



September 11, 2024

## **Input from GLAAD for Oversight Board on Gender Identity Debate Videos cases (2024-046-FB-UA, 2024-047-IG-UA)**

As the world's largest LGBTQ media advocacy organization and as leading experts in LGBTQ tech accountability, GLAAD's [Social Media Safety program](#) provides ongoing key stakeholder guidance with regard to LGBTQ safety, privacy, and expression to social media platforms, including Meta's Facebook, Instagram, and Threads. In addition to the following specific guidance to the Oversight Board on this case, we urge the Oversight Board to refer to the [2024 edition](#) of our annual [Social Media Safety Index report](#) for additional context.

This public comment from GLAAD is specifically addressing the Oversight Board's request for input on the following three main bullet points, with regard to the two videos (2024-046-FB-UA, 2024-047-IG-UA) under consideration by the Board:

- 1) "The impacts of Meta's Hate Speech and Bullying and Harassment policies on freedom of expression around gender identity issues, and the rights of transgender people, including minors."
- 2) "Technical challenges in enforcing bullying and harassment policies at-scale, the effectiveness of self-reporting requirements and their impacts on people targeted by bullying or harassment, and comparisons to alternative enforcement approaches".
- 3) "The sociopolitical context in the United States concerning freedom of expression and the rights of transgender people, especially for access to single-sex spaces and participation in sporting events."<sup>1</sup>

### **A Prefatory Note**

While most likely it is just an unintended awkwardness of phrasing, the Oversight Board's unfortunate titling of this case as the "Gender Identity *Debate* Videos case" warrants a quick prefatory reminder: While there are many important discussions to be had about transgender people, trans civil rights, and trans-related issues in society — trans, nonbinary, and gender non-conforming people are people. Just as the legitimacy of gay and lesbian people and their identities was a subject of "debate" decades ago (with framings of sexual orientation as "a phase" or questioning whether gay people were simply confused, mentally ill, or misguided),

---

<sup>1</sup> <https://www.oversightboard.com/news/new-cases-to-consider-debate-around-gender-identity>

trans people's *existence* and *identities* are not a subject for others to "debate." Specious rhetorical constructs seeking to delegitimize who we are (such as right-wing efforts to mischaracterize trans identity as an ideology<sup>2</sup>) are well-known strategies<sup>3</sup> that go hand-in-hand with other types of fear-mongering, scapegoating, and manufacturing of false narratives to legitimize violence<sup>4</sup> and other harms against historically marginalized groups. It is vitally important to keep this larger context in mind and to remember that framing gender identity itself as a "debate" is an *a priori* dehumanizing construct that threatens to erode the fundamental personhood of transgender people and even lurches towards an agenda of potential removal of protected characteristic status.

### **On Bullet Point 1**

*"The impacts of Meta's Hate Speech and Bullying and Harassment policies on freedom of expression around gender identity issues, and the rights of transgender people, including minors."*

The phrasing of this item seems to be asking whether certain kinds of speech related to so-called "gender identity issues" should be considered "free speech" rather than hate speech. While certainly this distinction is often a difficult one to discern (and obviously there are many examples of speech on these topics that are *not* in violation of Meta policies), it is instructive to remember the intent of social media platform hate speech policies (including Meta's) in protecting historically marginalized groups. To illuminate this point, if we extend this framing to discussions and posts around "race issues" (if you will), hopefully we all have a clear understanding that while some people may have opinions or thoughts related to different racial groups, *their* freedom of expression does not extend to speech that targets people with animus on the basis of race, or that questions their legitimacy and existence. Meta has extensive detailed policies prohibiting several specific subcategories of hateful expression — which apply to *all* protected characteristic groups. Trans, nonbinary, and gender non-conforming adults and minors are included in lists of protected characteristics groups because, like other historically marginalized groups of people, we are disproportionately targeted with hate and harassment. As Meta acknowledges in [its Community Standards](#), online hate "creates an environment of intimidation and exclusion, and in some cases may promote offline violence." According to a 2024 report from the ADL, 63% of transgender respondents reported that they had experienced online harassment in the last year, which happened most frequently on Meta's Facebook.<sup>5</sup> And according to the Trevor Project, 27% of transgender and nonbinary young people in the U.S. reported that "they have been physically threatened or harmed in the past year due to their gender identity."<sup>6</sup>

Like other groups, we depend on Meta's policies to protect *our* rights to be free from such attacks. As Samidh Chakrabarti, Facebook's former head of civic engagement, said in a (now

---

<sup>2</sup> [Breaking the Buzzword: Fighting the "Gender Ideology" Myth. Human Rights Watch](#)

<sup>3</sup> [History of the Anti-Gay Movement Since 1977 | Southern Poverty Law Center](#)

<sup>4</sup> [How Anti-LGBTQ+ Rhetoric Fuels Violence. Scientific American](#)

<sup>5</sup> [Online Hate and Harassment: The American Experience 2024](#)

<sup>6</sup> [2023 U.S. National Survey on the Mental Health of LGBTQ Young People](#)

deleted) April 2022 tweet: "Effective moderation is not inherently in conflict with free speech. It is *required* for people to feel free to speak."<sup>7</sup>

According to GLAAD's just-released 2024 *Accelerating Acceptance* report, 89% of non-LGBTQ Americans believe that LGBTQ people (including transgender people) should have the freedom to live their lives and not be discriminated against.<sup>8</sup> And yet, right-wing politicians and anti-LGBTQ commentators continue to promote inflammatory mischaracterizations of trans healthcare, baseless false assertions of LGBTQ people and our allies being threats to children, and malicious framings of LGBTQ people and identities as "ideologies" — all as an obvious blatant political strategy for attacking our basic equality and civil rights.<sup>9</sup> Such dangerous fear-mongering and scapegoating of marginalized groups is part of why social media platforms created their hate speech policies in the first place: to protect not only these groups but *all* platform users, as well as advertisers, from being surrounded by false and harmful content.

While there are legitimate conversations to be had that are related to transgender people and issues in society, such expression should not be predicated on dehumanizing rhetoric, misinformation, and outright lies.

## **On Bullet Point 2**

*"Technical challenges in enforcing bullying and harassment policies at-scale..."* [The second part of this item ("*effectiveness of self-reporting requirements and their impacts on people targeted by bullying or harassment*") will be addressed in the next section below].

Meta's customers depend on the company's enforcement of its own policies to keep them safe.<sup>10</sup> Indeed our society as a whole (including people who don't use social media) is consequentially impacted by the mass circulation of hate and disinformation on Meta's platforms — as evidenced in the extraordinary failures of the company to mitigate "Stop the Steal" Facebook groups and content that played an enormous role in facilitating the January 6th attack on the U.S. Capitol.<sup>11</sup> (Just to name one of many examples). There is no question that such enforcement at-scale must be enormously complex and difficult. As a trillion-dollar<sup>12</sup> corporation, Meta has met this challenge by laying off trust and safety staff<sup>13</sup> and backsliding on commitments to user safety and platform integrity,<sup>14</sup> while continuing to publicly assert to the consumers of its products that: "we don't allow hate speech on Facebook, Instagram, or Threads."<sup>15</sup> Meta must take active steps to mitigate violative content in accordance with its applicable policies and sub-policies. If such policies require evaluation by human moderators to

---

<sup>7</sup> [Elon Musk's free-speech Twitter dream: We've seen this before](#)

<sup>8</sup> [Accelerating Acceptance 2024 | GLAAD](#)

<sup>9</sup> [Online Hate = Offline Harm – 2024 Social Media Safety Index | GLAAD](#)

<sup>10</sup> ['Stop Lying': Muslim Rights Group Sues Facebook Over Claims It Removes Hate Groups : NPR](#)

<sup>11</sup> [Facebook Knows It Was Used To Help Incite The Capitol Insurrection](#)

<sup>12</sup> [Meta passes \\$1 trillion in market cap](#)

<sup>13</sup> [The Unbearably High Cost of Cutting Trust & Safety Corners | TechPolicy.Press](#)

<sup>14</sup> [Big Tech Backslide: How Social-Media Rollbacks Endanger Democracy Ahead of the 2024 Elections | Free Press](#)

<sup>15</sup> [Hate Speech | Transparency Center](#)

be accurately enforced “at-scale,” then the company should prioritize and assign adequate resources to achieve the timely and thorough review of such content and fulfill the commitments in its public-facing policies. Meta should also work with civil society to adequately understand how hate and harassment manifests on its platforms.

### **On Bullet Point 3**

*“The sociopolitical context in the United States concerning freedom of expression and the rights of transgender people, especially for access to single-sex spaces and participation in sporting events.”*

Please see additional notes below on Instagram Video Post #2. With regard to Video Post #2, in Meta’s argument for allowing it to stand the company says that: “Transgender people’s access to bathrooms that correspond to their gender identity is the subject of considerable political debate in the United States.” While these *topics* themselves are “newsworthy” — there is an important distinction to be made between legitimate discourse on a topic versus engaging in hate speech and harassment. For instance, while issues related to immigration and women’s rights are newsworthy — xenophobic and misogynist content targeting these protected characteristic groups with hate speech violates Meta’s hate and harassment policies.

With regard to the sociopolitical context in the U.S., please review these two citations below:

As noted in GLAAD’s *Debunking the “Bathroom Bill” Myth*<sup>16</sup> report:

"Public discussions about protecting LGBTQ people under federal, state, and local nondiscrimination laws are often sensationalized. While LGBTQ-inclusive nondiscrimination protections usually cover employment, housing, and public accommodations, opponents of these protections typically focus on generating fears about bathrooms, falsely claiming that such laws will make it legal for sexual predators to enter women’s restrooms. They imply or overtly claim that transgender women are not women, and therefore they should not be allowed to use the women’s restroom. (It is also often falsely implied that transgender women are in some way deviant and predatory.)<sup>17</sup> This is despite a lack of evidence to support their claims that transgender people put anyone in danger while in the restroom that aligns with the gender they live every day.<sup>18</sup>

These claims are simply untrue. It is important to note that nondiscrimination protections for transgender people do not change long-standing laws that make it illegal for anyone to enter a public restroom for the purpose of harassing or harming another person, or invading their privacy."

---

<sup>16</sup> [Debunking the ‘Bathroom Bill’ Myth – Accurate Reporting on LGBT Nondiscrimination: A Guide for Journalists | GLAAD](#)

<sup>17</sup> [STUDY: Houston Media Help Misrepresent Equal Rights Ordinance](#)

<sup>18</sup> [15 Experts Debunk Right-Wing Transgender Bathroom Myth | Media Matters for America](#)

As noted in GLAAD's *11th Annual Media Reference Guide* section on LGBTQ People and Sports<sup>19</sup> (created as a collaboration between GLAAD, Athlete Ally, the National Center for Lesbian Rights, and the Inclusion Playbook):

“Anti-LGBTQ activists are targeting transgender athletes, especially youth. In 2021, nine states (Alabama, Arkansas, Florida, Mississippi, Montana, South Dakota, Tennessee, Texas, West Virginia) passed laws banning transgender children from participating in sports. Major medical associations, including the [American Medical Association](#), denounced such legislation as harmful to the physical and social-emotional health of all young people, especially for trans youth already vulnerable to being ostracized at school and in society. There are risks to cisgender athletes as well, as some bills require [invasive anatomy screenings](#) and costly testing. There have been documented cases of [false accusations](#) against girls based on unfounded suspicions about their bodies, or haircuts.

Transgender athletes face uninformed opposition. Negative stereotypes and feelings about so-called ‘advantages’ transgender women have are not based in science, facts or evidence. Sport governing organizations like the International Olympic Committee (IOC), and the National College Athletics Associate (NCAA), as well as the Women’s Sports Foundation (WSF), have looked at the science associated with medical transition and made clear statements in support of the right of transgender athletes to participate in a way that is fair, equitable, and respectful to all.”

### **On Instagram Video Post #2 (2024-047-IG-UA)**

As described by the Oversight Board, an Instagram account posted a video of a trans girl athlete, and in the post’s text the account “refers to her as a boy, questioning whether they are female.”

The main aspects highlighted by the Oversight Board in the case description grapple with the question of whether this post would fall under policies related to “calls for exclusion.” We will address that aspect with some observations below, but first want to call attention to a different relevant clause in Meta’s policies.

The Oversight Board case description notes that “the Hate Speech policy does not include misgendering as a form of prohibited ‘attack’” and states that the Bullying and Harassment policy “does not consider misgendering a person to be a cognizable attack or call for exclusion.”

In Meta’s Transparency Center post about this case, Meta states that: “Meta determined that neither video violated our policies on [Hate Speech](#) or [Bullying and Harassment](#), as laid out in

---

<sup>19</sup> [LGBTQ People and Sports | GLAAD](#)

our Facebook Community Standards and [Instagram Community Guidelines](#), and left both pieces of content up.”<sup>20</sup>

For some reason, Meta doesn’t include any mention of the clause in their Bullying and Harassment policy<sup>21</sup> which states that: “... all private minors, private adults (who must self-report), and minor involuntary public figures **are protected from ... Claims about** romantic involvement, sexual orientation or **gender identity**.” Clearly this policy is applicable to cover targeted misgendering (as well as targeted deadnaming) — which is a “claim about a person’s gender identity.” (Specifically here, the account *denies* the minor’s gender identity by asserting that she is a boy).

There are several necessary slight digressions here, thanks for your patience. Firstly, this distinction (denying a person’s gender identity) is hugely significant. As a creative form of hate speech, targeted misgendering generally feigns to be harmlessly humorous or cleverly contrarian. (“*You say you’re a girl, I say you’re a boy. This is not hate speech, it’s just my opinion.*”) But, as is true of other creative forms of hate speech,<sup>22</sup> this practice seeks to convey extreme animus while pretending not to, and also is specifically crafted to evade existing hate and harassment policies. (This is why many platforms have developed policies specifically addressing the problem, which, like the creative form of antisemitism known as Holocaust denial, is also not an easy thing to moderate at-scale but is, like Holocaust denial, an extremely serious form of hate speech).<sup>23</sup>

One of the most unique aspects of targeted misgendering and deadnaming is that its rhetorical goal is a denial of the existence of both who *an individual person is* — as it relates to *their* gender identity (a protected characteristic); *and* as a denial of the existence of *the entire protected characteristic* of gender identity. The practice of targeted misgendering and deadnaming targets the individual and the larger groups of people protected by that category — primarily trans, nonbinary, and gender non-conforming people.<sup>24</sup>

In GLAAD’s 2023 report, [All Social Media Platform Policies Should Recognize Targeted Misgendering and Deadnaming as Hate Speech](#), we explain how this practice has emerged in recent years as one of the most prevalent modalities for expressing contempt and hate toward trans and nonbinary people across social media. The report notes that: “As is true with other hate-driven tropes, the far-right pundits, politicians, and accounts who engage in the practice tend to disingenuously feign that such expression is not hate speech, even as their animus is

---

<sup>20</sup> As a side note, in a 2021 Newsroom post about addressing bullying and harassment Meta refers to these as being inter-related policies, stating that: “we’ve [updated our AI technology](#) to train across three different but related violations: bullying and harassment, hate speech and violence and incitement.”

<sup>21</sup> [Bullying and Harassment | Transparency Center](#)

<sup>22</sup> [Malign Creativity: How Gender, Sex, and Lies are Weaponized Against Women Online | Wilson Center](#)

<sup>23</sup> [All Social Media Platform Policies Should Recognize Targeted Misgendering and Deadnaming as Hate Speech | GLAAD](#)

<sup>24</sup> It is also important to note thorough that gender identity related hate and disinfo impacts everyone, this is particularly evident in the maliciously creative anti-trans trope of so-called “transvestigation,” which targets cis public figures with dehumanizing fake conspiracy theories asserting that they are trans ([Transvestigation: What to Know About the Bizarre and Transphobic Conspiracy Theory | Teen Vogue](#)).

unmistakable, both from the posts themselves and from their clearly stated positions opposing the basic rights and dignity of trans and LGBTQ people. To be very clear — this isn't about accidentally getting someone's pronouns wrong or mistakenly mentioning their old name. This is hate speech pure and simple, designed to evade existing hate speech policies."

To return to the case of Instagram Video Post #2, possibly there is some arcane unexplained reason that Meta is not applying the "claims about gender identity" policy. However there is no particular explanation in the Community Standards for why this policy would not be relevant here.

Additionally, the Oversight Board case description notes that the girl is "a minor who Meta considers to be an involuntary public figure." So, again, in this case (the Tier 3 "claims about ... gender identity" policy), there is no self-reporting requirement or other loophole involved, and this policy should indeed be fully applicable and enforced. There is no apparent visible indication in Meta's policy pages of why this policy clause would not apply.

For the sake of thoroughness here, and because these are all extremely important distinctions to think through — please consider this analysis of one of the logical flaws of the self-reporting and public figure policies. Please see footnote 25 for further information.<sup>25</sup>

This case example also raises the question of whether Meta's moderators are fully aware of (and are evaluating potentially violative content in relation to) all of the company's relevant policies. Presumably Meta's moderators are familiar with the above-mentioned clause in Tier 3 of its Bullying and Harassment policy, which states that people "are protected from ... **Claims about** romantic involvement, sexual orientation or **gender identity**." But the fact that no mention

---

<sup>25</sup> With regard to whether the minor being targeted is a public figure and/or whether she is an adult, regardless of these aspects, cases like this illuminate the problematic limitations of these often arcane exceptions. In practice, they serve as loopholes to allow egregious hateful content that violates Meta's hate speech policies. In other words, if it is not okay to target a private minor, private adult, or minor involuntary public figure with an example of clearly identified malicious, hateful expression (in this case: "a claim about someone's gender identity"), why is it okay to target an adult or minor public figure with the same speech? Specifically when such speech is clearly not an example of legitimate "*critical commentary of people* who are featured in the news or who have a large public audience" (per Meta's policy), but *is* (we know because it will be recognized as such when directed at these other categories of private/minor public people) hate speech/harassment. Further, the requirement for self-reporting for private adults — but not for others — is also an absurd qualification that is similarly nonsensical in that if the platform can understand that, for instance, "a claim about... gender identity" when made with regard to a random private teenager can be mitigated by the platform without the need for that teen to self-report (this is indeed recognized as a form of bullying and harassment), then why would the platform similarly not be able to apply this to a private adult? This loophole means that transgender people must then personally monitor and report such attacks rather than the platform protecting them as is done for others. To walk through the whole logic: If the reason for requiring self-reporting is that the platform's position is that they do not know the person's gender and therefore are not qualified to determine whether a violation has occurred, then how is it that they are sufficiently confident to make that determination in the case of minors and involuntary public figures? The self-reporting requirement places undue burden on individuals being targeted (in this case trans adults) and also requires those targeted to expose themselves further to bullying and harassment (in not just the content but also the comments) during the process of monitoring and reporting.

is made of this policy in Meta’s Transparency Center statement about this case is concerning to say the least.

Lastly, regarding the Video Post #2 case: While targeted misgendering (i.e. intentionally referring to a person as a different gender than what they actually are — as here, saying a trans girl is a boy) is not something that non-trans people commonly experience as a form of bigotry (and so it may be difficult to understand or empathize with), this is an extremely common, and horrible, form of hateful harassment that trans and nonbinary people endure, especially on social media.<sup>26</sup> This practice should also be evaluated under several other clauses, including, the clauses of Meta’s Hate Speech policy that prohibit: “Statements **denying existence** (including but not limited to: ‘[protected characteristic(s) or quasi-protected characteristic] do not exist’, ‘no such thing as [protected characteristic(s) or quasi-protected characteristic];’” *and* “**Harmful stereotypes** historically linked to **intimidation, exclusion, or violence on the basis of a protected characteristic...**” (As we have articulated in previous comments to the Oversight Board,<sup>27</sup> targeted misgendering and deadnaming is now, indeed, posing the “harmful stereotype” that trans, nonbinary, and gender nonconforming people are not who we say we are. This dehumanizing construction attempts to invalidate the existence of a group of people on the basis of their protected characteristic, their gender identity — and it is very clearly “linked to intimidation and exclusion.”)<sup>28</sup> Lastly, yet another relevant violation exists in this clause: “**Content attacking concepts, institutions, ideas, practices, or beliefs associated with protected characteristics, which are likely to contribute to imminent physical harm, intimidation or discrimination against the people associated with that protected characteristic.** Meta looks at a range of signs to determine whether there is a threat of harm in the content. **These include but are not limited to:** content that could **incite imminent violence or intimidation**; whether there is a **period of heightened tension** such as an election or ongoing conflict; and **whether there is a recent history of violence against the targeted protected group.** In some cases, we may also consider whether the speaker is a public figure or occupies a position of authority.”

#### **On Facebook Video Post #1 (2024-046-FB-UA)**

*“In the first case, a Facebook user in the United States posted a video of a woman confronting a transgender woman for using the women’s bathroom. The post refers to the person being confronted as a man and asks why it is permitted for them to use a women’s bathroom.”*

Based on the description of the post, it would appear that the primary relevant policy to be applied is the “claims about gender identity” clause above (since in the post the Facebook user refers to this woman as a man — intentionally misgendering her). The account also advocates for her exclusion from a public place, and invites viewers to partake in the harassment.

---

<sup>26</sup> [The Relentless Misgendering of Dr. Rachel Levine, Pennsylvania Health Secretary, Is Violence](#)

<sup>27</sup> [GLAAD SUBMITS PUBLIC COMMENT TO OVERSIGHT BOARD ON FACEBOOK’S ANTI-TRANS HATE CONTENT CASE](#)

<sup>28</sup>



Lastly here, Meta says in its Transparency Center post about this case that: “Transgender people’s access to bathrooms that correspond to their gender identity is the subject of considerable political debate in the United States.” One would hope the company would recognize that such “political debate” should *not* extend to violations of its own Community Standards, including all of the notes outlined above in relation to Post #2. Such expressions are *not* debate but hate — which is exactly why the company delineates all of these things in its Bullying and Harassment and Hate Speech policies.

### **Conclusion**

Targeted misgendering is a (creative) form of hate speech. With malicious intent, it seeks to mock, denigrate, and dehumanize transgender, nonbinary, and gender non-conforming people in violation of Meta’s Hate Speech and Bullying and Harassment policies. It should be mitigated in accordance with all of Meta’s applicable policies and sub-policies; if such policies require evaluation by human moderators to be accurately enforced, then the company should prioritize and assign adequate resources to achieve the timely and thorough review of such content and fulfill the commitments it states in its public-facing policies.

### **About the GLAAD Social Media Safety Program**

As the leading national LGBTQ media advocacy organization GLAAD is working every day to hold tech companies and social media platforms accountable, and to secure safe online spaces for LGBTQ people. The GLAAD [Social Media Safety \(SMS\) program](#) researches, monitors, and reports on a variety of issues facing LGBTQ social media users — with a focus on safety, privacy, and expression. The SMS program has consulted directly with platforms and tech companies on some of the most significant LGBTQ policy and product developments over the years. In addition to ongoing advocacy work with platforms (including TikTok, X/Twitter, YouTube, and Meta's Facebook, Instagram, Threads, and others), and issuing the highly-respected annual [Social Media Safety Index \(SMSI\) report](#), the SMS program produces [resources](#), [guides](#), [publications](#), and [campaigns](#), and actively works to educate the general public and raise awareness in the media about [LGBTQ social media safety issues](#), especially anti-LGBTQ hate and disinformation.