2025 Social Media Safety Index Platform Scorecard — Indicators & Elements

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

- **Q1.1:** Do the company's policies prohibiting hate, harassment, and violence include prohibitions against content that targets people ***on the basis of protected characteristics***?
- **Q1.2:** Are public figures protected by this policy?
- **Q1.3:** Do the company's policies prohibiting hate, harassment, and violence expressly state that they ***include sexual orientation***?
- **Q1.4:** Do the company's policies prohibiting hate, harassment, and violence expressly state that they ***include gender identity***?
- **Q1.5:** Do the policies prohibiting hate, harassment, and violence contain a detailed list of harmful, hateful, harassing, and/or violent content and behaviors that fall under these policies?
- **Q1.6:** Does the company state that LGBTQ and human rights organizations can apply to receive priority consideration when flagging content to be evaluated for policy violations (e.g. through a "trusted flagger" program)?
- **Q1.7:** Do the platform's policies contain an explicit acknowledgement and exception for LGBTQ users' ***self-expressive usage*** of otherwise derogatory LGBTQ-related terms?

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.

- **Q2.1:** Does the company provide users with a dedicated field to ***add*** gender pronouns to their user profiles?
- **Q2.2:** Does the company state that users can ***change*** the gender pronouns on their profiles at any time?
- **Q2.3:** Does the company provide options for users to ***customize the audience*** of their gender pronouns through the platform's privacy and visibility settings?

Indicator 3a: The company should have a public-facing policy that prohibits *targeted misgendering*[1] on the basis of gender identity.

- **Q3a.1:** Do the platform's community guidelines or other policies include a prohibition against ***targeted misgendering***?
- **Q3a.2:** Are public figures protected by this policy?

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

- **Q3a.3:** Does the company have a policy that explains the ***processes and technologies*** that it uses to identify content and accounts that violate this policy (e.g. via human and/or automated content moderation)?
- **Q3a.4:** Does the company state ***that*** users *can report* a violation of the company's policy against targeted misgendering?
- **Q3a.5:** Does the company explain ***how*** users *can report* a violation of the company's policy against targeted misgendering?
- **Q3a6:** Does the company state that users can ***provide additional context*** when reporting a violation of the company's policy against targeted misgendering?
- **Q3a.7:** Does the company's policy state that it does ***not require self-reporting*** by the targeted user?
- **Q3a.8:** Does the company explain its process for enforcing this policy once violations are detected?

Indicator 3b: The company should have a public-facing policy that prohibits *targeted deadnaming*[2] on the basis of gender identity.

- **Q3b.1:** Do the platform's community guidelines or other policies include a prohibition against ***targeted deadnaming***?
- **Q3b.2:** Are public figures protected by this policy?
- **Q3b.3:** Does the company have a policy that explains the ***processes and technologies*** that it uses to identify content and accounts that violate this policy (e.g. via human and/or automated content moderation)?
- **Q3b.4:** Does the company state ***that*** users *can report* a violation of the company's policy against targeted deadnaming?
- **Q3b.5:** Does the company explain ***how*** users *can report* a violation of the company's policies against targeted deadnaming?
- **Q3b.6:** Does the company state that users can ***provide additional context*** when reporting a violation of the company's policy against targeted deadnaming?
- **Q3b.7:** Does the company's policy state that it does ***not*** require self-reporting by the targeted user?
- **Q3b.8:** Does the company explain its process for enforcing this policy once violations are detected?

Indicator 4: The company should have a public-facing policy that prohibits content *promoting so-called "conversion therapy."*[3]

---

[2] 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/

[3] "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."

- **Q4.1:** Do the company's community guidelines or other policies contain a prohibition against content ***promoting so-called "conversion therapy***?"
- **Q4.2:** Does the company state that, at least once per year, it engages with LGBTQ and human rights organizations on best practices around identifying harmful "conversion therapy" content?
- **Q4.3:** Does the company publicly explain the ***processes and technologies*** that it uses to identify content and accounts that violate this policy (e.g. via human and/or automated content moderation)?
- **Q4.4:** Does the company state ***that*** users *can report* a violation of the company's policy prohibiting harmful "conversion therapy" content?
- **Q4.5:** Does the company explain ***how*** users *can report* a violation of the company's policy against harmful "conversion therapy" content?
- **Q4.6:** Does the company explain its process for enforcing this policy once violations are detected?

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*.

- **Q5a.1:** Does the company's policy state that users have control over the company's ***collection of user information*** related to their sexual orientation?
- **Q5a.2:** Does the company's policy state that users have control over whether the company can ***attempt to infer*** the sexual orientation of users based on metadata and other signals related to their behavior?
- **Q5a.3:** Does the company's policy state that users, without deleting their account, ***can delete*** each type of user information related to their sexual orientation that the platform has collected?
- **Q5a.4:** Does the company's policy state that it does ***not*** use information related to a user's sexual orientation for the development of algorithmic systems — including but not limited to AI classifiers and machine learning models — unless the user has ***proactively opted in***?
- **Q5a.5:** Does the company state that it provides users with ***options to control*** how their information related to their sexual orientation is used for the development of algorithmic systems, including but not limited to AI classifiers and machine learning models?

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*.

- **Q5b.1:** Does the company's policy state that users have control over the company's ***collection of user information*** related to their gender identity?
- **Q5b.2:** Does the company's policy state that users have control over whether the company can ***attempt to infer*** the gender identity of users based on metadata and other signals related to their behavior?

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

- **Q5b.3:** Does the company's policy state that users, without deleting their account, ***can delete*** each type of user information related to their gender identity that the platform has collected?
- **Q5b.4:** Does the company's policy state that it does ***not*** use information related to a user's gender identity for the development of algorithmic systems — including but not limited to AI classifiers and machine learning models — unless the user has ***proactively opted in***?
- **Q5b.5:** Does the company state that it provides users with ***options to control*** how their information related to their gender identity is used for the development of algorithmic systems, including but not limited to AI classifiers and machine learning models?

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*.

- **Q6.1:** Does the company's policy state that it does ***not*** recommend content based on a user's disclosed or inferred sexual orientation or gender identity, unless the user has ***proactively opted in***?
- **Q6.2:** Does the company's policy explain ***how*** users can ***opt in*** to seeing recommended content based on their disclosed or inferred sexual orientation or gender identity at any time?
- **Q6.3:** Does the company's policy state ***that*** users can ***opt out*** of seeing recommended content based on their disclosed or inferred sexual orientation or gender identity at any time?
- **Q6.4:** Does the company's policy explain ***how*** users can ***opt out*** of seeing content based on their disclosed or inferred sexual orientation or gender identity?

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*.

- **Q7.1:** Do the company's policies state that it does ***not*** permit advertisers to ***target*** users based on their disclosed or inferred sexual orientation or gender identity, unless the user has ***proactively opted in***?
- **Q7.2:** Do the company's policies state that it does ***not*** permit advertisers to ***exclude*** users from seeing ads based on their disclosed or inferred sexual orientation or gender identity?
- **Q7.3:** Do the company's policies state that it provides users with options to ***control*** how their sexual orientation or gender identity are used for targeted advertising?
- **Q7.4:** Do the company's policies explain ***how*** users can ***opt in*** to being targeted with ads by third-party advertisers based on their sexual orientation or gender identity?
- **Q7.5:** Do the company's policies state ***that*** users can ***opt out*** of being targeted with ads by third-party advertisers based on their disclosed or inferred sexual orientation or gender identity at any time?

- **Q7.6:** Do the company's policies explain ***how*** users can ***opt out*** of being targeted with ads by third-party advertisers based on their disclosed or inferred sexual orientation or gender identity at any time?
- **Q7.7:** Does the company explain the ***processes and technologies*** that it uses to identify advertisers that violate these rules (e.g. via human and/or automated content moderation)?

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.

- **Q8.1:** Does the company state that it prohibits ***advertising content*** that promotes hate, harassment, and violence against LGBTQ individuals on the basis of protected characteristics?
- **Q8.2:** Does this policy contain a prohibition against ***ads promoting*** *so-called "conversion therapy*?"
- **Q8.3:** Does the company state that, at least once per year, it engages with LGBTQ and human rights organizations on best practices around identifying harmful "conversion therapy" content in advertising?
- **Q8.4:** Does the company explain the ***processes and technologies*** that it uses to identify advertising content or accounts that promote hate, harassment, and violence against LGBTQ people on the basis of protected characteristics (e.g. via human and/or automated content moderation)?

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.

- **Q9.1:** Does the company publish data on the number of ***pieces of content*** restricted based on violation of policies protecting LGBTQ people from hate, harassment, and violence?
- **Q9.2:** Does the company publish data on the number of ***accounts*** restricted based on violation of policies protecting LGBTQ people from hate, harassment, and violence?
- **Q9.3:** Does the company publish data on the number of ***pieces of content subsequently unrestricted*** — after content was wrongfully restricted for violation of policies designed to protect LGBTQ people from hate, harassment, and violence on the platform?
- **Q9.4:** Does the company publish data on the number of ***accounts subsequently unrestricted*** — after accounts were wrongfully restricted for violation of policies designed to protect LGBTQ people from hate, harassment, and violence on the platform?
- **Q9.5:** Does the company publish this data at least four times per year?

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.

- **Q10.1:** Does the company explain the ***concrete steps*** it takes to minimize wrongful demonetization and removal of legitimate content and accounts related to LGBTQ topics and issues?
- **Q10.2:** Does the company state that it ***regularly meets*** with LGBTQ content creators, or stakeholders advocating on their behalf, to address wrongful removals, suspensions, or demonetization on the platform?
- **Q10.3:** Does the company ***have a policy*** that prohibits targeted and malicious reporting of LGBTQ users and content?

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.

- **Q11.1:** Does the company publish data on the ***number of pieces*** of content related to LGBTQ topics and issues ***wrongly removed, filtered, demoted, or demonetized*** for policy violations?
- **Q11.2:** Does the company publish data on the ***number of accounts*** that produce content related to LGBTQ topics and issues ***wrongly removed, filtered, demoted, or demonetized*** for policy violations?
- **Q11.3:** Does the company publish this data at least four times per year?

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.

- **Q12.1:** Does the company state that it conducts annual, mandatory content moderator trainings that are dedicated to the ***safety, privacy, and expression of protected characteristic groups***?
- **Q12.2:** Does the company state that these trainings address ***LGBTQ safety, privacy, and expression***?

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.

- **Q13.1:** Does the company state that it has a dedicated LGBTQ policy lead?
- **Q13.2:** Does the company state that, at least once per year, it solicits key stakeholder guidance from LGBTQ rights organizations?
- **Q13.3:** Does the company state that, at least once per year, it trains all relevant policy, product, and Trust & Safety staff on how to best ensure the fulfillment of its commitments to LGBTQ safety, privacy, and expression?

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.

- **Q14.1:** Does the company express a public commitment to taking ***proactive steps*** to diversify its workforce?
- **Q14.2:** Does the company state it has an internal HR reporting mechanism that allows employees to ***voluntarily*** self-disclose their sexual orientation and gender identity?
- **Q14.3:** Does the company publicly report this voluntarily disclosed data in its ***workforce numbers***?
- **Q14.4:** Does the company's data on diversity in its workforce break out these ***numbers by different teams*** (e.g. engineering and product teams, policy teams)?
- **Q14.5:** Does the company publish this data at least once per year?