***The Case of Post in Polish targeting trans people***

**Case Analysis**

***Meta-Data*:**

* **Case Number**:
* **Date of decision**:
* **Featured case**: N/A
* **Region**:
* **Country**:
* **Type of expression**:
* **Judicial Body**:
* **Type of law**:
* **Main Themes**:
* **Outcome**:
* **Status**:
* **Tags**:

***Analysis:***

* **Summary and Outcome**:

On January 16, 2024, the Oversight Board overturned Meta’s decision not to remove a post depicting curtains bearing the colors of the Trans Pride flag with text overlay saying “Curtains that hang themselves", and, "spring cleaning <3" from a user whose biography stated, “I am a transphobe”. The Board noted that Meta’s practices in this case depicted challenges in enforcement as it found the policies to be clear and precise. The Board found the content violating of the Hate Speech policy as it was violent speech targeting transgender people and violating the Suicide and Self-injury policy as it encouraged and celebrated the high suicide rates of transgender people. The Board recommended Meta modify its internal guidance for reviewers to clarify that flag-based visuals to depict of gender identity even without any human figures.

\*The Oversight Board is a separate entity from Meta and will provide its independent judgment on both individual cases and questions of policy. Both the Board and its administration are funded by an independent trust. The Board has the authority to decide whether Facebook and Instagram should allow or remove content. These decisions are binding, unless implementing them could violate the law. The Board can also choose to issue recommendations on the company’s content policies.

* **Facts**:

In April 2023, a polish Facebook user posted an image of a curtain in the colors of the transgender flag with Polish overlay text stating. “New technology. Curtains that hang themselves” and “spring cleaning <3”. The post received less than 50 reactions, with the “Haha” reaction being the most used. The user described themselves as a transphobe in their biography.

The LGBTQIA+ community faces high levels of persecution in Poland. The community has been increasingly targeted with discriminatory speech from elected officials as well as enacting anti-LGBTIA+ legislations, such as “LGBT-free zones” in 2019. As a result 51% of LGBTQIA+ people in Poland avoid certain locations for fear of being assaulted. Transgender people have been specifically targeted, for example the Chairman of Poland’s ruling party referred to transgender people as “abnormal” and the Minister of Justice asked the Supreme Court to consider requiring the permission of transgender people’s children and spouses in addition to their parents if they want to access Legal Gender Recognition. Transgender people are physically and sexually attacked double that of other LGBTQIA+ groups in Poland.

Moreover, linguistic experts explained that the phrase “curtains that hang themselves” with the curtains being in transgender flag colors was a play on words to say, “to commit suicide by hanging” and an implicit transphobic slur. As for the “spring cleaning” phrase, experts concluded that in certain contexts it refers to getting rid of all unwanted items and/or people.

Between April and May 2023, 11 different users reported the content 12 times. 10 of these reports were not prioritized for human review due to “low severity and virality scores. However, two of the reports under the Suicide and Self-injury policy were sent for human reviewers who deemed the content non-violating and did not escalate it further. None of the reports made under the Hate Speech policy were sent for human review.

Three users appealed Meta’s decision to keep the content on Facebook. Only one appeal was reviewed by a human moderator who upheld the decision that the content did not violate Meta’s Suicide and Self-injury policy. The other two appeals, made under the Hate Speech policy were not sent for human review as Meta deduplicate multiple reports on the same piece of content to strengthen consistency in decisions and enforcement actions.

One of the users who originally reported the content appealed the decision to the Oversight Board. After the Board selected the case, Meta determined that the content violated the Hate Speech policy, and the Suicide and the Self-injury policy and removed the post. Moreover, Meta disabled the author’s account in August 2023 due to accumulation of several violations of Facebook’s Community Standards.

**Decision Overview**:

The main issue before the Board was whether Meta’s original decision to keep a post targeting transgender people with speech advocating for the suicide of members of the group was compatible with Meta’s content policies and human rights obligations.

In their appeal to the Board, the reporting user noted that the person who posted the content had a history of harassing transgender people online and had created a new account after being suspended from Facebook. Moreover, the user emphasized the high rate of suicide in the transgender community.

Meta eventually removed the post under Tier 1 of the Hate Speech policy as it found the content to be violent speech targeting people based on a protected characteristic. The policy’s internal guidelines instruct reviewers to remove violent content that includes calls to action, expressions of intent to cause harm, or statements that endorse or support causing death, disease, or injury, whether in written or visual form. Furthermore, the guidelines instruct reviewers to take in consideration all visual element in deciding whether the content targets a group of people based on a protected characteristic.

Meta noted that the previous assessments of the content as non-violating aligned with a strict application of the internal guidelines. Meta elaborated that the curtains resembling the Trans Pride flag would be interpreted as an attack on a flag, concept or institution rather than a group of people. However, Meta reached its subsequent decision of removal due to the phrase “curtains which hang themselves” which implicitly referred to the suicide rates in the transgender community.

After revising the Hate Speech policy, Meta transferred the rule banning violent speech towards groups with protected characteristics to the Violence and Incitement policy, and informed the Board that the content in this case was still violating.

Furthermore, Meta explained that the user’s biography, “I am a transphobe,” violated the Hate Speech policy as it admitted to intolerance based on a protected characteristic. This assessment further helped clarify the user’s intent in posting the content at hand.

In response to the Board’s question whether the content violated the Suicide and Self-Injury policy, Meta confirmed that the content was violating by encouraging suicide. Additionally, Meta clarified that the policy doesn’t differentiate between content encouraging suicide targeted at a specific person or a group of people.

The Board asked 13 questions related to Meta’s content moderation concerning transgender and LGBTQIA+ topics, focusing on the link between Hate Speech and Suicide and Self-Injury Community Standards, as well as the process used by moderators to evaluate "humor" and "satire" in the context of hate speech, the significance of "virality" and "severity" ratings in determining which content is prioritized for human revie is outlined, and Meta's strategies for prioritizing content reported multiple times by users for human review. Meta responded comprehensively to all 13 questions.

*Compliance with Meta’s Content Policies*

1. Content Rules

Hate Speech

The Board found the content in this case in violation of the Hate Speech policy as it included a call for death by suicide of a group based on a protected characteristic. The Board agreed that the reference to hanging in the post targeted transgender people and not a concept. Furthermore, the Board considered the broader context, and online and offline harms members of the LGBTQIA+ community face in Poland. The post was advocating for the suicide of transgender individuals which promoted hate and exclusion, potentially leading to physical harm. The post, which used dehumanizing language and imagery, was seen as exacerbating the mental health crisis among the transgender community, who already face higher risks of suicidal behavior. The inclusion of the transgender flag and specific references made it clear that the post targeted transgender people, with phrases like "spring cleaning" and a heart emoji indicating support for their death. The Board concluded that the content violated the platform's Hate Speech policy, which prohibits expressions suggesting that a protected group should not exist.

The Board noted that the Hate Speech policy and its internal guideline should be more responsive to malign creativity which refers to the employment of ambiguous terminology, recurring visual and textual memes that depend on context, and various strategies are used to evade detection on social media platforms. The Board saw that malign creativity was present in the case at hand through the user’s usage of coded references to suicide, “curtains that hang themselves” and “spring cleaning”, alongside visual depiction of the transgender flag to encourage suicide.

The Board drew similarities between this case and the *“*[*Armenians in Azerbaijan*](https://globalfreedomofexpression.columbia.edu/cases/oversight-board-case-of-armenians-in-azerbaijan/)*”* case where the Board highlighted the importance of context in determining whether speech targeted people based on a protected characteristic. While the Board recognized that the context of war prevailed in the *“Armenians in Azerbaijan”* case, it underlined the harm and threats transgender people suffer in Poland.

The Board expressed concern than Meta’s reviewers did not consider these contextual clues within the content which is why they found the content non-violating. While the Board recommended revisions to the internal guidelines of the Hate Speech policy, it underlined that the post violated the policy as it was written at the time of posting as both statements present in the post supported death of transgender people by suicide in addition to the user’s self-identification as a transphobe in their biography. The user’s biography, in itself, violated the Tier 2 of the Hate Speech policy which prohibits self-admissions to intolerance based on protected characteristics. The Board stressed Meta must improve the accuracy of enforcement of its policies on hate speech towards the LGBTQIA+ community, especially when posts include image or text constituting malign creativity.

The Board expressed further concern over Meta’s statement that the original decision aligned with a strict application of internal guidelines. The Board saw that such statement reflected that the internal guideline failed to sufficiently address the way text and images combine in a social media post to target a group identified by their gender identity disallowing reviewers from reaching the correct enforcement decisions. The Board recommended that Meta modify its guidance to ensure visual depictions of gender identity were analyzed when assessing content for attacks.

Suicide and Self-injury

The Board found the content violating of the Suicide and Self-injury policy which prohibits content promoting or encouraging suicide or self-injury. The internal guidelines further define promotion as “speaking positively of”. The Board agreed with Meta’s subsequent conclusion that the content in this case encourages suicide of a group based on their protected-characteristic.

Additionally, the Board noted that the Suicide and Self-injury policy should explicitly prohibit content encouraging the suicide of an identifiable group of people as opposed to a person in that group, rather than not differentiating between these two forms of content, given the challenges reviewers faced in identifying the content in this case as violating. Furthermore, Meta should clarify this on its public-facing policy and internal guidelines to reviewers.

1. Enforcement Action

The Board found that Meta’s automated review prioritization systems significantly impacted enforcement in this case as 10 out of the 12 user reports of the post were closed by the automated systems. Furthermore, two out of the three users appealing against Meta’s decisions had their appeals closed automatically. The Board raised concern over Meta’s policies being adequately enforced.

The Board noted that majority of user reports were closed due to Meta’s practice with multiple reports on the same piece of content. The Board acknowledged that deduplication was a reasonable practice for content moderation at scale, however, there must be more pressure on the initial decision on a report as it would determine the fate of all other reports grouped with it.

The Board stressed that it was crucial for Meta to improve its automated systems that enforce content policies and prioritize content for review, especially for content that might impact the LGBTQIA+ community to understand coded language and context-based images. The Board highlighted the user’s biography as a relevant signal to determine the severity of the content and/or to take an enforcement action. This signal could supplement other behavioral and analyses Meta employed to assess potentially violating content.

Furthermore, the Board expressed concern over Meta’s approach to LGTBQIA+ community-related content under which reviewers assessing appeals seem to have the same level of expertise as those conducting the first assessment. The Board underlined the importance of Meta’s investment in developing and training classifiers to prioritize hate speech affecting the LGBTQIA+ community for review. The Board advised Meta to extend its recommendations to improve enforcement against hateful content affecting all protected-characteristic groups.

Moreover, the Board noted that the challenges present in this case were a result of enforcement of the policies rather than the language of the policies. The Board highlighted five indicators of harmful content in this case: (1) the post's references to "self-hanging curtains"; (2) the post's reference to "spring cleaning <3"; (3) the user's self-description as a "transphobe" in a country context where high levels of hostility towards the LGBTQIA+ community are reported; (4) the number of user reports and appeals on the content; and (5) the number of reports and appeals relative to the virality of the content. The Board raised concern over Meta missing these indicators and noted this suggested the policies were underenforced.

*Compliance with Meta’s human-rights responsibilities*

1. Legality (Clarity and Accessibility of Rules)

In international human rights law, the principle of legality requires rules to be clear to those charged with implementing them, and those subject to them, and publicly accessible. The Board found Meta’s prohibition on "violent speech or support" against groups with protected characteristics, whether in writing or visually, on statements claiming that a protected characteristic should not exist, and on speech advocating or inciting suicide and self-harm, were adequately defined.

However, the Board noted that Meta should improve the enforcement accuracy of the policies through clarifying their guidance to reviewers. The Board stated that Meta should clarify that visual depictions of gender identity, such as through flags, should be enough to constitute an attack under the Hate Speech policy and that a call for a group to commit suicide violate the Suicide and Self-injury policy.

1. Legitimate aim

 For a restriction to be valid, it must pursue one of the legitimate aims stipulated in the International Covenant of Civil and Political Rights, including “rights of others”. The Board concluded in the [*“Knin Cartoon”*](https://globalfreedomofexpression.columbia.edu/cases/oversight-board-case-of-knin-cartoon/) decision, among others, that the Hate Speech policy pursued the legitimate aim of protecting people from harm caused by hate speech. Furthermore, the Board found the Suicide and Self-injury policy to pursue the legitimate aims of protecting people's right to the enjoyment of the highest attainable standard of physical and mental health, the right to life and right to equality and non-discrimination.

1. Necessity and proportionality

The principle of necessity and proportionality mandates that limitations on freedom of expression must be suitable for fulfilling their protective purpose, represent the least intrusive option available to achieve this goal, and be proportionate to the significance of the interest being safeguarded. The Board employed the six-factor test stipulated in the Rabat Plan of action to assess the risks imposed by the violent content at hand. The Board found removing the content was necessary and proportionate to protect the rights of the LGBTQIA+ community in Poland.

The Board recognized its previous stance of reclaiming derogatory terms for LGBTQIA+ people (*“*[*Reclaiming Arabic Words”*](https://globalfreedomofexpression.columbia.edu/cases/oversight-board-case-of-reclaiming-arabic-words/)decision) and noted that this was not the case here. Furthermore, the Board stated that while the use of curtains was not a recurring coded language targeting transgender people, malign creativity was often employed to target them, and in the Board’s view the content at hand fits with the malign creativity trend. Furthermore, while the Board recognized the content might be humorous to some people, humor and satire could be used for hate speech, which was evident by the post celebrating the high suicide rates among transgender people. Moreover, the Board underlined that the content did not contain political nor newsworthy expression contrary to the *“*[*Colombia Protests*](https://globalfreedomofexpression.columbia.edu/cases/oversight-board-case-of-colombian-protests/)*”* decision.

When assessing the intent of the user, the Board considered their biography where they admit to being a transphobe. Additionally, the post described suicide of transgender individuals as “spring cleaning” alongside a heart emoji, thus the Board found that the user had the intent to incite discrimination and violence. Furthermore, the Board highlighted that the post did not only encourage transgender people to harm themselves but also incited others to discriminate and act violently towards transgender people.

Finally, the Board underscored the significant offline risks that the Polish LGBTQIA+ community faced and that content like the post at hand would contribute to a harmful environment that has previously led to higher rates of suicide among transgender people. Moreover, violent content against transgender people contributed to the ongoing mental-health crisis affected transgender people and offline violence against the community.

*Policy Advisory Statement*

1. Content Policy

The Board recommended Meta modifies the Suicide and Self-injury policy to clearly prohibit content promoting or encouraging suicide targeting and identifiable group of people.

1. Enforcement

The Board recommended Meta updates its internal guidelines for reviewers to ensure they understand that flag imagery representing gender identity, even without human figures, symbolizes a group defined by that identity.

***Direction:***

* **Outcome**: Contracts Expression/Mixed Outcome/Expands Expression
	+ This should be based on international standards
	+ However, if you have knowledge of national standards, and can provide insights into how the decision impact precedent nationally, please do so.
* **Explanation for why and how it contracts or expands expression or has a mixed outcome.** You can also provide additional context about the case here.

***Perspective***:

* **Related International and/or regional laws**:

**Example:**

[ECHR, art. 10](https://www.echr.coe.int/Documents/Convention_ENG.pdf);

[ECHR, art. 11](https://www.echr.coe.int/Documents/Convention_ENG.pdf);

* **National law or jurisprudence**:
* **Example:** [Sp. Constitution art. 14](http://www.congreso.es/portal/page/portal/Congreso/Congreso/Hist_Normas/Norm/const_espa_texto_ingles_0.pdf);

**Other national law or jurisprudence**:

* **List here any references to national case law outside the Court’s jurisdiction**. For instance, if a UK Court relies on Canadian or Australian case law, it would be listed here.

***Significance***:

* **Binding or persuasive precedent within jurisdiction**; **Decision establishes influential or persuasive precedent outside jurisdiction**; **Explanation**:

Standard I: The decision establishes a binding or persuasive precedent within its jurisdiction.

Information: i.e. Judgments of the European Court of Human Rights are binding upon parties to the decision.

Standard II: Decision (including concurring or dissenting opinions) establishes influential or persuasive precedent outside its jurisdiction.

Information: i.e. The decisions of the European Court of Human Rights have precedential value on the interpretation of the right to freedom of expression for other States Parties to the European Convention for the Protection of Human Rights and Fundamental Freedoms.

* **Related Cases**: Self-generated
* **Date updated**: N/A

***Docs***:

* **Official Case Documents**:

**Examples:**

Judgment (in English) [Attached]

Press Release issued by …. (in English) [Attached]

* **Reports, Analysis, and News Articles**:
	+ Please list any titles and URLS of interesting articles or analyses about the case that might be of interest to readers