

IG-7THR3SI1

## Case Decision 2020-004-IG-UA

### Case Summary

The Oversight Board has overturned Facebook’s decision to remove a post on Instagram. After the Board selected this case, Facebook restored the content. Facebook’s automated systems originally removed the post for violating the company’s Community Standard on Adult Nudity and Sexual Activity. The Board found that the post was allowed under a policy exception for “breast cancer awareness” and Facebook’s automated moderation in this case raises important human rights concerns.

### About the case

In October 2020, a user in Brazil posted a picture to Instagram with a title in Portuguese indicating that it was to raise awareness of signs of breast cancer. The image was pink, in line with “Pink October,” an international campaign to raise awareness of this disease. Eight photographs within the picture showed breast cancer symptoms with corresponding descriptions. Five of them included visible and uncovered female nipples, while the remaining three photographs included female breasts, with the nipples either out of shot or covered by a hand. The post was removed by an automated system enforcing Facebook’s Community Standard on Adult Nudity and Sexual Activity. After the Board selected the case, Facebook determined this was an error and restored the post.

### Key findings

In its response, Facebook claimed that the Board should decline to hear this case. The company argued that, having restored the post, there was no longer disagreement between the user and Facebook that the content should stay up, making this case moot.

The Board rejects Facebook’s argument. The need for disagreement applies only at the moment the user exhausts Facebook’s internal appeal process. As the user and Facebook disagreed at that time, the Board can hear the case.

Facebook’s decision to restore the content also does not make this case moot, as the company claims. On top of making binding decisions on whether to restore pieces of content, the Board also offers users a full explanation for why their post was removed. The incorrect removal of this post indicates the lack of proper human oversight which raises human rights concerns. The detection and removal of this post was entirely automated. Facebook’s automated systems failed to recognize the words “Breast Cancer,” which appeared on the image in Portuguese, and the post was removed in error. As Facebook’s rules treat male and female nipples differently, using inaccurate automation to enforce these rules disproportionately affects women’s freedom

of expression. Enforcement which relies solely on automation without adequate human oversight also interferes with freedom of expression.

In this case, the user was told that the post violated Instagram's Community Guidelines, implying that sharing photos of uncovered female nipples to raise breast cancer awareness is not allowed. However, Facebook's Community Standard on Adult Nudity and Sexual Activity, expressly allows nudity when the user seeks to "raise awareness about a cause or educational or medical reasons" and specifically permits uncovered female nipples to advance "breast cancer awareness." As Facebook's Community Standards apply to Instagram, the user's post is covered by the exception above. Hence, Facebook's removal of the content was inconsistent with its Community Standards.

### **The Oversight Board's decision**

The Oversight Board overturns Facebook's original decision to remove the content and requires that the post be restored. The Board notes that Facebook has already taken action to this effect.

The Board recommends that Facebook:

- Inform users when automated enforcement is used to moderate their content, ensure that users can appeal automated decisions to a human being in certain cases, and improve automated detection of images with text-overlay so that posts raising awareness of breast cancer symptoms are not wrongly flagged for review. Facebook should also improve its transparency reporting on its use of automated enforcement.
- Revise Instagram's Community Guidelines to specify that female nipples can be shown to raise breast cancer awareness and clarify that where there are inconsistencies between Instagram's Community Guidelines and Facebook's Community Standards, the latter take precedence.

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

## **Full Case Decision**

### **1. Decision Summary**

The Oversight Board has overturned Facebook's original decision to take down the content, noting that Facebook restored the post after the Board decided to hear this case. Facebook's decision to reinstate the content does not exclude the Board's authority to hear the case.

The Board found that the content was allowed under a policy exception for "breast cancer awareness" in Facebook's Community Standard on Adult Nudity and Sexual Activity.

The Board has issued a policy advisory statement on the relationship between content policies on Instagram and Facebook, as well as on the use of automation in content moderation and the transparency of these practices.

## 2. Case Description

In October 2020, a user in Brazil posted a picture to Instagram with a title in Portuguese indicating that it was to raise awareness of signs of breast cancer. The image was pink, in line with “Pink October,” an international campaign popular in Brazil to raise breast cancer awareness. Eight photographs within a single picture post showed breast cancer symptoms with corresponding descriptions such as “ripples,” “clusters,” and “wounds,” underneath. Five of the photographs included visible and uncovered female nipples. The remaining three photographs included female breasts, with the nipples either out of shot or covered by a hand. The user shared no additional commentary with the post.

The post was detected and removed by a machine learning classifier trained to identify nudity in photos, enforcing Facebook’s Community Standards on Adult Nudity and Sexual Activity, which also applies on Instagram.

The user appealed this decision to Facebook. In public statements, Facebook has previously said that it could not always offer users the option to appeal due to a temporary reduction in its review capacity as a result of COVID-19. Moreover, Facebook has stated that not all appeals will receive human review.

The user submitted a request for review to the Board and the Board decided to take the case. Following the Board’s selection and assignment of the case to a panel, Facebook reversed its original removal decision and restored the post in December 2020. Facebook claims the original decision to remove the post was automated and subsequently identified as an enforcement error. However, Facebook only became aware of the error after it was brought to the company’s attention through the Board’s processes.

## 3. Authority and Scope

The Board has authority to review Facebook’s decision under Article 2 (Authority to Review) of [the Board’s Charter](#) and may uphold or reverse that decision under Article 3, Section 5 (Procedures for Review: Resolution of the Charter). Facebook has not presented reasons for the content to be excluded in accordance with Article 2, Section 1.2.1 (Content Not Available for Board Review) of [the Board’s Bylaws](#), nor has Facebook indicated that it considers the case to be ineligible under Article 2, Section 1.2.2 (Legal Obligations) of the Bylaws.

While Facebook publicly welcomed the Board’s review of this case, Facebook proposed that the Board should decline to hear the case in its filings before the Board because the issue is now moot.

Facebook argues that, having restored the content, there is no disagreement that it should stay on Instagram and that this is a requirement for a case to be heard, according to Article 2, Section 1 of the Board’s Charter:

*in instances where people disagree with the outcome of Facebook’s decision and have exhausted appeals, a request for review can be submitted to the Board.*

The Board disagrees, and interprets the Charter to only require disagreement between the user and Facebook at the moment the user exhausts Facebook's internal process. This requirement has been met. The Board's review process is separate from, and not an extension of Facebook's internal appeals process. For Facebook to correct errors the Board brings to its attention *and thereby exclude cases from review* would integrate the Board inappropriately to Facebook's internal process and undermine the Board's independence.

While Facebook reversed its decision and restored the content, irreversible harm still occurred in this case. Facebook's decision to restore the content in early December 2020 did not make up for the fact that the user's post was removed for the entire "pink month" campaign in October 2020.

Restoring the content in this case is not the only purpose of the remedy the Board offers. Under Article 4 (Implementation) of the Board's Charter, and Article 2, Section 2.3.1 (Implementation of Board Decisions) of the Bylaws, Facebook is committed to take action on "identical content with parallel context". Thus, the impact of the Board taking decisions extends far beyond the content in this case.

Moreover, a full decision, even where Facebook complies with its outcome in advance, is important. The Board's process offers users an opportunity to be heard and to receive a full explanation for why their content was wrongly removed. Where content removal is performed entirely through automation, the content policies are essentially embedded into code and may be considered inseparable from it and self-enforcing. Hearing the case allows the Board to issue policy advisory statements on how Facebook's content moderation practices are applied, including with the use of automation.

For these reasons, the Board finds that its authority to review this case is not affected by Facebook's decision to restore the content after the Board selected the case. The Board proceeds with its review of the original decision to remove the content.

#### **4. Relevant Standards**

The Board considered the following standards in its decision:

##### *1. Facebook's Content Policies:*

The Community Standard on Adult Nudity and Sexual Activity's policy rationale states that Facebook aims to restrict the display of nudity or sexual activity because some people "may be sensitive to this type of content" and "to prevent the sharing of non-consensual or underage content." Users should not "post images of real nude adults, where nudity is defined as [...] uncovered female nipples except in the context of [...] health-related situations (for example, post-mastectomy, breast cancer awareness [...])."

Instagram's Community Guidelines state a general ban on uncovered female nipples, specifying some health-related exceptions, but do not specifically include "breast cancer awareness." The Community Guidelines link to Facebook's Community Standards.

## II. Facebook's Values:

The Facebook values relevant to this case are outlined in the introduction to the Community Standards. The first is "Voice", which is described as "paramount":

*The goal of our Community Standards has always been to create a place for expression and give people a voice. [...] We want people to be able to talk openly about the issues that matter to them, even if some may disagree or find them objectionable.*

Facebook limits "Voice" in service of four values. The Board considers that two of these values are relevant to this decision:

Safety: *We are committed to making Facebook a safe place. Expression that threatens people has the potential to intimidate, exclude or silence others and isn't allowed on Facebook.*

Privacy: *We are committed to protecting personal privacy and information. Privacy gives people the freedom to be themselves, and to choose how and when to share on Facebook and to connect more easily.*

## III. 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. The Board's analysis in this case was informed by UN treaty provisions and the authoritative guidance of UN human rights mechanisms, including the following:

- The right to freedom of expression: International Covenant on Civil and Political Rights (ICCPR), Article 19; General Comment No. 34, Human Rights Committee (2011) ([General Comment 34](#)); UN Special Rapporteur on freedom of opinion and expression, reports: [A/HRC/38/35](#) (2018); [A/73/348](#) (2018); and [A/HRC/44/49](#) (2020).
- The right to health: International Covenant on Economic, Social and Cultural Rights (ICESCR), Article 12; General Comment No. 14, the Committee on Economic, Social and Cultural Rights, [E/C.12/2000/4](#) (2000).
- The right to effective remedy: ICCPR, Article 2; General Comment No. 31, the Human Rights Committee, [CCPR/C/21/Rev.1/Add. 13](#) (2004).
- The right to privacy: ICCPR, Article 17.
- The right to non-discrimination: ICCPR Article 2 (para. 1); Convention on the Elimination of All Forms of Discrimination against Women (CEDAW), Article 1.
- The rights of the child: Convention on the Rights of the Child (CRC), Article 6; General Comment No. 13, the Committee on the Rights of the Child, [CRC/C/GC/13](#) (2011).

## 5. User Statement

The user states that the content was posted as part of the national “Pink October” campaign for breast cancer prevention. It shows some of the main signs of breast cancer, which the user says are essential for early detection of this disease and save lives.

## **6. Explanation of Facebook’s Decision**

Facebook clarified that its original decision to remove the content was a mistake. The company explained to the Board that the Community Standards apply to Instagram. While the Community Standards generally prohibit uncovered and visible female nipples, they are allowed for “educational or medical purposes,” including for breast cancer awareness. Facebook restored the content because it fell within this exception.

Facebook claims that allowing this content on the platform is important for its values of “Voice” and “Safety.” The company states that the detection and original enforcement against this content was entirely automated. That automated process failed to determine that the content had clear “educational or medical purposes.” Facebook also claims that it is not relevant to the Board’s consideration of the case whether the content was removed through an automated process, or whether there was an internal review to a human moderator. Facebook would like the Board to focus on the outcome of enforcement, and not the method.

## **7. Third party submissions**

The Oversight Board considered 24 public comments for this case: eight from Europe; five from Latin American and Caribbean, and 11 from the United States and Canada. Seven were submitted on behalf of an organization. One comment was submitted without consent to publish.

The submissions covered the following themes: whether the post complied with Facebook’s Community Standards and values; the importance of breast cancer awareness in early diagnosis; critique on the over-sexualization and censorship of female nipples compared to male nipples; Facebook’s influence on society; over-enforcement due to automated content moderation, as well as feedback for improving the public comment process.

## **8. Oversight Board Analysis**

### **8.1 Compliance with Facebook content policies**

Facebook’s decision to remove the user’s Instagram post did not comply with the company’s content policies.

According to Facebook, the Community Standards operate across the company’s products, including Instagram. The user in this case was notified that the content violated Instagram’s Community Guidelines, which were quoted to the user. The differences between these rules warrants separate analysis.

#### *1. Instagram’s Community Guidelines*

The “short” Community Guidelines summarize Instagram’s rules as: “Respect everyone on Instagram, don’t spam people or post nudity.” Taken on their own, these imply that the user’s post violates Instagram’s rules.

The “long” Community Guidelines go into more detail. Under the heading “post photos and videos that are appropriate for a diverse audience,” they state:

*[F]or a variety of reasons, we don’t allow nudity on Instagram [...] It also includes some photos of female nipples, but photos of post-mastectomy scarring and women actively breastfeeding are allowed.*

This explanation does not expressly allow photos of uncovered female nipples to raise breast cancer awareness. While Instagram’s Community Guidelines include a hyperlink to Facebook’s Community Standard on Adult Nudity and Sexual Activity, the relationship between the two sets of rules, including which takes precedence, is not explained.

## *II. Facebook’s Community Standards*

The Community Standard on Adult Nudity and Sexual Activity, under Objectionable Content, states that the display of adult nudity, defined to include “uncovered female nipples,” as well as sexual activity, is generally restricted on the platform. Two reasons are given for this position: “some people in our community may be sensitive to this type of content” and “to prevent the sharing of non-consensual or underage content.”

The Community Standard specifies that consensual adult nudity is allowed when the user clearly indicates the content is “to raise awareness about a cause or for educational or medical reasons.” The “do not post” section of the Community Standard lists “breast cancer awareness” as an example of a health-related situation where showing uncovered female nipples is permitted.

The Board finds that the user’s post, while depicting uncovered female nipples, falls squarely within the health-related exception for raising breast cancer awareness. Accepting Facebook’s explanation that the Community Standards operate on Instagram, the Board finds that the user’s post complies with them.

Facebook’s decision to remove the content was therefore inconsistent with the Community Standards. The Board acknowledges Facebook has agreed with this conclusion.

## **8.2 Compliance with Facebook Values**

Facebook’s values are outlined in the introduction to the Community Standards but are not directly referenced in Instagram’s Community Guidelines.

Facebook’s decision to remove the user’s content did not comply with Facebook’s values. The value of “Voice” clearly includes discussions on health-related matters and is especially valuable for raising awareness of the symptoms of breast cancer. Images of early breast cancer

symptoms are especially valuable to make medical information more accessible. Sharing this information contributes to the “Safety” of all people vulnerable to this disease. There is no indication that the pictures included any non-consensual imagery. Therefore, “Voice” was not displaced by “Safety” and “Privacy” in this case.

### **8.3 Compliance with international human rights standards**

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

Facebook’s decision to remove the post also did not comply with international human rights standards on freedom of expression (Article 19, ICCPR). Health-related information is particularly important (A/HRC/44/49, para. 6) and is additionally protected as part of the right to health (Article 12, IESCR; E/C.12/2000/4, para. 11). In Brazil, where awareness raising campaigns are crucial to promote early diagnosis of breast cancer, the Board emphasizes the connection between these two rights.

This right to freedom of expression is not absolute. When restricting freedom of expression, Facebook should meet the requirements of legality, legitimate aim, and necessity and proportionality. Facebook’s removal of the content failed the first and third parts of this test.

#### **a. Legality**

Any rules restricting expression must be clear, precise, and publicly accessible (General Comment 34, para. 25). Facebook’s Community Standards permit female nipples in the context of raising breast cancer awareness, while Instagram’s Community Guidelines only mention post-mastectomy scarring. That Facebook’s Community Standards take precedence over the Community Guidelines is also not communicated to Instagram users. This inconsistency and lack of clarity is compounded by removal notices to users that solely reference the Community Guidelines. Facebook’s rules in this area therefore fail the legality test.

#### **b. Legitimate aim**

Any restriction on freedom of expression must be for a legitimate aim, which are listed in Article 19, para. 3 of the ICCPR. Facebook claims its Adult Nudity and Sexual Activity Community Standard helps prevent the sharing of child abuse images and non-consensual intimate images on Facebook and Instagram. The Board notes that both content categories are prohibited under separate Community Standards and are not subject to the exceptions that apply to consensual adult nudity. These aims are consistent with restricting freedom of expression under international human rights law to protect “the rights of others” (Article 19, para. 3, ICCPR). These include the right to privacy of victims of non-consensual intimate image sharing (Article 17 ICCPR), and the rights of the child to life and development (Article 6, CRC), which are threatened in cases of sexual exploitation (CRC/C/GC/13, para. 62).

#### **c. Necessity and proportionality**

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; they must be proportionate to the interest to be protected” (General Comment 34, para. 34).

The Board finds that removing information that serves a public interest without cause cannot be proportionate.

The Board is concerned that the content was wrongfully removed by an automated enforcement system and potentially without human review or appeal. This reflects the limitations of automated technologies to understand context and grasp the complexity of human communication for content moderation (UN Special Rapporteur on freedom of expression, A/73/348, para. 15). In this case, these technologies failed to recognize the words “Breast Cancer” that appear at the top left of the image in Portuguese. The Board accepts automated technologies are essential to the detection of potentially violating content. However, enforcement which relies solely on automation, in particular when using technologies that have a limited ability to understand context, leads to over-enforcement that disproportionately interferes with user expression.

The Board recognizes that automated enforcement may be needed to swiftly remove non-consensual intimate images and child abuse images, in order to avoid immediate and irreparable harm. However, when content is removed to safeguard against these harms, the action should be premised on the applicable policies on sexual exploitation, and users notified that their content was removed for these purposes. Regardless, automated removals should be subject to both an internal audit procedure explained under section 9.2 (I) and appeal to human review should be offered (A/73/348, para. 70), allowing enforcement mistakes to be repaired.

Automated content moderation without necessary safeguards is not a proportionate way for Facebook to address violating forms of adult nudity.

#### d. Equality and non-discrimination

Any restrictions on expression must respect the principle of equality and non-discrimination (General Comment 34, paras. 26 and 32). Several public comments argued Facebook’s policies on adult nudity discriminate against women.

Given that Facebook’s rules treat male and female nipples differently, the reliance on inaccurate automation to enforce those rules will likely have a disproportionate impact on women, thereby raising discrimination concerns (Article 1 CEDAW; Article 2 ICCPR). In Brazil, and in many other countries, awareness raising of breast cancer symptoms is a matter of critical importance. As such, Facebook’s actions jeopardize not only women’s right to freedom of expression but also their right to health.

#### *II. Right to remedy (Article 2 ICCPR)*

The Board welcomes that Facebook restored the content. However, the negative impacts of that error could not be fully reversed. The post, intended for breast cancer awareness month in October, was only restored in early December. Restoring the content did not make this case moot: as the Board had selected this case, the user had a right to be heard and to receive a fully reasoned decision.

The UN Special Rapporteur on freedom of opinion and expression identified the responsibility to provide remedy as one of the most relevant aspects of the UNGPs as they relate to business enterprises that engage in content moderation (A/HRC/38/35, para. 11). Facebook's over-reliance on automated enforcement, if there was no appeal, failed to respect the user's right to an effective remedy (Article 2, ICCPR; CCPR/C/21/Rev.1/Add. 13, para. 15) or meet its responsibilities under the UN Guiding Principles (Principles 29 and 31). The Board is especially concerned that Facebook does not inform users when their content is enforced against through automation, and that appeal to human review might not be available in all cases. This reflects a broader concern at Facebook's lack of transparency on its use of automated enforcement, and circumstances where internal appeal might not be available.

## **9. Oversight Board Decision**

### **9.1 Content Decision**

The Oversight Board overturns Facebook's original decision to take down the content, requiring the post to be left up. The Board notes Facebook has already taken action to this effect.

### **9.2 Policy Advisory Statement**

#### *1. Automation in enforcement, transparency and the right to effective remedy*

The Board recommends that Facebook:

- Improve the *automated detection* of images with text-overlay to ensure that posts raising awareness of breast cancer symptoms are not wrongly flagged for review.
- Ensure that users are always notified of the reasons for the enforcement of content policies against them, providing the specific rule within the Community Standard Facebook based its decision on.
- Inform users when automation is used to take enforcement action against their content, including accessible descriptions of what this means.
- Ensure users can appeal decisions taken by automated systems to human review when their content is found to have violated Facebook's Community Standard on Adult Nudity and Sexual Activity. Where Facebook is seeking to prevent child sexual exploitation or the dissemination of non-consensual intimate images, it should enforce based on its Community Standards on Sexual Exploitation of Adults and Child Sexual Exploitation, Abuse and Nudity, rather than rely on over-enforcing policies on adult nudity. Appeals should still be available in these cases, so incorrect removals of permitted consensual adult nudity can be reversed.

- Implement an internal audit procedure to continuously analyze a statistically representative sample of automated content removal decisions to reverse and learn from enforcement mistakes.
- Expand transparency reporting to disclose data on the number of automated removal decisions per Community Standard, and the proportion of those decisions subsequently reversed following human review.

These recommendations should not be implemented in a way which would undermine content moderators' right to health during the COVID-19 pandemic.

## *II. The relationship between the Community Standards and the Community Guidelines:*

The Board recommends that Facebook:

- Revise the "short" explanation of the Instagram Community Guidelines to clarify that the ban on adult nudity is not absolute;
- Revise the "long" explanation of the Instagram Community Guidelines to clarify that visible female nipples can be shown to raise breast cancer awareness;
- Clarify that the Instagram Community Guidelines are interpreted in line with the Facebook Community Standards, and where there are inconsistencies the latter take precedence.

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

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

## **Policies Cited**

### **Adult Nudity and Sexual Activity**