 34, para. 34).

In order to assess the risks posed by violent or hateful content, the Board is typically guided by the six-factor test described in the [Rabat Plan of Action](#), which addresses advocacy of national, racial or religious hatred that constitutes incitement to hostility, discrimination or violence. In this case, the Board finds that, despite the context of ongoing armed conflict and the charged cultural references employed by the user, it is unlikely that the post – a warning against a cycle of violence – would lead to harm. The Board concludes that the initial content removal was not necessary. Additionally, a majority of the Board finds that the warning screen was also not necessary, whereas a minority of the Board finds that it was both necessary and proportional.

Considering the relevant factors here, the Board concludes that despite the context of Russia's unlawful invasion of Ukraine where potentially inflammatory speech could increase tensions, the evident intention of the user (raising awareness around the war and its

concerns, the evident intention of the post (raising awareness around the war and its consequences), the reflective tone adopted in quoting a war poem, and the proliferation of other communications regarding the horrific events in Ukraine, mean the content is not likely to contribute significantly to the exacerbation of violence.

A majority of the Board concludes that the use of a warning screen inhibits freedom of expression and is not a necessary response in this instance, as the photographic image lacks clear visual indicators of violence, as described in Meta's internal guidelines to content moderators, which would justify the use of the warning screen. Social media companies should consider a range of possible responses to problematic content to ensure restrictions are narrowly tailored ( [A/74/486](#), para. 51). In this regard, the Board considers that Meta should further develop customizations tools so that users are able to decide on whether to see sensitive graphic content with or without warnings on Facebook and on Instagram.

A minority of the Board finds that the warning screen was a necessary and proportionate measure that was appropriately tailored to encourage participation and freedom of expression. This minority believes that, in consideration of the dignity of deceased persons, especially in the context of an armed conflict, and the possible effects of images depicting death and violence on a great number of users, Meta may err on the side of prudence by adding warning screens over content such as the one under analysis.

## **9. Oversight Board decision**

The Oversight Board overturns Meta's original decision to take down the content, and Meta's subsequent determination that the Violent and Graphic Content policy was violated, which led the company to affix a warning screen to the photographic image in the post.

## **10. Policy advisory statement**

### Content policy

1. Meta should add to the public-facing language of its Violence and Incitement Community Standard that the company interprets the policy to allow content containing statements with "neutral reference to a potential outcome of an action or an advisory warning," and content that "condemns or raises awareness of violent threats." The Board expects that this recommendation, if implemented, will require Meta to update the public-facing language of the Violence and Incitement policy to reflect these inclusions.

2. Meta should add to the public-facing language of its Violent and Graphic Content Community Standard detail from its internal guidelines about how the company determines whether an image “shows the violent death of a person or people by accident or murder.” The Board expects that this recommendation, if implemented, will require Meta to update the public-facing language of the Violent and Graphic Content Community Standard to reflect this inclusion.

### Enforcement

3. Meta should assess the feasibility of implementing customization tools that would allow users over 18 years old to decide whether to see sensitive graphic content with or without warning screens, on both Facebook and Instagram. The Board expects that this recommendation, if implemented, will require Meta to publish the results of a feasibility assessment.

### **\*Procedural note:**

The Oversight Board’s decisions are prepared by panels of five Members and approved by a majority of the Board. Board decisions do not necessarily represent the personal views of all Members.

For this case decision, independent research was commissioned on behalf of the Board. An independent research institute headquartered at the University of Gothenburg and drawing on a team of over 50 social scientists on six continents, as well as more than 3,200 country experts from around the world. The Board was also assisted by Duco Advisors, an advisory firm focusing on the intersection of geopolitics, trust and safety, and technology. The company Lionbridge Technologies, LLC, whose specialists are fluent in more than 350 languages and work from 5,000 cities across the world, provided linguistic expertise.