



# 2025 Social Media Safety Index Platform Scorecard

## — Research Guidance

The 2025 SMSI **Platform Scorecard** consists of 14 indicators<sup>1</sup> that draw on best practices from the [Ranking Digital Rights \(RDR\)](#) Big Tech Scorecard, the annual evaluation of the world’s most powerful digital platforms on their policies and practices affecting people’s rights to freedom of expression and privacy.<sup>2</sup>

The Scorecard evaluates six major social media platforms:

- Facebook, Instagram, and Threads (whose parent company is Meta)
- TikTok (parent company - ByteDance)
- YouTube (parent company - Alphabet/Google), and
- X

In advance of the 2025 research cycle, several revisions were made to the index methodology. Most notably, given the prevalence of content and ads promoting harmful so-called “conversion therapy” services online, we crafted one new indicator (indicator 4) to evaluate the public-facing policies of social media companies, to help better understand their role in circulating and amplifying “conversion therapy” services online. While this new indicator addresses “conversion therapy” content in organic user content, we also added corresponding elements to indicator 8, which addresses advertising content that would otherwise violate community guidelines. We also added several elements to the Platform Scorecard that address other emerging threats to LGBTQ safety, privacy, and expression that have gained traction since we first developed the inaugural Platform Scorecard.

---

<sup>1</sup> In 2022, GLAAD released the inaugural Platform Scorecard evaluating five major social media platforms. The methodology behind the Platform Scorecard was developed in collaboration with Goodwin Simon Strategic Research (GSSR) and Ranking Digital Rights (RDR). After developing 12 draft indicators, the research team revised and refined the indicators based on feedback from RDR, interviews with five expert stakeholders working at the intersections of technology and human rights, and input from the SMSI advisory committee. Additional methodological considerations were identified during the subsequent policy analysis and company research. During the 2024 research cycle, GLAAD also added an evaluation of Threads to the Platform Scorecard.

<sup>2</sup> Note: GLAAD’s SMSI Platform Scorecard is narrowly focused on a select number of digital rights issues related to LGBTQ expression, privacy, and safety. While our evaluation may have implications for other protected characteristic groups, indicators narrowly focus on LGBTQ expression, privacy, and safety due to methodological constraints. When considering results from the annual evaluation, GLAAD strongly encourages readers to consider the SMSI’s findings in the context of RDR’s [Big Tech Scorecard](#). RDR’s scorecard evaluates and ranks the world’s leading online platforms across 58 indicators on their commitments to protecting user rights in critical areas such as terms of service enforcement, processing of user information, and algorithmic accountability.

Each item below describes in detail what best practices companies should be aspiring to. All evaluations are conducted looking at the public-facing policies of each platform. While results have global implications for LGBTQ safety, privacy, and expression online, our analysis is based on the English-language versions of public-facing policies and documents covering the U.S. market.<sup>3</sup>

**The Scorecard does not include indicators that rate platforms on *enforcement of their policies*.** Guidance for future researchers interested in applying the methodology of the SMSI Platform Scorecard can also be found below. For each indicator, we provide select examples that illustrate company performance, and whether platforms receive full, partial, or no credit on various elements. In addition, we also include guidance on the specific company policies that typically contain relevant disclosures addressing each indicator.

**Indicator 1: The company should have public-facing policies that *protect LGBTQ people from hate, harassment, and violence on the platform*.**

LGBTQ people and other protected characteristic groups are frequent targets of hate, harassment, and violence online. We therefore expect companies to have a policy in place that protects LGBTQ people from hate, harassment, and violence on the platform (Element 1). This policy should also cover public figures<sup>4</sup> (Element 2).

Companies should also state that this policy includes both sexual orientation (Element 3) and gender identity (Element 4) as protected characteristics. In order to give users a clear understanding of the types of content and behaviors that are prohibited on the platform, the company's policy should contain a detailed list of prohibited content and behaviors and provide illustrative examples (Element 5).

The company should also state that it has a "trusted flagger" program that allows LGBTQ and human rights organizations to receive priority consideration when flagging content to be evaluated for policy violations (Source 6). In order to receive full credit on

---

<sup>3</sup> In order to receive credit, companies have to publish relevant policy commitments on their company's website. These disclosures can be found in various places, including platforms' community guidelines, privacy policy, or terms of service agreements. Relevant disclosures may also be found in companies' help pages, newsrooms, and blogs. For indicators related to targeted advertising, we also consider advertising policies in our evaluation. In some cases, relevant policy disclosures may also be found in companies' corporate social responsibility reports or workforce diversity reports. Social media posts or other external sources do not count for credit.

<sup>4</sup> A public figure is an "individual who has assumed roles of especial prominence in the affairs of a society or thrust themselves into the forefront of particular public controversies to influence the resolution of the issues involved. Public figures also include individuals who have achieved pervasive fame or notoriety." Source: [https://www.law.cornell.edu/wex/public\\_figure#:~:text=Primary%-20tabs.resolution%20-of%20the%20issues%20involved](https://www.law.cornell.edu/wex/public_figure#:~:text=Primary%-20tabs.resolution%20-of%20the%20issues%20involved)

this element, platforms should also indicate that organizations can proactively apply to the “trusted flagger” program, as compared to the company pre-selecting participants.

Companies should also acknowledge the LGBTQ community’s history of reappropriating derogatory terms and disclose an explicit acknowledgement and exception of self-expressive usage of otherwise derogatory LGBTQ-related terms by LGBTQ users (Element 7).

#### Examples from our 2025 research:

One example for a “protected groups” policy can be found in TikTok’s Community Guidelines. In the section "[Safety and Civility](#)" of the guidelines, the company provides the following on "Hate Speech and Hateful Behavior": "TikTok is enriched by the diversity of our community. Our differences should be embraced, rather than a cause for division. We do not allow any hate speech, hateful behavior, or promotion of hateful ideologies. This includes explicit or implicit content that attacks a protected group. When there are discussions about social issues on TikTok, we want them to be respectful. Content may be ineligible for the FYF when it indirectly demeans protected groups...Hateful ideologies are systems of beliefs that exclude, oppress, or otherwise discriminate against individuals based on their protected attributes. Protected groups means individuals or communities that share protected attributes. Protected attributes mean personal characteristics that you are either born with, are immutable, or it would cause severe psychological harm if you were forced to change them or were attacked because of them." The company also lists "sexual orientation" and "gender identity" as protected attributes.

TikTok’s public-facing policies also contain an example of a “trusted flagger” program. In [a Newsroom post](#), the company provides the following on its “Community Partner Channel”: "We have ramped up our efforts to onboard partners to our Community Partner Channel - a direct avenue for trusted flaggers around the world...to report content to us for review, which sits alongside our in-app reporting function." While this statement indicates that the company has a "trusted flagger" program, it is not clear whether LGBTQ and human rights organizations can proactively apply to participate in the program, or if they are pre-selected by the company. Therefore, the company receives partial credit.

An example of an explicit exception of self-expressive usage of an otherwise derogatory term can be found on a page explaining Meta’s "[Hateful Conduct](#)" policy: “In other cases, speech, including slurs, that might otherwise violate our standards is used self-referentially or in an empowering way. We allow this type of speech where the speaker’s intention is clear. Where intention is unclear, we may remove content.”

Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found in platforms' *Community Guidelines* or *Hate Speech policies*.

**Indicator 2: The company should have a public-facing policy that states it provides users with a dedicated field to *add and change gender pronouns* on their user profiles.**

Gender pronouns are integral to users' self-expression and identification—this is particularly true for transgender, nonbinary, and gender non-conforming users. We therefore expect companies to provide users with a dedicated feature that allows them to add their pronouns to their profiles (Element 1). Users should also have the ability to change the pronouns on their profiles at any time (Element 2). In order to strike a balance between user expression and privacy and safety, companies should also give users granular options to customize the audience that can view their pronouns (Element 3).

Examples from our 2025 research:

According to Meta's [Gender Identity Policy and User Tools](#) policy, Instagram has a feature that allows users to add up to four pronouns to their profiles. In the same document, the company also explains how Instagram users can edit their pronouns. However, the company falls short of full credit as this policy indicates that the feature may not be available for all users. In addition, Instagram provides only limited options for users to customize who can see their pronouns. In this regard, users have the option to show their pronouns to followers only, but it is not clear from the policy whether they also have more granular options to restrict visibility of their pronouns, e.g., displaying pronouns to a list of pre-selected followers only.

Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found on platforms' *Help pages* or *policy portals addressing LGBTQ-specific policies and resources*.

**Indicator 3a: The company should have a public-facing policy that prohibits *targeted misgendering*<sup>5</sup> on the basis of gender identity.**

---

<sup>5</sup> Targeted misgendering is a form of hate speech that involves the intentional use of the wrong gender and/or gender pronouns when referring or speaking to a transgender, nonbinary, or gender

Transgender, nonbinary, and gender non-conforming users are among the most vulnerable when it comes to online abuse and harassment.<sup>6</sup> Therefore, companies should have a policy in place that contains a clear prohibition against targeted misgendering (Element 1). This policy should also cover public figures (Element 2). Companies should clearly explain the processes and technologies that they use to identify content and accounts violating this policy (Element 3). We also expect companies to have a reporting mechanism in place that allows users to report a violation of the company's policy against targeted misgendering (Element 4), and to clearly explain how users can report these types of violations (Element 5). The company should also clearly state that users can provide additional context when reporting violations to the policy (Element 6). In order to make sure that all users can report violations to the policy, the platform should clearly state that it does not require self-reporting by the targeted individual (Element 7).<sup>7</sup>

Prohibiting targeted misgendering is not enough. In order to make this policy effective, companies also should explain their processes for enforcing this policy once violations to the policy are detected, including providing details of how it decides what may represent violating content, and the actions it may take in response to content and accounts violating this policy (Element 8).

*Element language for Elements 3 and 8 directly draw on element language on terms of service enforcement developed by RDR.*

Examples from our 2025 research:

Meta's [Gender Identity Policy and User Tools](#) policy contains a prohibition against targeted misgendering on the platform. However, the company falls short of fully protecting transgender, nonbinary, and gender non-conforming users from targeted misgendering as it is clear from the policy that it does not apply to public figures. In addition, Meta states it needs to hear from targeted individuals in order to determine whether a policy violation has occurred, effectively requiring users to self-report violations to the policy. The company's policy does not explain whether it also employs human review and/or automated content moderation to identify violations to the policy, and also fails to clearly explain whether users can provide additional context when reporting instances of targeted misgendering.

---

non-conforming person. Source: <https://glaad.org/releases/glaad-responds-twitters-roll-back-long-standing-lgbtq-hate-speech-policy/>

<sup>6</sup> Online Hate and Harassment: The American Experience 2023 | ADL

<sup>7</sup> For more information on how self-reporting requirements complicate the enforcement of targeted misgendering and deadnaming policies, please see GLAAD's post "[All Social Media Platform Policies Should Recognize Targeted Misgendering and Deadnaming as Hate Speech.](#)"

### Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found in platforms' *Community Guidelines* or *Hate Speech policies*.

### **Indicator 3b: The company should have a public-facing policy that prohibits *targeted deadnaming*<sup>8</sup> on the basis of gender identity.**

In addition to a clear prohibition of targeted misgendering, companies should also have a policy that protects transgender, nonbinary, and gender non-conforming people from targeted deadnaming (Element 1), and clearly state that this policy also applies to public figures (Element 2). Companies should also clearly explain the processes and technologies that they use to identify content and accounts violating this policy (Element 3). We also expect companies to state that they have a reporting mechanism that allows users to report violations to the policy (Element 4), and to clearly explain how users can report instances of targeted deadnaming (Element 5). The company should also state that users can provide additional context when reporting violations to the policy (Element 6). In order to make sure that all users can report violations to the policy, the platform should clearly state that it does not require self-reporting by the targeted individual (Element 7).

In addition, companies should explain their processes for enforcing this policy once violations to the policy are detected, including providing details of how it decides what may represent violating content, and the actions it may take in response to content and accounts violating this policy (Element 8).

*Element language for Elements 3 and 8 directly draw on element language on terms of service enforcement developed by RDR.*

### Examples from our 2025 research:

TikTok is one of only two platforms evaluated in the SMSI Platform Scorecard which has a public-facing policy that prohibits targeted deadnaming. In its [Inclusion and Belonging Guide](#), the company lists several content policies protecting LGBTQ people. According to this policy, TikTok also prohibits "intentional misgendering or deadnaming." It is also

---

<sup>8</sup> Targeted deadnaming is a form of hate speech whereby a person intentionally "reveal[s] a transgender person's former name without their consent – often referred to as 'deadnaming' – [which] is an invasion of privacy that undermines the trans person's true authentic identity, and can put them at risk for discrimination, even violence."

Source: <https://glaad.org/releases/glaad-responds-twitters-roll-back-long-standing-lgbtq-hate-speech-policy/>

clear from TikTok's [Community Guidelines](#) that its content policies apply to public figures: "We recognize that public figures are in a position of public attention, have ways to counter negative speech, and that some content related to them may be in the public interest to view. We do allow some negative or critical comments or images about public figures. However, we still remove content that violates other policies (such as violent threats, hate speech, or sexual exploitation), as well as serious forms of harassment (such as doxxing or expressing a desire that someone experience serious physical harm)."

However, the company falls short of full credit as it is not clear from TikTok's policies whether violations can be reported by users other than the targeted individual. In addition, it is not clear whether users can provide additional context when reporting violations to the policy.

#### Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found in platforms' *Community Guidelines* or *Hate Speech policies*.

#### **Indicator 4: The company should have a public-facing policy that prohibits content *promoting so-called "conversion therapy."***<sup>9</sup>

Social media companies continue to play a central role in circulating and amplifying dangerous "conversion therapy" content online, a practice that has been widely condemned by leading medical and psychological associations.<sup>10</sup> We therefore expect companies to have a policy in place that prohibits content promoting so-called "conversion therapy" on the platform (Element 1).<sup>11</sup> Complicating efforts to minimize the amplification of "conversion therapy" content online, purveyors of this dangerous practice continue to promote it using alternate labels such as "leaving homosexuality" and "unwanted same-sex attraction." Therefore, we expect companies to state that at least once per year, they engage with LGBTQ and human rights organizations on best practices around identifying harmful "conversion therapy" content (Element 2).

---

<sup>9</sup> "Conversion therapy" is a widely condemned practice that involves any psychological or religious intervention aimed at changing an LGBTQ person's sexual orientation, gender identity, or gender expression. Complicating efforts to address the amplification of harmful "conversion therapy" content online, its purveyors also promote this dangerous practice under alternate labels such as "leaving homosexuality" and "unwanted same-sex attraction."

Sources: <https://glaad.org/reference/conversion-therapy>;  
<https://globalextrremism.org/reports/conversion-therapy-online-the-ecosystem-in-2023/>

<sup>10</sup> Sources: <https://glaad.org/reference/conversion-therapy>;  
<https://globalextrremism.org/reports/conversion-therapy-online-the-ecosystem-in-2023/>

<sup>11</sup> Note: While this new indicator addresses "conversion therapy" content in organic user content, we also added corresponding elements to indicator eight, which addresses advertising content harmful to LGBTQ people.



Companies should also clearly explain the processes and technologies they use to identify content and accounts violating this policy, including any human and automated content moderation they employ (Element 3). Platforms should also have a reporting mechanism that allows users to report content and accounts violating this policy (Element 4), and clearly explain how users can report these violations (Element 5). We also expect companies to clearly explain their processes for enforcing this policy. In this regard, companies' policies should provide details on how platforms decide what may represent violating content and accounts, and the actions they may take in response to content and accounts violating this policy (Element 6).

*Element language for Elements 3 and 6 directly draw on element language on terms of service enforcement developed by RDR.*

Examples from our 2025 research:

In its [Inclusion and Belonging Guide](#), TikTok prohibits "content that promotes or advertises 'conversion therapy' programs or services." On various company pages, TikTok explains in detail how it identifies content and accounts violating this policy, and how it enforces this policy once policy violations are detected. However, the platform falls short of full credit as it is not clear from the company's policies whether TikTok engages with LGBTQ and human rights organizations on best practices around identifying harmful "conversion therapy" content.

Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found in platforms' *Community Guidelines* or *Hate Speech policy*.

**Indicator 5a: The company should have a public-facing policy that explains what options users have to *control or limit* the company's collection, inference, and use of data and information related to their *sexual orientation*.**

Companies collect vast amounts of data that allow them to make inferences about users' sexual orientation and gender identity. Ranking Digital Rights and other civil society groups have called for greater transparency and user control around data collection and processing of this information. Companies should also give users control over the collection and inference of information related to their sexual orientation (Elements 1 and 2). We also expect companies to publicly state that users have the ability to delete each type of user information related to their sexual orientation, without having to delete their entire account, that the company has collected about them (Element 3). Platforms should also state that they do not use information related to



users' sexual orientation for the development of algorithmic systems, unless users have proactively opted in (Element 4), and should give users options to control how this information is used for the development of algorithmic systems (Element 5).

Examples from our 2025 research:

The platforms evaluated in the SMSI continue to provide insufficient transparency regarding LGBTQ users' control over user information related to their sexual orientation. In this regard, TikTok provides limited information in its public-facing policies. For example, TikTok is the only platform that clearly states that it does not collect sexual orientation information. In this context, the company provides the following in its Inclusion and Belonging Guide: "TikTok does not collect sexual orientation information." The same page also indicates that TikTok users can delete this information. However, TikTok falls short of full credit as it does not clearly state that users have control over whether the company can attempt to infer their sexual orientation. In addition, TikTok provides no information on whether users have options to control how data related to their sexual orientation is used for the development of algorithmic systems.

Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found in the company's *Privacy Policy* or *Transparency Center*.

**Indicator 5b: The company should have a public-facing policy that explains what options users have to *control or limit* the company's collection, inference, and use of data and information related to their *gender identity*.**

In addition to giving users control over companies' collection and inference of user information related to their sexual orientation, users should have options to control the collection and inference of user information related to their gender identity (Elements 1 and 2). Users should also have the ability to delete each type of user information related to their gender identity, without having to delete their entire account (Element 3). Platforms should state that they do not use information related to users' gender identity for the development of algorithmic systems, unless users have proactively opted in (Element 4), and give users clear options to control how this information is used for the development of algorithmic systems (Element 5).

Examples from our 2025 research:

The platforms evaluated in the SMSI continue to provide insufficient transparency about LGBTQ users' control over information related to their gender identity. Limited relevant

information can be found in YouTube's policies.<sup>12</sup> In this context, Alphabet's user-facing policies provide limited information on users' ability to control the company's collection of data related to their gender identity. For example, in its [privacy policy](#), the company says the following about "activity controls": "Decide what types of activity you'd like saved in your account. For example, if you have YouTube History turned on, the videos you watch and the things you search for are saved in your account so you can get better recommendations and remember where you left off..." In addition, the page "[Privacy tools that put you in control](#)" provides that "you can pause specific types of data from being saved in your account – like your Search and browsing activity, YouTube History, or Location History." However, the company does not explain whether users can control the company's collection of all data related to their gender identity.

For creators, YouTube discloses a "[Creator Demographics](#)" survey that allows "artists" and YouTube channel creators to voluntarily disclose their demographic information. According to YouTube, this allows the company to "ensure YouTube is inclusive and works for everyone." Under the section "Edit or delete your response info," the company explains how creators and "artists" can edit and delete this information. However, it is not clear from this disclosure what types of demographics the company collects from creators.

The company also provides limited information on users' and creators' ability to delete various types of information related to their gender identity that Alphabet has collected and inferred about them, and provides limited options for creators to control how information related to their gender identity is used for the development of algorithmic systems.

#### Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found in the company's *Privacy Policy* or *Transparency Center*.

**Indicator 6: The company should have a public-facing policy that states that it *does not recommend content* to users based on their disclosed or inferred sexual orientation or gender identity, unless a user has *proactively opted in*.**

LGBTQ users should have full control over the information they see on their social media feeds, and recommendation of content based on their disclosed or inferred sexual orientation and gender identity should be off by default (Element 1). Companies should also clearly explain how users can opt in to seeing recommended content based

---

<sup>12</sup> Note: In their evaluation of YouTube, researchers evaluated user-facing policies as well as policies for content-creators for this indicator.

on their sexual orientation and gender identity (Element 2). Similarly, we expect companies to state that users can opt out of seeing recommended content based on information related to their sexual orientation and gender identity at any time (Element 3), and clearly explain in their policies how they can do so (Element 4).

#### Examples from our 2025 research:

Companies continue to provide insufficient information regarding the options users have to control the recommended content they see on their feeds, with none of the platforms stating that recommended content based on a user's disclosed or inferred sexual orientation or gender identity is off by default. Like its peers, YouTube's public-facing policies contain only limited relevant information. For example, the page "[Recommended Videos](#)" explains that "you can pause, edit, or delete your YouTube watch and search history whenever you want." Additional information can be found on the YouTube Help pages "[Manage your recommendations and search results](#)" and "[Browse YouTube while incognito on mobile devices](#)." The company falls short of full credit as users have some control over the content they see, but we did not locate any public-facing information that indicates users can *opt in* or *opt out* of seeing recommended content based on their sexual orientation or gender identity at any time.

#### Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found in the company's *Privacy Policy* or *Transparency Center*.

**Indicator 7: The company's public-facing policies should state that it *does not allow third-party advertisers to target users with, or exclude them from, seeing content or advertising based on their disclosed or inferred sexual orientation or gender identity, unless the user has proactively opted in.***

Ranking Digital Rights and other civil society organizations have long called attention to the harms caused by the targeted advertising-driven business models of social media companies that rely on the collection of large amounts of user data. Targeted advertising based on protected characteristics raises additional concerns for user privacy and safety, and there is an acute need for users to have full control over how their data is used for this purpose.

Companies should not target LGBTQ users with targeted advertising unless they have proactively opted in (Element 1). In order to ensure LGBTQ users are not *excluded* from economic, financial, and other opportunities, companies should make a commitment not to exclude LGBTQ users from advertising (Element 2). LGBTQ users should also have control over how their own data is used for targeted advertising (Element 3-6). In order

to give insight into how companies detect content and accounts violating these rules, they should explain the processes and technologies that they use to identify advertisers who are in violation of these policies (Element 7).

Companies that have a policy that clearly prohibits advertisers from targeting users with advertising based on their sexual orientation and gender identity receive full credit on Element 1. For these companies, Elements 3-6 are not applicable.

*Element language for Element 7 directly draws on element language on targeted advertising developed by RDR.*

Examples from our 2025 research:

Meta’s Advertising Standards page “[Discriminatory Practices](#),” which covers Facebook and Instagram, contains a clear prohibition against both wrongful targeting and exclusion of LGBTQ users from ad services (targeted advertising is currently not available on Threads): “Ads must not discriminate or encourage discrimination against people based on personal attributes such as race, ethnicity, color, national origin, religion, age, sex, sexual orientation, gender identity, family status, disability, medical or genetic condition. Meta prohibits advertisers from using our ads products to discriminate against people. This means that advertisers may not (1) use our audience selection tools to (a) wrongfully target specific groups of people for advertising (see advertising policy on Targeting), or (b) wrongfully exclude specific groups of people from seeing their ads; or (2) include discriminatory content in their ads.” In addition, Meta also clearly explains the processes and technologies—including human and automated content moderation—that it uses to identify advertisers violating these rules.

Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found in the company’s *Advertising* policies.

**Indicator 8: The company should have a public-facing policy that prohibits *advertising content* that promotes hate, harassment, and violence against LGBTQ individuals on the basis of protected characteristics.**

Companies should have a policy that prohibits advertising content that promotes hate, harassment, and violence against LGBTQ individuals (Element 1). Given the continued prevalence of so-called “conversion therapy” content on social media platforms, companies’ public-facing policies should also contain a clear prohibition of content promoting “conversion therapy” services or practices (Element 2). We also expect

companies to state that, at least once per year, they engage with LGBTQ and human rights organizations on best practices around identifying harmful “conversion therapy” content (Element 3). The company should explain the processes and technologies it uses to identify advertisers and ad content violating these rules (Element 4).

Examples from our 2025 research:

YouTube parent company Alphabet’s advertising policies contain a page titled “[Inappropriate content](#)” which prohibits: “Content that incites hatred against, promotes discrimination of, or disparages an individual or group on the basis of their race or ethnic origin, religion, disability, age, nationality, veteran status, sexual orientation, gender, gender identity, or any other characteristic that is associated with systemic discrimination or marginalization.” In addition, YouTube’s Advertising Policies Help page “[Misrepresentation: Unreliable claims](#)” prohibits “content promoting harmful health claims, or content that relates to a current, major health crisis and contradicts authoritative scientific consensus” and explicitly lists “gay conversion therapy” as one example of prohibited content.

Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found in the company’s *Advertising* policies.

**Indicator 9: The company should regularly *publish data* about the actions it has taken to *restrict content and accounts* that violate policies protecting LGBTQ people.**

In order to provide insight into how company policies are enforced, the company’s transparency report should publish the number of pieces of content restricted for violating platform policies protecting LGBTQ people (Element 1). This includes content removals, but also other types of enforcement actions the company may take (e.g., hiding content, labeling content with a warning to the user). The company should also report on the number of accounts restricted for violating policies protecting LGBTQ people (Element 2).

As the wrongful removal of content and accounts can have significant implications for freedom of expression and human rights, we expect companies to report on the number of pieces of content (Element 3) and accounts (Element 4) that were restored after being wrongfully restricted for violating policies designed to protect LGBTQ people. Drawing on RDR best practices, this data should be published four times per year (Element 5).

*Element 5 directly draws on element language on transparency reporting developed by RDR.*

Examples from our 2025 research:

The platforms evaluated in the SMSI Platform Scorecard continue to fall short of providing comprehensive data on the number of pieces of content and accounts restricted for violating policies designed to protect LGBTQ people from hate, harassment, and violence. For example, TikTok's [Community Guidelines Enforcement](#) report contains data on the number of pieces of content and accounts restricted for violations of its Community Guidelines, which also prohibit hate, harassment, and violence against LGBTQ people. The company's transparency report contains the number of pieces of content that were reinstated after they had been wrongfully removed for violating TikTok's Community Guidelines. However, the platform falls short of full credit as it does not provide the same data on wrongful account restrictions.

Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found in the company's *Transparency Report*.

**Indicator 10: The company's public-facing policies should *explain the proactive steps* it takes to *stop demonetizing and/or wrongfully removing* legitimate content and accounts related to LGBTQ topics and issues.**

LGBTQ people and other protected characteristic groups are frequent targets of wrongful censorship and demonetization on social media, not only depriving them of tools for expression, but also exacerbating economic and financial inequities. Companies should explain the concrete steps they take to address wrongful removal and demonetization of content and accounts related to LGBTQ issues and topics (Element 1). Companies should also state that they regularly meet with LGBTQ content creators or stakeholders advocating on their behalf to solicit expert input on wrongful removals and demonetization on the platform (Element 2). Given that reporting tools are frequently abused to wrongfully target content and accounts related to LGBTQ topics and issues, companies should also have a policy that prohibits targeted and malicious reporting of LGBTQ users and content (Element 3).

Examples from our 2025 research:

While legitimate content and accounts related to LGBTQ issues and topics are frequent targets of wrongful censorship and demonetization, social media companies'

public-facing policies do not sufficiently explain the steps they take to mitigate this issue. For example, LGBTQ creators and advocates frequently raise concerns about the wrongful removal, suspension, and demonetization of LGBTQ-related content and creators from YouTube. However, Alphabet (YouTube's parent company) continues to provide limited transparency in this regard, and discusses only piecemeal solutions rather than a comprehensive plan outlining the steps it takes to address demonetization and removal of LGBTQ content and creators on the platform.

In addition, Alphabet's policies do not sufficiently explain whether the company regularly meets with LGBTQ creators or other relevant stakeholders to address wrongful removals, suspensions, or demonetization on YouTube. In this regard, the company states the following on a page explaining its [Creator Demographics survey](#): "We also conduct research with creators like focus groups, in-person feedback sessions, surveys, and other types of research. Through this work, we can bring creators' points of view to our product development teams. The info from Creator Demographics will allow us to extend research invitations to more creators that reflect the diversity of communities on YouTube." However, it is not clear from this statement whether research invitations are extended to LGBTQ creators, resulting in partial credit on this element.

#### Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found in companies' *Newsroom posts* or *LGBTQ portal*.

**Indicator 11: The company should regularly *publish data* about the actions it has taken to *stop demonetizing and/or wrongfully removing* legitimate content and accounts related to LGBTQ topics and issues.**

In order to provide insight into the removal and demonetization of legitimate content and accounts related to LGBTQ issues and topics, the company should publish data on the number of pieces of LGBTQ-related content and accounts removed, demoted, or demonetized for violating the company's policies (Elements 1 and 2). The company should publish this data at least four times per year (Element 3).

*Element 3 directly draws on element language on transparency reporting developed by [RDR](#).*

#### Examples from our 2025 research:



Currently, none of the companies evaluated in the SMSI Platform Scorecard provide relevant data, giving no concrete insight into the state of censorship and demonetization of LGBTQ content and voices on many of the world's leading social media platforms.

Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found in companies' *Transparency Report* or *Newsroom posts*.

**Indicator 12: The company should publicly commit to providing *mandatory training for content moderators*, including those employed by contractors, focused on LGBTQ safety, privacy, and expression on the platform.**

To ensure that content moderators have a strong understanding of the unique challenges that LGBTQ people and other protected characteristic groups face online, companies should publicly commit to providing mandatory annual training for moderators that trains them on the needs of protected characteristic groups (Element 1), including LGBTQ people (Element 2).

Examples from our 2025 research:

Platforms continue to provide insufficient information on their protocols for content moderator training, particularly when it comes to the unique challenges faced by LGBTQ people and other protected characteristic groups. For example, Meta receives partial credit based on a statement in its [Gender Identity Policy and User Tools](#) policy. According to the policy, Meta's human reviewers around the world "have undertaken specific training on gender identity policy enforcement in 2022. We give reviewers more explicit and detailed internal guidance about when to consider a trans, non-binary or genderfluid person to be attacked on the basis of gender identity. This helps us better enforce our policy at scale for the 2.8 billion people who use our technologies, across every country and language where we operate. It involves providing guidance on the language used by the LGBTQ+ community to identify indicators for gender identity for trans, genderfluid, non-binary and gender nonconforming people (such as the Trans Pride flag)."

However, it is not clear from the company's statement whether it has conducted similar training since 2022. Given that Meta does not specify whether training for content moderators is conducted on an annual basis, the company receives partial credit.

Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found in companies' *Annual Reports* or *Newsroom posts*.

**Indicator 13: The company should have a public-facing policy that explains its *internal structures* to best ensure the fulfillment of its commitments to overall LGBTQ safety, privacy, and expression on the platform.**

The company should state that it has an LGBTQ policy lead who advises policy and product teams on how its policies, products, and services may impact LGBTQ safety, privacy, and expression on the platform (Element 1). The potential risks that LGBTQ people may face online are constantly evolving. Therefore, the company should also state that at least once per year, it engages with LGBTQ rights organizations to solicit expert input on key issues related to LGBTQ safety, privacy, and expression online (Element 2). The company should also state that it has a formal training in place that trains employees at different levels of the company about the needs of LGBTQ users (Element 3).

Examples from our 2025 research:

As in previous years, platforms' policies do not sufficiently explain how their commitments to LGBTQ safety, privacy, and expression are implemented within the company. Notably, none of the platforms evaluated in the SMSI states it has an LGBTQ policy lead. In this regard, TikTok previously publicly stated it had an LGBTQ policy lead, but we did not locate a similar statement during this year's research cycle. While the company states that it engages with LGBTQ rights organizations, it is not clear from TikTok's public-facing policies whether meetings with these organizations take place at least once per year.

Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found in companies' *Annual Reports* or *Newsroom posts*.

**Indicator 14: To create products that better serve all of its users, the company should make a public commitment to *continuously diversify its workforce*, and ensure accountability by periodically publishing voluntarily self-disclosed data on the number of LGBTQ employees across all levels of the company.**

The company should make a public commitment to taking proactive steps towards diversifying its workforce (Element 1). The company should also publicly state that it has an internal reporting mechanism that allows employees to voluntarily self-disclose their

sexual orientation and gender identity (Element 2). This voluntarily disclosed data should be published in the company's workforce numbers (Element 3) and should be broken out by different teams (Element 4). For the purpose of this research, we are focused on DEI employment data for a company's U.S.-based workforce. The company can only receive full credit on this indicator if it publishes this data at least once a year (Element 5).

Examples from our 2025 research:

Besides X, all of the platforms express a public commitment to taking proactive steps towards building a more diverse workforce. During the 2025 SMSI research cycle, Meta and Alphabet also published partial data showing progress towards their diversity and inclusion goals.

Potential sources for data collection:

Relevant policy commitments addressing this indicator can typically be found in companies' *Workforce Diversity Report* or *Annual Report*.