



Oversight Board
September 27, 2023

Re: Post in Polish Targeting Trans People 2023-023-FB-UA

GLAAD is the world’s largest lesbian, gay, bisexual, transgender, and queer (LGBTQ) media advocacy organization. GLAAD’s Social Media Safety Program actively monitors, documents, and reports anti-LGBTQ hate speech and harassment to help advance online safety, privacy, and expression.

This submission serves as a response to Meta’s Oversight Board’s request for public comments on the above mentioned case.¹ The Board selected this case to assess the accuracy of Meta’s enforcement of its Hate Speech policy and to better understand how Meta approaches content that falls between hate speech and the promotion of suicide or self-injury.

The Oversight Board has requested public comments that address the state of anti-LGBTQ hate speech that may be described as humorous or satirical; the risks associated with that hate speech on social media and Meta’s human rights responsibilities; the state of anti-LGBTQ commentary on social media and public discourse in Poland; and Meta’s account level enforcement, policies, and practices to stop anti-LGBTQ statements that encourage death by suicide as a form of hate speech.

GLAAD’s 2023 Social Media Safety Index (SMSI) shows that Meta, on a systemic level, continues to fail to protect LGBTQ users, notably transgender (trans) people, from online abuse.² Largely ignoring expertise and guidance from civil society, the company is failing to remedy longstanding issues, contributing to real-world violence and other harms.

Current State of Anti-LGBTQ Online Hate and Connections to Offline Harm

Online hate speech targeting LGBTQ, especially trans, people has risen significantly in recent years.³ Prominent right-wing media outlets, pundits with monetary and political motives, and opportunist political figures are using social media to incite and encourage fear, hate, and violence against LGBTQ people.⁴

Hate content, seeded over time to foster dehumanizing narratives in politics and society, often violates Meta’s Community Guidelines but uses disingenuous rhetoric (including satire and humor) to circumvent safeguards.⁵ Meta is failing to adequately confront the issues of “malign creativity” that allow for unmitigated hate speech as bad actors adapt to moderation policies

¹ Oversight Board, [Oversight Board announces Post in Polish Targeting Trans People case](#) (Sep. 2023).

² GLAAD, [Social Media Safety Index 2023](#) (June 2023).

³ Hum an Rights Campaign, [An Epidemic of Violence 2022](#); GLAAD, [Drag events faced at least 124 protests and significant threats in 2022](#) (2022).

⁴ GLAAD, [Social Media Safety Index 2023](#) (June 2023).

⁵ United Nations, *Report: Online hate increasing against minorities* (March 2021); Harel, Tal Orian, et al. [“The Normalization of Hatred: Identity, Affective Polarization, and Dehumanization on Facebook in the Context of Intractable Political Conflict.”](#) *Social Media + Society*, vol. 6, no. 2, Apr. 2020, p. 205630512091398.

in a dynamic process.⁶ Even worse, Meta has failed to consistently apply its policies during heightened periods of violence toward minority communities, who pay the price.⁷

LGBTQ people globally are facing an alarming rise in targeted violence.⁸ In the US, armed Proud Boys and neo-Nazis with swastika flags outside pride events have become a regular occurrence.⁹ Demonstrations have included violent chants such as “pedophiles get the rope” and references to “groomers,” a derogatory term that falsely characterizes LGBTQ people as pedophiles and threats to children.¹⁰

The online popularity of the “groomer” trope in particular has contributed to real-world violence. From June 2022-May 2023, GLAAD and the Anti-Defamation League (ADL) documented at least 356 anti-LGBTQ incidents in the US, “with at least 191 anti-LGBTQ+ incidents of harassment, vandalism and assault making explicit references to ‘grooming’ or ‘pedophilia.’”¹¹ The report found an additional 145 incidents during June 2023 alone.¹² These attacks are following the horrifying trajectory of illiberal states such as Poland, Hungary, and Russia that are cracking down on LGBTQ rights.¹³

As highlighted in GLAAD’s SMSI, Meta’s Facebook and Instagram are largely failing to mitigate anti-LGBTQ hate and disinformation, despite such content violating its own policies.¹⁴ The SMSI also recommends that Meta and others better train moderators on the safety needs of LGBTQ users, and enforce policies around anti-LGBTQ content across all languages, cultural contexts, and regions.¹⁵ A 2021 Council of Europe report shows “increasing discrimination and attacks on LGBTI people in Poland and Europe as a whole,” illuminating the role of social media platforms in facilitating offline harm.¹⁶ Further, in 2022, Amnesty International called on authorities to “stop hateful rhetoric against LGBTI people and act to protect them from violence and discrimination.”¹⁷

In a 2023 ILGA-Europe report, Poland ranked 42nd out of 49 countries in the EU for LGBTQ equality. ILGA-Europe’s primary recommendation to the Polish government is to introduce hate speech and hate crime laws that explicitly cover all anti-LGBTQ bias-motivated crimes.¹⁸ The report notes anti-trans hate in Poland is rising, and online hate speech is a serious issue.

⁶ Malign Creativity: [How Gender, Sex, and Lies are Weaponized Against Women Online](#), Wilson Center.

⁷ Elliott, Vittoria. “[Meta Isn’t Enforcing Its Own Political Ads Policy, While the 2024 US Election Looms.](#)” *Wired*, 1 Sept. 2023; Paul, Katie, and Ayanat Mersie. “[Meta Accused in Lawsuit of Allowing Posts That Inflamed Ethiopia Conflict.](#)” *Reuters*, 15 Dec. 2022.

⁸ Feleke, Larry Madowo, Bethlehem. “[Uganda Passed One of the World’s Harshest Anti-Gay Laws. LGBTQ People Describe Living There as ‘Hell.’](#)” *CNN*, 29 June 2023; Camut, Nicolas. “[Anti-LGBTQ+ Violence in Europe Hits Decade-High, Report Finds.](#)” *POLITICO*, 20 Feb. 2023.

⁹ ADL, [Weekend of Anti-LGBTQ+ Extremism Highlights Heightened Threat to Community](#), Dec. 4, 2022).

¹⁰ [What is “Grooming?” The Truth Behind the Dangerous, Bigoted Lie Targeting the LGBTQ+ Community](#), ADL Center on Extremism; [Nazis protest a Wisconsin Pride event, shouting homophobic slurs](#), *NBC News* (2023).

¹¹ [Year in Review: Anti-LGBTQ+ Hate & Extremism Incidents, 2022 – 2023](#), ADL.

¹² [GLAAD, Anti-LGBTQ Hate & Extremism Spike During Pride 2023.](#)

¹³ Tony Wesolowsky, [The Worrying Regression Of LGBT Rights In Eastern Europe](#), Radio Free Europe/Radio Liberty.

¹⁴ [Hate Speech | Transparency Center](#). Meta, 2023.

¹⁵ GLAAD, [Social Media Safety Index 2023](#) (June 2023).

¹⁶ “[New Reports Show Increasing Discrimination and Attacks on LGBTI People in Poland and Europe as a Whole.](#)” Council of Europe, 10 Feb. 2021.

¹⁷ “[Poland: Authorities Must Stop Hateful Rhetoric against LGBTI People and Act to Protect Them from Violence and Discrimination.](#)” Amnesty International, 20 July 2022.

¹⁸ [Annual Review of the Human Rights Situation of Lesbian, Gay, Bisexual, Trans, and Intersex People in Poland](#). Rainbow Europe.

Meta’s Failure to Confront “Malign Creativity” and Adversarial Adaptation to Moderation

Social media users aiming to promote hate and harassment against marginalized groups (those with protected characteristics according to Meta’s Community Guidelines)¹⁹ often employ new slurs, tropes, and dog whistles that are only understood by intended audiences. Even when hate is more explicit, bad actors often defend a post or meme as humorous or satirical, when it is actually anti-LGBTQ hate or harassment. The Wilson Center coined this type of tactic as “malign creativity” in a report documenting the phenomenon.²⁰ Bad actors often develop novel means of targeting the LGBTQ community, and countering this requires continuous cross-functional adaptations from platforms in consultation with civil society.²¹

Meta’s policies prohibit content targeting people with hate based on protected characteristics, such as gender identity. Yet the company consistently fails to enforce its policies when reviewing reports on content that employs “malign creativity,” like disingenuously framing trans people as an “ideology.”²² For example, in Michael Knowles’s viral March 2023 CPAC (Conservative Political Action Conference) speech, he stated “Transgenderism must be eradicated from public life entirely,” which he uploaded to Instagram and Facebook.²³ By employing “transgenderism,” Knowles disingenuously asserts that his call for “eradication” refers to an *ideology*, not to trans *people*;²⁴ this is a gross dehumanization of trans people in its false assertion that being trans is an ideology rather than an identity. Influential accounts devoted to anti-trans hate continue to use similar dehumanization tactics to spread fear mongering, and even genocidal, rhetoric about minority groups. Meta’s failure to adequately differentiate between good-faith criticisms of ideas and thinly veiled (or even overt) hate speech has allowed the latter to flourish on its platforms.

Statements that Encourage or Applaud Death by Suicide as a Form of Hate Speech

Death by suicide is an acute public health crisis that disproportionately impacts LGBTQ, especially trans, people. US data indicates that 82% of trans people have considered killing themselves and over 40% of trans people have attempted suicide.²⁵ The crisis is especially acute among LGBTQ youth. Death by suicide is the second leading cause of death for young people between the ages 10–24,²⁶ and the suicide rate for LGBTQ youth is four times as high as their non-LGBTQ peers.²⁷ Anti-LGBTQ hate actors have weaponized these statistics online, telling trans people to kill themselves using the phrase “41% yourself,” a reference to the

¹⁹ [Hate Speech | Transparency Center](#). Meta, 2023.

²⁰ Nina Jankowicz et al., [Malign Creativity: How Gender, Sex, and Lies Are Weaponized Against Women Online](#), Wilson Center, (2021),

²¹ GLAAD documents and monitors the use of tropes and changes in slurs across major social media platforms in its Guide to Anti-LGBTQ Online Hate and Disinformation. GLAAD, [Guide to Anti-LGBTQ Online Hate and Disinformation](#) (2023),

²² Fitzsimons, Tim. [“Transgender Facebook Content Dominated by Right-Wing Sources. Study Finds.”](#) *NBC News*, 20 July 2020.

²³ GLAAD, [Social Media Safety Index 2023](#) (June 2023),

²⁴ Wade, Peter, and Patrick Reis. [“CPAC Speaker Calls for Eradication of ‘Transgenderism’ -- and Somehow Claims He’s Not Calling for Elimination of Transgender People.”](#) *Rolling Stone*, 6 Mar. 2023,

²⁵ Austin, A., Craig, S. L., D’Souza, S., & McInroy, L. B. (2022). [Suicidality Among Transgender Youth: Elucidating the Role of Interpersonal Risk Factors](#). *Journal of interpersonal violence*, 37(5-6), NP2696–NP2718.

²⁶ CDC, [Disparities in Suicide](#).

²⁷ Johns, M. M., Lowry, R., Haderxhanaj, L. T., et al. (2020). Trends in violence victimization and suicide risk by sexual identity among high school students — Youth Risk Behavior Survey, United States, 2015–2019. *Morbidity and Mortality Weekly Report*, 69,(Suppl-1):19–27.

reported attempted suicide rate of trans people. This should be considered a violation of Meta's Hate Speech and promotion of Suicide and Self-injury policies.

Data on suicide rates for LGBTQ people in Poland is limited, but a 2015-2016 report found that 44.2% of LGBTQ respondents have had suicidal thoughts. LGBTQ respondents under the age of 18 said they "thought about death much more frequently — more than two thirds of this group (69.4%) considered suicide."²⁸

Researchers have identified online hate and harassment as a leading contributor to high suicide rates amongst LGBTQ people.²⁹ These studies frequently identify targeted messages as especially harmful.³⁰ Meta states that it does not "allow people to intentionally or unintentionally celebrate or promote suicide or self-injury" and that it "removes any content that encourages suicide or self-injury, including fictional content such as memes or illustrations."³¹

In violation of Meta policies, the content at the center of this case encourages death by suicide by mocking the high suicide rate among trans people.³² Even though the content took the form of a meme and did not involve explicit language, it should be considered a violation of Meta's policies. Meta's Community Guidelines are broad enough to cover posts that implicitly promote self-injury by using hateful iconography, dog whistles, or other allusions. The policy's emphasis that even "unintentionally" celebratory content is rule-breaking and its prohibition on fictionalized content "such as memes or illustrations" is indicative of the policy's broad intent.

Meta's current moderation systems and practices are inadequate to address the harm from content that glorifies suicide or targets users with "humor" urging self-harm — allowing users, like the one in this case, to circumvent platform policies. Meta can help solve this problem by more broadly interpreting its prohibition on content that celebrates death by suicide and self-injury. Additionally, Meta could follow other moderation models that allow users to add comments when reporting content.³³ These comments serve two crucial functions. First, they allow users to provide additional context. Secondly, collecting information on reported content would enable Meta to better train its (increasingly automated) content moderation tools.³⁴

Meta's current reporting processes only allow users to select one reason for reporting content. The post in question was both hateful and advocating for death by suicide (i.e. harassment). This forces users, who generally have limited insight into Meta's review policies, to choose what could be effective in getting content reviewed. The users who reported the post under Hate Speech did not have their reports prioritized for human review, whereas the users who

²⁸ Świder, Magdalena, and Mikołaj Winiewski. [Situation of LGBTQ+ Persons in Poland](#). Campaign Against Homophobia, 2016,

²⁹ Reisner, S. et. al. Gender minority social stress in adolescence: disparities in adolescent bullying and substance use by gender identity. *Journal of sex research*, 52(3), 243-256. (2015).

³⁰ "2023 [U.S. National Survey on the Mental Health of LGBTQ Young People](#)." The Trevor Project, 2023,

³¹ [Meta, Suicide Prevention](#).

³² See e.g. Balt, E. et al. Social media use of adolescents who died by suicide: lessons from a psychological autopsy study. *Child Adolescent Psychiatry Mental Health* 17, 48 (2023).

³³ Twitter, [Report a Post](#).

³⁴ Fried, Ina. *Facebook's Content-Review Black Hole*. Axios, 22 June 2023,

reported the content for violating the Suicide and Self-injury policy did. This illuminates the gap in Meta’s ability to recognize and address dual- and multi-threat content.

Meta’s Practices for Reviewing Multiple Reports and Repeat Offenders

Meta currently does not consider the number of reports for priority human review. Meta should review this policy, and determine how effective the practice is for protecting targeted users from hate and harassment.

Meta should also consider users’ history of engaging in anti-LGBTQ hate and harassment when evaluating content. Many high-profile anti-LGBTQ accounts, such as Libs of TikTok (LoTT), amass followings by peddling hateful stereotypes.³⁵ Users behind these accounts frequently have a sophisticated understanding of content moderation practices which enables them to skirt the line between permissible and violating content, while purposefully ginning up hate amongst their followers.³⁶ Meta should consider the totality of an account’s activity, its audience’s size and nature, and offline context when evaluating content. This would help lead to a more robust and accurate assessment of its impact. For example, in 2022, LoTT posted false but permissible content about Boston Children’s Hospital’s trans healthcare program — and bomb and death threats followed.³⁷ The content could have been judged more accurately if Meta had considered the larger context, including LoTT’s history of promoting anti-LGBTQ violence.³⁸ Instead, this account continues to use Meta’s platforms to incite real-world harassment against schools, healthcare providers, and everyday LGBTQ people, threatening public safety (September 2023 has seen bomb threats tied to LoTT posts — in Tulsa, Oakland and Highland Park, IL).³⁹

Conclusion

Meta’s Hate Speech policy prohibits content targeting people based on protected characteristics, such as gender identity. The post in question is clearly asserting the horrific, and violative, sentiment that trans people should kill themselves. Meta’s content moderators should have accurately enforced its policies in the first place. It is a serious problem that the post was only removed after the Oversight Board alerted Meta. This case powerfully illuminates highly consequential systemic failures with the company’s moderation practices that have broad implications for all anti-LGBTQ hate content, as well as for content that targets *all* historically marginalized groups. Such moderation may be more complex than recognizing basic slurs, but this is why the company must provide adequate training and guidance to its moderators on recognizing anti-trans hate. Meta is fully capable of implementing such training yet continues to fail to prioritize it, resulting in epidemic levels of anti-LGBTQ hate across its platforms.

³⁵ Taylor Lorenz et. al., Washington Post, [Twitter account Libs of TikTok blamed for harassment of children’s hospitals.](#)

³⁶ GLAAD, [Meta and Twitter Refuse Action on ‘Libs of TikTok’ Posts As Doctors and Staff of Boston Children’s Hospital Receive Death Threats](#) (2022)./; Taylor Lorenz et. al., Washington Post, [Twitter account Libs of TikTok blamed for harassment of children’s hospitals.](#)

³⁷ Jessica Bartlett Globe Staff et al., [Boston Children’s Faced Months of Death Threats and Hate Mail. Here’s How Staff Members Are Coping.](#) The Boston Globe,

³⁸ Sultan, Reina. “[How Libs of TikTok Became an Anti-LGBTQ+ Hate Machine.](#)” Them, 30 Sept. 2022,

³⁹ Nikki McCann Ramirez, [Tulsa School District Bombarded With Bomb Threats After Right-Wing Influencer Post](#), Rolling Stone (Aug. 30, 2023),