SOCIAL MEDIA SAFETY INDEX

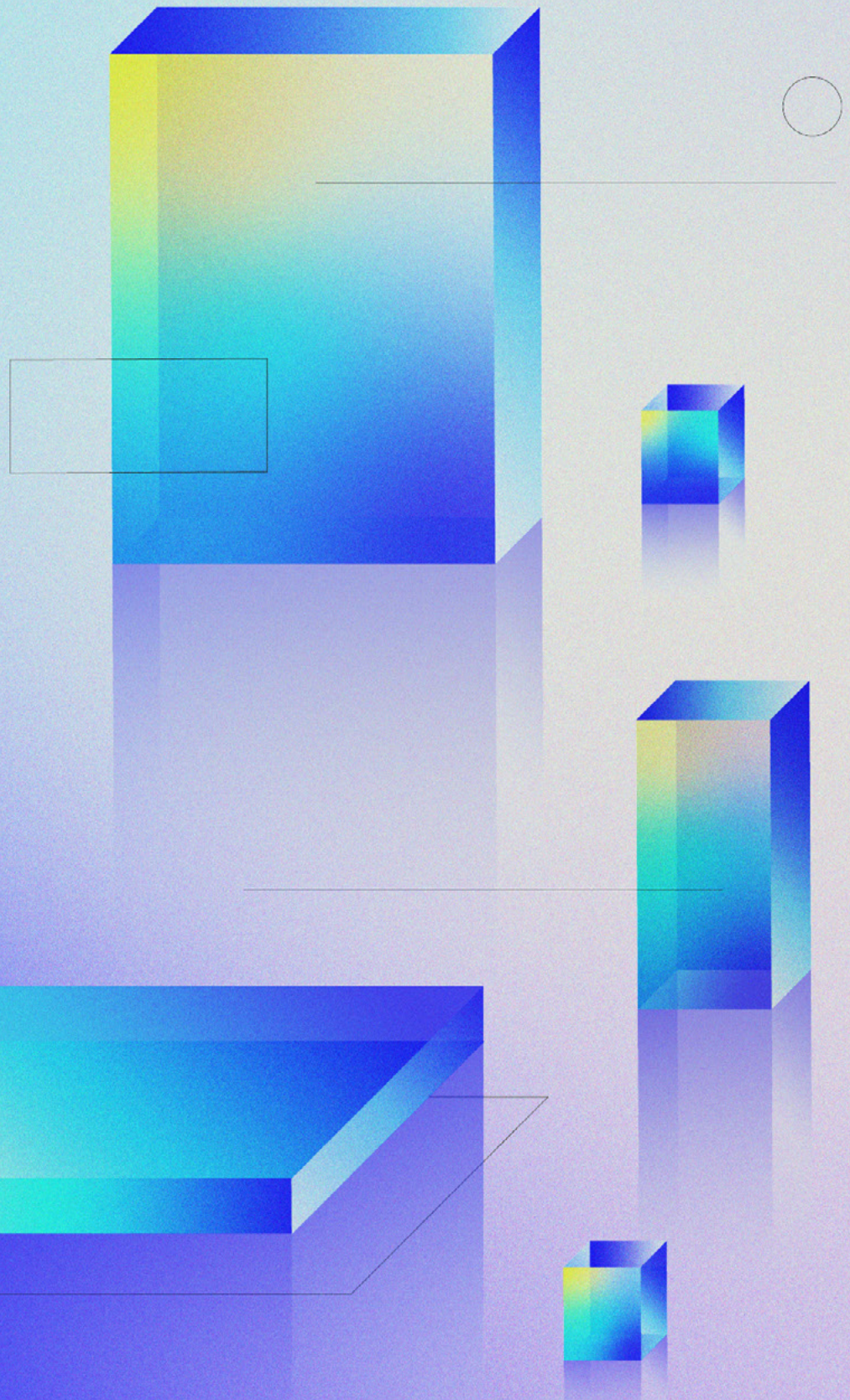## 2023

glaad

TABLE OF CONTENTS

LETTER FROM GLAAD PRESIDENT & CEO

SARAH KATE ELLIS

GLAAD tracks and reports dangerous conditions for LGBTQ Americans. I want to be clear: despite record-high support for LGBTQ equality, we currently live in an unsafe America for LGBTQ people, especially for transgender people. From countless examples of physical violence (including the horrific attack at the Club Q nightclub in Colorado Springs) to the more than 500 anti-LGBTQ/anti-trans bills attempting to retract our basic rights, we are in the midst of an epidemic plaguing our nation: a culture of anti-LGBTQ rhetoric and violence from politicians and political extremists. In my December 2022 testimony to the US House Oversight Committee I spelled out the long and terrible list of these harms.[1] As we have seen over and over again — there is a direct line from dangerous words to violent behavior against the LGBTQ community. The nexus and vehicle for so much of this rhetoric is the major social media companies. And we have seen — over and over again — how these companies fail to protect LGBTQ users and fail to enforce their own policies, which assert that hate speech, bullying, and harassment are not allowed on their platforms.

For more than 35 years, GLAAD has been the leader in creating safe and inclusive environments in Hollywood, journalism, and across our culture. Our founders were visionaries who understood that what people see and hear in the media affects the decisions made in schools, offices, living rooms, courtrooms and ballot boxes. Because of GLAAD's media work — and the work of so many content creators and media industry leaders — the world came to know lesbian, gay, bisexual, transgender and queer (LGBTQ) people and to accept us. By ensuring LGBTQ people were included and represented in fair and accurate ways, GLAAD's work changed hearts and minds and LGBTQ acceptance grew.

GLAAD has continued to innovate to keep step with the rapidly and ever-changing media landscape. Since its 2021 launch, our *Social Media Safety Index (SMSI)* has created the industry's first standard for tackling online anti-LGBTQ hate and intolerance and increasing safety for LGBTQ social media users. The Index provides targeted industry recommendations on LGBTQ user safety across the five major social platforms: Facebook, Instagram, Twitter, YouTube, and TikTok. While we have seen important achievements this past year in our efforts to create safer online spaces for LGBTQ people, an enormous amount of work lies ahead to better protect LGBTQ users.

The viral dehumanizing anti-LGBTQ content on these platforms negatively impacts public understanding of LGBTQ people — driving hatred, and even violence, against our community. All of the platforms have policies protecting against anti-LGBTQ hate speech, but effective enforcement requires groups like GLAAD to hold them accountable.

The LGBTQ community is under attack, steadily having our basic rights stripped away in state after state, not to mention rising physical violence and threats. And as is true of the targeting of *any* vulnerable group — this all must also be understood as an attack on *everyone*. It is impacting our entire society in countless ways. Along with the ongoing weaponizing of other forms of bigotry and hate-driven disinformation (attacks on CRT, falsehoods about climate change, anti-vax/mask, etc.) this anti-LGBTQ, and especially anti-trans, hate will continue to be one of the greatest dangers we face as a society — and will no doubt continue to escalate leading up to the 2024 election. These attacks on our community must be seen for what they are: hate and lies. It is time for our allies to stand up and speak out against the whole array of these attacks: from the toxic and disgusting anti-LGBTQ "groomer" conspiracy theory which continues to rage out of control to the targeting of LGBTQ books and curriculum and the attacks (physical and legislative) on drag and drag events and LGBTQ Pride.

Here at GLAAD we will continue to be on the front line in demanding that social media platforms make their products safe for LGBTQ users; as we also provide expert guidance, via our Social Media Safety Program and publications like the SMSI, to assist these companies in achieving these goals.

Along with other advocacy organizations, activists, colleagues, and allies **GLAAD is calling upon the major social media platforms to step up and #StopLGBTQHate now** and to make your products safe for LGBTQ users — and for us all.

Sarah Kate Ellis
President & CEO, GLAAD

---

[1]   Sarah Kate Ellis addresses the US House Oversight Committee

EXECUTIVE SUMMARY

With our Social Media Safety Program — and this annual Social Media Safety Index (SMSI) and Platform Scorecard — GLAAD is working every day to hold tech companies and social media platforms accountable, and to secure safe online spaces for LGBTQ people.

Created in partnership with Goodwin Simon Strategic Research and the noted Big Tech accountability watchdog group, Ranking Digital Rights, the SMSI Platform Scorecard offers an evaluation of LGBTQ safety, privacy, and expression on five major platforms (Facebook, Instagram, TikTok, YouTube, and Twitter) based on 12 LGBTQ-specific indicators.

Problems identified in the Scorecard, and in the SMSI report in general, include: inadequate content moderation and enforcement (including issues with both failure to action on anti-LGBTQ hateful content *and* over-moderation/censorship of LGBTQ users); harmful and polarizing algorithms; 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 problems are exacerbated for those who are members of multiple communities (BIPOC, women, immigrants, people with disabilities, people of historically marginalized faiths, etc.). Social media platforms should be safe for everyone, in all of who we are.

Key findings in the 2023 SMSI include:

- Anti-LGBTQ rhetoric on social media translates to real-world offline harms.

- The problem of anti-LGBTQ hate speech and disinformation continues to be an alarming public health and safety issue.

- Platforms are largely failing to successfully mitigate this dangerous 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 true transparency reporting from the platforms.

Select Core Recommendations:

- **Strengthen and enforce existing policies** that protect LGBTQ people and others from hate, harassment, and mis- dis- and malinformation (MDM), 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.[2]

- **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.[3]

- **Stop violating privacy/respect data privacy.** To protect LGBTQ users from surveillance and discrimination, platforms should reduce the amount of data they collect and retain. They should implement end-to-end encryption by default on all private messaging to protect LGBTQ people from persecution, stalking, and violence.[4] And cease the practice of targeted surveillance advertising, including the use of powerful algorithms to recommend content,[5] potentially outing users.[6]

---

[2]   The EU tells Twitter to hire more human content moderators amid concerns of rise of illegal content | Euronews

[3]   Transparency is the best first step towards better digital governance
("The time has come for governments to act, and transparency measures are the best first step. Governments should require social media companies to disclose more information about how they operate and how they amplify, restrict and remove content on their systems. Transparency rules such as these are a traditional way to put pressure on companies to act in the public interest and to protect consumers without burdensome mandates setting out exactly how they should conduct themselves.")

[4]   OPEN LETTER: Make DMs Safe

[5]   Disinformation, Radicalization, and Algorithmic Amplification: What Steps Can Congress Take?

[6]   For more context see this Accountable Tech overview which shows that 81% of Americans support banning companies from collecting people's personal data and using it to target them with ads. Also see this Tech Policy Press interview with Dr. Nathalie Maréchal, former Policy Director at Ranking Digital Rights and Miami University associate professor, Dr. Matthew Crain. As Maréchal explains: "Banning surveillance advertising will protect individual privacy, reduce corporate incentives to maximize invasive data collection, and spur innovation by unleashing the potential of the digital contextual advertising sector that has been held back by the dominant surveillance advertising platforms."

- **Promote civil discourse** and proactively message expectations for user behavior (including actually respecting platform hate and harassment policies).[7]

While many of these problems overlap, complicating efforts toward mitigation of harm, it is clear that social media companies can — and must — do better. As part of our ongoing monitoring, rapid response and advocacy work, GLAAD repeatedly encounters failures in enforcement of community guidelines across every platform. Too often, when reports are filed on content that clearly violates these guidelines, GLAAD researchers and advisors are informed that no enforcement or mitigation action will be taken, or even that content will not be evaluated at all. As recently as April 2023, Meta has closed (without evaluating) our reporting of extreme anti-trans hate posted by known anti-LGBTQ Instagram accounts, giving the following explanation via their automated in-app reporting system: "Because of the high volume of reports we receive, our team hasn't been able to review this post." This is gravely concerning. As a new 2023 report from The Leadership Conference on Civil and Human Rights observes: "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."[8]

In the year ahead, we will continue to spotlight new and existing issues facing LGBTQ users in real time, both directly to the platforms and to the public and the press. GLAAD also looks forward to maintaining ongoing dialogue about LGBTQ user safety, privacy, and expression among colleagues in the field of platform accountability, as we work together shoulder to shoulder to advocate for change. 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.

In addition to consulting this year's Appendix of Articles and Reports for further context and analysis, please also refer to the 2022 and 2021 SMSI reports which remain substantial and valuable resources on this topic.

---

[7]   For a Civil Internet – How the tone of online conversations can build trust | Impact of Social Sciences

[8]   Cause for Concern 2024: The State of Hate

A NOTE TO OTHER COMPANIES: WE'RE LOOKING AT YOU, TOO!

While this report is focused on the five 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 these companies and others to make the safety of their LGBTQ customers and users an urgent priority, both in their policy development *and* in their policy enforcement.

INTRODUCTION AND METHODOLOGY

Last year's [2022 edition of the Social Media Safety Index](#) garnered an unprecedented level of attention, shining the spotlight on the alarming and steadily growing epidemic of anti-LGBTQ hate and disinformation on social media platforms — with a particular focus on the high-follower hate accounts and right wing figures who continue to seed and amplify most of this activity.[9] The devastating impact of mis- dis- and malinformation (MDM) on society continues to be one of the most consequential issues of our time, with hate-driven and politically-motivated false narratives and conspiracy theories running rampant and causing real-world harm to our collective public health and safety. Not surprisingly such rhetoric puts historically marginalized groups, including LGBTQ people, in the crosshairs.[10] In addition to these egregious levels of inadequately moderated anti-LGBTQ material we are also seeing a 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[11] and similar suppression of LGBTQ content[12] (an area of concern which directly parallels current anti-

democratic trends across the country — such as the array of book ban efforts, Florida's "Don't Say Gay" bill, various so-called "Parents' Rights" campaigns, etc. — that seek to retract basic rights and freedoms of LGBTQ people).[13]

There are several areas of GLAAD's Social Media Safety program advocacy work. The most significant of these involves monitoring and documentation of harms, and direct confidential communications with teams and departments at the platforms. We are grateful to the platforms and to the individual colleagues at these companies who work to evaluate and implement our guidance. Each of the platforms can take pride (pun intended) in the extensive ongoing work they do in support of the LGBTQ community. Our chief emphasis in this report is on laying out the state of LGBTQ safety with regard to the products themselves, the fact that we are not devoting time to showcasing the *many* positive initiatives of the platforms is not for lack of appreciation of those efforts.

As with previous editions of the SMSI, in preparing this year's report, GLAAD reviewed thought leadership, research, journalism, and findings across the field of social media safety and platform accountability — as well as consulting with our GLAAD SMSI advisory committee and many other organizations and leaders in technology and social justice. As reflected in this 2023 SMSI [Articles and Reports Appendix](#) there are continual ongoing developments regarding the real-world impact of social media platforms on individual user safety and on public health and safety as a whole.

The centerpiece of this year's report is our Platform Scorecard. Developed in partnership with [Ranking Digital Rights](#) (RDR) and [Goodwin Simon Strategic Research](#) (GSSR), the 2023 Social Media Safety Index Platform Scorecard looks at twelve LGBTQ-specific indicators and evaluates each of the five major platforms utilizing RDR's standard methodology to generate numeric ratings for each product with regard to LGBTQ safety. Researchers interested in digging deeper into these results may explore this 2023 [Research Guidance](#) document.

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.

---

[9] [Online Amplifiers of Anti-LGBTQ+ Extremism | ADL](#)

[10] [What is "Grooming?" The Truth Behind the Dangerous, Bigoted Lie Targeting the LGBTQ+ Community | ADL](#)

[11] [Shadowbanning is real: How social media decides who to silence - The Washington Post](#)

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

[13] [Free speech groups condemn suppression of LGBTQ expression in Florida - National Coalition Against Censorship](#)
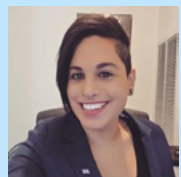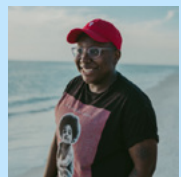
ADVISORY COMMITTEE

ALOK
(they/them)
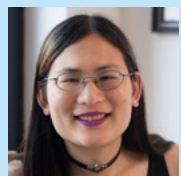Internationally acclaimed writer, performer, and media personality

Lucy Bernholz
(she/her)
Director, Digital Civil Society Lab at Stanford University

Alejandra Caraballo, Esq.
(she/her)
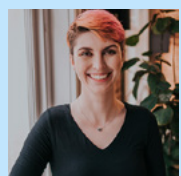Clinical Instructor, Cyberlaw Clinic, Berkman Klein Center for Internet & Society, Harvard Law School

Jelani Drew-Davi
(they/them)
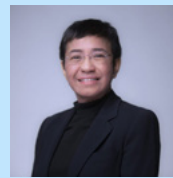Director of Campaigns, Kairos
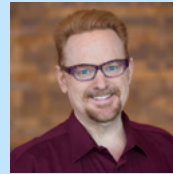
Liz Fong-Jones
(she/her)
Field CTO, Honeycomb

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

WE ALL DESERVE TO FEEL SAFE ON SOCIAL MEDIA — AND IN THE WORLD

"This [anti-LGBTQ] rhetoric and behavior reveals desperation against the growing acceptance of LGBTQ people and the broader awareness about gender diversity and expression. It's fomented by extremists who seek control by rejecting expertise and lived experience, and by social media companies that profit from fear, anger, and polarized politics."[14]

— GLAAD President and CEO Sarah Kate Ellis in *Time Magazine*

Right wing media outlets, hate-driven grifter pundits, and opportunist political figures continue to target LGBTQ people and the rights and dignity of other historically marginalized groups. In doing so, they make special use of social media, in many cases flagrantly violating the community guidelines and hate speech policies set in place by platforms to protect users. As these multi-billion dollar platforms continue to profit from such hate — specifically by refusing to meaningfully enforce those policies — our community should be rightfully furious with these companies as well.

It may sound naive to say, but here at GLAAD we firmly believe: *We all deserve to feel safe on social media — and in the world*. Indeed, even amidst so much hate expressed against us, this optimal vision of social media platforms *does* also currently exist: as a place where LGBTQ people and our allies express ourselves and connect — finding resources, community, activism, culture, and joy. But the threats encroaching on this vision have vastly increased.

In this past year we have seen an unprecedented surge of hateful, violent, and false rhetoric hurled at LGBTQ people both on and offline. From weaponized patently false conspiracy theories of LGBTQ people as "groomers" or threats to children, to lies and disinformation about gender-affirming care for trans youth, this toxic content is widely circulated on social media platforms, causing real-world harm to our community.

This is all paralleled by the alarming onslaught of legislative attempts to roll back our fundamental rights, and to restrict our freedom of expression and freedom of association. Amidst all of this, the prevalence of anti-LGBTQ hate, harassment, and disinformation on social media platforms continues to stand out as one of the most urgent problems facing our community.

According to the Anti-Defamation League (ADL), in their June, *2022 Online Hate and Harassment Report: The American Experience*: 66% of LGBTQ+ respondents reported experiencing harassment to-date (e.g., lifetime harassment) — a rate disproportionately higher than any other identity group (the rate for non-LGBTQ+ respondents was 38%). This 66% figure is up from 64% in the previous 2021 survey. Further,

[14]　No One Should Be Surprised by the Mass Shooting at Club Q | Time

54% of LGBTQ+ respondents also reported experiencing *severe harassment* to-date (defined as physical threats, sustained harassment, stalking, sexual harassment, doxing, or swatting). This figure is also a far higher rate than any other group, and is twice as high as the 26% rate for non-LGBTQ+ respondents. Note that the 2023 ADL report has not yet been released. An additional notable figure in the 2022 report is that 38% of all respondents reported occurrence of harassment due to their identity, up from 33% in 2021. The ADL report further specifies the online locations of these hate and harassment incidents: 68% of respondents reported any harassment to date taking place on Facebook (down from 75% in 2021), followed by 26% on Instagram (about the same as 2021, 24%), 23% on Twitter (24% in 2021), 20% on YouTube (21% in 2021), and 14% on TikTok (9% in 2021).
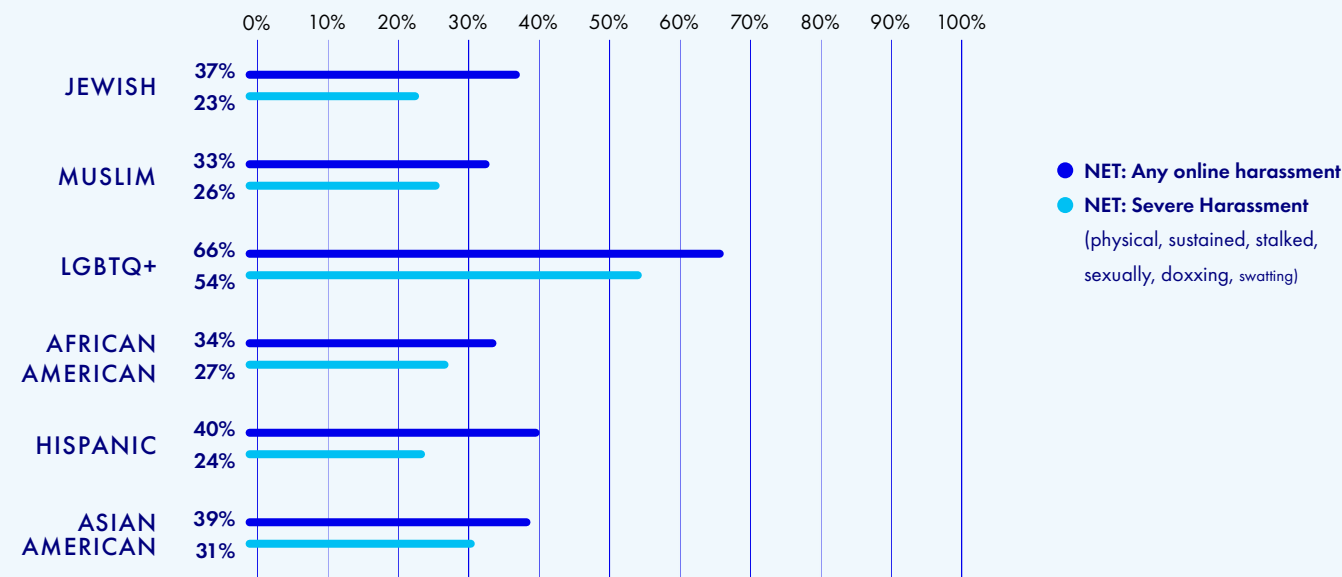
These numbers are absolutely unacceptable.

While Facebook, Instagram, Twitter, YouTube, TikTok and others must balance concerns around free expression, it cannot be stated strongly enough that social media platforms *must* take substantive, meaningful, and far more aggressive action to prioritize the safety of their LGBTQ users and to combat the spread of hate and extremism. These efforts must also address the growing prevalence of what has come to be known as MDM: misinformation (false, but not created or shared with the intention of causing harm), disinformation (deliberately created to mislead, harm, or manipulate), and malinformation (based on fact, but used out of context to mislead, harm, or manipulate).[15]

---

[15]   Foreign Influence Operations and Disinformation | Cybersecurity and Infrastructure Security Agency CISA

ONLINE HARASSMENT vs **SEVERE HARASSMENT**
EVER EXPERENCED: TARGETED GROUPS
Which, if any, of the following have happened to you, personally, ONLINE?

Group	NET: Any online harassment	NET: Severe Harassment
JEWISH	37%	23%
MUSLIM	33%	26%
LGBTQ+	66%	54%
AFRICAN AMERICAN	34%	27%
HISPANIC	40%	24%
ASIAN AMERICAN	39%	31%

● **NET: Any online harassment**
● **NET: Severe Harassment**
(physical, sustained, stalked, sexually, doxxing, swatting)

Source: ADL, Online Hate and Harassment Report: The American Experience 2022[16]

---

[16]   "Online Hate and Harassment Report: The American Experience 2022." ("The annual online hate and harassment survey of 2,330 American adults is conducted on behalf of ADL by YouGov, a public opinion and data analytics firm. The survey examines American adults' experiences with and views of online hate and harassment... Surveys were conducted from January 26th – February 14th 2022.")

GLAAD and other LGBTQ civil society groups engaged in advocacy work with social media companies repeatedly come up against loopholes in platform policies. These loopholes allow even the most egregious expressions of anti-LGBTQ hate to be evaluated as non-violative.

One of the most significant examples of this kind of loophole is Meta's distinction between public figures and private individuals. In Meta's newly created "Gender Identity Policy and User Tools" (launched in March 2023) the company explains, "We distinguish between public figures and private individuals because we want to allow discussion, which often includes critical commentary of people who are featured in the news or who have a large public audience. For public figures, we remove attacks that use derogatory terms related to sexual activity, calls for sexual assault or exploitation, calls for mass harassment, and threats to release private information. For private individuals, our protection goes further. We remove content that's meant to degrade or shame someone for their sexual orientation or gender identity, among other protections."[17]

This policy distinction leaves LGBTQ public figures unprotected from a broad range of truly shocking anti-LGBTQ hate (especially the longstanding practice of malicious targeted misgendering and deadnaming — directed specifically at prominent transgender public figures). Moreover, a recent report from GLAAD, UltraViolet, Kairos, and the Women's March shows that 60% of LGBTQ people feel harmed not only from harassment and hate directed against them personally, but also from witnessing hate and harassment of other LGBTQ community members such as celebrities and public figures.[18] Clearly, directing hate against LGBTQ notables is being used as a vehicle for expressing general anti-LGBTQ bigotry and hate. When companies maintain policy loopholes that allow such hate, this perpetuates harm against the entire community.

Another type of anti-LGBTQ hate that warrants specific note is the phenomenon of networked/stochastic harassment, in which extremist figures and accounts engage in posting characterizations of LGBTQ people which stoke animus and fear in their followers and incorporate a combination of vague and specific threats (ranging from dog whistle incitements to violence like "you know what to do," to explicit statements like "transgenderism [sic] must be eradicated from public life entirely"). When real-world harms and violence eventually occur (including the *threat* of physical violence), such events cannot be precisely tracked back to a specific source.

As the ADL explains, using Twitter as an example: "Influential Twitter users do not need to engage directly in hate and harassment to cause harm. Because Twitter moderation focuses on holding individual accounts responsible for harmful content, it frequently misses how influential accounts with large followings actually operate. This can be referred to as "stochastic harassment": weaponizing talking points that incite others to harassment without being a harasser."

Gays Against Groomers Instagram post (May 2023)

---

[17]   LGBTQ+ Safety

[18]   From URL to IRL: The Impact of social Media on People of Color, Women, and LGBTQ+ Communities - We Are UltraViolet

The ADL feature spotlights the "Gays Against Groomers" accounts to illustrate this phenomenon. With its publicly stated mission of opposing: "Propagandizing youth with LGBTQ+ media; Drag and pride events involving children; Queer and Gender Theory being taught in the classroom; and The mutilation and sterilization of minors,"[19] Gays Against Groomers as an entity is devoted to targeting LGBTQ people with dangerous, dehumanizing, and intentionally false anti-LGBTQ bigotry and to inciting animus in their followers. It has been identified by the ADL as "an anti-LGBTQ extremist coalition." The ADL's full article on networked harassment can be read here. Their exploration of Online Amplifiers of Anti-LGBTQ+ Extremism provides important additional context for other such accounts.

Briefly on the topic of Twitter, the company's increasingly worrying policy and product decisions[20] include the April 2023 removal of protections against targeted misgendering and deadnaming for trans and nonbinary users[21] and CEO Elon Musk's own personal extensive anti-LGBTQ posts and comments on the platform.[22] As Free Press co-CEO and Stop Toxic Twitter campaign co-lead Jessica Gonzalez puts it, Musk is sending the platform into "a death spiral as hundreds of advertisers have joined the mass exodus from the platform."[23] Accountable Tech executive director Nicole Gill concurs, telling *Axios* that "Any company cutting a check to Twitter is endorsing the surging racism, disinformation, and hate Musk has encouraged on his platform."[24] And *The Atlantic* succinctly concludes in a May 2023 feature headline that Twitter Is a Far-Right Social Network.

Social media companies often argue that moderation of such content is too difficult or expensive or somehow not possible. We have to ask what solutions they might discover if they were compelled to devote a greater portion of their billions of dollars in revenue toward maximizing product safety rather than maximizing profits. One recent illuminating example of this question of resource prioritization is Meta's March 2023 implementation of a solution to give advertisers control so that their ads won't appear near harmful content. As Reuters explains, "Marketers have long advocated for greater control over where their ads appear online, complaining that big social media companies do too little to prevent ads from showing alongside hate speech, fake news and other offensive content."[25]

This is another example of a new product development feature in which Meta prioritizes maximizing revenue over protecting user safety, and indeed public safety as a whole. This same emphasis on brand safety over user safety is shared by other platforms — including Twitter, which made a similar move in January,[26] and YouTube, which made similar updates in May.[27]

In fact there are many methods platforms already employ to curb anti-LGBTQ conduct and content, including adding context or links to posts (in the same way that platforms add an official voter information link to posts that include the word "vote" or "election"); removing, demoting, or limiting sharing of content; demonetizing posts or suspending accounts (some platforms apply a "three strikes and you're out" policy); and banning/de-platforming (individuals or organizations will not be allowed to create new accounts or pages on a given platform).

There are also numerous strategies — like speed-bumps or circuit breakers that throttle harmful viral content — that have been used effectively to *slow* the spread of misinformation, including anti-LGBTQ hateful content. These kinds of strategies have been implemented mainly in relation to public health and safety issues, especially most recently around COVID-19 and vaccine misinformation. It's worth noting that, as researcher James Laxa writes in the March 2023 issue of the *Journal of Public Health*: "it now appears that disinformation or, more specifically its consumption, is a public health issue in so much as it dramatically affects the wellbeing of individuals and social order."[28]

On the one hand, LGBTQ individuals are *vulnerable to hate speech* and other manifestations of online homophobia, biphobia, and transphobia. On the other hand, we are also *vulnerable to censorship* and disproportionate limitations of free expression related to our identities. In addition to being yet another thread in a social fabric of marginalization, bias, and oppression, the suppression of our speech (whether by human content moderators or by AI systems) creates real harms and obstructions for LGBTQ people — including

impacting our right to freely organize online, to access information, and to exercise our economic, social, and political rights. We know that LGBTQ youth especially need and depend on access to LGBTQ community online and on major social media platforms for "peer connection, identity development and management, and social support."[29]

After years of monitoring and reporting anti-LGBTQ content and appealing to these companies to enforce their own policies it is increasingly obvious that significant improvements will only be forthcoming when compelled by regulatory oversight (more on this shortly). In the meantime organizations like GLAAD and our colleagues in the field will continue to advocate for change and work to create a world where we all feel safe — on social media and in the world.

28   The consumption of disinformation as a health crisis

29   Social Media and Youth Mental Health - U.S. Surgeon General's Advisory, May 2023

19   Gays Against Groomers website

20   Evaluating Twitter's Policies Six Months After Elon Musk's Purchase | ADL

21   Twitter removes hateful conduct protections for transgender users

22   Musk's extremist tweets, via Eric Diesel

23   Civil-Rights Groups Give Twitter Advertisers a Deadline to Stop Funding Musk's Toxic Twitter | Free Press

24   Big tech still advertising on Twitter despite rise in hate speech

25   Meta rolls out long-sought tools to separate ads from harmful content | Reuters

26   Twitter Tries to Lure Advertisers Back With New Brand Safety Partnerships

27   IAS Enhances YouTube Brand Safety and Suitability Measurement Offering
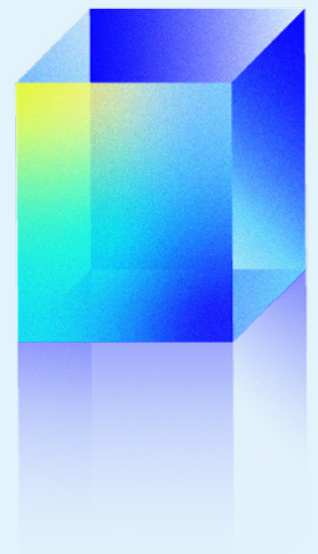
2023 SOCIAL MEDIA SAFETY INDEX PLATFORM SCORECARD

The 2023 SMSI **Platform Scorecard** consists of twelve indicators that draw on best practices and guidelines from the Ranking Digital Rights (RDR) Big Tech Scorecard, the annual evaluation of the world's most powerful digital platforms — benchmarking companies with indicators that set high but achievable standards for corporate transparency and policies that align with internationally recognized human rights standards.[30] The Scorecard evaluates the five major social media platforms: Twitter, Facebook and Instagram (whose parent company is Meta), YouTube (parent company: Alphabet/Google), and TikTok (parent company: ByteDance).

For last year's inaugural 2022 Scorecard, after developing a first set of draft indicators in close collaboration with GLAAD, the Goodwin Simon Strategic Research (GSSR) team revised and refined the indicators based on feedback from RDR, interviews with five expert stakeholders working at the intersections of technology and human rights, and input from the SMSI advisory committee. Additional methodological considerations were identified during the subsequent policy analysis and company research.

In creating the SMSI 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, companies are invited to provide written feedback as well as additional source documents. Of all the companies evaluated, only Twitter declined to participate in reviewing initial scorecard findings to provide feedback.

Note that these twelve 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 2023 Research Guidance. The full scoring sheets are also available here.

---

30   RDR Research Lab - Ranking Digital Rights

An Extremely Important Note About The Scorecard Ratings:

While the five platforms each have general policies prohibiting hate and harassment on the basis of sexual orientation and gender identity and/or expression, the Scorecard below does not include an indicator to rate them on *enforcement* of those policies. GLAAD and other monitoring organizations repeatedly encounter failures in enforcement of community guidelines across every platform. 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.

SMSI Platform Scorecard LGBTQ-specific indicators

1. The company should disclose a policy commitment to protect LGBTQ users from harm, discrimination, harassment, and hate on the platform.

2. The company should disclose an option for users to add pronouns to user profiles.

3. The company should disclose a policy that expressly prohibits targeted deadnaming and misgendering of other users.

4. 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.

5. 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.

6. 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.

7. The company should disclose that it prohibits advertising content that could be harmful and/or discriminatory to LGBTQ individuals.

8. The company should disclose the number of accounts and pieces of content it has restricted for violations of policies protecting LGBTQ individuals.

9. 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.

INSTAGRAM

63% SMSI SCORE

In the 2023 SMSI Platform Scorecard, Instagram receives a score of 63, a 15 points improvement from its score in the 2022 Scorecard. Notably, Meta has adopted a prohibition against targeted misgendering (though it requires direct reporting by the individual user and does not apply to public figures, and the company still does not disclose a similar disclosure related to targeted deadnaming). In its newly disclosed "Gender Identity Policy and User Tools" policy, the company also discloses a training for content moderators that educates them about gender identity policy enforcement. According to the policy, Meta "give(s) reviewers more explicit and detailed internal guidance about when to consider a trans, non-binary or genderfluid person to be attacked on the basis of gender identity...It involves providing guidance on the language used by the LGBTQ+ community to identify indicators for gender identity for trans, genderfluid, non-binary and gender nonconforming people (such as the Trans Pride flag)." Among other changes, the company has also made improvements to its Targeted Advertising policy, which now also prohibits wrongful ad targeting of users based on their gender identity (In last year's index, the company fell short of full credit on this indicator as a similar disclosure was only found related to sexual orientation).

However, the company continues to fall short of providing adequate transparency in several other key areas. Notably, Instagram does not have a policy in place that expressly protects transgender, gender non-conforming, and non-binary 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. Instagram also discloses only limited information regarding the options users have to control the company's collection and inference of user information related to their sexual orientation and gender identity.

KEY RECOMMENDATIONS

- **Make an express policy commitment to protecting transgender, non-binary, and gender non-conforming users from targeted deadnaming:** Similar to its prohibition against targeted misgendering, the company should adopt a policy that protects users from targeted deadnaming. Such policies should not require direct reporting by individual users and should also apply to public figures.

- **Provide all users with tools to express their gender identity:** The company should make its feature allowing users to add their preferred gender pronouns to their user profiles available to all users and provide more options for users to customize who can see their gender pronouns.

- **Give users full control over the company's collection and inference of user information related to their sexual orientation and gender identity:** The company should provide users with clear options to control the company's collection and inference of information related to their sexual orientation and gender identity.

FACEBOOK

61% SMSI SCORE

In the 2023 SMSI Platform Scorecard, Facebook receives a score of 61, a 15 point increase from last year's Scorecard. Notably, Meta has adopted a prohibition against targeted misgendering (though it requires direct reporting by the individual user and does not apply to public figures, and the company still has not disclosed a similar disclosure related to targeted deadnaming). In its "Gender Identity Policy and User Tools" policy, the company also discloses a training for content moderators that educates them about gender identity policy enforcement. The company has also made improvements to its Targeted Advertising policy, which now also prohibits wrongful ad targeting of users based on their gender identity. (In last year's index, the company fell short of full credit on this indicator as a similar disclosure was only found related to sexual orientation.)

However, the company's policies fail to adequately protect LGBTQ users in several other key areas. Facebook currently has no policy expressly protecting users from targeted deadnaming. In addition, the company discloses only limited options for users to control the company's collection and inference of user information related to their sexual orientation and gender identity. The company also publishes only limited data on the actions it takes to restrict content and accounts that violate policies protecting LGBTQ people.

KEY RECOMMENDATIONS

- **Make an express policy commitment to protect transgender, non-binary, and gender non-conforming users from targeted deadnaming:** The company should adopt a policy that protects users from targeted deadnaming. Such policies should not require direct reporting by individual users and should also apply to public figures.

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

- **Implement commitment to LGBTQ expression and privacy across the company:** Meta should disclose that it has an LGBTQ policy lead who advises policy and product teams on how the companies' policies, products, and services may impact the online privacy, expression, and safety of LGBTQ users.

TIKTOK

57% SMSI SCORE

In the 2023 SMSI Platform Scorecard, TikTok earns a score of 57, a 14 points increase from its score in the 2022 Scorecard. Notably, TikTok is currently the only company evaluated in the SMSI that protects transgender, gender nonconforming, and non-binary users from both targeted deadnaming and misgendering, and the only company to provide comprehensive information on how it detects violations to this policy. The company also made several improvements to other policies protecting LGBTQ users. For example, the company discloses limited information on options users have to control how their information related to sexual orientation and gender identity is used for targeted advertising. The company also discloses different trainings for content moderators educating them about the needs of LGBTQ users and other vulnerable communities. On the page "Combatting hate and violent extremism," the company also discloses it has a policy or program lead who focuses on the needs of LGBTQ users.

However, the company should provide greater transparency on other key issues. Rather than an outright ban of targeted advertising based on users' sexual orientation and gender identity, ad targeting limitations currently depend on local laws. The company also does not give users full control over the company's collection and inference of user information related to their sexual orientation and gender identity. While TikTok discloses a commitment to diversifying its workforce, it currently does not disclose data on its LGBTQ workforce.

KEY RECOMMENDATIONS

- **Give users full control over the company's collection and inference of user information related to their sexual orientation and gender identity:** TikTok should provide users with clear options to control the company's collection and inference of their user information, including information related to their sexual orientation and gender identity.

- **Ban targeted advertising based on users' sexual orientation and gender identity:** The company should implement an outright ban of third party advertising targeting users based on their sexual orientation and gender identity.

- **Follow up on commitment to diversify TikTok's workforce by publishing diversity numbers:** The company should publish data on its LGBTQ workforce, and break out this data by different teams.

YOUTUBE

54% SMSI SCORE

In the 2023 SMSI Platform Scorecard, YouTube earns a score of 54, a score improvement of nine points. YouTube discloses a training for content moderators that educates them about LGBTQ-related issues. The company has also made improvements to its diversity reporting, which discloses a voluntary reporting mechanism allowing Alphabet employees to disclose whether they identify as LGBTQ+. YouTube also discloses it has an "Inclusion Working Group" which helps "teams embed equity into the foundation of the platform and business." However, it is not clear whether it has a policy lead who addresses the needs of LGBTQ users more specifically.

The company also continues to fall short of providing adequate transparency in several other key areas. Notably, YouTube has no policy in place that expressly protects users from targeted deadnaming and misgendering. Alphabet continues to provide only limited information regarding the steps it takes to address demonetization, filtering, and removal of LGBTQ creators. In this context, the company's transparency reports also provide no data giving insights into removal and demonetization of LGBTQ creators from ad services.

KEY RECOMMENDATIONS

- **Make an express policy commitment to protect transgender, non-binary, and gender non-conforming users from targeted deadnaming and misgendering:** The company should adopt a policy that protects users from targeted deadnaming and misgendering.

- **Give users control over their own data:** YouTube should give users full control over their own data, including options to control the company's collection and inference of information related to users' sexual orientation and gender identity.

- **Show greater commitment to addressing demonetization and wrongful 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. This should also include comprehensive data on the wrongful removal of LGBTQ creators and their content and accounts.

TWITTER

33% SMSI SCORE

In the 2023 SMSI Platform Scorecard, Twitter earns a score of 33, a 12 points decrease from its score in the 2022 Scorecard. Notably, the company's "Hateful Conduct" policy no longer contains a prohibition against targeted deadnaming and misgendering. While the company makes a public commitment to diversifying its workforce, it appears that the company no longer publishes a Diversity Report disclosing data on its LGBTQ workforce. The company improved its disclosure regarding the options users have to control the recommended content they see on their Feeds based on their disclosed or inferred sexual orientation or gender identity.

Twitter falls short of providing adequate transparency in other key areas. The company currently does not disclose a feature that gives users an option to add their gender pronouns to their profiles. The company also does not disclose options for users to control the company's collection and inference of information related to their sexual orientation and gender identity. Further, Twitter does not disclose whether it has an LGBTQ policy lead who addresses the needs of LGBTQ users, or whether it has a formal training in place that educates all employees on the needs of LGBTQ users.

KEY RECOMMENDATIONS

- **Reinstate a policy commitment to protect transgender, non-binary, and gender non-conforming users from targeted deadnaming and misgendering:** The company should reinstate a policy expressly protecting users from targeted deadnaming and misgendering.

- **Implement commitment to LGBTQ expression and privacy across the company:** Twitter should disclose that it has an LGBTQ policy lead who ensures that the needs of LGBTQ users are implemented across the company's products, policies, and services.

- **Follow up on commitment to diversify Twitter's workforce:** In order to follow up on its commitment to taking proactive steps to diversify its workforce, Twitter should publish data on its LGBTQ workforce.

FROM URL TO IRL: 2022–2023 REPORTS ON LGBTQ SOCIAL MEDIA SAFETY
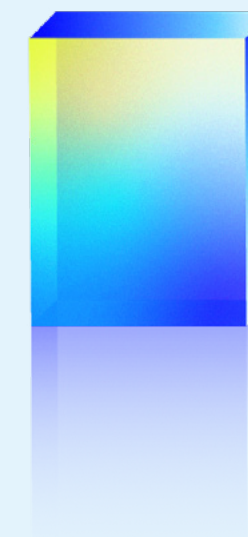
In 2021 the inaugural GLAAD Social Media Safety Index report offered the first-of-its-kind dedicated analysis on LGBTQ safety and social media platforms. There are now so many powerful reports and studies devoted to these issues (including how anti-LGBTQ rhetoric *online* results in anti-LGBTQ violence *offline*) we wanted to showcase a few of them here.

The dedicated year-round work of Media Matters for America bears special mention for being so extensive and prolific. Their dozens of dispatches and reports from their LGBTQ program can be found here. Highlights from the past year include: Daily Wire hosts have repeatedly violated YouTube and Spotify terms of service with statements of anti-LGBTQ hate (March 2023), Grifter Gays: How conspiracy theorists and right-wing operatives created Gays Against Groomers (February 2023), Meta is still profiting off ads that use the anti-LGBTQ "groomer" slur, despite the platform's ban (October 2022). Thanks to a new partnership between ADL and GLAAD focused on countering anti-LGBTQ+ extremism and hate there have been several powerful short-form reports released in the past six months including these three important overviews: At CPAC 2023, Anti-Transgender Hate Took Center Stage; Online Amplifiers of Anti-LGBTQ+ Extremism; and Antisemitism & Anti-LGBTQ+ Hate Converge in Extremist and Conspiratorial Beliefs. This December 2022 ADL Center for Tech and Society feature is an extremely important exploration of networked/stochastic harassment, an increasingly dangerous phenomenon. Please also see more info below about the October 2022 ADL report: Meta Profits Off Hateful Advertising.

From URL to IRL: The Impact of Social Media on People of Color, Women, and LGBTQ+ Communities

UltraViolet, Women's March, Kairos, and GLAAD — November 2022

This November 2022 report commissioned by UltraViolet, GLAAD, Kairos, and Women's March shows that women, people of color, and LGBTQ+ people experience higher levels of harassment and threats of violence on social media than other users. Among other key findings, the report shows that 57% of people have seen posts calling for physical violence based on a person's race, gender, or sexuality. Additionally, LGBTQ+ people and women respondents report higher rates of harassment than other groups. The study further shows that 60% of LGBTQ people feel harmed not only from direct harassment and hate, but from witnessing harassment against other LGBTQ community members such as celebrities and public figures (compared to only 24 percent of the base sample). We know that high-follower hate accounts show a pattern of directing hateful content against LGBTQ celebrities as a vehicle for expressing general anti-LGBTQ bigotry. We also know that social media companies maintain policy loopholes that permit hateful content against public figures to remain on their platforms; this perpetuates harm against entire communities. If platforms truly believe in making their products safe for LGBTQ people, these loopholes should be re-evaluated.

Digital Hate: Social Media's Role in Amplifying Dangerous Lies About LGBTQ+ People

Human Rights Campaign and the Center for Countering Digital Hate — August 2022

"Extremist politicians and their allies engineered an unprecedented and dangerous anti-LGBTQ+ misinformation campaign that saw discriminatory and inflammatory 'grooming' content surge by over 400% across social media platforms. [Content that] platforms not only failed to crack down on, but also profited from[...]. In a matter of mere days, just ten people drove 66% of impressions for the 500 most viewed hateful 'grooming' tweets — including Gov. Ron DeSantis's press secretary Christina Pushaw, extremist members of Congress like Marjorie Taylor Greene and Lauren Boebert, and pro-Trump activists like 'Libs of TikTok' founder Chaya Raichick. On Facebook and Instagram, 59 paid ads promoted the same narrative. Despite similar policies prohibiting anti-LGBTQ+ hate content on both social media platforms, only one ad was removed."

Fact Sheet: Anti-LGBT+ Mobilization on the Rise in the United States

Armed Conflict Location & Event Data Project — November 2022

"[In 2022] Acts of political violence targeting the LGBT+ community have more than tripled compared to 2021. With the role of social media platforms in the dissemination of mis-/disinformation — aggravated by ineffective content moderation policies and failures to quell the spread of false claims and conspiracy theories — the anti-LGBT+ narrative has reached far beyond the areas that have seen the highest concentration of offline activity." Also see the Anti-LGBT+ Mobilization section of ACLED's December 2022 report "From the Capitol Riot to the Midterms: Shifts in American Far-Right Mobilization Between 2021 and 2022."

Toxic Twitter: How Twitter Makes Millions from Anti-LGBTQ+ Rhetoric

Center for Countering Digital Hate — March 2023

"Twitter is making millions of dollars as anti-LGBTQ+ 'grooming' rhetoric jumps 119% under Elon Musk. Often targeting educators, pride events, or drag story hour events, the 'grooming' narrative demonizes the LGBTQ+ community with hateful tropes, using slurs like 'groomer' and 'pedophile.' The hateful 'grooming' narrative online is driven by a small number of influential accounts with large followings. Now new estimates from the Center show that just five of these accounts are set to generate up to $6.4 million per year for Twitter in ad revenues. These five accounts promote online hate that has been reported to have real-world violence, like harassment and threats, including some bomb threats."

Meta Profits Off Hateful Advertising

ADL — October 2022

"ADL's analysis found Meta has accepted large sums of money for ads on hateful topics such as antisemitism and transphobia [...]. Despite Meta forbidding baseless accusations of 'grooming' that target the LGBTQ+ community because they violate its hate speech policy, the company continues to profit off political ads promoting such hateful messages. An estimated 2.9 billion people use at least one of Meta's platforms daily, including Facebook and Instagram. Meta states, 'we have a responsibility to promote the best of what people can do together by keeping people safe and preventing harm.' Yet Meta regularly fails its users by profiting from ads that promote antisemitism and homophobia. Meta is not only providing a platform that allows these hateful messages to reach thousands of users, the company is also giving these narratives a dangerous level of credibility with its audience."

A Snapshot of Anti-Trans Hatred in Debates around Transgender Athletes

ISD Global — January 2022

"According to Twitter and Meta policies, transgender individuals, together with all members of the LGBTQIA+ community, are a protected group that should be safeguarded from hate speech on their platforms. However, new ISD research has found that these policies are poorly enforced and still suffer from gaps in implementation."

Digital Targeting and Its Offline Consequences for LGBT People in the Middle East and North Africa

Human Rights Watch — February 2023

"The targeting of LGBT people online is enabled by their precarious legal status[...]. In the absence of protection by laws or sufficient digital platform regulations, both security forces and private individuals have been able to target LGBT people with impunity. Under the United Nations Guiding Principles on Business and Human Rights, social media companies have a responsibility to respect human rights, including the rights to nondiscrimination, privacy, and freedom of expression. Digital platforms, such as Meta (Facebook, Instagram), and Grindr, are not doing enough to protect users vulnerable to digital targeting[...]. Digital platforms should invest in content moderation, particularly in Arabic, by quickly removing abusive content as well as content that could put users at risk. Platforms should conduct human rights due diligence that includes identifying, preventing, ceasing, mitigating, remediating, and accounting for potential and actual adverse impacts of digital targeting on human rights."

Report On Google Civil Rights Audit

WilmerHale — March 2023

In this civil rights audit released by Google in March 2023, the most substantial LGBTQ guidance (which continues to be a recommendation of the Social Media Safety Index) is this: "unless violative content is covered within its existing hate speech, harassment, and cyberbullying policies, YouTube's policies do not on their face prohibit intentional misgendering or deadnaming of individuals. Both acts have the potential to create an unsafe environment for users and real-world harm. We recommend Google review its policies to ensure it is appropriately addressing issues such as the intentional misgendering or deadnaming of individuals and continue to regularly review its hate and harassment policies to adapt to changing norms regarding protected groups." Additional recommendations include that YouTube expand "mandatory unconscious bias and LGBTQ cultural sensitivity training" for its moderators; that they continue to "evaluate how YouTube's products and policies are working for creators and artist communities of different races, ethnicities, gender identities, and sexual orientations;" and, "For ads that may consider gender for targeting purposes, Google should prioritize implementation of inclusive gender identity options for users and ensure targeting features respect those declarations."

Please see the 2023 SMSI Articles and Reports Appendix for links to dozens of other reports of interest. Important research and work on tech accountability is being done by numerous other organizations including: Access Now, ACLED, Accountable Tech, Amnesty International USA, the Anti-Defamation League, AVAAZ, BotSentinel, Center for American Progress, Center for Democracy & Technology, Color of Change, Consumer Reports, Dangerous Speech Project, Data & Society, Decode Democracy, EFF, Fight for the Future, Free Press, Global Project Against Hate & Extremism, Human Rights Watch, Institute for Strategic Dialogue, International Women's Media Foundation, Just Security, Kairos, the Leadership Conference on Civil and Human Rights, Media Justice, Media Matters for America, MPower Change, Muslim Advocates, the Oversight Board, PEN America, Ranking Digital Rights, Reframe Health and Justice, Southern Poverty Law Center, UltraViolet, Women's March, Woodhull Freedom Foundation, World Economic Forum. And many others.

REAL WORLD HARMS: HOW ONLINE HATE TURNS INTO OFFLINE VIOLENCE

"In recent years, the world has witnessed several mass atrocities. In many of these cases, hate speech was identified as a 'precursor to atrocity crimes, including genocide.' While the use of social media and digital platforms to spread hatred is relatively recent, the weaponization of public discourse for political gain is unfortunately not new. As history continues to show, hate speech coupled with disinformation can lead to stigmatization, discrimination and large-scale violence."[31]

— United Nations, Understanding Hate Speech

"We know that there is a direct correlation between being mentioned on this account [Libs of TikTok] and the targets of those tweets experiencing harassment, threats, and even acts of violence."[32]

— Sarah Moore, Anti-LGBTQ+ Extremism Analyst for ADL and GLAAD

Anti-LGBTQ rhetoric, and especially anti-trans rhetoric, has exploded across social media platforms. This horrible bigotry is promoted by high-follower accounts devoted to spreading hate-driven narratives (anti-LGBTQ, antisemitism, racism, misogyny, COVID dis- and misinformation, and miscellaneous conspiracy theories) meant to provoke outrage, drive engagement, and incite animus. These bad actors push increasingly extreme right wing political agendas while also seeking to consolidate power, mobilize voters, sow discord, and generate revenue from a combination of subscriptions, merchandise sales, donations, and ad revenue (including via platform creator monetization opportunities).

There are very real resulting harms to LGBTQ people *online*, including a chilling effect on LGBTQ freedom of expression for fear of being targeted, and the sheer traumatic psychological impact of being relentlessly exposed to slurs and hateful conduct (whether it is targeted at oneself directly or at LGBTQ celebrities or other members of the community as a vehicle for general anti-LGBTQ animus). There is also a direct line that can be drawn from this dangerous online content to physical real world violence and harms. These harms impact LGBTQ people — *and our entire society*.[33] Researchers have been documenting this phenomenon for years in relation to other

forms of online hateful rhetoric targeting other communities.[34] From the Christchurch massacre to the Myanmar genocide we know there are myriad instances of social media hate and disinformation leading to extreme violence and harm (specifically due to the negligence of social media companies and their willful refusal to effectively maintain the safety of their products).[35] In a recent fact sheet report, the Armed Conflict Location & Event Data (ACLED) Project revealed that, "Acts of political violence targeting the LGBT+ community have more than tripled [in 2022] compared to 2021."[36]

In this past year we've seen innumerable examples of *actual* physical violence and *threats* of physical violence — which have had devastating impacts on our communities. There is ample documentation of instances where such violence has been *prompted* by anti-LGBTQ social media content. The Documentation of Anti-LGBTQ Real World Harms spreadsheet below offers a powerful snapshot of these harms. Here is one concise example from a March 2023 Vice article:

"In August she [Chaya Raichik and Libs of TikTok] tweeted over a dozen times about Boston Children's Hospital and its gender-affirming care facilities, falsely claiming that they were providing gender-affirming hysterectomies to minors. As a result, doctors and nurses received death threats and the hospital received a bomb threat. In the days and weeks after these threats were made, Raichik made false claims about numerous other hospitals, including Akron Children's Hospital, the Children's National Hospital in Washington, D.C, and at least half a dozen others. After Raichik tweeted about them, staff at those facilities were threatened and harassed."[37]

31   Hate speech and real harm | United Nations

32   Transphobic Influencers Are Driving a Violent Groomer Conspiracy | Decade of Hate

33   Incendiary Speech That Spurs Violence is Rising in US, But Tools Exist to Shrink It

34   How online hate turns into real-life violence - The Washington Post

35   Hate Speech on Social Media: Global Comparisons | Council on Foreign Relations

36   UPDATE | Fact Sheet: Anti-LGBT+ Mobilization on the Rise in the United States

37   The GOP Is Weaponizing LibsOfTikTok's Anti-'Woke' Hate

Documentation of
Anti-LGBTQ Real World Harms

HEADLINE	SOURCE	DATE
After bomb threats, Iowa City school district removes book targeted by anti-LGBTQ Twitter account	Iowa City Press-Citizen	03-30-23
The Influencers Driving a Violent Groomer Conspiracy: "This is how the 'groomer' conspiracy theory is reshaping right-wing politics and how this can lead directly to violence and murder."	Vice News	02-28-23
Physical attacks track spikes in hate speech on Twitter	Washington Post	01-22-23
Does Online Hate Mongering Have Real Impact? Half of all targets of Libs of TikTok's digital harassment experienced real-world consequences within 5 days of the tweet.	Task Force Butler	12-22-22
How anti-trans hate speech online leads to real-world violence	Marketplace	12-21-22
Anti-LGBTQ hate thrives online, spurs fears of more violence	AP	12-17-22
Report details online harassment of trans health care providers	Axios	12-13-22
Groomer discourse intensifies and neo nazis celebrate in wake of Colorado Springs attack	ISD Global	11-23-22
Colorado Springs: Far-Right Influencers Made LGBTQ People Into Targets	SPLC	11-22-22
Doxxed Doctors, Library Bomb Threats, and Attacks on Pride Centers: A Week in Escalating Anti-LGBTQ Violence	The New Republic	09-28-22
Where Online Hate Speech Can Bring the Police to Your Door	The New York Times	09-23-22
Twitter account Libs of TikTok blamed for harassment of children's hospitals	Washington Post	09-02-22
Will it take another death to stop the spread of anti-trans hate online?	The LA Blade	08-31-22

HEADLINE	SOURCE	DATE
Libs of TikTok vows to continue targeting hospitals after Twitter suspension for hateful conduct	Media Matters	08-30-22
Children's hospitals targeted by anti-LGBTQ activists online	NPR	08-26-22
Meta and Twitter Refuse Action on 'Libs of TikTok' Posts As Doctors and Staff of Boston Children's Hospital Receive Death Threats	GLAAD	08-19-22
Boston Children's Hospital warns employees over far-right online harassment campaign	NBC News	08-16-22
Monkeypox and 'Groomers': How Twitter Facilitated a Hate-Riddled Public Health Disinformation Campaign	ISD Global	08-11-22
Followers of right-wing forum Libs of TikTok harass a Tosa first-grade educator for teaching pronouns at past job	Milwaukee Journal Sentinel	07-20-22
A GOP Congressman Wrote an Anti-Trans Bill Based Off YouTube Videos About Trans People	Them	07-14-22
Anti-LGBTQ threats, fueled by internet's far right 'machine,' shut down trans rights and drag events	NBC News	06-17-22
Hate speech and real harm	United Nations	The United Nations	06-10-22
The real victims in the "Libs of TikTok" discourse are the teachers and LGBTQ people harassed because of the account	Media Matters	04-19-22

ENRAGEMENT EQUALS ENGAGEMENT: BIG TECH MAKES BIG BUCKS ON ANTI-LGBTQ HATE

"The fact that Meta willingly and knowingly profits from ads which falsely and maliciously slander LGBTQ people as threats to children, in violation of its own hate speech policies, should make LGBTQ and ally employees of Meta ashamed and angry."[38]

— GLAAD spokesperson to NY Daily News, February 2023

As dangerous anti-LGBTQ content and disinformation saturate social media platforms, it remains bewildering that such material is so poorly moderated given the existence of platform policies which would seem to prohibit it. As researchers have observed for many years, it is not that the companies are doing a bad job at this enforcement, but rather that this is exactly how their products and policies are intentionally designed (as the title of this 2021 *AI Ethics* paper succinctly conveys: "Facebook's ethical failures are not accidental; they are part of the business model").

Organizations like Check My Ads, campaigns like The Facebook Logout, and coalitions like Stop Hate for Profit, Change the Terms, and Stop Toxic Twitter have shown how social media companies earn enormous profits from such hate and disinformation. Much of this content is generated by high-follower profit-driven "superspreaders"[39] and amplifiers[40] seeking *engagement* — driven by *enragement* — who generate click-throughs, shares, likes, and ad impressions (advertising being the primary source of revenue for social media company business models). The posting and circulation of anti-LGBTQ, and especially anti-trans, hate and disinformation has risen to extraordinary levels this past year. This epidemic is characterized by a combination of unfiltered bigotry and harassment; inflammatory and false mischaracterizations of LGBTQ-inclusive education, of trans healthcare, and of drag performers and events; as well as disingenuous rhetorical feints such as the shocking March 2023 call for anti-trans violence from right-wing extremist Michael Knowles. Knowles proclaimed in a speech to the CPAC (Conservative Political Action Coalition) conference[41] (in a speech that he previewed just days before on YouTube) that "transgenderism [sic] must be eradicated from public life entirely."[42]

It is hoped that there may be some improvements forthcoming by the end of the year as platforms face the need to comply with imminent global regulations such as the EU's Digital Services Act and are faced with corresponding financial penalties as they fail to abide by those regulations.

It is obvious that Meta, YouTube, Twitter, and TikTok have an inherent conflict of interest when it comes to enforcing (or *not enforcing*) their hate speech policies. The decision to *allow* anti-LGBTQ hate on their platforms not only benefits the grifters and bigots who post it, it also benefits the companies themselves. Reporting and research from the past year illuminates this phenomenon: Facebook is making millions off Matt Walsh's transphobic documentary (*Xtra Magazine*, Oct 2022); Meta has profited from over 200 ads using the anti-LGBTQ "groomer" slur, even though the platform claims it prohibits the term (Media Matters for America, Feb 2023); Toxic Twitter: How Twitter Makes Millions from Anti-LGBTQ+ Rhetoric (Center for Countering Digital Hate, March 2023); Meta has profited from millions in ad revenue from The Daily Wire's anti-trans campaign (Media Matters for America, March 2023).

2022 Annual Advertising Revenue By Platform:

Facebook & Instagram (Meta) ($31.2 billion)[43]

YouTube ($29.24 billion)[44]

TikTok ($11.4 billion)[45]

Twitter (unclear since Twitter is now a private company, but earnings have reportedly fallen by 40% since Elon Musk's acquisition; 2021 ad revenue was $5.08 billion[46], *Insider Intelligence* projects that 2023 ad revenue will plummet to $2.98 billion.[47])

[38] Meta has profited from more than 60 ads using homophobic slur 'groomer' since December

[39] Social Media's Role in Amplifying Dangerous Lies about LGBTQ+ People

[40] Online Amplifiers of Anti-LGBTQ+ Extremism | ADL

[41] At CPAC 2023, Anti-Transgender Hate Took Center Stage | ADL

[42] CPAC speaker: 'Transgenderism must be eradicated from public life'

[43] META Earnings: All you need to know about Meta Platforms' Q4 2022 earnings results | AlphaStreet

[44] YouTube: share of Google net ad revenues 2022 | Statista

[45] TikTok Ad Revenue (2019–2024) [Jan 2023 Update] | Oberlo

[46] Twitter revenue, earnings reportedly fell 40% shortly after Musk buyout | Ars Technica

[47] Twitter's ad business will bring in $2 billion less in 2023 than previously projected - Insider Intelligence Trends, Forecasts & Statistics

UNDERSTANDING ANTI-LGBTQ CONSPIRACY THEORIES AND DISINFORMATION

"LGBTQ people in America are under attack like never before."

— Under Fire: The War on LGBTQ People in America, Movement Advancement Project (February 2023)[48]

"Movements grounded in attempts to whitewash history and deny the rights of the LGBTQ+ community have turned hate into campaign platforms. Contributing to this climate are social media companies that have not internalized the lessons of the past and have set the stage for a 2024 election year that is at least as toxic online as past elections. Platforms have policies in place that curb and prevent the spread of hate and voting disinformation, but they do not consistently enforce them. Furthermore, major platforms have cut back or eliminated their trust and safety staff and hollowed out protections against hate incitements on their platforms."

— Cause for Concern 2024: The State of Hate, The Leadership Conference Education Fund (April 2023)[49]

There have been three particularly alarming and consequential tropes in anti-LGBTQ hate speech over the past year which have served as the most popular fodder of high-follower anti-LGBTQ hate accounts across the major social media platforms. The increasingly widespread "Groomer" libel is the vicious, false, and baseless assertion that LGBTQ people are a threat to children. This dangerous trope has been accompanied by an equally popular strain of extremist rhetoric baselessly and falsely mischaracterizing safe, effective, evidence-based gender affirming healthcare for trans youth with terms like "child abuse, mutilation, sterilization, etc." This inflammatory mischaracterization, which has gained traction with right wing extremists leading to bomb threats on hospitals and targeting of health care providers, is a very clear form of anti-trans hate speech intended to foment animus and violence against trans people and their allies. The other top trope of the past several years, targeted misgendering and deadnaming of trans and nonbinary people, continues to be an extremely widespread mode of hate speech[50] across all platforms. (In response to being relentlessly targeted with this hate in July 2020, Admiral Rachel Levine described such attacks as: "intolerance and discrimination against LGBTQ individuals, and specifically transgender individuals.") The practice is specifically

utilized to bully and harass prominent public figures while simultaneously expressing hatred and contempt for trans and non-binary people in general.[51]

These kinds of conspiracy theories, tropes, dog whistles, and harmful rhetoric have terrible real world impacts. Taking the "Groomer" libel as an example, as a recent Institute for Strategic Dialogue (ISD) Global report summarizes: "Around the world today, the use of the term 'groomer' is used to justify hate, discrimination and violence against the LGBTQ+ community. In the US particularly, the use of this language, along with conspiratorial thinking around queer people, has led to legislation preventing the discussion of LGBTQ+ issues in schools and preventing trans children from accessing gender affirming healthcare, and has motivated attacks on LGBTQ+ individuals."[52]

The ISD report further explains: "Part of the success of this mainstreaming lies in the ability of fringe actors to manipulate the general public's lack of knowledge of

---

48   Movement Advancement Project | Under Fire Series: The War on LGBTQ People in America

49   Cause for Concern 2024: The State of Hate

50   The Relentless Misgendering of Dr. Rachel Levine, Pennsylvania Health Secretary, Is Violence

51   Currently, TikTok and Pinterest explicitly prohibit targeted misgendering and deadnaming in their hate and harassment policy, while for both Facebook and Instagram, Meta's "Gender Identity Policy and User Tools" policy states that they "prohibit violent or dehumanizing content directed against people who identify as LGBTQ+ and remove claims about someone's gender identity upon their request." This policy can be interpreted as a prohibition against targeted misgendering (though it requires direct reporting by the individual user who is being misgendered, and does not apply to public figures, a loophole which leaves transgender notables unprotected and allows anti-trans hate accounts to target them as a way of expressing general anti-trans animus). This same policy can also be potentially interpreted as a prohibition against targeted deadnaming, since deadnaming involves a "claim about someone's gender identity." The company has not made any public statements with regard to this interpretation of the policy. In 2021, GLAAD joined Media Matters for America and 18 other organizations in calling for YouTube to also align with this as a best practice in their hate speech policy. Twitter first enacted its policy against targeted misgendering and deadnaming in 2018 but quietly removed it in April 2023.

52   The 'Groomer' Slur - ISD

queer culture and particularly their insensitivity to the plight of trans people. This has been coupled with the most potent fear — that of people harming children, which has been used to justify hatred and irrationality for centuries. In reality, the 'groomer' slur harms those children who are most in need of support — queer and gender non-conforming children. According to The Trevor Project's 2023 National Survey on LGBTQ Youth Mental Health, 41 percent of LGBTQ+ youth have seriously considered attempting suicide in the past year."

Note that Meta's hate speech policies specifically prohibit anti-LGBTQ "groomer" content ("Do not post: Content targeting a person or group of people [...] on the basis of their aforementioned protected characteristic(s) with: Dehumanizing speech or imagery in the form of comparisons, generalizations, or unqualified behavioral statements (in written or visual form) to or about: [...] **Sexual predators**.")[53] and the company has made public statements underscoring this.[54] And yet the company consistently does not enforce the policy, including in their allowance of multiple Gays Against Groomers accounts which continue to exist across Facebook and Instagram despite containing the violative language in their account names and being almost exclusively devoted to perpetuating this bigoted and dangerous lie.[55] This is even

more heinous given how Meta itself characterizes the harms of such hate in its own hate speech policy (and reiterates in its newly created document: "Gender Identity Policy and User Tools"): "We don't allow hate speech on Facebook. It creates an environment of intimidation and exclusion, leads to dehumanization and in some cases can promote offline violence."

As documented in GLAAD's Guide to Anti-LGBTQ Online Hate Speech, there is a long history and a wide array of additional examples of these types of disingenuous rhetorical strategies, arising from a variety of sources and driven by various motivations. The newest of these tropes is the anti-trans concept of "transgenderism."

---

[53]   Hate Speech | Transparency Center

[54]   A political memes subreddit kicked off an internet-wide call to get baseless 'groomer' claims classified as hate speech

[55]   Hate Speech | Transparency Center "Do not post: Content targeting a person or group of people [...] on the basis of their aforementioned protected characteristic(s) with: Dehumanizing speech or imagery in the form of comparisons, generalizations, or unqualified behavioral statements (in written or visual form) to or about: [...] Sexual predator."

HATE IS HATE AND LIES ARE LIES — CASE STUDY EXAMPLE: "TRANSGENDERISM"

Countless right-wing bad actors employ disingenuous rhetorical trolling as a strategy for asserting that their anti-LGBTQ hate content is *not* anti-LGBTQ hate, when it very clearly is. The ongoing refusal of platforms to confront this pernicious strategy lies at the heart of many of the worst, high-follower anti-LGBTQ hate accounts across social media. Accounts like Gays Against Groomers and Libs of TikTok; the stables of hate-driven grifters at far-right media entities like The Daily Wire, Blaze Media, Turning Point USA; figures like Tim Pool, Liz Wheeler, et al., and so many others, express vicious anti-LGBTQ animus while feigning innocence. When companies accept such rhetorical gaslighting at face value — especially when there is such an obvious financial conflict of interest — they become complicit in anti-LGBTQ hate, discrimination, and violence.

In a quintessential example of this malicious strategy, after his March 2023 CPAC speech (in which he stated that: "Transgenderism [sic] must be eradicated from public life entirely") Michael Knowles immediately responded to critiques of his statement by asserting that his beyond-the-pale genocidal rhetoric was referring to an ideology and not to trans people (note that the current usage of "transgenderism" as a term arises from anti-trans extremists and, like phrases such as "gender ideology," is crafted to delegitimize trans people by implying that being trans is an ideology rather than an identity). Or as Parker Molloy summarizes it, in a terrific piece called On the Right's Call

to "Eradicate Transgenderism" (It Means Exactly What You Think It Means): "A reminder that words like 'transgenderism' and 'gender ideology' are almost exclusively used by anti-trans activists to obscure the fact that trans people are simply people who just happen to be trans. It's not a belief system."

It is also worth noting that Knowles posted a version of his speech (featuring this and other extreme anti-trans rhetoric)

on YouTube just a few days prior to his CPAC appearance[56] and then another (featuring the same statements) immediately after.[57] Variations of the same rhetoric continue to saturate all of his social media accounts,[58] and the use of the term amongst other far-right media figures skyrocketed in the weeks following his speech.

From a social media strategy standpoint, the term has also served these bad actors to disingenuously evade the hate speech and harassment policies of the platforms — which prohibit speech targeting *people* with hate on the basis of protected characteristics (while speech targeting *ideologies* is allowed). Despite the fact that the term is a dog whistle expressing contempt and hatred of transgender people, the platforms refuse to recognize it as a breach of their own rules.

---

56   "The Trans Card" Is A Weapon For Libs And Criminals | Ep. 1192

57   Fake News Must Be Eradicated From Public Life | Ep. 1196

58   How To Eradicate Transgenderism From Public Life - Facebook

"TECH COMPANIES WILL NOT POLICE THEMSELVES" — THE TIME FOR REGULATORY OVERSIGHT OF SOCIAL MEDIA PLATFORMS IS NOW

"[Elon] Musk's Twitter is simply a further manifestation of how self-regulation by tech companies will never work, and it highlights the need for genuine oversight. We need well-staffed and well-resourced teams working inside tech companies to ensure that algorithmic harms do not occur, but we also need legal protections and investment in external auditing methods. **Tech companies will not police themselves**, especially not with people like Musk in charge. We cannot assume — nor should we ever have assumed — that those in power aren't also part of the problem."[59]

— Rumman Chowdhury, Responsible AI Fellow, Berkman Klein Center, Harvard University

As things stand currently for the major social media companies, anti-LGBTQ hate, harassment, and malicious disinformation continue to flow freely on their platforms (while legitimate voices are often stifled), and they design their algorithms and implement their AI to maximize revenue without sufficient concern for negative impacts on users and public health and safety.

There is an urgent need for effective regulatory oversight of the entire tech industry — and especially social media companies — with the goal of protecting LGBTQ people (and everyone) from the dangerous impacts of an industry that continues to prioritize private profits over the public interest. An especially notable related example is the realm of generative AI, as evidenced by an April 2023 study from the Center for Countering Digital Hate which discovered that Google's "Bard," its AI chatbot product, had been generating extreme hate and misinfo, including anti-LGBTQ material, antisemitism, racism, and misogyny.[60] This is merely one of many such alarming examples.

As is the case with other major industry sectors (agriculture, banking, energy, telecommunications, etc.) "regulatory oversight" is about implementing transparency and accountability mechanisms for protecting public safety.[61] While such solutions may be complex and thorny, it is essential that thoughtfully crafted approaches proceed apace. As tech accountability experts Nathalie Maréchal, Rebecca MacKinnon and Jessica Dheere concisely observe in a seminal 2020 report from Ranking Digital Rights:

---

59   I Watched Elon Musk Kill Twitter's Culture From the Inside - The Atlantic

60   Misinformation on Bard, Google's new AI chat - Center for Countering Digital Hate

61   Transparency is essential for effective social media regulation

"Holding companies accountable to the public interest is the responsibility of lawmakers, in consultation with civil society."[62] It is also important to note that bills being proposed to create accountability for tech companies must be carefully crafted so as not to create unintentional negative impacts on LGBTQ users, especially LGBTQ youth (a serious issue with many recent legislative proposals attempting to address children's online safety).[63]

It is also far past time for the implementation of true data privacy[64] protections,[65] including reining in the practice of targeted surveillance advertising, and demanding algorithmic transparency.[66] The need for data privacy for LGBTQ users is illuminated in a May 2023 *Wall Street Journal* article on TikTok's tracking of users who watched LGBTQ content which points out that, "No comprehensive U.S. privacy law regulates the practice of collecting sensitive data. Seven states have passed privacy laws, including some that require companies to protect certain categories of sensitive data, including gender and sexual identity."[67] There are many unique potential ramifications for LGBTQ users, including unintentional outing of LGBTQ people (including youth), and the potential for government requests of such

data in ways that can endanger people, as the *WSJ* article notes: "This data can essentially create a list of vulnerable users in parts of the world where some LGBT people face harassment and violence."

Congress, the White House, existing federal agencies and others can and must drive forward the development and implementation of such regulatory oversight — working in partnership with civil society, tech policy experts, and others, and looking to other work being done around the world in this field, for example the EU's Digital Service Act (DSA) and Digital Markets Act (DMA).[68]

As online hate content is fomenting anti-LGBTQ violence and leading to real-world harm against our community (including bomb threats targeting children's hospitals that offer healthcare for trans youth,[69] attacks on drag events[70] and Pride festivals,[71] and the current wave of draconian legislative attacks retracting basic rights, especially targeting transgender people[72]) clearly this is truly one of the most important issues facing the LGBTQ movement today.

As noted media personality, writer, performer, and SMSI advisor ALOK eloquently expressed at the 2023 Upfront Summit tech conference: "Let's hold Big Tech accountable for violence against LGBTQ people. Let's redefine the agenda of tech. Let's harness the innovation that this industry has to actually create a more generous and kind world."[73]

62   Getting to the Source of Infodemics: It's the Business Model - Ranking Digital Rights

63   How far should the government go to control what your kids see online?

64   State Bills Aren't Enough: The Case for National Legislation on Data Privacy and Civil Rights

65   U.S. data privacy laws to enter new era in 2023 | Reuters

66   The Slow Death of Surveillance Capitalism Has Begun | WIRED UK

67   TikTok Tracked Users Who Watched Gay Content, Prompting Employee Complaints - WSJ

68   The internet is about to get a lot safer | MIT Technology Review

69   Boston Children's Hospital warns employees over far-right online harassment campaign

70   https://www.glaad.org/blog/anti-drag-report

71   Pride events targeted in surge of anti-LGBTQ threats, violence

72   Anti-Trans Legislative Risk Map - by Erin Reed

73   Jamie Lee Curtis Interviews ALOK on the World Beyond the Gender Binary | 2023 Upfront Summit

MITIGATING ONLINE ANTI-LGBTQ HATE AND DISINFORMATION, SUPPRESSION OF LGBTQ VOICES, AND THE NEED FOR TRANSPARENCY

"Tech companies control online information flows on their platforms through proprietary rules and Terms of Service, giving them significant power with little accountability. Communities already facing discrimination are also at risk of having their content removed online through discriminatory flagging campaigns or biased moderation processes, and thus face being doubly silenced. Wrongful action taken on content can have a disproportionate impact on already-vulnerable populations, such as members of ethnic or religious minorities, LGBTQ+ people, and women. It also routinely affects journalists, political activists, and human rights defenders operating in repressive environments."

— Santa Clara Principles on Transparency and Accountability in Content Moderation[74]

It is an unfortunate state of affairs that private companies are the gatekeepers of so much of our public discourse. And yet, here we are. Clearly, we face a long ongoing journey towards some solutions, with the harms from these products

continuing to impact us all. In addition to the need for mitigation of anti-LGBTQ hate and disinformation, companies must also address the problem of suppression of LGBTQ voices — and, as in previous editions of the SMSI, we urge the platforms to improve transparency across multiple realms. This includes the need to provide additional data points in their transparency reports (also known as terms of service enforcement or content removal reports), as well as visibility into their appeals processes.

It's important to be mindful that transparency reports need to share granular data about *all* take-downs and user appeals in order to also make visible the problem of the over-moderation, suppression, demonetization, or deplatforming of legitimate posts and accounts (for instance, there is extensive documentation of the systemic deplatforming of sex workers and sex educators even when they are not violating terms of service[75] — and specifically there are double standards in this regard for non-LGBTQ vs. LGBTQ people). As noted in a 2022 article in *Prism*: "The users most censored and deplatformed are people of color, LGBTQ+ folks, and people of other marginalized identities."[76]

GLAAD continues to advocate for approaches in line with best practices in the field, including the Santa Clara Principles

74   Santa Clara Principles

75   https://www.tandfonline.com/doi/abs/10.1080/23268743.2021.2005907?journalCode=rprn20

76   Sex workers say legislation is needed to prevent censorship on online platforms

to convey anti-trans hate and lies, and Walsh can point viewers to other revenue-generating sources like subscriptions, merchandise, and his own independent sponsors and advertisers). Another recent mitigation is Meta's AP Fact Check overlay applied to the Liz Wheeler "LGBTQS for Satanism" video on her Facebook and Instagram pages (which of course still conveys a vicious, dangerous anti-LGBTQ lie — that is 100% in violation of Meta's hate speech policy — once you click past the label). And then there's pre-Musk Twitter's suspension of Jordan Peterson's account for his aggressive targeted anti-trans harassment of Elliot Page (though the account has now been restored by Elon Musk)[78] as well as TikTok's outright suspension of Gays Against Groomers and Libs of TikTok for their incessant anti-LGBTQ animus and incitement of networked harassment campaigns.

On the one hand these mitigations are commendable examples of enforcing basic hate speech policies. The most important takeaway from these instances of enforcement is that Meta, YouTube, Twitter, and TikTok could decide to implement mitigations of *huge quantities* of this blatant hate speech and harassment. They could choose to protect LGBTQ people (and everyone, really) from the high-follower accounts which post dangerous and consequential anti-LGBTQ hate — material that clearly violates their own policies. But they don't. They actively and intentionally decide thousands of times a day to interpret such material as being allowable. Which steadily dilutes the purpose of their hate speech policies and ultimately serves to *encourage* hate accounts to push the boundaries further and further to increase account engagement and revenue.

It is worth noting that TikTok does often appear, at least anecdotally, to have a lower tolerance for such overt anti-LGBTQ hate, and as a result the content posted by at least some of these far-right figures on their TikTok accounts is not as egregiously hateful as the content they post on other platforms.

It is also very much worth noting that *many* other sites and services have suspended their relationships with these high-follower hate-driven figures and accounts for being in such clear violation of their terms of service. For example, in addition to being suspended five times from (pre-Musk) Twitter for promoting "violence, threats or harassment against others based on their sexual orientation or other factors such as race or gender,"[79] the far-right anti-LGBTQ account Libs Of TikTok was removed entirely from TikTok for violating community guidelines,[80] suspended temporarily from Instagram, and permanently suspended from both Slack and Linktree.[81] Similarly, the anti-LGBTQ extremist coalition Gays Against Groomers has been suspended from (pre-Musk)

Twitter four times.[82] While Venmo,[83] Printful[84] and Wix.com[85] have all suspended or cut ties with GAG entirely due to it being in violation of the hate speech clauses in their terms of service. Also, TikTok terminated the account entirely in the fall of 2022.[86]

Removing violative content isn't the only way for platforms to address anti-LGBTQ hate and disinfo. As evidenced from the various examples above, companies have a wide array of proactive and reactive options, tools, and modalities. A striking example of this ability to implement mitigation is illustrated in the damning November 2020 *New York Times* story, "Facebook Struggles to Balance Civility and Growth," which describes how, in the days after the 2020 US Presidential election, the platform implemented an algorithm to demote posts it had determined were: "bad for the world," but that, because of the resulting reduction in site engagement (and corresponding negative impact on revenue), the decision was made to "less stringently demote such content."

---

on Transparency and Accountability in Content Moderation. As referenced above, the need for thoughtful regulatory solutions, while extraordinarily challenging, is urgent.[77]

With regard to mitigation of anti-LGBTQ hate, it is interesting to see some of the recent examples where platforms have decided to apply their terms of service to either remove, demonetize, or filter content or accounts. To name a few: YouTube's April 2023 demonetization of the channel of far-right anti-trans media personality Matt Walsh (although his videos have been demonetized they nonetheless continue

77   Facebook, Twitter and other social media companies need to be treated like Big Tobacco

78   YouTube demonetized the corollary video that Peterson posted there (it has more than 3.4 million views). The same video, in which Peterson relentlessly misgenders and deadnames Page in a bigoted 15-minute rant, currently has more than 1.3 million views on Peterson's official Facebook page and 600,000 views on Instagram. He did not post it to his TikTok account, presumably knowing that the platform would not have allowed it. It's evident that Peterson is targeting Page as a means for expressing contempt, abuse, insults, and hate towards transgender people in general.

79   Libs of TikTok - Wikipedia

80   Meta and Twitter Refuse Action on 'Libs of TikTok' Posts As Doctors and Staff of Boston Children's Hospital Receive Death Threats | GLAAD

81   https://linktr.ee/status/blocked?username=libsoftiktok

82   Gays Against Groomers - Wikipedia

83   Google Bans Anti-Trans Hate Group Gays Against Groomers

84   Printful

85   https://web.archive.org/web/20220818005126/https://twitter.com/againstgrmrs/status/1560064757645955072

86   https://web.archive.org/web/20220706151645/https://www.gaysagainstgroomers.com/

ACKNOWLEDGMENTS

WE KEEP US SAFE

The emphasis of the GLAAD Social Media Safety program is on platform accountability — advocating for all social media companies, apps, and websites to prioritize the safety of their LGBTQ users. In addition to this advocacy we are also working to provide resources to LGBTQ people directly, to help empower everyone to be safer online.

These companies themselves should be doing a better job making their products safe, and the onus for safety should not fall so much on individual users. However, given the dangers in the current social media landscape and our community's need for assistance GLAAD has pulled together a new LGBTQ Digital Safety Guide featuring some easy tips to help protect yourself and our community.