FOOTAGE OF MOSCOW TERRORIST ATTACK

2024-038-FB-UA, 2024-039-FB-UA, 2024-040-FB-UA

Summary

The Board has overturned Meta's decisions to remove three Facebook posts showing footage of the March 2024 terrorist attack in Moscow, requiring the content to be restored with "Mark as Disturbing" warning screens.

While the posts violated Meta's rules on showing the moment of designated attacks on visible victims, removing them was not consistent with the company's human rights responsibilities. The posts, which discussed an event on front page news worldwide, are of high public interest value and to be protected under the newsworthiness allowance, according to the majority of the Board. In a country such as Russia with a closed media environment, accessibility on social media of such content is even more important. The posts each contain clear language condemning the attack, showing solidarity with or concern for the victims, with no clear risk of them leading to radicalization or incitement.

Suppressing matters of vital public concern based on unsubstantiated fears it could promote radicalization is not consistent with Meta's responsibilities to free expression. As such, Meta should allow, with a "Mark as Disturbing" warning screen, third-party imagery of a designated event showing the moment of attacks on visible but not identifiable victims when shared for news reporting, condemnation and raising awareness.

About the Cases

The Board has reviewed three cases together involving content posted on Facebook by different users immediately after the March 22, 2024, terrorist attack at a concert venue and retail complex in Moscow.

The first case featured a video showing part of the attack inside the retail complex, seemingly filmed by a bystander. While the attackers and people being shot were visible but not easily identifiable, others leaving the building were identifiable. The caption asked what is happening in Russia and included prayers for those impacted.

The second case featured a shorter clip of the same footage, with a caption warning viewers about the content and stating there is no place in the world for terrorism.

The third case involved a post shared on a Facebook group page by an administrator. The group's description expresses support for former French presidential candidate Éric Zemmour. The post included a still image from the attack, which could have been taken from the same video, showing armed gunmen and victims. Additionally, there was a short video of the retail complex on fire, filmed by someone driving past. The caption stated that Ukraine had said it had nothing to do with the attack, while pointing out that nobody had claimed responsibility. The caption also included a statement of support for the Russian people.

Meta removed all three posts for violating its Dangerous Organizations and Individuals policy, which prohibits third-party imagery depicting the moment of such attacks on visible victims. Meta designated the Moscow attack as a terrorist attack on the day it happened. According to Meta, the same video shared in the first two cases had already been posted by a different user and then escalated to the company's policy or subject matter experts for additional review earlier on in the day. Following that review, Meta decided to remove the video and added it to a Media Matching Service (MMS) bank. The MMS bank subsequently determined that the content in the first two cases matched the banked video that had been tagged for removal and automatically removed it. In the third case, the content was removed by Meta following human review.

The attack carried out on March 22, 2024 in Moscow's Crocus City Hall claimed the lives of at least 143 people. An affiliate of the Islamic State, ISIS-K, claimed responsibility soon after the attack. According to experts consulted by the Board, tens of millions of Russians watched the video of the attack on state-run media channels, as well as Russian social media platforms. While Russian President Vladimir Putin

claimed there were links to Ukraine and support from Western intelligence for the attack, Ukraine has denied any involvement.

Key Findings

While the posts were either reporting on, raising awareness of or condemning the attacks, Meta does not apply these exceptions under the Dangerous Organizations and Individuals policy to "third-party imagery depicting the moment of [designated] attacks on visible victims." As such, it is clear to the Board that all three posts violate Meta's rules.

However, the majority of the Board finds that removing this content was not consistent with Meta's human rights responsibilities, and the content should have been protected under the newsworthiness allowance. All three posts contained subject matter of pressing public debate related to an event that was front page news worldwide. There is no clear risk of the posts leading to radicalization or incitement. Each post contains clear language condemning the attack, showing solidarity with or concern for the victims, and seeking to inform the public. In combination with the lack of media freedom in Russia, and the fact the victims are not easily identifiable, this further moves these posts in the direction of the public interest.

Suppressing content on matters of vital public concern based on unsubstantiated fears it could promote radicalization is not consistent with Meta's responsibilities to free expression. This is particularly the case when the footage has been viewed by millions of people and accompanied by allegations that the attack was partly attributable to Ukraine. The Board notes the importance of maintaining access to information during crises particularly in Russia, where people rely on social media to access information or to raise awareness among international audiences.

While, in certain circumstances, removing content depicting identifiable victims is necessary and proportionate (e.g., in armed conflict when victims are prisoners of war), as the victims in these cases are not easily identifiable, restoring the posts with an age-gated warning screen is more in line with Meta's human rights responsibilities. Therefore, Meta should amend its policy to allow third-party imagery of visible but not

personally identifiable victims when clearly shared for news reporting, condemnation or awareness raising.

A minority of the Board disagrees and would uphold Meta's decisions to remove the posts from Facebook. For the minority, the graphic nature of the footage and the fact that it shows the moment of attack and, in this case, death of visible victims, makes removal necessary for the dignity of the victims and their families.

In addition, the Board finds that the current placement of the rule on footage of violating violent events under the Dangerous Organizations and Individuals policy creates confusion for users. While the "We remove" section implies that condemnation and news reporting is permissible, other sections state that perpetrator-generated imagery and third-party imagery of moment of attacks on visible victims is prohibited and does not specify that Meta will remove such content even if it condemns or raises awareness of attacks.

The Oversight Board's Decision

The Oversight Board overturns Meta's decisions to remove the three posts, requiring the content to be restored with "Mark as Disturbing" warning screens.

The Board also recommends that Meta:

- Allow, with a "Mark as Disturbing" warning screen, third-party imagery of a designated event showing the moment of attacks on visible but not personally identifiable victims when shared in the contexts of news reporting, condemnation and raising awareness.
- Include a rule under the "We remove" section of the Dangerous Organizations and Individuals Community Standard on designated violent events. It should also move the explanation of how Meta treats content depicting designated events out of the policy rationale section and into this section.

* Case summaries provide an overview of cases and do not have precedential value.

Full Case Decision

Section 1: Case Description and Background

The Oversight Board has reviewed three cases together involving content posted on Facebook by different users immediately after the March 22, 2024, terrorist attack at a concert venue and retail complex in Moscow. Meta's platforms have been blocked in Russia since March 2022, when a government ministry labeled the company an "extremist organization." However, Meta's platforms remain [accessible](#) to people through Virtual Private Networks (VPNs).

In the first case, a Facebook user posted a short video clip on their profile accompanied by a caption in English. The video showed part of the attack from inside the retail complex, with the footage seemingly taken by a bystander. Armed people were shown shooting unarmed people at close range, with some victims crouching on the ground and others fleeing. The footage was not high resolution. While the attackers and people being shot were visible but not easily identifiable, others leaving the building were identifiable. In the audio, gunfire could be heard, with people screaming. The caption asked what is happening in Russia and included prayers for those impacted. When Meta removed the post within minutes of it being posted, it had fewer than 50 views.

In the second case, a different Facebook user posted a shorter clip of the same footage, also accompanied by an English caption, which warned viewers about the content, stating there is no place in the world for terrorism. When Meta removed the post within minutes of it being posted, it had fewer than 50 views.

The third case involves a post shared on a group page by an administrator. The group's description expresses support for former French presidential candidate Éric Zemmour. The post included a still image from the attack, which could have been taken from the same video, showing armed gunmen and victims. Additionally, there was a short video of the retail complex on fire, filmed by someone driving past. The French caption included the word "Alert" alongside

commentary on the attack, such as the reported number of fatalities. The caption also stated that Ukraine had said it had nothing to do with the attack, while pointing out that nobody had claimed responsibility for it. The caption concluded with a comparison to the Bataclan terrorist attack in Paris and a statement of support for the Russian people. When Meta removed the post the day after it was posted, it had about 6,000 views.

The company removed all three posts under its Dangerous Organizations and Individuals Community Standard, which prohibits sharing all perpetrator-generated content relating to designated attacks as well as footage captured by or imagery produced by third parties (e.g., bystanders, journalists), depicting the moment of terrorist attacks on visible victims. Meta designated the Moscow attack as a terrorist attack on the same day it happened. According to Meta, the same video shared in the first two cases had already been posted by a different user and then escalated to the company's policy or subject matter experts for additional review earlier on in the day. Following that review, Meta decided to remove the video and added it to a Media Matching Service (MMS) bank. The MMS bank subsequently determined that the content in the first two cases matched the banked video that had been tagged for removal and automatically removed it. Meta did not apply a strike or a feature limit to the users' profiles as the bank was configured to remove content without imposing a strike. In the third case, the content was removed by Meta following human review, with the company applying a strike that resulted in a 30-day feature limit. The feature limit applied to the user prevented them from creating content on the platform, creating or joining Messenger rooms, and advertising or creating live videos. It is unclear why the MMS system did not identify this content.

In all three cases, the users appealed to Meta. Human reviewers found each post violating. After the Board selected these cases for review, Meta confirmed its decisions to remove all three posts were correct but removed the strike in the third case.

The Board notes the following context in reaching its decision.

The [attack](#) carried out on March 22, 2024 in Moscow's Crocus City Hall claimed the lives of at least 143 people. An affiliate of the Islamic State, ISIS-K, claimed responsibility soon after the attack. Russian investigators quickly charged four men. [Russian](#) officials stated they had 11 people in custody, including the four alleged gunmen, and claimed to have found a link between the attackers and Ukraine although Ukraine has denied any involvement.

ISIS-K emerged in 2015 from disaffected fighters of the Pakistani Taliban. The group has been fighting the Taliban in Afghanistan, as well as carrying out targeted attacks in Iran, Russia and Pakistan. According to [reporting](#), the group has "released a flood of anti-Russian propaganda, denouncing the Kremlin for its interventions in Syria and condemning the Taliban for engaging with the Russian authorities decades after the Soviet Union invaded Afghanistan."

According to experts consulted by the Board, tens of millions of Russians watched the video of the attack on state-run media channels, as well as Russian social media platforms. Russian President Vladimir Putin claimed there were links to Ukraine and support from Western intelligence for the attack. [According](#) to a public opinion survey conducted by the Levada Center in Russia from April 18-24, almost all respondents said they knew of the attack and were following the story closely, while half believed that the Ukrainian intelligence services were involved.

According to research commissioned by the Board, the video shared in these cases was circulated widely online, including by Russian and international media accounts. Researchers found some posts on Facebook with the footage and isolated instances of accounts possibly affiliated with or supportive of ISIS celebrating the attack. [Researchers](#) report that social media platforms with less rigorous content moderation contain significantly more perpetrator-generated content.

In 2024, VK, WhatsApp and Telegram were the most widely used platforms in Russia. The government exerts significant control of the media environment, with direct or indirect authority over "all national television networks and most radio and print outlets." Since the invasion of Ukraine, the "government also began restricting access to [a] wide variety of websites, including those of domestic and foreign news outlets. More than 300 media outlets have been forced to suspend their activities." The government also severely restricts reporting access for foreign media outlets and has subjected affiliated journalists to false charges, arrests and prison.

Section 2: User Submissions

The users in all three cases appealed to the Board. In their statements, they explained that they shared the video to warn people in Russia to stay safe. They said that they condemn terrorism, and that Meta should not prevent them from informing people of real events.

Section 3: Meta's Content Policies and Submissions

I. Meta's Content Policies

Dangerous Organizations and Individuals Community Standard

The Dangerous Organizations and Individuals policy rationale states that, in an effort to prevent and disrupt real-world harm, Meta does not allow organizations or individuals that proclaim a violent mission or are engaged in violence to have a presence on its platforms. The Community Standard prohibits "content that glorifies, supports, or represents events that Meta designates as violating violent events," including terrorist attacks. Nor does it allow "(1) glorification, support or representation of the perpetrator(s) of such attacks; (2) perpetrator-generated content relating to such attacks; or (3) *third-party imagery depicting the moment of such attacks on visible victims*," (emphasis added). The Community Standard provides the following examples of violating violent events: "terrorist attacks,

hate events, multiple-victim violence or attempted multiple-victim violence, serial murders, or hate crimes." However, it does not provide specific criteria for designation or a list of designated events.

According to internal guidelines for reviewers, Meta removes imagery depicting the moment of attacks on visible victims "regardless of sharing context."

Violent and Graphic Content Community Standard

The [Violent and Graphic Content](#) policy rationale states that the company understands people "have different sensitivities with regard to graphic and violent imagery," and that Meta removes the most graphic content, also adding a warning label to other graphic content to warn people. This policy allows, with a "Mark as Disturbing" warning screen, "imagery (both videos and still images) depicting a persons' violent death (including their moment of death or the aftermath) or a person experiencing a life threatening event." The policy prohibits such imagery when they depict dismemberment, visible innards, burning or throat slitting.

Newsworthiness Allowance

In certain circumstances, the company will allow content that may violate its policies to remain on the platform if it is "[newsworthy](#) and if keeping it visible is in the public interest." When making the determination, "[Meta will] assess whether that content surfaces an imminent threat to public health or safety, or gives voice to perspectives currently being debated as part of a political process." The analysis is informed by country-specific circumstances, considering the nature of the speech and political structure of the country affected. "For content we allow that may be sensitive or disturbing, we include a warning screen. In these cases, we can also limit the ability to view the content to adults, ages 18 and older. Newsworthy allowance can be 'narrow,' in which an allowance applies to a single piece of content or 'scaled,' which may apply more broadly to something like a phrase."

Meta found all three posts violated its Dangerous Organizations and Individuals policy prohibiting third-party imagery depicting the moment of such attacks on visible victims. Meta finds "removing this content helps to limit copycat behaviors and avoid the spread of content that raises the profile of and may have propaganda value to the perpetrator." Additionally, the company aims to "protect the dignity of any victims who did not consent to being the subject of public curiosity and media attention." According to Meta, as with all policy forums, the company will consider a range of sources in making a decision, including academic research, external stakeholder feedback, and insights from internal policy and operational teams.

Meta also explained that it will allow such violating content under the newsworthiness allowance on a limited basis. However, in these three cases, the company did not apply the allowance as it concluded that the public interest value of permitting the content to be distributed did not outweigh the risk of harm. Meta considered the fact that the footage exposed visible victims and was shared shortly after the attacks. In its view, displaying this footage was not necessary to condemn or raise awareness.

Meta recognizes that removing this kind of content regardless of context "can risk over-enforcement on speech and may limit information and awareness about events of public concern, particularly when coupled with commentary condemning, raising awareness, or neutrally discussing such attacks." The current default approach is that the company configures MMS banks to remove all content that matches banked content, regardless of caption, without applying a strike. The approach prevents the distribution of the offending content without applying a penalty, recognizing that many users may be sharing depictions of a crisis for legitimate reasons or without nefarious motives. Meta conducted a formal policy development process regarding designated violent attack imagery, including videos depicting terrorist attacks.

That process concluded this year, after the content in these three cases was posted. As a result of this process, Meta adopted the following approach: after an event is designated, Meta will remove all violating event imagery (perpetrator-generated or third-party showing moment of attacks on victims) without strikes in all sharing contexts for longer periods than the current protocol. After this period, only imagery shared with glorification, support or representation will be removed and receive a severe strike. The company stated that this approach is the least restrictive means available to mitigate harms to the rights of others, including the right to privacy and protecting the dignity of the victims and their families.

The Board asked Meta questions on whether the company considered the impact in countries with closed media environments of prohibiting all perpetrator and third-party imagery of moment of attacks on visible victims; whether there are policy levers in the Crisis Policy Protocol relevant to designated events; and the outcome of Meta's policy development process on imagery of designated events. Meta responded to all questions.

Section 4: Public Comments

The Oversight Board received six public comments that met the terms for submission. Five of the comments were submitted from the United States and Canada and one from West Africa. To read public comments submitted with consent to publish, click here.

The submissions covered the following themes: risks of overenforcement; use of graphic videos by designated entities and risk of radicalization; the psychological harms from proliferation of graphic content; the challenge of distinguishing between perpetrator-produced and third-party footage; the importance of social media for timely information during crises; the value of such content for documentation by the public, journalists and researchers; the option of age-gated warning screens; and the need to clarify definitions in the Dangerous Organizations and Individuals and Violent and Graphic Content policies.

Section 5: Oversight Board Analysis

The Board analyzed Meta's decisions in these cases against Meta's content policies, values and human rights responsibilities. The Board also assessed the implications of these cases for Meta's broader approach to content governance.

5.1 Compliance With Meta's Content Policies

I. Content Rules

It is clear to the Board that all three posts (two videos and one image) violate Meta's prohibition on "third-party imagery depicting the moment of [designated] attacks on visible victims." Meta designated the March 22 attack in Moscow under its Dangerous Organizations and Individuals policy before the posts were shared. The rule, as set out in the Community Standard and further explained in the internal guidelines, prohibits all such footage of attacks regardless of the context in or the caption with which it is shared (see [Hostages Kidnapped From Israel](#) decision).

The video showed armed individuals shooting unarmed people at close range, with some victims crouching on the ground and others fleeing. The video included audio with gunfire and sounds of people screaming. The third post captured the same event in a still image. While all three posts were either reporting on, raising awareness about or condemning the attacks, Meta does not apply its exception for these purposes under the prohibition on third-party imagery of moment of attacks on visible victims.

However, the majority of the Board finds that all three posts are at the heart of what the newsworthiness allowance aims to protect. The content depicts an event that was front page news worldwide. Each piece of content was shared soon after the attack and included information intended for the public. During this time, when facts about what had happened, who might be responsible and how the Russian government was responding were all subjects of pressing

debate and discussion, the public interest value of this content was especially high. Images and videos such as these allow citizens the world over to form their own impressions of events without having to rely entirely on content filtered through governments, media or other outlets. The Board considers the lack of media freedom in Russia and its impact on access to information relevant to its analysis, given that they underscore the importance of content that can help to facilitate an informed public. The fact that victims are visible, but not identifiable, in all three posts helps to further tilt this content in the direction of the public interest, as weighed against the privacy and dignity interests at stake. For additional analysis and the minority view, relevant to the Board's decision, see the human rights section below.

II. *Transparency*

According to Meta, the company has a set of Crisis Policy Protocol levers to address "over-enforcement as needed in crisis situations." However, it did not use these levers as the attack in Moscow was not designated as a crisis under the protocol. Meta created the Crisis Policy Protocol in response to a recommendation from the Board that the company should develop and publish a policy that governs Meta's response to crises or novel situations ([Former President Trump's Suspension](#), recommendation no. 19). The Board then called on Meta to publish more information about the Crisis Policy Protocol ([Tigray Communication Affairs Bureau,](#) recommendation no. 1). In response, Meta [published](#) this explanation on its Transparency Center but still declined to publicly share the protocol in full. The Board finds the short explanation shared publicly is not sufficient to allow the Board, users and the public to understand the Crisis Policy Protocol. The Board has already stressed the importance of such a protocol for ensuring an effective and consistent response by Meta to crises and conflict situations. A 2022 "[Declaration of principles](#) for content and platform governance in times of crisis" – developed by NGOs Access Now, Article 19, Mnemonic, the Center of Democracy and Technology, JustPeace Labs, Digital Security Lab Ukraine, Center for Democracy and Rule of Law (CEDEM) and the

Myanmar Internet Project – identifies the development of a crisis protocol as a key tool for effective content governance during crisis. The Board and the public are in the dark, however, as to why the Crisis Policy Protocol was not applied in this case, and how the treatment of the content might have differed if it had been. Therefore, greater transparency is necessary about when and how the protocol is used, results of the audits and assessments the company carries out about the effectiveness of the protocol and any changes to policies or systems that address identified shortcomings. In accordance with the UN Guiding Principles on Business and Human Rights (UNGPs), companies should "track the effectiveness of their [mitigation measures]" (Principle 20) and "communicate this externally" (Principle 21). Without such disclosures it is impossible for the Board, the Meta user base or civil society to understand how well the protocol is working or how its efficacy might be enhanced.

5.2 Compliance With Meta's Human Rights Responsibilities

The Board finds that although the posts do violate Meta's Dangerous Organizations and Individuals policy, removing this content was not consistent with Meta's policies, its commitment to the value of voice or its human rights responsibilities.

Freedom of Expression (Article 19 ICCPR)

On March 16, 2021, Meta announced its Corporate Human Rights Policy, in which it outlines its commitment to respecting rights in accordance with the UN Guiding Principles on Business and Human Rights (UNGPs). The UNGPs, endorsed by the UN Human Rights Council in 2011, establish a voluntary framework for the human rights responsibilities of private businesses. These responsibilities mean, among other things, that companies should "avoid infringing on the human rights of others and should address adverse human rights impact with which they are involved," (Principle 11, UNGPs). Companies are expected to: "(a) Avoid causing or contributing to adverse human rights impacts through their own activities, and address such impacts when they occur;

(b) Seek to prevent or mitigate adverse human rights impacts that are directly linked to their operations, products or services by their business relationships, even if they have not contributed to those impacts," (Principle 13, UNGPs).

Meta's content moderation practices can have adverse impacts on the right to freedom of expression. Article 19 of the [International Covenant on Civil and Political Rights](#) (ICCPR) provides broad protection for this right, given its importance to political discourse, and the Human Rights Committee has noted that it also protects expression that may be "deeply offensive," ([General Comment No. 34](#), paras. 11, 13 and 38). When 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). These requirements are often referred to as the "three-part test." The Board uses this framework to interpret Meta's voluntary human rights commitments, in relation both to the individual content decisions under review and to Meta's broader approach to content governance. As the [UN Special Rapporteur on freedom of opinion and expression](#) has stated, although "companies do not have the obligations of Governments, their impact is of a sort that requires them to assess the same kind of questions about protecting their users' right to freedom of expression," ([A/74/486](#), para. 41).

I.   *Legality (Clarity and Accessibility of the Rules)*

The principle of legality requires rules limiting expression to be accessible and clear, formulated with sufficient precision to enable an individual to regulate their conduct accordingly (General Comment No. 34, para. 25). Additionally, these rules "may not confer unfettered discretion for the restriction of freedom of expression on those charged with [their] execution" and must "provide sufficient guidance to those charged with their execution to enable them to ascertain what sorts of expression are properly restricted and what sorts are not," (*Ibid.*). The UN Special Rapporteur on freedom of expression has stated that when applied to private actors' governance of online speech, rules should be clear and specific (A/HRC/38/35, para. 46). People using Meta's platforms should

be able to access and understand the rules and content reviewers should have clear guidance regarding their enforcement.

The Board finds the current placement of the rule on footage of violating violent events under the Dangerous Organizations and Individuals policy likely creates confusion for users. The "We remove" section of the policy states: "We remove Glorification of Tier 1 and Tier 2 entities as well as designated events. For Tier 1 and designated events, we may also remove unclear or contextless references if the user's intent was not clearly indicated." The line specifically explaining the prohibition on perpetrator-generated and third-party imagery (which is a separate policy from the above) appears in the "Policy rationale" and the section marked "Types and tiers of dangerous organizations" under the Community Standard. The language in the "We remove" section implies that condemnation and news reporting is permissible, whereas the language in the other sections (policy rationale and types/tiers) states that perpetrator-generated imagery and third-party imagery of moment of attacks on visible victims is prohibited, and does not specify that Meta will remove such content regardless of the motive or framing with which the content is shared (e.g., condemnation or awareness raising). The placement of the rule and lack of clarity in the scope of applicable exceptions creates unnecessary confusion. Meta should move the rule on footage of designated events under the "We remove" section, creating a new section for violating violent events.

II. *Legitimate Aim*

Meta's Dangerous Organizations and Individuals policy aims to "prevent and disrupt real-world harm." In several decisions, the Board has found that this policy pursues the legitimate aim of protecting the rights of others, such as the right to life (ICCPR, Article 6) and the right to non-discrimination and equality (ICCPR, Articles 2 and 26), because it covers organizations that promote hate, violence and discrimination as well as designated violent events motivated by hate. See [Referring to Designated Dangerous Individuals as "Shaheed,"](#) [Sudan's Rapid Support Forces Video Captive](#), [Hostages Kidnapped from Israel](#) and [Greek](#)

2023 Elections Campaign decisions. Meta's policies also pursue the legitimate aim of protecting the right to privacy of identifiable victims and their families (see Video After Nigeria Church Attack decision).

III. Necessity and Proportionality

Under ICCPR Article 19(3), necessity and proportionality requires that restrictions on expression "must be appropriate to achieve their protective function; they must be the least intrusive instrument amongst those which might achieve their protective function; they must be proportionate to the interest to be protected," (General Comment No. 34, para. 34).

In these three cases, the majority of the Board finds there is no clear and actual risk of these three posts leading to radicalization and incitement. Each post contains clear language condemning the attack, showing solidarity with or concern for the victims, and seeking to inform the public. The videos were posted immediately after the attack, with the caption for the first post explicitly showing support for the victims and indicating that the person who posted the content was doing so to share information to better understand what had happened. The person who posted the second post expressed solidarity with the victims in Russia, condemning the violence. And the third post provided information along with a still image and a brief video, reporting that nobody had yet claimed responsibility and that Ukraine had stated it had nothing to do with the attack; content contradicting propaganda widely disseminated by Russian state media. Suppressing content on matters of vital public concern based upon unsubstantiated fears that it could promote radicalization is not consistent with Meta's free expression responsibilities, especially when the same footage has been viewed by millions of people accompanied by allegations that the attack was partly attributable to Ukraine. The Board takes note of the importance of maintaining access to information during crises and the closed media environment in Russia, where people rely on social media to access information or to raise awareness among international audiences.

Allowing such imagery with a warning screen, under Meta's Violent and Graphic Content Community Standard, provides a less restrictive means of protecting the rights of others (see the less restrictive means analysis in full below). That policy allows, with a Mark as Disturbing warning screen, "imagery (both videos and still images) depicting a persons' violent death (including their moment of death or the aftermath) or a person experiencing a life threatening event."

Additionally, as the Board has previously held, when victims of such violence are identifiable in the image, the content "more directly engages their privacy rights and the rights of their families," (see Video After Nigeria Church Attack decision). In that decision, in which the content showed the gruesome aftermath of a terrorist attack, the majority of the Board decided that removing the content was neither necessary nor proportionate, restoring the post with an age-gated warning screen. The footage at issue in these three posts is not high resolution, and the attackers and people being shot are visible but not easily identifiable. In certain circumstances, removal of content depicting identifiable victims will be the necessary and proportionate measure (e.g., in armed conflict when victims are prisoners of war or hostages subject to special protections under international law). However, in these three cases, given that victims are not easily identifiable, or seen in a humiliating or degrading manner, restoring the posts with an age-gated warning screen is more in line with Meta's human rights responsibilities.

A minority of the Board disagrees and would uphold Meta's decision to remove the three posts from Facebook. The minority agrees that the content in this case, captured by someone at the venue and shared to report on or to condemn an attack, is not likely to incite violence or promote radicalization. However, for the minority, the graphic nature of the footage with sounds of gun fire and victims' screams, and as it shows the moment of attack and, in this case, death of visible if not easily identifiable victims, means the privacy and dignity of the victims and their families make removal necessary. In the aftermath of terrorist attacks, when footage of violence spreads quickly and widely, and can re-traumatize survivors and the families of the deceased, the minority believes that Meta is

justified in prioritizing the privacy and dignity of the victims and their families above the public interest value of allowing citizens access to newsworthy content. For the minority, the newsworthiness of the content counts against it remaining on the platform. The minority maintains that the attack of March 22 was widely covered in Russia as well as by international media. Therefore, in the view of the minority, allowing this footage on Meta's platforms was not necessary to ensure access to information about the attack. Users who wished to comment on the attack or challenge the government's narrative attributing it to Ukraine, could have done so without sharing the most graphic moments of the footage.

The Board understands that in developing and adopting its policy on imagery of terrorist attacks during the recent policy development process, Meta has erred on the side of safety and privacy, adopting reasoning similar to that of the minority on the latter. The company explained there is a risk of adversarial behavior, for example, the repurposing of third-party footage by violent actors, and there are enforcement challenges in terms of moderating content at-scale that mean that a more permissive approach would increase these risks. The company also highlighted the risks to the privacy and dignity of victims of these attacks and their families, when victims are visible. A public comment submitted by the World Jewish Congress highlights similar considerations to those articulated by Meta. Referring to the online proliferation of videos of the October 7, 2023, attack by Hamas on Israel, the submission notes that "in such events, the understanding of who is a 'bystander' or 'third party' is problematic, as many accomplices were filming and distributing terrorist content," (PC-29651).

The Board acknowledges that, in the digital age, videography and photography are tools employed by some terrorists in order to document and glorify their acts. But not all video of attacks involving designated entities is created with this purpose, calibrated to yield this effect or seen as such by viewers. Imagery that is not produced by perpetrators or their supporters is not created for the purpose of glorification or promotion of terrorism. When recorded by a bystander, a victim, an independent journalist or through a CCTV camera, the imagery itself is

not intended to and generally less likely to sensationalize and fetishize violence (i.e., footage recorded through a headcam of the perpetrator is different than footage captured by a CCTV camera or a bystander). It will capture the horror of violence but may not in its presentation trivialize or promote it. While there are risks of imagery of attacks being repurposed to encourage glorification of violence or terrorism and copycat behavior, absent signs of such recasting, a blanket ban overlooks the potential for video documenting violent attacks to trigger sympathy for victims, foster accountability and build public awareness of important events, potentially steering anger or contempt towards the perpetrators, and putting the public on notice about the brutal nature of terrorist groups and movements.

In the policy advisory opinion on Referring to Designated Dangerous Individuals as "Shaheed," the Board noted several UN Security Council resolutions calling on states to address incitement to terrorist acts and raising concerns about the use of the internet by terrorist organizations. See UN Security Council Resolution 1624 (2005), UNSC Resolution 2178 (2014) and UNSC Resolution 2396 (2017). Meta's approach may be understood as an effort to address these concerns. However, as the Board also noted in that policy advisory opinion, the UN Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism has warned against adopting overbroad rules and spoke of the impact of focusing on the content of speech rather than the "causal link or actual risk of the proscribed result occurring," (Report A/HRC/40/52, para 37). See also Joint Declaration on the Internet and on Anti-Terrorism Measures of the UN Special Rapporteur on freedom of expression, the OSCE Representative on freedom of the media and the OAS Special Rapporteur on freedom of expression (2005).

The Board agrees there is a risk that creating exceptions to the policy could lead to underenforcement of content depicting terror attacks, and of footage being reused for malign purposes that Meta will not be able to identify and remove effectively. The Board commends Meta for seeking to address the risk of its

platforms being used by violent actors to recruit and radicalize individuals, and to address the harms to the privacy and dignity of victims. However, as the three posts in these cases demonstrate, images of attacks can serve multiple functions and there are risks to freedom of expression, access to information and public participation from a policy that errs on the side of overenforcement, when less restrictive means are available to enable a more proportionate outcome.

When Meta applies an age-gated warning screen, content is not available to users under the age of 18, other users have to click through to view the content, and the content is then removed from recommendations to users who do not follow the account (see Al-Shifa Hospital and Hostages Kidnapped From Israel decisions). Meta can rely on MMS banks to automatically apply a "Mark as Disturbing" warning screen to all content that contains identified imagery. These measures can mitigate the risks of content going viral or reaching particularly vulnerable or impressionable users who have not sought it out. A warning screen thus lessens the likelihood that the content will provide unintended inspiration for copycat acts. A warning screen does not fully mitigate risks of footage being repurposed by bad actors. However, once the risk of virality is mitigated, Meta has other, more targeted, tools to identify repurposing by bad actors and remove such content from the platform (e.g., internal teams proactively looking for such content and Trusted Partner channels). A more targeted approach will undoubtedly require additional resources. Given the extent of Meta's resources, and the impact on expression and access to information of the current approach, a more targeted approach is warranted.

Images of attacks can communicate and evoke moral outrage, create a sense of solidarity with victims and provide a mechanism for sharing information with those on the ground or international audiences. There are also some indications that there is a greater tendency to help or a stronger emotional response from people when they can see a picture or a video of a specific victim versus when the information is presented through abstract description or mere numbers. In a country with a closed media environment where the government exerts

significant control over what the people see and how information is presented, the accessibility on social media of content with strong public awareness interest and political salience is even more important. The majority concludes that the prohibition and removal of all third-party imagery of attacks on visible but not personally identifiable victims, when shared for news reporting, awareness raising and condemnation is not a necessary nor a proportional measure. When the video/image is perpetrator-generated, shows personally identifiable victims in degrading circumstances or depicts particularly vulnerable victims (e.g., hostages or minors), or lacks a clear awareness-raising, reporting or condemning purpose, it may be appropriate for Meta to err on the side of removal. But a rule prohibiting all third-party imagery of attacks on visible victims, regardless of the reason for and context in which the post is shared, eschews a more proportionate and less restrictive approach, when it is not clear that such a heavy-handed approach is necessary.

Meta should allow, with a "Mark as Disturbing" interstitial, third-party imagery showing moment of attacks on visible but not identifiable victims when shared in news reporting and condemnation contexts. This would be in line with the Dangerous Organizations & Individuals policy rationale, which states: Meta's policies are designed to allow room for … references to designated organizations and individuals in the context of social and political discourse [including] content reporting on, neutrally discussing or condemning dangerous organizations and individuals and their activities." However, given the different types of violent attacks that are eligible for designation and that the context of a given situation may present especially high risks of copycat behavior or malicious use, Meta should utilize expert human review in evaluating specific situations and enforcing the policy exception recommended by the Board.

For the minority, Meta's current policy prohibiting all imagery of designated attacks depicting visible victims is in line with the company's human rights responsibilities and the principles of necessity and proportionality. When graphic footage of an attack depicts visible victims, even where victims are not

easily identifiable, the aim of protecting the right to privacy and dignity of survivors and victims far outweighs the value of voice, in the view of the minority. Even content recorded by a third-party can harm the privacy and dignity of victims and their families. And applying a warning screen to content showing the death of a person, as the majority recommends, does not protect the privacy or dignity of the victims or their families from those who opt to move past the screen. As the minority in the [Video After Nigeria Church Attack](#) decision stated, when terrorist attacks occur, videos of this nature frequently go viral, compounding the harm and increasing risk of re-traumatization. Meta should act quickly and at-scale in order to prevent and mitigate the harms to the human rights of victims, survivors and their families. This also serves the broader public purpose of countering the widespread terror that perpetrators of such attacks seek to instill, knowing that social media will amplify their psychological impacts. Additionally, as the Board has indicated in prior decisions, Meta could ease the burden on users and mitigate risks to privacy by providing users with more specific instructions or access within its products to, for instance, face-blurring tools for videos depicting visible victims of violence (see [News Documentary on Child Abuse in Pakistan](#) decision).

Section 6: The Oversight Board's Decision

The Oversight Board overturns Meta's decisions to take down the three posts, requiring the content to be restored with "Mark as Disturbing" warning screens.

Section 7: Recommendations

<u>Content Policy</u>

1. To ensure its Dangerous Organizations and Individuals Community Standard is tailored to advance its aims, Meta should allow, with a "Mark as Disturbing" warning screen, third-party imagery of a designated event showing the moment of attacks on visible but not personally identifiable victims when shared in news reporting, condemnation and awareness-raising contexts.

The Board will consider this recommendation implemented when Meta updates the public-facing Dangerous Organizations and Individuals Community Standard in accordance with the above.

2. To ensure clarity, Meta should include a rule under the "We remove" section of the Dangerous Organizations and Individuals Community Standard and move the explanation of how Meta treats content depicting designated events out of the policy rationale section and into this section.

   The Board will consider this recommendation implemented when Meta updates the public-facing Dangerous Organizations and Individuals Community Standard moving the rule on footage of designated events to the "We remove" section of the policy.

*Procedural Note:

- The Oversight Board's decisions are made by panels of five Members and approved by a majority vote of the full Board. Board decisions do not necessarily represent the views of all Members.

- Under its [Charter](), the Oversight Board may review appeals from users whose content Meta removed, appeals from users who reported content that Meta left up, and decisions that Meta refers to it (Charter Article 2, Section 1). The Board has binding authority to uphold or overturn Meta's content decisions (Charter Article 3, Section 5; Charter Article 4). The Board may issue non-binding recommendations that Meta is required to respond to (Charter Article 3, Section 4; Article 4). When Meta commits to act on recommendations, the Board monitors their implementation.

- For this case decision, independent research was commissioned on behalf of the Board. The Board was assisted by Duco Advisors, an advisory firm focusing on the intersection of geopolitics, trust and safety, and technology. Memetica, a digital investigations group providing risk advisory and threat intelligence services to mitigate online harms, also provided research.