

SOCIAL MEDIA SAFETY INDEX

2024

TABLE OF CONTENTS

We Are All Stronger Together: Letter from GLAAD President and CEO Sarah Kate Ellis	2
Executive Summary, Key Conclusions and Recommendations, and Methodology	3
2024 Social Media Safety Index Platform Scorecard:	9
TikTok	11
Facebook	12
Instagram	13
YouTube	14
Threads	15
X/Twitter	16
Rising Hate, Rising Profits: The Anti-LGBTQ Hate Industrial Complex	17
Improving Policies, Creating Accountability, and Raising Awareness: The Year in GLAAD Social Media Safety Program Advocacy Work (2023-2024)	19
Anti-LGBTQ Online Hate Tropes and Connections to Offline Harm	23
Fighting Hate and Lies: Recent Progress On Two Key LGBTQ Safety Protections	25
LGBTQ Content Suppression Case Study: Men Having Babies (Instagram)	27
Spotlight on Data Protection	28
Focus on AI: Risks for LGBTQ People	29
Solutions for All: Legislative and Regulatory Approaches to Social Media and Tech Accountability	31
Showcase of 2023-2024 Reports on LGBTQ Social Media Safety	33
Acknowledgements	40
<u>2024 SMSI Articles and Reports Appendix</u>	
<u>2024 Bibliography of Anti-LGBTQ Online To Offline Real World Harms</u>	
<u>2024 Appendix of Policies, Guidelines and Reporting Links</u>	

WE ARE ALL STRONGER TOGETHER.



A LETTER FROM GLAAD PRESIDENT AND CEO SARAH KATE ELLIS

"Now in its fourth year, GLAAD's Social Media Safety Index (SMSI) continues to stand as the leading analysis of anti-LGBTQ online hate and disinformation, with a determined focus on increasing safety for LGBTQ social media users. In this report, and in the daily work of our Social Media Safety program, we illuminate the dire landscape of LGBTQ safety across the major social media platforms, as well as shining the light on how these online harms translate to offline harms — from the hundreds of anti-LGBTQ legislative attacks attempting to retract our basic rights to the alarming current state of anti-LGBTQ physical threats and violence. (A chilling 2023 report documented more than 700 anti-LGBTQ hate and extremism incidents from November 2022 - November 2023). Such real-world harms include hate-crime homicides and assaults, bomb threats, death threats, vandalism, burning of LGBTQ Pride flags and other community symbols, and an array of other anti-LGBTQ violence. (Another similar report documented more than 60 incidents of anti-LGBTQ incidents targeting religious institutions from June 2022 - January 2024.)

As we have seen over and over again — there is a direct line from dangerous online rhetoric and targeting to violent offline behavior against the LGBTQ community. Sadly, we also see — over and over again — how social

media companies refuse to enforce their own rules to protect LGBTQ people and other marginalized groups despite their assertions that hate speech, bullying, and harassment are not allowed on their platforms.

Amidst this frightening landscape, GLAAD's Social Media Safety program continues our monitoring, documenting, reporting, and advocacy work. The expectation of product safety is a fundamental aspect of every business and industry. It should be no different for tech companies and social media platforms. GLAAD urgently calls on all of these companies to improve the safety of their products — for the sake of their LGBTQ users and the invaluable creators whose labor they profit from, for the sake of their advertisers, and for everyone.

While we have seen important achievements this past year in our efforts, an enormous amount of work lies ahead as we advocate for platforms to fulfill their commitments to LGBTQ safety, privacy, and expression. Weaponized anti-LGBTQ hate and disinformation, and especially anti-trans hate, will continue to be an extraordinarily harmful and dangerous problem, and will no doubt escalate across social media leading up to the U.S. election in November.

At GLAAD we continue to believe that companies, brands, and institutions across civil society — including social media platforms — can stand up for the pluralistic values we all share. As documented in GLAAD's 2023 Advertising Visibility Index report, we know that a commitment to such priorities is also good for business. It is incumbent upon us all to stand up — together — and say these relentless attacks on our community (rhetorical, legislative, judicial, and physical) must stop.

GLAAD PRESIDENT AND CEO

SARAH KATE ELLIS

A handwritten signature of Sarah Kate Ellis in dark ink, written in a cursive style.

EXECUTIVE SUMMARY, KEY CONCLUSIONS AND RECOMMENDATIONS, AND METHODOLOGY

“When social media platforms use algorithms that amplify hate, fail to enforce their own policies against hate, and profit off the targeting of communities, people suffer — and democracy is undermined.” — The Leadership Conference on Civil and Human Rights ¹

In addition to the annual Platform Scorecard ratings and recommendations, this year’s Social Media Safety Index (SMSI) report provides an overview of the current state of LGBTQ social media safety, privacy, and expression, including sections on: the economy of hate and disinformation, predominant anti-LGBTQ tropes, policy best practices, suppression of LGBTQ content, the connections between online hate and offline harm, regulation and oversight, AI, data protection and privacy, and more.

In the 2024 SMSI Platform Scorecard, while some platforms have shown improvements in their scores and others have fallen, overall the scores remain abysmal, with all platforms other than TikTok receiving F grades (TikTok reached a D+).

An Important Note About The Scorecard

Ratings: While the six platforms all have policies prohibiting hate and harassment on the basis of sexual orientation, gender identity and/or expression, and other protected characteristics, **the SMSI Scorecard does not include indicators to rate them on enforcement of those policies.** GLAAD and other monitoring organizations repeatedly encounter failures in enforcement of community guidelines across major platforms. However, given the difficulty involved in assessing enforcement methodologically — which is further complicated by a relative lack of transparency from the companies — these failures are not quantified in the scores below.

D+

TIKTOK

receives a score of 67%,
(+10 points from 2023)

F

FACEBOOK

receives a score of 58%
(-3 points from 2023)

F

YOUTUBE

receives a score of 58%
(+4 points from 2023)

F

INSTAGRAM

receives a score of 58%
(-5 points from 2023)

F

THREADS

receives a score of 51%
(new 2024 rating)

F

X (TWITTER)

receives a score of 41%
(+8 points from 2023)

Specific LGBTQ safety, privacy, and expression issues identified in the Platform Scorecard, and in the SMSI report in general, include: inadequate content moderation and problems with policy development and enforcement (including issues with both failure to mitigate anti-LGBTQ content and over-moderation/suppression of LGBTQ users); harmful algorithms and lack of algorithmic transparency; inadequate transparency and user controls around data privacy; and an overall lack of transparency and accountability across the industry, among many other issues — all of which disproportionately impact LGBTQ users and other marginalized communities who are uniquely vulnerable to hate, harassment, and discrimination.

These areas of concern are exacerbated for those who are members of multiple communities, including people of color, women, immigrants, people with disabilities, religious minorities, and more. Social media platforms should be safe for everyone, in all of who we are.

Like the [2023 Social Media Safety Index](#) (SMSI), this year's report illuminates the epidemic of anti-LGBTQ hate, harassment, and disinformation across the major social media platforms: TikTok, X, YouTube, and Meta's Instagram,² Facebook, and newly added Threads. The report especially makes note of the high-follower hate accounts and right-wing figures who continue to manufacture and circulate most of this activity. The devastating impact of hate, disinformation, and conspiracy theory content continues to be one of the most consequential issues of our time, with hate-driven and politically-motivated false narratives running rampant online and offline and causing real-world harm to our collective [public health](#), [safety](#), and [democracy](#). As a major January 2024 [Online Extremism](#) report from the U.S. Government Accountability Office (GAO) notes: "Research suggests the occurrence of hate crimes is associated with hate speech on the internet [and] suggests individuals radicalized on the internet can perpetrate violence as lone offenders."

Targeting historically marginalized groups, including LGBTQ people, with fear-mongering, lies, and bigotry is both an intentional strategy of bad actors for attempting to consolidate political power, as well as being a lucrative enterprise (for the right-wing figures and groups

that drive such campaigns and tropes, and for the multi-billion dollar tech companies that host them). It's clear that regulatory oversight of the entire industry is necessary to address, as the Global Disinformation Index puts it: "the perverse incentives that drive the corruption of our information environment."³

In addition to egregious levels of inadequately moderated anti-LGBTQ material across platforms (for example see GLAAD's recent report, [Unsafe: Meta Fails to Moderate Extreme Anti-trans Hate Across Facebook, Instagram, and Threads](#)), we also see the corollary problem of over-moderation of legitimate LGBTQ expression — including wrongful takedowns of LGBTQ accounts and creators, mis-labeling of LGBTQ content as "adult," unwarranted demonetization of LGBTQ material under such policies, shadowbanning, and similar suppression of LGBTQ content. Meta's recent policy change limiting⁵ algorithmic eligibility of so-called "political content" ([partly defined by Meta](#) as: "social topics that affect a group of people and/or society large") is especially concerning.⁷

There's nothing unusual or surprising about the fact that companies prioritize their corporate profits and bottom line over public safety and the best interests of society (which is why we have [regulatory agencies](#) to oversee major industries). Unfortunately social media companies are currently woefully under-regulated. And platform safety concerns have risen, particularly over the past year, as so many of the world's largest social media companies have slashed content moderation teams. In an NBC News feature about the wave of layoffs an anonymous X staffer reflected that, "hateful conduct and potentially violent conduct has really increased."⁸

Such downsizing is also negatively impacting the fight against disinformation on social media platforms, as outlined at length in the 2023⁹ Center for Democracy and Technology report [Seismic Shifts](#). The impacts of these business decisions on our information environment will no doubt continue to worsen as major social media companies have shifted towards policies allowing highly-consequential, known false information and hate-motivated extremism to proliferate on their platforms.¹⁰

2 [Online Amplifiers of Anti-LGBTQ+ Extremism](#)

3 [Disinformation as Adversarial Narrative Conflict](#)

4 [Social Media Platforms Target, Restrict and Silence LGBTQ+ Content, Experts Say](#)

5 [Shadowbanning is real: How social media decides who to silence](#)

6 [Oversight Board presses Meta to revise 'convoluted and poorly defined' nudity policy](#)

7 [Instagram's shift away from political content raises questions about what counts as 'political'](#)

8 [Tech layoffs hit 'trust and safety' teams, raising fears of backsliding efforts to curb online abuse](#)

9 [Seismic Shifts: How Economic, Technological, and Political Trends are Challenging Independent Counter-Election-Disinformation in the United States](#)

10 [Following Musk's lead, YouTube and Facebook are giving up on policing conspiracies](#)

KEY CONCLUSIONS:

- Anti-LGBTQ rhetoric and disinformation on social media translates to real-world offline harms.
- Platforms are largely failing to successfully mitigate dangerous anti-LGBTQ hate and disinformation and frequently do not adequately enforce their own policies regarding such content.
- Platforms also disproportionately suppress LGBTQ content, including via removal, demonetization, and forms of shadowbanning.
- There is a lack of effective, meaningful transparency reporting from social media companies with regard to content moderation, algorithms, data protection and data privacy practices.



CORE RECOMMENDATIONS:

- Strengthen and enforce existing policies that protect LGBTQ people and others from hate, harassment, and misinformation/disinformation, and also from suppression of legitimate LGBTQ expression.
- Improve moderation including training moderators on the needs of LGBTQ users, and moderate across all languages, cultural contexts, and regions. This also means not being overly reliant on AI.¹¹
- Be transparent with regard to content moderation, community guidelines, terms of service policy implementation, algorithm designs, and enforcement reports. Such transparency should be facilitated via working with independent researchers.¹²
- Stop violating privacy/respect data privacy. To protect LGBTQ users from surveillance and discrimination, platforms should reduce the amount of data they collect, infer, and retain. They should cease the practice of targeted surveillance advertising, including the use of algorithmic content recommendation.¹³ And should implement end-to-end encryption by default on all private messaging to protect LGBTQ people from persecution, stalking, and violence.¹⁴
- Promote civil discourse and proactively message expectations for user behavior, including respecting platform hate and harassment policies.¹⁵

THE CHIEF EMPHASIS OF THIS REPORT

is on the state of LGBTQ safety with regard to the platforms; while it should be acknowledged that there are many positive initiatives and activities these companies continue to implement to protect their LGBTQ users, they simply must do so much more. It is also vitally important to remember the crucial role social media platforms play for LGBTQ people. As noted in our section on regulatory oversight, proposed legislative social media safety solutions should be mindful of not causing unintended harm to LGBTQ users, especially LGBTQ youth. As a US-based organization GLAAD's focus is primarily domestic; however there are enormous global implications of this work, and we call upon platforms to take responsibility for the worldwide impacts and safety of their products.

¹² [Transparency is the best first step towards better digital governance](#)

¹³ 81% of Americans support [banning companies](#) from collecting people's personal data for ad targeting.

¹⁴ [OPEN LETTER: Make DMs Safe](#)

¹⁵ [For a Civil Internet – How the tone of online conversations can build trust](#)

METHODOLOGY

IN PREPARING THIS YEAR'S REPORT,

GLAAD reviewed thought leadership, research, journalism, and findings across the field of social media safety and platform accountability, consulting with the GLAAD SMSI advisory committee and other organizations and leaders in technology and social justice. The [2024 SMSI Articles and Reports Appendix](#) features links to important relevant topics and developments in the field of LGBTQ social media safety (from the implications of Meta's recent Instagram/Threads "[political content](#)" [policy](#) implementation, to anti-trans hate trope trends like "[transvestigation](#)," to the super-charging of disinformation-for-profit via financially-incentivized [anonymous accounts on X](#)). Please also refer to the [2023](#), [2022](#), and [2021](#) SMSI reports and the [2024 Bibliography of Anti-LGBTQ Online To Offline Real World Harms](#), which remain substantial and valuable resources on these topics.

The centerpiece of this year's report is the Platform Scorecard. Spearheaded by GLAAD's independent researcher Andrea Hackl, working in partnership with [Ranking Digital Rights](#) (RDR), the 2024 Social Media Safety Index Platform Scorecard looks at 12 LGBTQ-specific indicators and evaluates each of the six major platforms drawing on RDR's standard methodology to generate numeric ratings for each product with regard to LGBTQ safety. Researchers interested in digging deeper into the 12 LGBTQ-specific indicators may explore this 2024 [Research Guidance](#) document.

While this report is focused on the six major social media platforms, we know that other companies and platforms — from Snapchat to Spotify, Amazon to Zoom — can benefit from these recommendations as well. We strongly urge all platforms and companies to make the safety of their LGBTQ customers and users an urgent priority.

ON THE FIREWALL BETWEEN FINANCIAL SPONSORSHIP AND GLAAD'S ADVOCACY WORK

Several of the companies that own products and platforms listed in this report are current financial sponsors of GLAAD, a 501(c)3 non-profit. A firewall exists between GLAAD's advocacy work and GLAAD's sponsorships and fundraising. As part of our media advocacy and work as a media watchdog, GLAAD has and will continue to publicly call attention to issues that are barriers to LGBTQ safety, as well as barriers to fair and accurate LGBTQ content and coverage — including issues originating from companies that are current financial sponsors.

SOCIAL MEDIA SAFETY INDEX ADVISORY COMMITTEE

The GLAAD SMSI advisory committee includes an array of respected leaders working at the intersections of tech accountability and LGBTQ social justice.

ALOK (THEY/THEM)

Internationally acclaimed writer, performer, and media personality

LUCY BERNHOLZ (SHE/HER)

Director, Digital Civil Society Lab at Stanford University

ALEJANDRA CARABALLO, ESQ. (SHE/HER)

Clinical Instructor, Cyberlaw Clinic, Berkman Klein Center for Internet & Society, Harvard Law School

JOAN DONOVAN PHD (SHE/HER)

Founder, Critical Internet Studies Institute and Assistant Professor of Journalism and Emerging Media Studies, Boston University

JELANI DREW-DAVI (THEY/THEM)

Senior Communications Specialist, Kairos

LIZ FONG-JONES (SHE/HER)

Field CTO, Honeycomb

EVAN GREER (SHE/THEY)

Director, Fight for the Future

LEIGH HONEYWELL (SHE/HER)

CEO and Co-founder, Tall Poppy

MARIA RESSA (SHE/HER)

Journalist and CEO, Rappler

TOM RIELLY (HE/HIM)

Founder, TED Fellows Program and Founder, PlanetOut.com

DR. SARAH T. ROBERTS (SHE/HER)

Faculty Director, UCLA Center for Critical Internet Inquiry and Associate Professor, Department of Gender Studies at UCLA

BRENNAN SUEN (HE/HIM)

Deputy Director of External Affairs, Media Matters for America

KARA SWISHER (SHE/HER)

Renowned journalist and Editor-at-Large at New York Magazine, and host of "On with Kara Swisher" and "Pivot" podcasts

MARLENA WISNIAK (SHE/HER)

Senior Advisor, Digital Rights, European Center for Not-for-Profit Law

2024 SOCIAL MEDIA SAFETY INDEX PLATFORM SCORECARD

The 2024 SMSI Platform Scorecard consists of twelve indicators that draw on best practices and guidelines from the Ranking Digital Rights (RDR) Big Tech Scorecard.¹⁶ The Scorecard evaluates six major social media platforms: Facebook, Instagram, and Threads (whose parent company is Meta), TikTok (parent company: ByteDance); X (formerly known as Twitter), and YouTube (parent company: Alphabet/Google).

THE SCORECARD

Was developed in 2022 in close collaboration with Goodwin Simon Strategic Research (GSSR), Ranking Digital Rights (RDR), expert stakeholders working at the intersections of technology and human rights, and input from the SMSI advisory committee. In creating the Scorecard, GSSR utilized RDR's evaluation, methodology, and scoring processes and guidelines. Full details can be found [here](#). In essence, the companies receive an average score of their performance across all the indicators evaluated. Each indicator has a list of elements, and companies receive credit (full, partial, or no credit) for each element they fulfill. The evaluation includes an assessment of disclosure for every element of each indicator. As part of the process, following RDR's methodology, all companies are invited to provide written feedback as well as additional source documents. For this year's process, all the companies evaluated actively participated in the feedback process.

Note that these 12 indicators only address some of the issues impacting LGBTQ users. Further, the recommendations below reflect only some of the important steps that companies should take. Much greater detail and analysis can be found in the [2024 Research Guidance](#). The full scoring sheets are also available [here](#).

AN IMPORTANT NOTE ABOUT THE SCORECARD RATINGS:

As mentioned above, the SMSI Scorecard does not include indicators to rate platforms on enforcement of their policies. GLAAD and other monitoring organizations repeatedly encounter failures in enforcement of community guidelines across major platforms. However, given the difficulty involved in assessing enforcement methodologically these failures are not quantified in the scores below.

SMSI PLATFORM SCORECARD

LGBTQ-SPECIFIC INDICATORS

01. The company should disclose a policy commitment to protect LGBTQ users from harm, discrimination, harassment, and hate on the platform.
02. The company should disclose an option for users to add pronouns to user profiles.
03. The company should disclose a policy that expressly prohibits targeted deadnaming and misgendering of other users.
04. The company should clearly disclose what options users have to control the company's collection, inference, and use of information related to their sexual orientation and gender identity.
05. The company should disclose that it does not recommend content to users based on their disclosed or inferred sexual orientation or gender identity, unless a user has opted in.
06. The company should disclose 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 opted in.
07. The company should disclose that it prohibits advertising content that could be harmful and/or discriminatory to LGBTQ individuals.
08. The company should disclose the number of accounts and pieces of content it has restricted for violations of policies protecting LGBTQ individuals.
09. The company should take proactive steps to stop demonetizing and/or wrongfully removing legitimate content related to LGBTQ issues in ad services.
10. The company should disclose training for content moderators, including those employed by contractors, that trains them on the needs of vulnerable users, including LGBTQ users.
11. The company should have internal structures in place to implement its commitments to protect LGBTQ users from harm, discrimination, harassment, and hate within the company.
12. The company should make a public commitment to continuously diversifying its workforce, and ensure accountability by periodically publishing voluntarily self-disclosed data on the number of LGBTQ employees across all levels of the company.

IN THE 2024 SMSI PLATFORM SCORECARD,

TikTok receives a score of 67, a 10-point increase from its 2023 score. The company has made several notable improvements to its policies. For instance, in its revised "[TikTok's Anti-Discrimination Ad Policy](#)," the company explicitly prohibits advertisers from wrongfully targeting or excluding users from seeing ads based on their sexual orientation or gender identity. The company has also improved its transparency regarding LGBTQ users' control over their own information. On a recently launched [portal](#) that contains policy disclosures and resources for LGBTQ users, the company provides that it does not collect users' sexual orientation information. Moreover, TikTok discloses on this page that users who share information related to their sexual orientation and/or gender identity can delete this information. Of the scored platforms, TikTok is one of only two with a policy expressly prohibiting both targeted misgendering and deadnaming.

It also continues to be the only company that does not require user self-reporting of a potential community guidelines violation in this area, and also employs technologies, human review, and/or reporting from other users to detect violations to the policy.

However, the company's policies fail to adequately protect LGBTQ users in other key areas. The company discloses only limited information regarding the proactive steps it takes to address wrongful demonetization and removal of LGBTQ creators and content from ad services on the platform. TikTok also does not disclose any data showing how many pieces of content and accounts related to LGBTQ issues have been wrongfully demonetized or removed from ad services. While the company makes a public commitment to diversifying its workforce, it does not publish any data on its LGBTQ workforce.

KEY RECOMMENDATIONS:

- 1 Give LGBTQ users greater control over the content they see: LGBTQ users should have more options to control the content they see, and recommendation of content based on their sexual orientation and gender identity should be off by default.
- 2 Be more transparent about the wrongful demonetization and removal of LGBTQ TikTokers: The company should publish comprehensive data on the wrongful demonetization and removal of LGBTQ creators and their content from ad services.

IN THE 2024 SMSI PLATFORM SCORECARD,

Facebook receives a score of 58, a three-point decrease from its 2023 score. Meta has a policy in place protecting transgender, nonbinary, and gender non-conforming users from targeted misgendering. However, the policy requires self-reporting, does not extend to public figures, and Facebook does not have a similar policy in place that protects users from targeted deadnaming. Facebook also has a comprehensive policy that prohibits advertisers from discriminating against users based on protected characteristics such as sexual orientation and gender identity. According to this policy, advertisers may not use Meta’s audience selection tools to “wrongfully target specific groups of people for advertising,” “wrongfully exclude specific groups of people from seeing their ads,” and “include discriminatory content in their ads.”

Meta’s policies fail to adequately protect LGBTQ users in several other key areas. Facebook currently does not have a policy prohibiting targeted deadnaming. While Meta’s “Gender Identity Policy and User Tools” policy discloses a training held in 2022 that educated content moderators on enforcement of the company’s gender identity policy, it is not clear whether the company has conducted a similar training in the last year. Facebook continues to provide little transparency regarding the options users have to control the company’s collection and inference of user data related to their sexual orientation and gender identity. The company’s transparency report continues to disclose limited data on the number of pieces of content and accounts restricted for violations to policies protecting LGBTQ users from hate, harassment, and discrimination.

KEY RECOMMENDATIONS:

Give LGBTQ users full control over the company’s handling of user information related to sexual orientation and gender identity: Meta should provide users with clear options to control the company’s collection and inference of user information related to their sexual orientation and gender identity.

- 2 Protect transgender, nonbinary, and gender non-conforming users (including public figures) from targeted deadnaming: The company should adopt a comprehensive policy that prohibits targeted deadnaming on Facebook, explain in detail how this policy is enforced, and should not require self-reporting (the company should also update its targeted misgendering policy to not require self-reporting, and to protect public figures). The company should also disclose that it employs various processes and technologies — including human and automated content moderation — to detect content and behaviors violating these policies.

- 3 Publish comprehensive data on how policies protecting LGBTQ users are enforced: The company should regularly publish data on the actions it has taken to enforce policies protecting LGBTQ users, and break out this data by different sub-policies.

INSTAGRAM

SMSI SCORE:
58%

F

IN THE 2024 SMSI PLATFORM SCORECARD,

Instagram receives a score of 58, a five-point decrease from last year's score. Instagram's [Community Guidelines](#) continue to prohibit hate, discrimination, and harassment against LGBTQ users on the platform. The company also discloses comprehensive information on the types of behaviors and content that are prohibited under its protected groups policy. While Meta continues to have a policy in place that protects transgender, nonbinary, and gender non-conforming users from targeted misgendering, this policy requires self-reporting,¹⁸ does not extend to public figures, and the company does not disclose a similar prohibition against targeted deadnaming.

The company also continues to fall short of providing adequate transparency in several other key areas. According to company disclosure, Instagram's feature allowing users to add pronouns to their profiles is currently not available for all users. In other important areas, the company has taken steps back when it comes to user transparency. While Meta's "[Gender Identity Policy and User Tools](#)" policy discloses that content moderators were trained on gender identity policy enforcement in 2022, it is not clear whether the company has conducted a similar training in the last year.

KEY RECOMMENDATIONS:

Provide all users with tools for self-expression: The company should make its feature allowing users to add their gender pronouns to their profiles available to all users, and give users more granular options to control who can see their pronouns.

Protect transgender, nonbinary, and gender non-conforming users (including public figures) from targeted deadnaming: The company should adopt a comprehensive policy that prohibits targeted deadnaming on Instagram, explain in detail how this policy is enforced, and should not require self-reporting (the company should also update its targeted misgendering policy to not require self-reporting, and to protect public figures). The company should also disclose that it employs various processes and technologies — including human and automated content moderation — to detect content and behaviors violating these policies.

YOUTUBE

SMSI SCORE:
58% **F**

IN THE 2024 SMSI PLATFORM SCORECARD,

YouTube earns a score of 58, a score improvement of four points. In a change from last year's evaluation, YouTube has launched a dedicated feature allowing YouTube creators to add gender pronouns to their channels. YouTube also provides creators with limited options to customize who can see their pronouns, allowing them to choose whether to display their pronouns to everyone or to their subscribers only. However, no disclosure was found that indicates that a similar feature is available for users. The company continues to make a public commitment to taking proactive steps to diversify its workforce, and publishes voluntarily self-disclosed data on the number of its LGBTQ employees.

The company continues to fall short of providing adequate policy protections for its LGBTQ users in several other key areas. Notably, YouTube is the only company evaluated in the SMSI that has no policy in place that protects transgender, nonbinary, and gender non-conforming users from targeted misgendering and deadnaming. While the company prohibits targeted advertising based on users' disclosed or inferred sexual orientation or gender identity, no similar disclosure was located that prohibits advertisers from excluding users from seeing ads based on their disclosed or inferred sexual orientation or gender identity. The company continues to provide only limited information regarding the steps it takes to address the wrongful demonetization, filtering, and removal of LGBTQ creators and content from ad services.

KEY RECOMMENDATIONS:

Make an express policy commitment to protect transgender, nonbinary, and gender non-conforming users (including public figures) from targeted misgendering and deadnaming: The company should adopt a policy that protects users from targeted misgendering and deadnaming that does not require self-reporting; this policy should also protect public figures, and YouTube should also disclose that it employs various processes and technologies — including human and automated content moderation — to detect content and behaviors violating these policies.

Prevent advertisers from wrongfully excluding LGBTQ users from seeing ads: The company should disclose that it does not permit advertisers to exclude users from seeing ads based on their disclosed or inferred sexual orientation and/or gender identity.

Show greater commitment to addressing wrongful demonetization and removal of LGBTQ creators and their content: The company should disclose comprehensive information on the concrete steps it takes to minimize wrongful demonetization and removal of legitimate content related to LGBTQ issues from ad services. YouTube should also publish comprehensive data on the wrongful removal of LGBTQ creators and their content.

THREADS

SMSI SCORE:
51%

F

LAUNCHED IN 2023,

Meta's app Threads is new to the SMSI Platform Scorecard, and receives a score of 51. Covered by Instagram's [Community Guidelines](#), Threads has a comprehensive protected groups policy that protects LGBTQ users from hate, harassment, and discrimination on the platform. In its "[Gender Identity Policy and User Tools](#)" policy, and in tier 3 of its [Bullying and Harassment policy](#), the company discloses a policy protecting transgender, nonbinary, and gender non-conforming users from targeted misgendering. However, the policy requires self-reporting,¹⁹ the protections do not extend to public figures, and the company does not disclose a similar disclosure related to targeted deadnaming. In its [2023 Responsible Business Practices Report](#), Meta also discloses a commitment to diversifying its workforce, and publishes voluntarily self-disclosed data on its LGBTQ employees.

However, Meta falls short of providing adequate policy protections for its LGBTQ users on several other important issues. Notably, Threads does not have a policy in place that expressly protects users from targeted deadnaming. While the company has a feature allowing users to add preferred pronouns to their user profiles, the company discloses that this option is currently not available to all users. The company also discloses only limited options for users to control who can see their gender pronouns. Moreover, Meta's most recent Community Guidelines Enforcement report does not disclose any data for Threads, and it is therefore not clear how many pieces of content or accounts were restricted for violations to the platform's policies protecting LGBTQ users.

- 1 Protect transgender, nonbinary, and gender non-conforming users (including public figures) from targeted deadnaming: The company should adopt a comprehensive policy that prohibits targeted deadnaming on Threads, explain in detail how this policy is enforced, and should not require self-reporting (the company should also update its targeted misgendering policy to not require self-reporting and to protect public figures). The company should also disclose that it employs various processes and technologies — including human and automated content moderation — to detect content and behaviors violating these policies.

Provide all users with tools for self-expression: The company should make its feature allowing users to add their gender pronouns to their user profiles available to all users, and provide them with more control over who can see their pronouns.

Be transparent about content and account restrictions: The company should publish transparency reporting for Threads showing the number of pieces of content and accounts restricted for violations to policies protecting LGBTQ users from hate, harassment, and discrimination.

¹⁹ For more information on how self-reporting requirements complicate the enforcement of targeted deadnaming and misgendering policies, please see GLAAD's post: "[All Social Media Platform Policies Should Recognize Targeted Misgendering and Deadnaming as Hate Speech](#)."

X (TWITTER)

SMSI SCORE:
41%

F

IN THE 2024 SMSI PLATFORM SCORECARD,

X/Twitter earns a score of 41, an eight-point increase from 2023. Notably, the company quietly revived its policy prohibiting targeted misgendering and deadnaming on the platform. In this context, the company is one of only two companies evaluated in the SMSI that prohibits both targeted misgendering and deadnaming. However, the company falls short of full credit as it effectively requires users to self-report instances of targeted misgendering and deadnaming (the company discloses it needs to hear from targeted individuals in order to determine whether a policy Violation has occurred),²⁰

and does not disclose whether it also employs human review or automated content moderation to identify violations to the policy.

The company also fails to adequately protect LGBTQ users in other important areas. X/Twitter is the only platform evaluated in the SMSI that does not disclose any information on whether it has a training in place that educates content moderators about the needs of LGBTQ people and other users in protected categories. To date, the company has also failed to renew its commitment to diversifying its workforce, and has not published any employment diversity data in the last year.

KEY RECOMMENDATIONS:

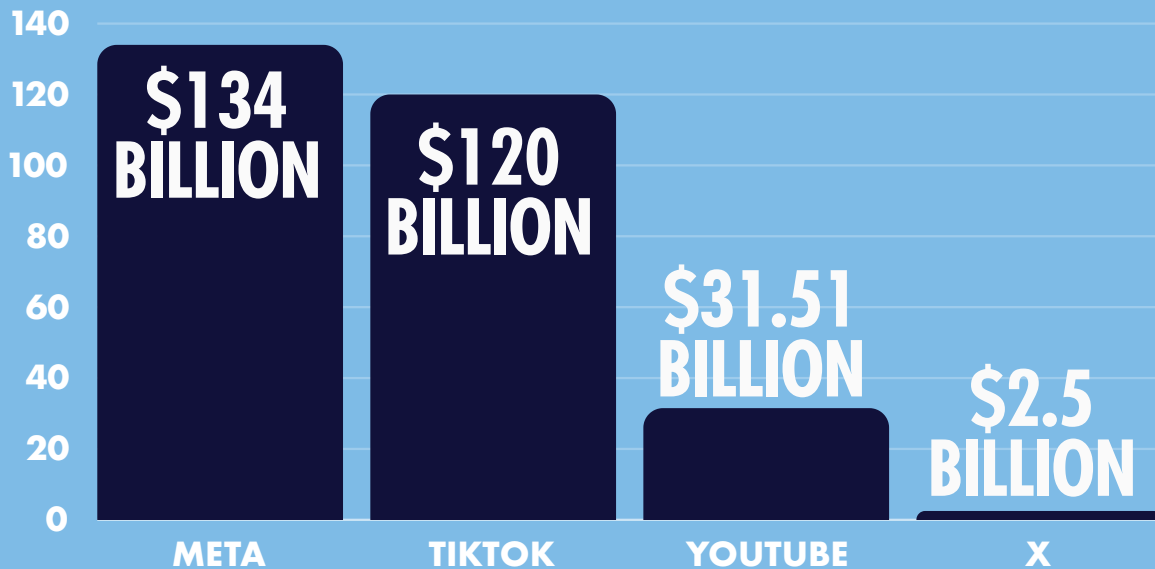
- 1 The company should not require self-reporting of targeted misgendering and deadnaming, and should also disclose that it employs various processes and technologies — including human and/or automated content moderation — to detect content and behaviors violating these policies. The policy should be enforced globally.
- 2 Give users tools for self-expression: The company should disclose a dedicated feature that allows users to add their gender pronouns to their profiles, and give users granular options to control who can see their pronouns.
- 3 Train content moderators on the needs of LGBTQ users: The company should disclose training for content moderators that trains them on the needs of LGBTQ people and other users in protected categories.
- 4 Express commitment to a diverse workforce: The company should make a renewed commitment to diversifying its workforce, and publish annual data showing its progress towards reaching diversity and inclusion goals.

²⁰ For more information on how self-reporting requirements complicate the enforcement of targeted deadnaming and misgendering policies, please see GLAAD's post: "[All Social Media Platform Policies Should Recognize Targeted Misgendering and Deadnaming as Hate Speech](#)."

²¹ Assessment of X/Twitter's policy on targeted misgendering and deadnaming is based on the January 2024 version of its "Abuse and Harassment" policy. In March 2024, the company updated this policy, which now states that the company will enforce its targeted misgendering and deadnaming policy "where required by local laws." As a best practice, policy protections against targeted misgendering and deadnaming should be enforced globally and not depend on local laws.

RISING HATE, RISING PROFITS: THE ANTI-LGBTQ HATE INDUSTRIAL COMPLEX

2023 ANNUAL ADVERTISING REVENUE BY PLATFORM:



FACEBOOK, INSTAGRAM, THREADS — META

According to Statista: "In 2023, Meta Platforms generated a revenue of over 134 billion U.S. dollars, up from 119 billion USD in 2022. The majority of Meta's profits come from its advertising revenue, which amounted to 131.9 billion U.S. dollars in 2022."²²

YOUTUBE

Statista notes that: "In 2023, YouTube's global advertising revenues amounted to approximately 31.51 billion U.S. dollars, up by almost eight percent from the 29.2 billion U.S. dollars in the preceding fiscal period. Whereas the owned online video platform does not generate the same amount of revenue as Google's key segment Search, it is nonetheless a significant money-maker for parent company Alphabet."²³

TIKTOK

Reuters reports that: "ByteDance's revenue of \$120 billion in 2023 was up about 40% from a year earlier, driven by TikTok's exploding growth, although China accounts for a big portion of the company's sales, the FT reported, citing five people with knowledge of the matter."²⁴

X (TWITTER)

Ars Technica reports that: "Twitter's revenue and adjusted earnings reportedly fell about 40 percent year over year in December 2022 amid an advertiser exodus following Elon Musk's takeover. Twitter no longer reports earnings publicly since Musk bought the company and took it private in late October. But Twitter reported the December 2022²⁵ revenue and earnings declines in an update to investors." Bloomberg reports that X: "is on track to bring in roughly \$2.5 billion in advertising revenue in 2023 — a significant slump from prior years."²⁶

²² [Meta: annual revenue and net income 2023](#)

²³ [TikTok's US revenue hits \\$16 bln as Washington threatens ban, FT reports](#)

²⁴ [YouTube global advertising revenues 2023](#)

²⁵ [Twitter revenue, earnings reportedly fell 40% shortly after Musk buyout](#)

²⁶ [Elon Musk's X 2023 Ad Revenue Projected to Slump to About \\$2.5 Billion](#)

LOOKING AT THE EXTRAORDINARILY

high levels of revenue generated by these companies alongside the extraordinarily high (and rising) levels of hate speech on their platforms (see below), it is simply unconscionable to not see a greater investment in product safety from them. Note that since X and TikTok are private companies, these cited revenue numbers have been reported by news outlets but are not confirmed by the companies themselves.

Most social media platforms and other online public spaces (ranging from gaming environments to website

comments fields) continue to inadequately mitigate toxic hate and harmful rhetoric. While these companies are profiting from such hate (refusing to meaningfully enforce their own policies and even financially incentivizing false and hateful content), as consumers we can, and must, convey our expectations and reasonable desires for them to fulfill their own terms of service and to create safe products that do not inflict harms on society and democracy. Though such mitigations are by no means simple, as is true of other industries, tech companies should be required to meet basic product safety standards.



IMPROVING POLICIES, CREATING ACCOUNTABILITY, AND RAISING AWARENESS:

THE YEAR IN GLAAD SOCIAL MEDIA SAFETY PROGRAM ADVOCACY WORK (2023-2024)

SOCIAL MEDIA SAFETY PROGRAM

2024

SOCIAL MEDIA SAFETY INDEX

2023



GLAAD'S SOCIAL MEDIA SAFETY (SMS) PROGRAM

actively 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.

Below are just a few highlights of the program's advocacy work, tackling some of the most urgent problems and issues in the field over the past year.

IF YOU WANT TO KEEP SEEING “POLITICAL CONTENT” FROM NEW AND DIFFERENT LGBTQ GROUPS AND CREATORS IN YOUR INSTAGRAM FEED:

- ➔ GO TO YOUR PROFILE
- ➔ TAP THE  ICON
- ➔ CONTENT PREFERENCES
- ➔ POLITICAL CONTENT
- ➔ DON'T LIMIT POLITICAL CONTENT



A PSA from GLAAD's Social Media Safety Program

IN APRIL 2024, GLAAD

implemented an Instagram [PSA campaign](#) to raise awareness about the alarming implications of Meta's new “political content” policy, as well as launching an [open letter advocacy campaign](#) with Accountable Tech and more than 200 Instagram creators. Meta's trio of problematic policy moves included: the decision to characterize “social topics that affect a group of people and/or society large” as “political content,” the corresponding decision to no longer algorithmically recommend such content to users, and the equally concerning implementation of the new policy — changing all user settings to limit such content by default, rather than making it an opt-in option. [Speaking out to The Washington Post](#), a GLAAD spokesperson explained, Categorizing ‘social topics that affect a group of people and/or society large’ as ‘political content’ is an appalling move. LGBTQ people's lives are simply that, our lives. Our lives are not ‘political content’ or political fodder. This is a dangerous move that not only suppresses LGBTQ voices, but decimates opportunities for LGBTQ people to connect with each other, and allies, as our content will be excluded from the algorithm.”

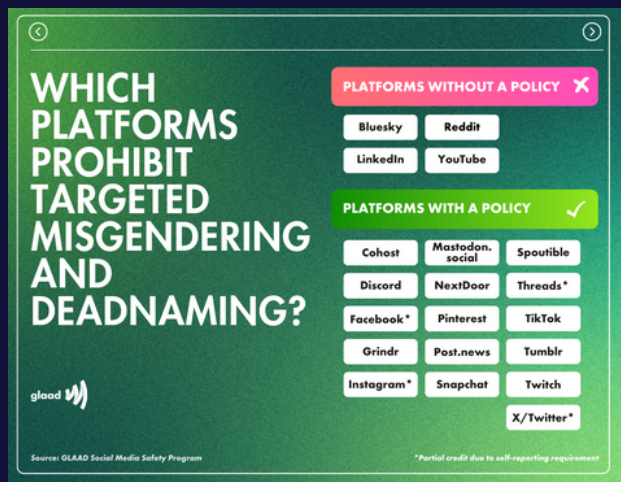
The Washington Post

Meta is failing to curb anti-trans hate, new report says

GLAAD catalogued dozens of instances of anti-trans posts, many by high-profile influencers and media outlets

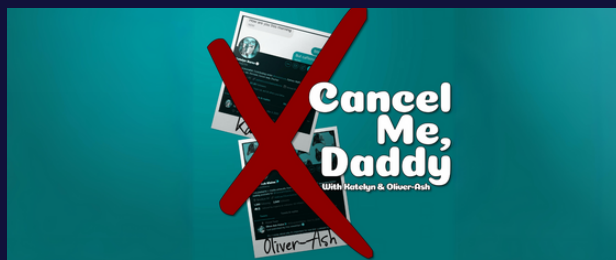
RELEASED IN MARCH 2024,

GLAAD's impactful report — [Unsafe: Meta Fails to Moderate Extreme Anti-trans Hate Across Facebook, Instagram, and Threads](#) — showcases dozens of examples of the shocking, dehumanizing anti-trans content that Meta actively allows on its platforms. All of the content examples in the report were submitted to Meta's standard reporting systems by GLAAD; Meta either replied that posts were not violative or simply did not take action on them. The Washington Post [broke the news with an exclusive feature](#), highlighting several of the content examples (“In one Instagram post cited in the report,” the Post story notes, “a trans person's body is shown twisted on the ground while being beaten to death with stones, which have been replaced with the laughing emoji. The caption reads: ‘[trans flag] people are devils.’”) The Post also notes that Meta's ombudsman Oversight Board had previously “lambasted the company's failure to enforce rules against anti-trans hate and threats.” Although multiple major media outlets requested comments from Meta on the report — including the [Washington Post](#), [Fast Company](#), [Engadget](#), [The Daily Dot](#), and [The Verge](#) — the company did not respond. In the aftermath of the report, seven of the 30 reported items were removed, though no explanation was given by Meta (23 of the egregiously harmful items remain live).



THE SMS PROGRAM

saw major victories over the past year in our advocacy work around two key LGBTQ safety policy protections — prohibiting targeted misgendering and deadnaming and prohibiting the promotion and advertising of harmful so-called “conversion therapy.” After consultation with GLAAD, companies adopting one or both of the policies in recent months include: Snapchat, Discord, Post, Spoutible, Grindr, IFTAS (the non-profit supporting the Fediverse moderator community), and Mastodon.social (Mastodon’s largest server). While these additions do not solve the extremely significant related issue of policy enforcement (which GLAAD is working on as well), these are excellent concrete examples of how the SMS program creates impact. In February 2024 GLAAD released two significant reports documenting how some platforms lag behind in these LGBTQ safety policy protections, but that numerous platforms and apps are increasingly adopting the policies that GLAAD’s SMS program advocates as best practices for the industry.



EDUCATING THE GENERAL PUBLIC

and raising awareness in the media about LGBTQ social media safety issues is another important part of GLAAD’s Social Media Safety program work. In January 2024

SMS program manager Leanna Garfield joined MSNBC contributor Katelyn Burns on her podcast “Cancel Me Daddy” to discuss how social media companies like Meta and Twitter are profiting from online hate speech against LGBTQ people, especially transgender people, and how GLAAD has been at the forefront of pushing these companies to work toward solutions. Listen to the episode here.



HAVING PREVIOUSLY SUBMITTED

an official September 2023 public comment to the important “Post in Polish Targeting Trans People” case, the SMS program was heartened by the January 2024 ruling from the Oversight Board — the body that makes non-binding but precedent-setting rulings about Meta content moderation cases — overturning Meta’s original repeated decisions to not take down a Facebook post targeting transgender people with violent speech, despite more than a dozen requests from community members. The post was an egregious example of anti-trans hate advocating for transgender people to commit suicide, featuring an image of a striped curtain in the blue, pink and white colors of the transgender flag with a text overlay in Polish saying: “New technology. Curtains that hang themselves.” The post was only removed after the Oversight Board alerted Meta. The case illuminates systemic failures with the company’s moderation practices, including widespread failure to enforce their own policies (as noted both by the Oversight Board and GLAAD). Following Meta’s response to the ruling in March 2024, the SMS program issued an additional statement condemning Meta’s continuing negligence and ongoing failure to moderate anti-trans hate. GLAAD’s advocacy work on this is ongoing.



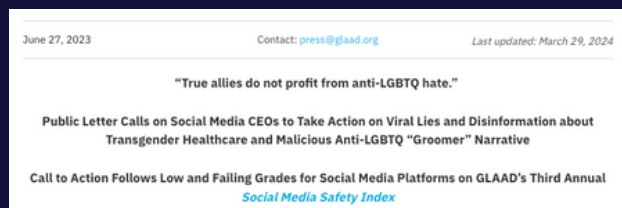
IN OCTOBER 2023,

on the one-year anniversary of Elon Musk's takeover of X/Twitter, GLAAD worked with SMSI advisory committee member [Sarah T. Roberts](#) (UCLA Associate Professor of Information Studies) to publish a reflection on the plummeting state of the platform: "[X is Neither Trusted Nor Safe](#)." Roberts noted that: "Generating blatant disinformation (lies) intended to propagate hate, fear, and dangerous conspiracy theories demonizing members of historically marginalized groups is a familiar strategy for consolidating power and distracting from other real issues. X's decreasing interest in enforcing their own policies, and their retraction of previous LGBTQ policy protections, resulted in a record-low score of 33% in GLAAD's [2023 Social Media Safety Index](#) report in July." X's platform safety has steadily declined — with ongoing [anti-LGBTQ \(and especially anti-trans\) behavior](#), rhetoric, and activity running rampant (much of it [spearheaded by Musk](#)). GLAAD continues to be a member of [Stop Toxic Twitter](#), a coalition of 60+ civil rights and civil society groups calling on X advertisers to demand a safer platform for their brands and for users. 75 out of the top 100 U.S. advertisers have now ceased their ad spending on X/Twitter [according to Sensor Tower](#), and traffic has steadily declined (down 30% according to [a study by Edison Research](#)). Musk's ongoing personal boosting of hate and extremism content and targeting of [civil society groups](#), [journalists](#), and [commentators](#) remains gravely concerning.



WORKING TOGETHER WITH

GLAAD's [Spanish-Language & Latine Media](#) program, September 2023 saw the launch of [Nos Mantenemos Seguros: Guía de seguridad digital LGBTQ](#), a new Spanish translation of our valuable resource: [We Keep Us Safe — Guide to Anti-LGBTQ Online Hate and Disinformation](#). For LGBTQ activists and organizations, journalists and public figures, and for anyone who spends time online, the guide includes tips and easy best practices everyone can implement to improve their digital safety.



IN JUNE 2023,

GLAAD and the Human Rights Campaign facilitated an open letter from [more than 250 LGBTQ celebrities, public figures, and allies](#) urging social media companies to address the epidemic of anti-transgender hate on their platforms. The list of signatories on the letter included such high-profile names as Elliot Page, Laverne Cox, Jamie Lee Curtis, Shawn Mendes, Janelle Monáe, Gabrielle Union, Judd Apatow, Ariana Grande, Jonathan Van Ness, and more. The letter received extensive national media coverage highlighting the plea for platforms to urgently create and share plans for addressing: Content that spreads malicious lies and disinformation about healthcare for transgender youth; accounts and postings that perpetuate anti-LGBTQ extremist hate and disinformation (including the anti-LGBTQ "groomer" trope), in violation of platform policies; dehumanizing, hateful attacks on prominent transgender public figures and influencers; and anti-transgender hate speech, including targeted misgendering, deadnaming, and hate-driven tropes.

ANTI-LGBTQ ONLINE HATE TROPES AND CONNECTIONS TO OFFLINE HARM

IN NOVEMBER 2022,

a gunman opened fire at Club Q, an LGBTQ nightclub in Colorado Springs, Colorado killing five and wounding 19. In the year following this horrific attack, more than 700 anti-LGBTQ incidents of hate and extremism in the US were counted. This includes acts of harassment, like bomb threats and protests; acts of vandalism, like graffiti and arson; and acts of physical assault. These incidents do not occur inside a vacuum — instead, they are driven by an ecosystem of anti-LGBTQ hate and extremism running rampant across social media platforms. Amidst this hateful rhetoric, our research and reporting over this past year has identified and explored three prominent anti-LGBTQ tropes and the real-world harm they have inspired.

FALSE ALLEGATIONS OF PEDOPHILIA / GROOMING BEHAVIORS

The first prevalent anti-LGBTQ trope pertains to false allegations of “pedophilia” and “grooming” used to demonize the LGBTQ community writ large. While grooming behaviors are a very real danger to young children, it is patently false to apply this dangerous label to the LGBTQ community as a whole. The people and accounts spreading this narrative claim that LGBTQ adults are preying on kids simply because they discuss issues related to sexual orientation and gender identity, or live authentically in their own non-cisgender and/or non-straight identities. Content that has been falsely labeled as “grooming” by anti-LGBTQ extremists include cases like a queer couple kissing on TV, showcasing trans athletes in sports, and offering school lesson plans that are LGBTQ-inclusive.

Despite the fact that these actions are in no way related to or inviting of the sexual abuse of minors, the terms “groomers” and “pedophiles” are being used as slurs for LGBTQ people overall — and inciting acts of violence.

According to a 2023 report, approximately 40% of anti-LGBTQ incidents tracked over an 11-month period specifically targeted drag shows and drag performers. The vast majority of these same incidents perpetuated the false “grooming” allegations, including claims that drag queens are attempting to sexualize children.

Extremist groups like Blood Tribe, Patriot Front, and the Proud Boys routinely protest drag shows across the country, where they call event-goers “pedos [pedophiles]” and hold signs with slogans like “Groom dogs, not kids.” In one particularly egregious case in March 2023, an individual associated with the white supremacist “White Lives Matter” network attempted to firebomb a church in Chesterland, Ohio to stop an upcoming drag show, claiming the act was to “save the children.”

“GENDER IDEOLOGY,” “TRANSGENDERISM,” AND FALSE ALLEGATIONS OF INDOCTRINATION

Another prominent anti-LGBTQ trope includes the use of anti-trans buzzwords like “gender ideology” and “transgenderism” to claim that the LGBTQ+ community and its allies aim to indoctrinate or brainwash kids into identifying as transgender. By comparing one’s gender identity to a political ideology, anti-LGBTQ influencers are attempting to skirt content moderation policies and counter critics by claiming they are targeting a belief system, rather than a specific group of people.

These false allegations of “indoctrination” often target specific school districts and educators that offer support to their LGBTQ students. Teachers who, for example, sponsor Gender & Sexuality Alliance (GSA) clubs, place Pride flags on bulletin boards, or include LGBTQ history lessons frequently face claims of “brainwashing” students with a “trans agenda.” Regardless of the fact that LGBTQ identities are inherent to each person and are not something that kids can be coerced into, anti-LGBTQ extremists have latched onto these claims to justify violence against supportive school districts.

In the above-mentioned report, approximately 10% of all anti-LGBTQ incidents nationwide targeted schools and universities. In many of these incidents, researchers observed individuals repeating these same false tropes regarding “indoctrination” and pushing “gender ideology,” including in the dozens of documented protests outside schools and school board meetings. Similarly, a USA Today report found that at least 12 separate schools, school districts, and teachers have been targeted with bomb threats following online targeting by the anti-LGBTQ extremist account Libs of TikTok via Meta and X (note that the account was permanently banned from TikTok in 2021).

FALSE AND MISLEADING ALLEGATIONS TARGETING GENDER-AFFIRMING CARE

Researchers have also tracked false and misleading allegations around gender-affirming care, which is a type of healthcare that supports transgender and gender non-conforming youth and adults. Gender-affirming care is supported by every major medical association and includes an array of services, from surgical and pharmaceutical treatment to mental health and physical healthcare.

Anti-LGBTQ influencers, however, continue to spread dangerous and misleading information regarding gender-affirming care online. Many of these conversations are centered on the false claims that this care amounts to the “castration,” “sterilization” and/or “mutilation” of children. The sensationalized word choice in these tropes is designed to elicit a strong reaction from parents and guardians of young kids — and also potentially inspires violence against healthcare providers who are characterized as perpetrating this perceived harm.

The June 2023 report found that approximately 6% of all anti-LGBTQ incidents nationwide specifically focused on healthcare facilities and providers, the vast majority of which were targeted due to their provision of gender-affirming care. The resulting incidents include numerous cases of bomb threats against hospital facilities, such as the spree of bomb threats targeting Boston Children’s Hospital, as well as doxxing and death threats targeting individual medical providers.

As this research demonstrates, perpetrators of hate and extremism incidents are frequently inspired by the false and sensationalized tropes circulating on social media platforms that are seeded and amplified by anti-LGBTQ influencers. It is imperative that we understand the connection between online hate and offline harm to better protect the safety and rights of LGBTQ people; LGBTQ people should not have to live in fear of such hate-fueled violence.

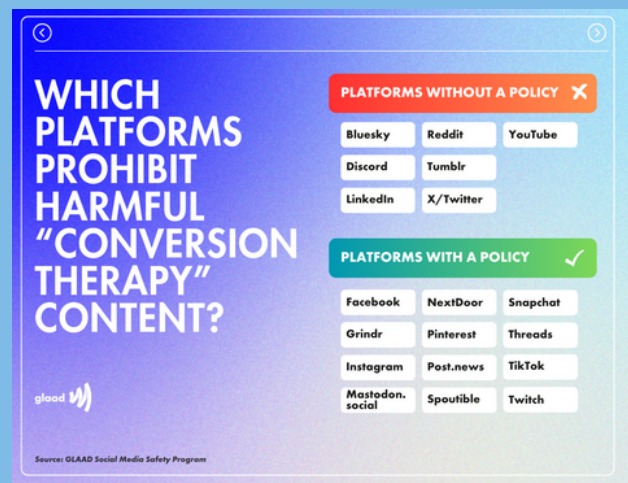
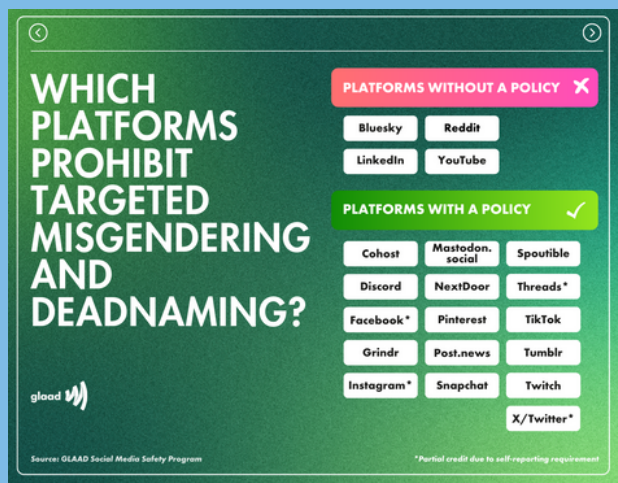
FIGHTING HATE AND LIES: RECENT PROGRESS ON TWO KEY LGBTQ SAFETY PROTECTIONS

SINCE THE FIRST SMSI WAS RELEASED,

in 2021, GLAAD's Social Media Safety Program has been advocating for platforms to update their policies to include additional protections for LGBTQ safety. While there are many areas for improvement, two specific policy additions (the prohibition of targeted misgendering and deadnaming, and banning the promotion of so-called "conversion therapy")²⁸ have been an especially significant area of focus. In 2022, GLAAD and UltraViolet worked with TikTok to have the platform add both of these protections. This past year, the Social Media Safety Program staff have worked extensively with many platforms, apps, and companies on these important policy areas. Companies adopting one or both of the policies in recent months include: Snapchat, Discord, Post, Spoutible, Grindr, IFTAS (the non-profit supporting the Fediverse moderator community), and Mastodon.social (Mastodon's largest server).

A pair of major GLAAD reports released in February 2024 document and explore the current state of these two important LGBTQ safety policy protections that GLAAD's Social Media Safety Program advocates as best practices for the industry.

A sequence of articles in The Advocate celebrated the policy adoptions as "pivotal," and in a February 2024 announcement, IFTAS explained the importance of the policies: "Due to the widespread and insidious nature of expressing anti-transgender sentiments in bad faith, it's imperative to have specific policy addressing this issue. This approach is considered a best practice for two key reasons: it offers clear guidance to users, and it assists moderators in recognizing and understanding the intent behind such statements. It's important to reiterate that the focus is not about accidentally getting someone's pronouns wrong. Rather, our concern centers on deliberate and targeted acts of hate and harassment."



ABOUT HARMFUL SO-CALLED “CONVERSION THERAPY” CONTENT

In its January 2024 report, [Conversion Therapy Online: The Ecosystem in 2023](#), the Global Project Against Hate & Extremism (GPAHE) observes that: “there is worldwide agreement among medical and psychological professionals that conversion therapy is dangerous and causes harm to LGBTQ+ individuals.” As noted above, there is broad consensus and building momentum toward protecting LGBTQ people, and especially [LGBTQ youth](#), from this dangerous practice. However the promotion of such services continues to be widespread on social media platforms, via both organic content and advertising. While the GPAHE report cites that some platforms have made improvements in their mitigations of this content over the past year, they note that “GPAHE found deficiencies on all of the tech and social media platforms in its 2023 research.” There are many examples of simple solutions that platforms can and should urgently implement. For instance, GPAHE’s recommendations include: investing more in non-English language moderation, expanding the use of contextual authoritative information panels that accurately characterize “conversion therapy,” and incorporating other common terms for such practices (such as: “unwanted same-sex attraction”) into moderation algorithms to identify and mitigate violative content.

ABOUT TARGETED MISGENDERING AND DEADNAMING

Civil society groups, including [GLAAD](#) and [Media Matters](#), have identified the practice of targeted misgendering and deadnaming as a form of hate speech. In recent years it has become one of the most common modalities of high-follower anti-LGBTQ accounts for expressing contempt and hate toward transgender and nonbinary people across major social media platforms. It is notably utilized to bully, mock, and harass prominent trans public figures (Geena Rocero, Admiral Rachel Levine, Dylan Mulvaney, 16-year-old Zaya Wade to name a few). Targeting well-known people serves to escalate visibility and engagement on posts, and express general hatred of all trans and nonbinary people and the LGBTQ community as a whole. A particularly egregious example of this practice is a July 2022 YouTube video in which far-right

ideologue Jordan Peterson (known for his [virulent anti-LGBTQ extremism](#)) [viciously misgenders and deadnames actor Elliot Page](#) dozens of times in the course of a 15-minute rant. YouTube [demonetized](#) — but didn’t remove — the video, which has 3.5 million views, for violating its “advertising policies around hateful and derogatory content.” At some point in the past few months YouTube added one of their standard contextual debunking info panels to a [similar Peterson video](#) (which also features egregious anti-trans rhetoric, including references to so-called “conversion therapy”).

The info panel, which is an example of one of GLAAD’s recommended best practices for mitigating harmful content and which GLAAD had repeatedly urged YouTube to add, explains that: “Conversion therapy refers to a range of dangerous and discredited practices aimed at changing one’s sexual orientation or gender identity or expression.” The [March 2024 Media Matters report](#) mentioned above reveals that “Right-wing YouTubers and Daily Wire personalities with millions of subscribers regularly misgender and deadname trans people in content with ads” and notes that both creators and YouTube continue to profit from such hate.

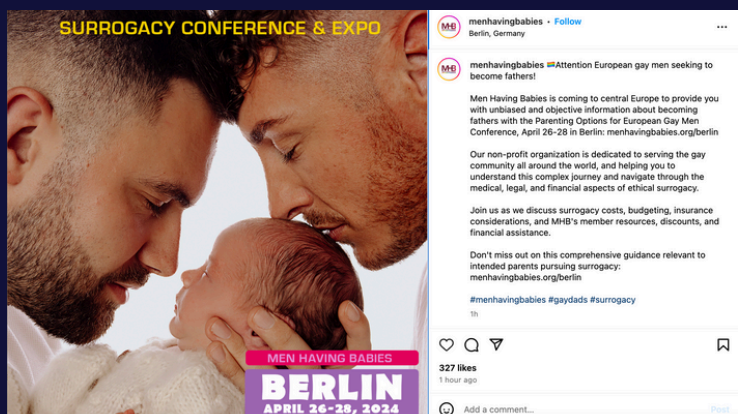
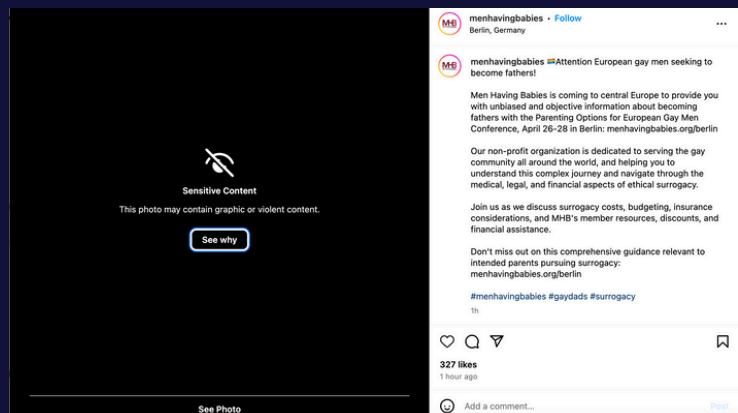
For a deeper understanding of the need for these policies, and to see more details on which platforms have adopted them, please read the full GLAAD reports: [All Social Media Platform Policies Should Recognize Targeted Misgendering and Deadnaming as Hate Speech](#) and [All Social Media Platforms Should Have Policy Prohibitions Against Harmful So-Called “Conversion Therapy” Content](#).

LGBTQ CONTENT SUPPRESSION CASE STUDY: MEN HAVING BABIES (INSTAGRAM)

IN MARCH 2024

the nonprofit [Men Having Babies](#) (MHB) [posted](#) a sweet photo of two gay dads holding their newborn on its Instagram. The organization, which supports gay couples pursuing ethical surrogacy, aimed to promote its annual conference with the post. Shortly after posting, MHB discovered that Instagram had censored the content with [an overlay](#) warning that read “Sensitive Content ... This photo may contain graphic or violent content.” When a user taps “See why,” the overlay reads: “This post doesn’t go against our Community Standards, but may contain images that some people might find upsetting. We cover graphic content so people can choose whether to see it.” As seen above, clearly the photo is neither graphic nor sensitive. Such instances of wrongful suppression of LGBTQ content are part of a larger, troubling trend: Research [shows](#) that LGBTQ content is [disproportionately censored](#) via [removal](#), [shadowbanning](#), [demonetization](#), or “[graphic content](#)” overlays like this one. This incident is particularly upsetting in light of Meta’s failures to mitigate the vast amount of [hate-fueled](#) and [violent content](#) that [researchers say](#) Meta has allowed to proliferate across its platforms.

After escalation from GLAAD to Meta the warning label was eventually removed. Commenting on the case, GLAAD Social Media Safety Program Manager Leanna Garfield [observed](#) to PinkNews: “Graphic content overlays are a valuable tool to mitigate extreme content — That shouldn’t include something as innocuous as a photo of two fathers with their newborn.”



Why this post is covered

- ✓ We use technology or a review team to identify content that should be covered.
- ⚠ This post doesn't go against our Community Standards, but may contain images that some people might find upsetting.
- 🚫 We cover graphic content so people can choose whether to see it.

[Learn more](#) about what kinds of content we cover in Community Guidelines.

SPOTLIGHT ON DATA PROTECTION

THE NEED FOR ROBUST, FEDERAL DATA PROTECTION — AND HOW IT WOULD HELP PROTECT LGBTQ PEOPLE ONLINE AND OFF

AS IN YEARS PRIOR,

the 2024 Social Media Safety Index Platform Scorecard highlights the importance of robust data safeguards. As [GLAAD](#) and other digital rights experts have [long noted](#), the tech industry is [largely failing](#) to self-regulate and continues to prioritize profit over [public safety](#) and [human rights](#); this includes a lack of adequate data protection. The Electronic Privacy Information Center (EPIC), for example, [writes](#): “social media companies harvest sensitive data about individuals’ activities, interests, personal characteristics, political views, purchasing habits, and online behaviors. In many cases this data is used to algorithmically drive user engagement and to sell behavioral advertising—often with distortive and discriminatory impacts.” This harms LGBTQ people online, and off.

When social media companies exploit our personal information for profit, platforms become less safe, especially for marginalized communities who are [disproportionately harmed](#) by [targeted surveillance](#) and [disinformation](#). This lack of data privacy protections helps fuel the hate and disinformation ecosystem — known as [surveillance capitalism](#) — particularly through [media manipulation campaigns](#) that drive polarization and animus. Platforms use our personal information to algorithmically propel user engagement and to sell advertising, including targeting people based on personal characteristics and behavior. In 2022, the United Nations [wrote](#): “Freedom of expression and the right to privacy are among the human rights most

impacted by the digital transformation. Interrelations between these rights have too been transformed.”

All of this is largely happening in the dark, as there is [little transparency](#) into data collection and usage practices, coupled with [convoluted](#) privacy policies. Platforms should [clearly disclose](#) user options to control the collection, inference, and use of information related to their sexual orientation and gender identity, and should not allow third-party advertisers to target people based on that data.

There are also salient connections here with regard to how LGBTQ people and other marginalized groups face disproportionately higher risks of state surveillance. In [the US](#) and [around the world](#), police have been known to engage in online targeting of LGBTQ people, including via their social media activity. Law enforcement authorities in countries including Egypt, Iraq, Jordan, Lebanon, and Tunisia have also reportedly used social media monitoring to [track](#) and [persecute](#) LGBTQ users. Human Rights Watch wrote about the disturbing trend in January 2024:

“Security forces have entrapped LGBT people on social media and dating applications, subjected them to online extortion, online harassment, doxxing, and outing; and relied on illegitimately obtained digital photos, chats, and similar information in prosecutions. In cases of online harassment, which took place predominantly in public posts on Facebook and Instagram, affected individuals faced offline consequences, which often contributed to ruining their lives.” In these cases, strong data protection regulations and enforcement, as well as continuous human rights impact assessments, would potentially help reduce harms.

Some companies (and regulatory agencies) are exploring meaningful solutions to improve industry standards. For example, decentralized social media platforms like [Mastodon](#) and [Bluesky](#) have opened industry possibilities for less exploitative data collection practices, offering users more control. In December 2023, following pressure from LGBTQ and reproductive rights [advocates](#), Meta enabled end-to-end encryption on Messenger, strengthening privacy on that platform. On the other hand, as Accountable Tech has

pointed out, Google is still failing to protect the location privacy of people who seek reproductive healthcare (a problem of potential significant concern for those seeking trans healthcare). Meta may also soon face a ban on the processing of personal data for targeted advertising in the EU, which implemented a national data protection regulation (known as the GDPR) in 2018.

The U.S. doesn't have a federal data protection law, something that civil society has long called for, prompting nearly a dozen U.S. states to pass consumer privacy laws in recent years (many of which tech companies have lobbied against). In April 2024, federal lawmakers introduced the American Privacy Rights Act (APRA), a sweeping, bipartisan proposal to adopt national online privacy protections and give consumers more power over their data. In 2022, the FTC also launched a rulemaking process on corporate surveillance and data security, and in 2023, the FCC created a privacy and data protection task force. To help truly protect everyone from Big Tech's business models, we need strong, and thoughtfully crafted, data protection oversight today.

“AI WILL ALWAYS FAIL LGBTQ PEOPLE.”
— MARY L. GRAY, SENIOR PRINCIPAL RESEARCHER, MICROSOFT RESEARCH

FOCUS ON AI: RISKS FOR LGBTQ PEOPLE

While in the last decade, artificial intelligence (AI) technology has had many positive aspects, as it has accelerated and become more widespread, it has also elicited concerns from human rights advocates about an array of dangers and risks particularly for marginalized communities. Experts such as Joy Buolamwini, Timnit Gebru, Latanya Sweeney, and others have sounded the alarm regarding large-scale issues of algorithmic bias and AI-facilitated disinformation campaigns, including deepfakes. In April 2024, more than 200 civil society organizations including Free Press, Color of Change, and GLAAD sent an open letter to the CEOs of 12 major tech companies, urging them to adopt more aggressive policies to mitigate dangerous, AI-fueled political propaganda leading up to the U.S. presidential election. The letter achieved some initial impact, with eight of the 12 companies issuing responses; committed to maintaining ongoing pressure, organizers were clear that much more action is needed as “platforms evade their responsibility to users around the world.” Various civil society efforts continue to hold companies to account.

While many fears circle AI's future capabilities, today, generative AI (which encompasses technologies like image generators and large language models) already poses unique risks for marginalized groups, including LGBTQ people. For one, several studies have illuminated fundamental bias baked into AI within natural language processing, facial and image recognition, and recommendation systems — which can have real-world implications. Access Now explains: “Companies and governments are already using AI systems to make decisions that lead to discrimination. When police or government officials rely on them to determine who they should watch, interrogate, or arrest — or even ‘predict’ who will violate the law in the future — there are serious and sometimes fatal consequences.”

Generative AI systems learn from existing data, which may also contain harmful stereotypes about LGBTQ people, misrepresenting diverse experiences within the community. (An April 2024 investigation by Wired, for example, found that many AI tools, like OpenAI's Sora, tended to portray LGBTQ people as white, young, and with purple hair.) In recent years, some companies have developed "automated gender recognition" (AGR) technology, which claims to predict a person's gender (often to sell them products via targeted advertising). Spotify's 2021 patent, for example, claims to be able to detect, among other things, "emotional state, gender, age, or accent" to recommend music. For trans, nonbinary, and gender non-conforming individuals, and others with genders that fall outside the binary, this technology is particularly problematic. Access Now's Daniel Leufer writes: "Research shows that AGR technology based on facial recognition is almost guaranteed to misgender trans people and inherently discriminates against non-binary people. As Os Keyes explains in their paper, The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition, approaches to AGR are typically based on a male-female gender binary and derive gender from physical traits; this means that trans people are often misgendered, while non-binary people are forced into a binary that undermines their gender identities."

In addition, social media platforms are automating content moderation processes, and many large platforms are using AI to decide whether content violates or doesn't violate their hate speech and harassment policies. In recent years, many social media companies have reduced vital trust and safety teams, often opting to contract with third-party vendors that fail to adequately recognize and understand harmful content targeting marginalized groups. In a 2021 study, for example, MIT Technology Review reported that "scientists tested four of the best AI systems for detecting hate speech and found that all of them struggled in different ways to distinguish toxic and innocuous sentences."

Some of the world's largest social media platforms have shown they are not equipped to handle the rise of AI-facilitated hate, harassment, and disinformation campaigns, including deepfakes and bots that can spew

hate-based imagery at massive scale. In a July 2023 paper, Mozilla researchers wrote: "This research shows scale degrades datasets further, amplifying bias and causing real-world harm." *The New York Times* also reports: "In the hands of anonymous internet users, A.I. tools can create waves of harassing and racist material. It's already happening on the anonymous message board 4chan." In one example, from May 2023, a deepfake video of President Biden in drag (with anti-trans overtones) went viral on Instagram and TikTok (many instances of the post are still live on both platforms). Another altered video in February 2023 showed Biden making transphobic remarks in a speech. (It was fake.)

Thankfully, some tech companies are starting to recognize that they must make at least some effort to address the risks of generative AI on trust and safety. In January 2024, Meta started requiring disclosures for AI-created or altered political ads, and in February 2024, the company said it would begin labeling AI-generated images, audio, and videos. Similarly, in March 2024, YouTube began requiring creators to disclose when they make realistic videos with AI. It remains unclear whether these measures will be effective at scale.

People are impacted by these tools, and addressing these risks requires collaboration between technologists, policymakers, LGBTQ advocates, and other stakeholders to pursue safeguards to help prevent malicious use and unintended consequences. Ensuring compliance with existing anti-discrimination laws and developing regulatory safeguards to address emerging risks will be essential for protecting the rights and well-being of LGBTQ people and other marginalized groups. As Tech Policy Press notes, in the US, there is no federal legislation close to becoming law. But in the past year, there has been a surge of AI laws proposed and passed, and some have already taken effect. In October 2023, the Biden Administration issued an executive order on "safe, secure, and trustworthy" AI, which provides an ambitious blueprint.

SOLUTIONS FOR ALL:

LEGISLATIVE AND REGULATORY APPROACHES TO SOCIAL MEDIA AND TECH ACCOUNTABILITY

THE EXPERTISE, WATCHDOG WORK,

and guidance of tech accountability groups, researchers, and civil society organizations like GLAAD — urging social media companies to voluntarily improve safety, privacy, and expression on their platforms for LGBTQ people — continues to be vitally necessary. What is even more urgently needed is accountability via external oversight and regulation of the entire industry, to protect the public interest from the harmful business practices and product designs of social media companies. As GLAAD has previously [noted](#), amidst the complexities of the many current U.S. legislative proposals intending to try to fix these problems, and especially the many proposals focused on youth safety, it is important that such approaches be carefully crafted lest they create potential harms and unintended negative impacts for LGBTQ people and other marginalized communities (and everyone). To preserve LGBTQ rights and safety, regulatory solutions should focus on addressing specific harmful business practices of tech companies, such as [surveillance advertising](#) and over-aggressive [collection and misuse of user data](#), versus proposals that would actually expand collection of data and potentially expose people to government surveillance or lead to suppression and censorship of LGBTQ material. The chilling impact of 2018’s misguided FOSTA-SESTA (Fight Online Sex Trafficking Act and Stop Enabling Sex Traffickers Act) [legislation](#) should be an object lesson that attempts to solve problems can create grave new impacts, especially for historically marginalized communities.

In the case of FOSTA-SESTA, the broad consensus from researchers, legal scholars, and even the Government Accountability Office is that, as historian Sascha Cohen [writes](#) in her in-depth overview in *The Nation*: “the law has been counterproductive at best and deadly at worst.” FOSTA-SESTA has especially impacted LGBTQ people and sex workers, resulting in a situation where, as Melissa Gira Grant [pointed out](#) in *The New Republic*, “The ‘solution’ became the problem.” (To learn more, read the aforementioned [feature](#) in *The Nation*). There are other approaches to addressing children’s safety online which should rightly put the burden on platforms to address these issues for all users.

Delving into why kids-only legislation won’t solve the problem of widespread online manipulation and harm, Nora Benevidez (lead author of the 2023 Free Press report, [Big Tech Backslide](#)) observes in a January 2024 *Tech Policy Press* [article](#): “Comprehensive privacy and civil rights protections are possible. They would also avoid the pitfalls of well-intentioned legislation that claims to reduce harm and remove harmful content but actually exposes everyone, including children and teens, to more invasive practices and government overreach. The most powerful step we can take now to rein in online manipulation is to introduce and pass robust federal data-privacy legislation that would limit the data collected about, and then used against, all of us.”

Of course social media companies engage in business practices that prioritize their corporate profits and bottom line rather than prioritizing the best interests of society. This is a fundamental reality of all industries, and not a surprising one. The EPA, FDA, SEC, and other regulatory agencies came into existence for this very reason. Indeed, creating guidelines and oversight to ensure the public health and safety of the American people is not a radical idea. Although tech and social media companies may be under-regulated in the US, thankfully there are existing public agencies which do have some jurisdiction over the industry.

For instance, the [Federal Trade Commission \(FTC\)](#) continues to protect the public interest through various mechanisms like its [ongoing antitrust case against Meta](#) or via imposing fines, as it has in [Meta's cases of repeated data privacy violations](#) such as the Facebook Cambridge Analytica case. In May 2023 Samuel Levine, director of the FTC's Bureau of Consumer Protection [stated that](#): "Facebook has repeatedly violated its privacy promises. The company's recklessness has put young users at risk, and Facebook needs to answer for its failures."

Delving into the array of regulatory situations across the globe — including the work of the [EU's Digital Services Act \(DSA\)](#), the [Australian eSafety Commissioner](#) and [others](#) — is beyond the scope of this report.

But it is illuminating to see that other governments around the world are taking rigorous approaches to protect the public interest over the business interests of corporations, via [regulation of tech and digital business models and practices](#).

While implementation of oversight may be a long and complex process, lawmakers must find solutions that do not create new problems, solutions that require companies to be accountable and transparent, solutions that protect us all.



SHOWCASE OF 2023-2024 REPORTS ON LGBTQ SOCIAL MEDIA SAFETY

IN 2021,

The inaugural GLAAD Social Media Safety Index report offered the first-of-its-kind dedicated analysis on LGBTQ safety and social media platforms. There are now many powerful reports and studies devoted to these issues, some of the most significant of the past year are showcased here. The report summaries below offer an invaluable overview of the current landscape for LGBTQ safety, privacy, and expression online. We urge everyone, especially platform leadership and executives, to read the full reports.

Important 2023-2024 research on anti-LGBTQ hate and violence includes the February 2024 report: [New Findings Show 60+ Anti-LGBTQ+ Incidents Targeting US Religious Institutions](#), and the November 2023 [documentation of more than 700 anti-LGBTQ hate and extremism incidents](#) in the year following the tragic attack at Club Q in Colorado Springs. The annual [Year in Review: Anti-LGBTQ+ Hate & Extremism Incidents](#) report tracked more than 350 anti-LGBTQ incidents in the US between June 2022 and April 2023.



TIMELINE: The impact of Libs of TikTok

MEDIA MATTERS — April 2024 - Present

(April 2022 - Present) An expansive, ongoing report from Media Matters for America documents how harassment and threats of violence against institutions, events, and individuals have occurred after incitement from [anti-LGBTQ extremist account Libs of TikTok](#) across X and Meta's platforms. As of the date of publication of this report, the current number of incidents is at least 48. A November 2023 USA Today [feature](#) on the report confirmed the research, and also explores the phenomenon of stochastic terrorism (in which public demonization of a person or group results in the incitement of a violent act, which is statistically probable but whose specifics cannot be predicted) and [networked harassment](#). The dedicated year-round work of Media Matters bears special mention for being so extensive and prolific. Their dozens of LGBTQ dispatches and reports can be found [here](#).



ANTI-TRANS DANGEROUS SPEECH DURING THE 2024 U.S. ELECTION

DANGEROUS SPEECH PROJECT — APRIL 2024

The [Dangerous Speech Project](#) has been tracking forms of harmful expression and advising tech companies since 2010. This new report from their Global Research Initiative offers a sobering analysis of the prevalence, intensity, harms, and dangers of anti-trans hate speech online, and concludes with examples of efforts to counter this

dangerous speech. From [the report](#): “As the 2024 United States election approaches, speech that can move people toward accepting violence, called “dangerous speech” (DS), is flourishing, and with it, there is a growing possibility of intergroup violence. Transgender, nonbinary, and gender nonconforming people have become a primary target of such speech. Influential politicians, media figures, and religious leaders are portraying transgender people as threats to children, other women, families, traditional values and public safety ... A defining feature of dangerous speech is that it often promotes fear, as much as it expresses or promotes hatred. For example, one can assert that another group is planning to attack one’s own group without expressing hatred, yet that message might easily convince people to condone or commit violence, ostensibly to fend off the attack. Violence would seem defensive, and therefore justified ... Combatting anti-trans rhetoric requires a multifaceted approach, including education, advocacy, and policy changes aimed at promoting inclusivity and equality. Understanding the rhetorical patterns used to advance dangerous narratives about trans people is one step toward designing better interventions to counter such speech.”



USER ERROR: LGBTQ

KAIROS FELLOWSHIP — MARCH 2024

From the respected racial justice and platform accountability organization Kairos Fellowship, this latest in the series of “[User Error](#)” reports looks at the history of LGBTQ people and the Internet, shows how platforms are complicit in the harm the queer community faces online, and envisions what an internet that values queerness and queer people could look like. From [the report](#): “The bottom line: Social media companies’ role in

spreading violent homophobia and transphobia is undeniable. On the surface, Big Tech may pretend to care about LGBTQ people but behind the scenes they will do anything to make a dollar, which includes allowing anti-LGBTQ content online for profit and unfairly taking down LGBTQ+ content. We must fight for a pro-queer future. This future is not just an internet that simply allows or tolerates LGBTQ+ people. It is one where everyone can exist without being targeted. It’s a world in which online platforms can’t be used by anti-LGBTQ+ groups or politicians to further a destructive agenda. It’s a future where tech works for and humanizes queer and trans people.”



UNSAFE

GLAAD — MARCH 2024

In June 2023, more than 250 LGBTQ public figures and allies [delivered an open letter to social media companies](#), pleading with them to develop and share a plan to address the epidemic of anti-trans hate on their platforms. Released nine months later, in March 2024, GLAAD’s “[Unsafe](#)” report showcases dozens of examples of the shocking, dehumanizing anti-trans content that Meta actively allows on Facebook, Instagram, and Threads. All of the content examples in the report were submitted to Meta’s standard reporting systems by GLAAD; Meta either replied that posts were not violative or simply did not take action on them. The Washington Post [covered the release](#) of the report, noting that posts “included calls for the violent extermination of transgender people as well as descriptions of trans and gender nonconforming people as ‘satanic,’ ‘sexual predators,’ ‘terrorists,’ ‘mentally ill’ and ‘perverts.’” The report calls for Meta to take urgent action as trans and LGBTQ people continue to experience real-world harms stemming from anti-LGBTQ accounts with large followings. Also see GLAAD’s February 2024 report: [All Social Media Platform Policies Should Recognize Targeted Misgendering and Deadnaming as Hate Speech](#).

CONVERSION THERAPY
ONLINE: THE ECOSYSTEM IN
2023 GLOBAL PROJECT ON
HATE & EXTREMISM
(GPAHE) — JANUARY 2024



This powerful and extremely detailed January 2024 report from the Global Project on Hate & Extremism (GPAHE) reveals that harmful so-called “conversion therapy” content remains widely available online, particularly in non-English languages. The report offers clear and easy to implement best practice guidance for social media platforms to mitigate such content. The practice of so-called “conversion therapy” has been condemned by all major medical, psychiatric, and psychological organizations as dangerous, and has been banned by dozens of countries and states. In addition to GLAAD’s efforts urging platforms to add prohibitions against the promotion of so-called “conversion therapy” to their community guidelines, GLAAD also urges these companies to effectively enforce these policies. This GPAHE report can serve as a useful practical guide in implementing solutions. Also see GLAAD’s February 2024 report: All Social Media Platforms Should Have Policy Prohibitions Against Harmful So-Called “Conversion Therapy” Content.



SECURE OUR SOCIALS
HUMAN RIGHTS WATCH — JANUARY 2024

This ambitious campaign is a follow-up companion to a February 2023 Human Rights Watch report (“[Digital Targeting and Its Offline Consequences for LGBT People in the Middle East and North Africa](#)”). The “Secure Our Socials” campaign calls on Meta to take concrete steps to make their platforms safer for those being targeted, pointing out how LGBTQ users “have faced online harassment, doxxing, outing, extortion, hacked accounts, and entrapment by security forces and private individuals on Instagram and Facebook.” The campaign’s open letter to Meta calls for: rapid, accountable, and transparent content moderation

practices; greater transparency and investments in user safety; investment in content moderators with proficiency in each Arabic dialect and in other regional languages; and also requests that the company introduce a one-step account lockdown tool allowing users to wipe all Meta content on a given device. The campaign also includes this downloadable “[LGBT People Should Feel Secure Online — Awareness Tips](#)” resource which offers guidance on how to prevent becoming a victim of digital targeting, and explains how to mitigate the harms and impact if one has been targeted.



COMBATING ANTI-LGBTQ+ PSEUDOSCIENCE
SOUTHERN POVERTY LAW CENTER —
DECEMBER 2023

From [the report](#): “In recent years, researchers, social scientists and historians have all shown how social media amplification of anti-LGBTQ+ disinformation that includes conspiracy theories about medical providers derived from pseudoscientific claims about LGBTQ+ identity fuels violent attacks targeting individual LGBTQ+ people and healthcare providers. Anti-LGBTQ+ organizations use pseudoscience to attack not only affirming medical practices, but almost any social, religious, commercial or governmental affirmation of LGBTQ+ identity and LGBTQ+ representation in popular culture. With the help of right-wing and extremist social media personalities ... anti-LGBTQ+ groups have peddled and continue to use pseudoscience for several reasons. Most importantly, their agenda is unpopular among the general public ... Also, they know that perpetuating harmful stereotypes to keep the ‘debate’ over LGBTQ+ rights alive is financially lucrative. Finally, many anti-LGBTQ+ groups see potential in dividing public opinion and policymakers over transgender rights. They see an opportunity to reinvigorate their anti-LGBTQ+, anti-abortion, ‘parental

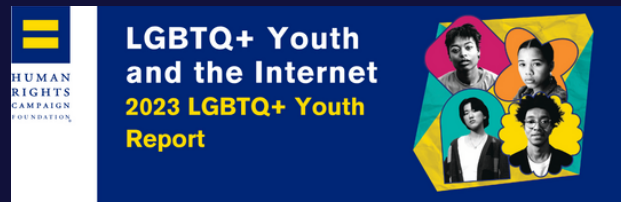
rights’ and ‘religious freedom’ agendas by promoting junk science about transgender people, gender-affirming health care and the experiences of LGBTQ+ young people, specifically. Anti-LGBTQ+ pseudoscience has become a prominent component of anti-LGBTQ+ policy networks who want to disguise their bigoted motivations with seemingly objective language. These networks help translate anti-LGBTQ+ pseudoscience into legislative and legal campaigns to limit bodily autonomy and LGBTQ+ people’s human rights.”



MONETIZING HATE: HOW 100+ MAJOR BRANDS ARE BANKROLLING ANTI-LGBTQ EXTREMISM ON YOUTUBE
EKŌ — NOVEMBER 2023

This hard-hitting expose from corporate accountability group Ekō details how ads for dozens of major brands, including Nike, J.Crew, and L’Oréal, are appearing next to videos inciting violence and hatred against LGBTQ people. The findings underscore significant failings in YouTube’s monetization and moderation systems, as well as reputational and business risks to some of the world’s biggest brands. The authors of the report call on advertisers to take a principled stand against funding hate and disinformation through a host of measures including demanding access to detailed information about ad placements, and enforcing business contracts regarding brand safety standards. The report authors also demand action from YouTube to bolster moderation and demonetize harmful channels, in addition to urging transparency interventions from U.S. policymakers. YouTube’s [Community Guidelines](#) prohibit anti-LGBTQ hate speech and harmful content; yet the report clearly shows those policies aren’t being properly enforced. Researchers analyzed 13 monetized videos by well-known, anti-LGBTQ figures seemingly breaching YouTube’s policies on hateful and derogatory content,

violent content, and incendiary and demeaning content. The report found that at least 104 brands that have expressed public support for the LGBTQ community were found to be funding anti-LGBTQ hate content.



2023 LGBTQ+ YOUTH REPORT (LGBTQ+ YOUTH AND THE INTERNET)
HUMAN RIGHTS CAMPAIGN (HRC) — NOVEMBER 2023

Illuminating the upsides and downsides of social media for LGBTQ youth, this annual report from the Human Rights Campaign (HRC) found that virtually all LGBTQ youth (95.3%) have used the internet to find information that helps them understand their LGBTQ identity. More than two-thirds (68.3%) of respondents said they participate in online communities to engage with other LGBTQ people in their age group, with the figure being even higher (73.3%) for transgender and gender-expansive youth. However, according to the report: “LGBTQ+ youth are inundated with anti-LGBTQ+ content online and are often cyberbullied.” In fact, 96% of LGBTQ youth have seen content on social media that was offensive or hurtful toward LGBTQ people; and half of LGBTQ young people have experienced cyberbullying “based on an identity of theirs’ in the last year (49% of LGBTQ+ youth and 52.9% of transgender and gender-expansive youth). And 15.2% of LGBTQ+ youth (and 16.7% of trans/gender-expansive youth) say they do not feel safe participating in online activities. Pointing to the need for better messaging and resources from platforms, the report also notes that 18.2% (and 17.5% of trans/gender-expansive youth) did not know how to report encountered cyberbullying or online hate speech to social media platforms; and 65.5% of LGBTQ youth (and 67.3% of trans/gender-expansive youth) were not confident that platforms would take steps to resolve the issue if they did report such problems.



REPORT: GENDERED DISINFORMATION AND FREEDOM OF EXPRESSION

UN SPECIAL RAPPORTEUR — OCTOBER 2023

This report to the UN General Assembly from Special Rapporteur Irene Khan explores the negative impact of gendered disinformation, especially on women and gender nonconforming persons, and its implications for the right to freedom of expression. Identifying gendered disinformation as both a strategy to silence the free expression of women and gender nonconforming persons, and as a threat to their safety and health, the report also importantly acknowledges the dangerous impact of [surveillance-based advertising](#).



FUELLING HATE REPORT **TRANS JUSTICE PROJECT — AUGUST 29, 2023**

The 2023 "Fuelling Hate" report — the largest-ever project investigating anti-trans hate in Australia — polled 3,000 respondents (including 1,300 trans people). The survey found that anti-trans hate is ubiquitous both online and offline — 94.55% of all participants reported witnessing online anti-trans hate in the previous year.

Approximately half (49.20%) of trans participants had experienced online anti-trans hate (abuse, harassment, or vilification) in the prior 12 months, with a similar rate of experiences of offline anti-trans hate (47.9%) — defined as in-person harassment, abuse, or violence. According to the report: "Trans people experience a range of online hate, such as deliberate misgendering, hate speech, bullying, doxing, stalking, threats of violence and sexual assault, death threats, incitement to commit suicide and incitement to genocide." Such anti-trans hate is concentrated in: "social media platforms, comment sections (especially on news websites) and dating apps. Perpetrators were sometimes well-known anti-trans figures but were often anonymous." The report also found that anti-trans hate is escalating over time (1 in 3 participants reported an increase in anti-trans hate since 2020). As in the U.S. (and around the world), the report cites the problem of "targeted disinformation campaigns from media outlets and anti-trans lobbyists," as key precipitating factors for the current wave of online and offline anti-trans hate. Amidst this dire landscape, the Australian eSafety Commissioner is [pursuing various efforts](#) to reign in social media platforms; its work also includes offerings such as this dedicated [LGBTIQ+ eSafety resource site](#).



ANTI-TRANS EXTREMISM: THE FAR RIGHT'S NEW STRATEGY TO SPREAD HATE **HUMAN RIGHTS FIRST — JUNE 2023**

This powerful fact-sheet simply and directly explains: "The rapid growth and normalization of anti-trans extremism in the United States threatens the trans and broader LGBTQ+ community, while also creating inroads

for attacks on other targeted communities. It amplifies bigotry, regularly inspires harassment and hate-fueled attacks, creates inequitable and unsafe environments for LGBTQ+ individuals and their families, and ultimately undermines democracy.” The fact-sheet showcases examples of false inflammatory anti-trans social media posts; offers a concise debunking of common anti-trans tropes; explains the systemic attacks of the far-right on trans rights at the federal, state, and local level; illuminates the destructive impact of anti-trans extremist attacks; and features simple practical guidance for how to respond to anti-trans extremism (including rejecting extremist narratives and supporting inclusive policies).



THE ONLINE PUBLIC DISCOURSE ABOUT LGBTQ PEOPLE - ANALYSIS & NUMBERS **AMNESTY INTERNATIONAL NORWAY & FRI — 2023**

This report illuminates how anti-LGBTQ “hate speech and negative discourses” on social media in Norway have highly consequential impacts for all of Norwegian society, noting that: “It is a serious democratic problem if minority groups refrain from participating in the public debate and expressing themselves publicly for fear of incitement, disparaging, or highly critical comments about their sexual orientation or gender identity.” Using a proprietary algorithm, the researchers undertook a quantitative and qualitative analysis of LGBTQ content on X/Twitter and Facebook from 2018 to 2022, finding that: “The number of tweets about queer topics increased from 6,457 to 44,620 in the period. The number of tweets about transgender people increased from 1,485 to 23,465 - almost 16 times the number of tweets.” While the report found a high percentage of anti-trans content on X/Twitter (47%), it also found a significant quantity of trans-affirming content (40%); though anti-trans content has risen over time, while supportive sentiment has dropped.

OUT OF BOUNDS:

Foreign and Digital Influence Targeting
LGBTI Rights in Ghana



OUT OF BOUNDS: FOREIGN AND DIGITAL INFLUENCE TARGETING LGBTI RIGHTS IN GHANA IPAS — 2023

“As this report clearly documents, foreign interests are the impetus for a bill driven by right-wing, mostly evangelical Christian groups in the U.S. that are on a strategic anti-rights mission around the world, and that, with the cooperation of its politicians, have set their sights on Ghana.” Similar to the 2023 Amnesty International Norway report, this fascinating analysis employs AI and machine-learning to look at large amounts of social media content in Ghana from 2018 to 2022, with a focus on identifying the foreign-influenced anti-LGBTQ drivers behind the country’s draconian anti-LGBTQ bill (passed by Parliament in February, but as of May 2024 still awaiting the action of Ghana’s president, who is expected to veto it). The bill would criminalize LGBTQ advocacy organizations and individuals, as well as anyone failing to report an LGBTQ person to authorities or to report anyone using a social media platform for such LGBTQ advocacy. (The bill would also place liability on social media companies such as Meta and X/Twitter should they allow on their platforms material that supports LGBTQ activities and rights.) Specifically illuminating the malign utilization of social media platforms by conservative foreign (especially U.S.) evangelical Christian anti-LGBTQ groups to seed, disseminate, and amplify anti-LGBTQ messaging, researchers found that: “While inside Ghana there has been significant public approval for the oppression of the LGBTI community, support for the LGBTI community — as indicated by engagement with pro-LGBTI content for the purpose of this report — has also increased.” One of the most significant determinations of the report is that: “It is

necessary to educate key stakeholders, politicians, advocates, and journalists on the difference between a real grassroots campaign with millions of supporters and an online petition or campaign that uses paid media and global supporters to create a facsimile of public outrage.” Interestingly, though anti-LGBTQ rhetoric and attacks on LGBTQ rights continue, the report concludes that: “The balance of public opinion has in fact shifted toward tolerance for the LGBTI community.”



**A YEAR OF HATE: UNDERSTANDING THREATS
AND HARASSMENT TARGETING DRAG SHOWS
AND THE LGBTQ+ COMMUNITY**
ISD — JUNE 2023

From the report: “Internationally, rising hate and extremism pose an existential threat to human rights and democratic freedoms. LGBTQ+ communities are often the first group to come under attack, and understanding the contours of these assaults matters both for the protection of these communities and to be better able to safeguard human rights and democracy more broadly. In new research by ISD, including four country profiles [the U.S., UK, Australia and France], we examine the trends in anti-LGBTQ+ hate and extremism with a particular focus on harassment targeting all-ages drag shows. This research draws on ethnographic monitoring of over 150 Telegram channels, Twitter profiles and Facebook groups, as well as external resources such as news reports, Armed Conflict Location & Event Data (ACLED) and Crowd Counting and previous reports on anti-drag by GLAAD and the Southern Poverty Law Center.”

**PLEASE SEE THE 2024 SMSI
ARTICLES AND REPORTS
APPENDIX FOR MANY OTHER
REPORTS OF INTEREST.**

ACKNOWLEDGMENTS

GLAAD is grateful to the many organizations and individuals with whom we do this important work. We especially want to acknowledge our SMSI advisory committee: ALOK; Lucy Bernholz; Alejandra Caraballo, Esq.; Joan Donovan; Jelani Drew-Davi; Liz Fong-Jones; Evan Greer; Leigh Honeywell; Maria Ressa; Tom Rielly; Dr. Sarah T. Roberts; Brennan Suen; Kara Swisher; and Marlena Wisniak.

Very special thanks to: Research consultant, Andrea Hackl, anti-LGBTQ extremism researcher Sarah Moore, and the team at Ranking Digital Rights. Thanks to GLAAD staff including Senior Art Director Crystal Stratford, Web Producer Heidi Spillum, Research Assistant Allison Bloom, Chief Communications Officer Rich Ferraro, Social Media Safety Program Manager Leanna Garfield and Senior Director of Social Media Safety Jenni Olson. And extra special thanks to anyone else we forgot to mention. The work of countless journalists, researchers, activists, and others continues to drive change forward. We are indebted to them all.

This report would not have been possible without the support of: Craig Newmark Philanthropies, the Gill Foundation, and Logitech. We are also grateful to David Hornik, Kara Swisher, and the Stanford Center on Philanthropy and Civil Society for early support of this project.

**IF YOU'D LIKE TO
SUPPORT THE WORK
OF THE SOCIAL MEDIA
SAFETY PROGRAM AND
THIS REPORT, PLEASE
DONATE TO GLAAD AT
GLAAD.ORG/DONATE**