SOCIAL MEDIA SAFETY INDEX

glaad

2022

TABLE OF CONTENTS

LETTER FROM GLAAD PRESIDENT & CEO SARAH KATE ELLIS

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.

The evolution of social media has significantly shifted the way that people relate to LGBTQ communities, and GLAAD has continued to innovate to keep step with the rapidly and ever-changing media landscape. Since its launch last year, our Social Media Safety Index (SMSI) is creating and sustaining the industry's first standard for tackling online hate and intolerance and increasing safety for LGBTQ social media users. The Index and our new Platform Scorecard provide targeted industry recommendations on LGBTQ user safety across all five major social platforms: Facebook, Instagram, Twitter, YouTube, and TikTok. While we've seen important progress from the platforms this past year (many arising from our advocacy efforts), there is an enormous amount of work that must still be done to better protect LGBTQ users. This work is urgent with our latest research with Community Marketing & Insights finding that 84% of LGBTQ adults agree there are not enough protections on social media to prevent discrimination, harassment, or disinformation. 40% of all LGBTQ adults, and 49% of transgender and nonbinary people, do not feel welcomed and safe on social media.

At GLAAD, we know that media — including social media — is an empathy machine that has the power to change hearts and minds. From an LGBTQ perspective, it is not enough for companies to post a rainbow during a Pride month marketing campaign or use LGBTQ

creators to make their brands seem diverse and inclusive, while failing to stand up for us and protect us in real-world ways. We often say to companies and brands: You have to join the movement, not just market to the moment.

LGBTQ people are under attack right now, all across the globe. Since the start of 2022, Republican lawmakers have proposed 325 anti-LGBTQ bills, 130 of which specifically target the rights of transgender people, especially trans youth.

Attacking vulnerable groups of people as a political strategy, and stoking fear and hatred about them, is something we've seen across history. It's a reprehensible practice — and the spread of such hate today is further facilitated by social media platforms. This type of rhetoric and "content" that dehumanizes LGBTQ people has real-world impact. These malicious and false narratives, relentlessly perpetuated by right wing media and politicians, continue to negatively impact public understanding of LGBTQ people — driving hatred, and violence, against our community.

At this point, after their years of empty apologies and hollow promises, we must also confront the knowledge that social media platforms and companies are prioritizing profit over LGBTQ safety and lives.

This is unacceptable.

Here at GLAAD we will continue to reach out to platforms in real time about content and policies that place LGBTQ users in jeopardy. This Index and our larger program will continue to demand that social media platforms make their products safe for LGBTQ users, and for everyone.

SARAH KATE ELLIS
President & CEO, GLAAD

EXECUTIVE SUMMARY

The goal of the GLAAD Social Media Safety Index (SMSI) report, and our Social Media Safety program as a whole, is to push the major social media platforms to be accountable and transparent — specifically to the LGBTQ users they serve, but also to society as a whole. The centerpiece of this second edition of the SMSI is our **Platform Scorecard**. Created in partnership with Goodwin Simon Strategic Research and the noted Big Tech accountability watchdog group, Ranking Digital Rights, the Scorecard offers an evaluation of LGBTQ safety, privacy, and expression on five major platforms (Facebook, Instagram, Twitter, TikTok, and YouTube) based on 12 LGBTQ-specific indicators. A full **RESEARCH GUIDANCE** document is included as an additional component of the Scorecard.

Problems identified in the Scorecard, and in the SMSI report in general, include: inadequate content moderation and enforcement (including issues with both 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 even more exacerbated for folks 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 they are.

A FEW OF OUR MOST IMPORTANT 2022 RECOMMENDATIONS ACROSS PLATFORMS INCLUDE:

Improve design of algorithms that currently circulate and amplify harmful content, extremism and hate.

Be transparent with regard to content moderation, community guidelines and terms of service policy implementation, and algorithm designs.

Train moderators to understand the needs of LGBTQ users, and to moderate across all languages, cultural contexts, and regions.

Strengthen and enforce existing community guidelines and terms of service that protect LGBTQ people and others.

Observe human rights frameworks and work with independent researchers to understand how to keep users safe.

Respect data privacy, especially where LGBTQ people are vulnerable to serious harms and violence. This includes ceasing the practice of targeted surveillance advertising, in which companies use powerful algorithms to recommend content to users in order to maximize profit.[1]

Confront the problems of bias in artificial intelligence (AI) which disproportionately impact LGBTQ people and other marginalized communities (this is also intricately connected with surveillance advertising).

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 action will be taken. This is gravely concerning.

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. Through our year-round presence and presentations at major technology and media conferences and events, GLAAD also looks forward to maintaining an ongoing dialogue about LGBTQ platform safety, privacy, and expression among industry colleagues, as we work together shoulder to shoulder to advocate for real platform accountability.

In addition to perusing last year's full 50-page SMSI report, which remains a substantial resource on this topic, please also explore this year's Appendix of Articles and Reports for invaluable deeper reading. A few recommended articles from the last few months include: "It's Time to Open the Black Box of Social Media," by Renée DiResta, Laura Edelson, Brendan Nyhan, Ethan Zuckerman (Time Magazine, April 28, 2022); "Reactions to the Banning Surveillance Advertising Act" by Justin Hendrix (Tech Policy Press, January 22, 2022); "OK, So Facebook Is Bad. Now What?" by Nick Robins-Early (Vice, November 15, 2021); and "Opinion: Elon Musk's Twitter takeover exposes the real threat to free speech: Big Tech monopolies" by Evan Greer (CNN Business, May 10, 2022).

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. We have seen over and over again that even when companies include protections in their policies, they frequently fail to enforce them. (Spotify's refusal to enforce their own Dangerous Content policy with regard to the extreme anti-LGBTQ hate expressed on their show, The Joe Rogan Experience, is a prime example — as outlined in the December 2021 Media Matters article, "Joe Rogan Wrapped: A year of COVID-19 misinformation, right-wing myths, and anti-trans rhetoric").

1. 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, 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."

GLAAD SOCIAL MEDIA SAFETY INDEX ADVISORY COMMITTEE

ALOK
(they/them)
Author, performer, &
media personality

Leigh Honeywell
(she/her)
CEO and Co-Founder,
Tall Poppy

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

Maria Ressa
(she/her)
Journalist & CEO,
Rappler

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

Tom Rielly
(he/him)
Founder,
TED Fellows program,
Digital Queers
& PlanetOut.com

Jelani Drew-Davi
(they/them)
Director of Campaigns,
Kairos

Brennan Suen
(he/him)
Deputy Director of
External Affairs, Media
Matters for America

Liz Fong-Jones
(she/her)
Principal Developer
Advocate for SRE
& Observability,
Honeycomb

Kara Swisher
(she/her)
contributing writer
& host of the
Sway podcast at
The New York Times

Evan Greer
(she/they)
Director,
Fight for the Future

"Close your eyes, for just a moment. And imagine the world as it should be. A world of peace, trust and empathy. Bringing out the best that we can be. Open your eyes. Now go. We have to make it happen. Please, let's hold the line. Together."

– MARIA RESSA,
2021 Nobel Prize Laureate
& SMSI Advisory
Committee Member

INTRODUCTION & METHODOLOGY

Since the May 2021 release of GLAAD's inaugural Social Media Safety Index (SMSI) report, the tremendously harmful and dangerous aspects of the major social media platforms have been exposed ever more prominently. Congressional hearings, alarming research findings, and massive media coverage have laid bare the urgent need for independent regulatory oversight of these companies — with virtually universal agreement about the need for industry-wide transparency and accountability.

Current content moderation and hate speech policies continue to be inadequate, failing to protect LGBTQ users and the LGBTQ community as a whole, while at the same time companies knowingly neglect to enforce the policies that do exist. Over the course of this past year, GLAAD has monitored and documented some of the harmful consequences of this failure, while also communicating these issues directly to each platform via meetings, calls, and emails. Our ongoing challenge to those in leadership at Facebook, Instagram, Twitter, YouTube, and TikTok

is to take urgent action to make their products safe: they must monitor violations and enforce their existing policies. Currently, they are failing to meet this challenge.

There are dozens of stories of such failures in the **APPENDIX OF ARTICLES & REPORTS**, such as this overview of anti-trans videos on YouTube from the researchers at Media Matters for America and this scathing May 2022 Media Matters report on Facebook and Instagram, which notes that: "Meta's hate speech policy insufficiently protects transgender and nonbinary users, as well as users who speak languages other than English. Media Matters identified nearly 1,000 violations of Meta's hate speech policy, including Instagram accounts promoting white supremacy and Facebook posts pushing anti-LGBTQ smears."

One of our major achievements from this past year is worth highlighting here in hopes that there will be universal adoption: Following the recommendation of the 2021 SMSI, TikTok stepped forward in March 2022 to protect

trans and non-binary people by adding to their community guidelines an explicit prohibition against targeted misgendering and deadnaming. (Misgendering is the practice of intentionally referring to a transgender person with the wrong gender. Deadnaming is referring to a trans person by a former name, usually one assigned to them prior to transitioning, without their consent.) Note that Twitter was the first major platform to take leadership on this, implementing this protection in 2018 (Pinterest also prohibits: "denial of an individual's gender identity"). GLAAD continues to urge all platforms to follow this as an industry best practice, a recommendation that remains an especially high priority in our current landscape, where anti-trans rhetoric and attacks are so prevalent, vicious, and harmful.

In preparing this year's report, GLAAD reviewed thought leadership, research, articles, reports 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 year's extremely hefty SMSI **APPENDIX OF ARTICLES & REPORTS,** 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 new **Platform Scorecard**. Developed in partnership with Ranking Digital Rights (RDR) and Goodwin Simon Strategic Research (GSSR), the 2022 Social Media Safety Index Platform Scorecard utilizes twelve LGBTQ-specific indicators to generate numeric ratings with regard to LGBTQ safety, privacy, and expression. The Scorecard evaluates five major social media platforms: Twitter, Facebook and Instagram (whose parent company is Meta), YouTube (parent company, Alphabet/Google), and TikTok (parent company, ByteDance).

ABOUT GLAAD'S WORK WITH THE MAJOR PLATFORMS

The most significant part of GLAAD's Social Media Safety program advocacy work involves providing policy recommendations to the platforms with regard to LGBTQ safety, privacy, and expression, as well as doing real time outreach and advocacy to hold platforms accountable around monitoring and enforcement of these policies.

This involves direct behind-the-scenes communications and advocacy with teams and departments at the platforms, and ongoing rapid response efforts. We are grateful to the platforms and to the individual colleagues at these companies who work to evaluate and implement recommendations from GLAAD and other organizations. Each of the platforms can take pride in the extensive ongoing work they do in support of the LGBTQ community in a variety of ways including the creation and promotion of LGBTQ content. Our chief emphasis in this report is on laying out the state of LGBTQ safety with regard to the products themselves and to advocate for stronger policies and enforcement of those policies. This Index does not showcase content production and the many positive initiatives of the platforms. This is not for lack of appreciation of those efforts, and GLAAD also regularly shares and promotes such important actions.

GLAAD's guidance and input to the platforms is provided via a variety of formal channels. GLAAD is an organizational member of Facebook's Network of Support, a coalition of LGBTQ organizations that advise on select policy and product updates. GLAAD is also an organizational member of the Twitter Trust & Safety Council and has been advising the platform on these matters since 2016. Lastly, GLAAD is also part of the TikTok Community Partner program reporting system, a direct line to flag hate and harassment.

TRANSPARENCY & ACCOUNTABILITY

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[...]. **The era of self-regulation for social media companies is over.** Policy makers are determined to establish a regulatory regime for these vital platforms for self-expression and commerce. Transparency, disclosures and openness are essential elements in this new regulatory structure.[2]
– Mark MacCarthy, Nonresident Senior Fellow, Governance Studies, Center for Technology Innovation at the Brookings Institution

In the current absence of regulatory oversight requiring transparency from social media companies, there are innumerable important things we simply do not know. How do the platforms even decide what content they will flag as anti-LGBTQ material? Who makes those decisions and what are the internal protocols? What about shadowbanning or demoting LGBTQ creators or content? How are decision makers trained to deal with LGBTQ material? As noted by platform accountability experts, current voluntary transparency reports offer limited visibility and are in many ways characterized by a relative *lack* of transparency. In a powerful April 2022 *Scientific American* article, "It's Time to Open the Black Box of Social Media," a team of leading scholars in the field observes: "After two decades of little regulation, it is time to require more transparency from social media companies[...]. The only way to understand what is happening on the platforms is for lawmakers and regulators to require social media companies to release data to independent researchers."

2. "Transparency is the best first step towards better digital governance," Brookings, May 10, 2022

ABOUT HATE
(AND LOVE)

Right wing media outlets, anti-LGBTQ -driven pundits and influencers, and opportunist political figures continue to target LGBTQ people and incite hatred against us, while also attacking the rights and dignity of countless other historically marginalized groups. (The GLAAD Accountability Project is one of our ongoing projects to track such figures and accounts, from Spotify's Joe Rogan to Texas Attorney General Ken Paxton.) 2022 has seen an unprecedented surge of hateful, violent, and false rhetoric hurled at LGBTQ people, especially targeting transgender youth. From maliciously characterizing LGBTQ people as "groomers" or pedophiles (which this April 2022 Vox piece, "The right's moral panic over 'grooming' invokes age-old homophobia clearly illuminates as a politically-motivated lie) to deceptive disinformation about gender affirming care for trans youth (this May 2022 *Scientific American* article is one of many helpful explanations of the facts: "Laws that ban gender-affirming treatment ignore the wealth of research demonstrating its benefits for trans people's health"), this kind of toxic and dangerous content is widely circulated on social media platforms. Amidst endless bad faith arguments about "free speech," the real world harms of these coordinated attacks are evident all around us. On the larger level, this coordinated right-wing fomenting of fear and hate against LGBTQ people is leveraged as a strategic component in a broader effort to undermine democracy.[3]

It is important to remember that the platforms themselves also bear responsibility. These companies have an inherent financial conflict of interest, which provides at least a partial explanation for their refusal to categorize certain content as harmful or to remove it from their platforms once it has been identified. There are countless examples to point to with regard to

"Prioritize your joy over their fear. Reclaim your body from their shame. Love so that you can truly live."

— ALOK

enforcement failures, one of the most illuminating examples of the past few years is the Steven Crowder-Carlos Maza YouTube case from 2019. As a 2021 Brennan Center report, "Double Standards in Content Moderation" points out: "YouTube had rules and enforcement mechanisms that it could use against Crowder, but it chose to exercise its discretion to allow him to continue violating the rules." The hateful content that targets LGBTQ people — as it drives clicks, comments, shares, and follows — also drives revenue and profits for these companies. To quote one of our advisory committee members, Kara Swisher, "enragement equals engagement."

WE WILL NOT LET HATE WIN!

In our exploration of the social media landscape for LGBTQ people, the prevalence and intensity of hate speech and harassment continues to stand out as an especially urgent problem.

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

ONLINE HARASSMENT vs SEVERE HARASSMENT EVER EXPERENCED: TARGETED GROUPS

Which, if any, of the following have happened to you, personally, ONLINE?

Group		
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

other identity group (the rate for non-LGBTQ+ respondents was 38%). This 66% figure is up from 64% in the previous 2021 survey. Further, 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. An additional notable figure, showing a rise in harassment of other individuals from historically marginalized groups, is that 38% of all respondents reported harassment due to their identity, up from 33% in 2021.[4] 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).

3. "Ron DeSantis is following a trail blazed by a Hungarian authoritarian," Vox, Apr 28, 2022

4. "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.")

The rates for younger LGBTQ social media users are similarly alarming. According to the 2021 "State of Play" report on identity-based Gen Z cyberbullying in the U.S.: 77% of respondents have seen someone bullied online for their gender identity; and 78% of respondents have seen someone bullied, abused or harassed online based on their sexual identity.

These figures vividly illuminate the severity and breadth of the problems outlined in this report, and are all the more alarming in light of recently released statistics from The Trevor Project reporting that 45% of LGBTQ youth had seriously contemplated suicide in the past year, with one in ten cisgender youth and nearly one in five trans/ non-binary youth attempting suicide in 2021.

While Facebook, Instagram, Twitter, YouTube, TikTok and other platforms must balance concerns around free expression, it cannot be stated strongly enough that social media platforms must take more meaningful and aggressive action to protect the safety of their LGBTQ users and to staunch the epidemic of hate, falsehoods, and extremism. These efforts must also address the spread of anti-LGBTQ disinformation (the intentional spreading of false or misleading information).

In our social media safety advocacy work we talk a lot about fighting anti-LGBTQ disinformation. To be clear though, a better word than disinformation is actually: Lies. And malicious conservatives telling lies about LGBTQ people is nothing new. It goes far back in our history, most notably to Anita Bryant, the right-wing Christian (and brand ambassador for the Florida Citrus Commission) who spouted many of the same politically-motivated homophobic lies in 1977 as Florida Governor Ron DeSantis is expressing 45 years later. As Michigan State Senator Mallory McMorrow put it earlier this year when she forcefully decried this malicious and cynical

practice of hurling accusations of pedophilia at LGBTQ people and our allies: "This is a hollow, hateful scheme... to target and marginalize already targeted and marginalized people."

As Senator McMorrow also proclaimed: "We will not let hate win!"

GLAAD calls on all of the platforms to find new and better ways to reduce the spread of this kind of deceitful anti-LGBTQ content (possible mitigations could include fact checks that label such false information in the same way that misinfo about COVID or elections is treated).

It is also extremely important to be mindful of how anti-LGBTQ hate, harassment, and dehumanizing language on platforms (slurs and lies, deadnaming and misgendering, etc.) essentially work to restrict the free expression of LGBTQ people. In her 2018 *New York Times* opinion piece reflecting on Twitter's newest policy addition at the time, "How Twitter's Ban on 'Deadnaming' Promotes Free Speech" writer Parker Molloy explains: "At *The Guardian*, Kenan Malik argued that banning misgendering will shut down debate on trans issues and strike a blow to free speech. But in fact, the content free-for-all chills speech by allowing the dominant to control the parameters of debate, never letting discussion proceed past the pedantic obsession with names and pronouns." In the current landscape of bad-faith vitriol hurled at us by troll-cum-pundits who we're not even going to name here, LGBTQ people (and especially trans people) know all too well this reality that Molloy describes. Of course this is true for other historically marginalized groups as well.

STRATEGIES TO MITIGATE HATE

There are many ways for platforms to curb and mitigate anti-LGBTQ conduct and content. Precedents for this abound, including adding context, warnings, or links (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; de-monetizing posts or suspending accounts (with a robust due process appeals process); and even banning/de-platforming (where accounts are taken down and individuals are not allowed to create new accounts or pages on a given platform).

There are also numerous strategies — like speed-bumps or circuit breakers that throttle viral content — that have been used effectively to slow the spread of misinformation, including anti-LGBTQ hateful content. These particular strategies have been developed mainly in relation to election integrity and public health issues, most recently of course around COVID and vaccine misinfo. It is worth noting that if policy makers were to reframe anti-LGBTQ dis- or misinformation and hate speech as a public safety/public health issue, then social media platforms might be forced to effectively address its harmful impacts. In fact, a considerable amount of anti-trans content falls in the category of health misinfo, especially malicious false postings about the safety of gender-affirming care for trans youth (for the record, the medical guidelines of gender-affirming care for trans youth are supported by every major medical association in the U.S.). The promotion of falsehoods about so-called "conversion therapy" also falls under the category of health disinformation, as such practices are denounced as harmful to LGBTQ mental health by all major medical and psychiatric associations (as is noted in the recent Open Democracy article, "Google, Facebook and Amazon turn blind eye to anti-gay disinformation." See also the separate section below on this topic).

AN IMPORTANT NOTE ABOUT DISINFORMATION & HATE

What social media companies define as "hate" is insufficient. An enormous amount of the anti-LGBTQ disinformation circulating amidst our current culture wars is indeed, quite simply, hate. Much of this material should rightly be evaluated against current existing policies for hate and harassment and dehumanizing speech, with corresponding enforcement. In the same way that these companies prohibit lies about COVID, the posting of inaccurate voting information, denial of the Holocaust — **the intentional posting of patently false disinformation intended as a targeted attack on LGBTQ people (and other historically marginalized groups) must be mitigated.**

Without meaningfully addressing this problem one can only conclude that although these companies say they care about us in word, this isn't manifest in practice. In the May 2021 Decode Democracy exposé of Facebook, "As Bills Sweep Statehouses, Facebook Profits From Harmful Anti-Trans Ads," activist and media strategist Gillian Branstetter concisely observes, "These narratives are harmful first and foremost because they are false." The article outlines examples of dozens of anti-trans ads on the platform which received hundreds of thousands of impressions (and generated substantial amounts of revenue for the company). Despite being alerted to all of these misleading hate-driven ads, Facebook took no action.

THE ISSUE OF CENSORSHIP

On the one hand, LGBTQ individuals are vulnerable to hate speech and other manifestations of online homophobia, biphobia, and transphobia — acts which have very real offline impacts and harms. On the other hand, we are also vulnerable to censorship and disproportionate limitations of free expression related to our identities.

The most succinct explanation of this vulnerability is that because our sexual orientation is a defining aspect of LGBTQ identity there are greater opportunities for these characteristics to be flagged. It is also the case that actual homophobia, biphobia, and transphobia can come into play (intentionally or not, the outcome is the same) on the part of AI and human content moderators and result in disproportionate suppression of LGBTQ expression (this could include things like the deplatforming or demonetization of sexual health materials, or reduced access to LGBTQ content for youth). Of course it is also the case that LGBTQ users are often targeted by bad actors, with false reports or brigading resulting in our accounts being suspended or legitimate, non-violative content restricted or taken down. Requirements around "real names" is another realm in which trans and non-binary users are often negatively impacted since their "real names" at times do not match what is on legal documents and IDs.

In addition to being yet another thread in a social fabric of marginalization, bias, and oppression, these examples of bias (whether in human content moderators or in AI systems) create 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 cultural rights.

ABOUT LOVE

GLAAD's approach to platform accountability advocacy work has two aspects: On the one hand, our community is experiencing unprecedented attempts to dehumanize and degrade us. We must forcefully denounce and fight against such hate, violence, and lies. At the same time, we passionately celebrate our LGBTQ lives, our dignity, and our human and civil rights. And we point to the much more powerful overarching truth of our progress and victories and love.

One notable story that powerfully counteracts the current tidal wave of anti-LGBTQ hate is the fact that according to a remarkable March 2022 survey from the Public Religion Research Institute (PRRI): **"the majority of Americans (across every subgroup of Americans, including race, age, religion, partisanship, and geography) strongly support LGBTQ protections against discrimination** in jobs, public accommodations, and housing. The PRRI survey polled more than 10,000 Americans and determined that even the groups least likely to support nondiscrimination protections — Republicans (62%) and white evangelical Protestants (62%) — show majority support... [only] a slim minority of Americans (7%) hold consistently unfavorable views toward LBGTQ policies."

While it is a loud, highly coordinated, and vicious group of people who seek to scapegoat LGBTQ people for political and financial gain, they are indeed a very slim minority. We must continue to illuminate the nature of these attacks — as we find common cause with the majority, in opposition to these blatant manifestations of hate and injustice.

And as we stand together to fight against hate, we stand also united in love — as a community: LGBTQ together.

2022 SOCIAL MEDIA SAFETY INDEX PLATFORM SCORECARD

The 2022 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 on their policies and practices affecting people's rights to freedom of expression and privacy. 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 the SMSI advisory committee. Additional methodological considerations were identified during the subsequent policy analysis and company research. The Scorecard evaluates five major social media platforms: Twitter, Facebook and Instagram (whose parent company is Meta), YouTube (parent company, Alphabet/Google), and TikTok (parent company, ByteDance).

In creating the SMSI Scorecard, GSSR utilized the RDR evaluation and scoring process. Full details can be found here, briefly though: "Companies receive an average score of their performance across all RDR Index indicators. 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." Guidance for future researchers interested in applying these indicators can be found in the **RESEARCH GUIDANCE**.

Note that these twelve indicators only address some of the issues impacting LGBTQ users. And the recommendations below are only some of the important steps that companies should take. Much greater detail and analysis can be found in the **RESEARCH GUIDANCE;** the full scoring sheets are also available here. Given the ever-evolving threats to LGBTQ and other historically marginalized users on social media platforms, GLAAD will continue to expand the SMSI methodology to include new indicators addressing additional issues — for instance diving deeper into the topic of anti-LGBTQ related disinformation (e.g., false assertions related to the best practice of gender affirming care for trans youth, the safety and effectiveness of PrEP for HIV prevention, the promotion of so-called "conversion therapy," etc.).

SOCIAL MEDIA SAFETY INDEX PLATFORM SCORECARD

		Instagram	Facebook	Twitter	YouTube	TikTok
	TOTAL	**48.38**	**46.3**	**44.7**	**45.11**	**42.51**
1	Policy commitment to protect LGBTQ users	100	100	100	100	100
2	Gender pronouns on user profiles	50	25	0	0	50
3	Targeted deadnaming and misgendering prohibition	0	0	91.67	0	100
4	Sexual orientation and gender identity data control	14.29	14.29	0	35.71	0
5	Recommendations based on sexual orientation or gender identity	37.5	37.5	0	37.5	37.5
6	Third party advertisers	64.29	64.29	100	83.33	0
7	Harmful advertising prohibition	100	100	100	100	75
8	Actions to restrict harmful content	35.71	35.71	21.43	50	57.14
9	Stop demonetizing or removing legitimate LGBTQ content	7.14	7.14	0	21.43	7.14
10	Content moderators training	25	25	0	0	50
11	Commitments to protect LGBTQ users from harm	66.67	66.67	33.33	33.33	33.33
12	Diverse workforce	80	80	90	80	0

AN IMPORTANT NOTE ABOUT THE SCORECARD RATINGS

While the five platforms all have general policies prohibiting hate and harassment on the basis of sexual orientation and gender identity, the Scorecard below does not include an indicator to rate them on enforcement of those policies. Mainly this is due to the fact that, in the absence of far greater transparency and data from the platforms, this is not a metric that can be measured by outside researchers.

2022 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 a **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.

FULL DETAILED SCORING SPREADSHEETS:
TWITTER, **YOUTUBE**, **TIKTOK**, **META** (FACEBOOK & INSTAGRAM)

INSTAGRAM

GLAAD SOCIAL MEDIA SAFETY INDEX SCORE:

48%
OUT OF 100%

In the 2022 SMSI Platform Scorecard, Instagram received a score of 48. In its Community Guidelines, Instagram has a comprehensive protected groups policy that protects users from threats, violence, hate speech, and harassment based on protected characteristics such as sexual orientation and gender identity. The company also discloses limited information on how users can opt out of seeing content based on their disclosed or inferred sexual orientation or gender identity. Meta was the only company to disclose limited information on how its Civil Rights team engages internally to advise policy and product teams on how company policies, products, and services may impact the online rights of marginalized users. The company does not disclose whether it has an LGBTQ policy lead.

However, the company falls short of providing adequate transparency in several key areas. Notably, Instagram has no policy protecting users from targeted deadnaming and misgendering. While the company has a feature allowing users to add 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. In its targeted advertising policies, Meta provides that it prohibits targeted advertising based on sensitive topics, including topics related to sexual orientation. However, no similar disclosure was found that indicates the company does not permit detailed targeting based on users' gender identity.

KEY RECOMMENDATIONS

- **Make a policy commitment to protecting 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.

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

- **Do not allow third party advertisers to target users based on their gender identity:** The company should make a commitment to not allowing third party advertisers to target users based on their gender identity.

GLAAD
SOCIAL
MEDIA
SAFETY
INDEX
SCORE:

46%
OUT OF 100%

In the 2022 SMSI Platform Scorecard, Facebook received a score of 46. In its Community Guidelines, Facebook discloses a comprehensive policy that protects users from threats, violence, hate speech, and harassment based on protected characteristics such as sexual orientation and gender identity. The company also discloses limited information on how users can control the content they see based on their disclosed or inferred sexual orientation or gender identity. Meta was one of only two companies to disclose information on its trainings for content moderators. The company's policies also contain a clear prohibition of advertising content that could be harmful and/or discriminatory to LGBTQ individuals.

However, the company's policies do not adequately protect LGBTQ expression and privacy in several key areas. Facebook currently has no policy protecting users from targeted deadnaming and misgendering. According to the company's policies, Facebook's feature to add gender pronouns to user profiles is only available for users who select a custom gender and select one or more genders. The company also 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. In its targeted advertising policies, Meta provides that it prohibits targeted advertising based on sensitive topics, including topics related to sexual orientation. However, no similar disclosure was found that indicates the company does not permit detailed targeting based on users' gender identity. The company only publishes limited data about the actions it has taken to restrict content and accounts that violate policies protecting LGBTQ individuals (independent researchers such as Media Matters have also pointed to problems related to enforcement of these policies).

KEY RECOMMENDATIONS

- **Make a 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.

- **Do not target users with advertising based on their gender identity:** The company should disclose that it does not allow third party advertisers to target users based on their gender identity.

- **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.

GLAAD
SOCIAL
MEDIA
SAFETY
INDEX
SCORE:

45%

OUT OF 100%

In the 2022 SMSI Platform Scorecard, Twitter earned a score of 45. Twitter was one of only two companies evaluated with an existing policy against targeted deadnaming and misgendering. Like the other companies evaluated in this year's index, Twitter also has a policy protecting LGBTQ users from attacks or threats based on their sexual orientation and gender identity. The company also discloses that it prohibits targeted advertising based on sensitive categories, including sexual orientation and gender idenity. Twitter also prohibits advertising content that could be harmful and/or discriminatory to LGBTQ individuals including "content that promotes claims or services attempting to change a person's sexual orientation, gender identity or gender expression. "The company also makes a public commitment to diversifying its workforce, and publishes data on its LGBTQ workforce that is based on voluntarily self-disclosed data.

However, the company falls short of providing adequate transparency in other key areas. The company 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 of information related to their sexual orientation and gender identity.

Twitter also does not disclose any options for users to control the company's collection and attempts to infer user information related to their sexual orientation and gender identity. Further, recommendation of content based on users' disclosed or inferred sexual orientation or gender identity is not off by default. The company also does not disclose any information on whether it requires content moderators to participate in trainings that educate them on the needs of LGBTQ people and other vulnerable users.

While the company discloses that it engages with organizations and groups that advocate on behalf of the LGBTQ community, it does not disclose that it has an LGBTQ policy lead or that it has a formal training in place to educate all employees on the needs of LGBTQ users.

KEY RECOMMENDATIONS

- **Provide additional tools for user expression:** The company should disclose a voluntary dedicated feature that allows users to add their preferred gender pronouns to their user profiles. In order to ensure user privacy, Twitter should also disclose an option for users to customize who can see their gender pronouns.

- **Give users more control over their own data:** Twitter should give users greater 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.

- **Implement commitment to LGBTQ expression and privacy across the company:** In order to ensure that the needs of LGBTQ users are implemented across the company's products and services, Twitter should have an LGBTQ policy lead and provide evidence of formal trainings that educate all employees to understand the needs of LGBTQ users.

▶ YOUTUBE

GLAAD SOCIAL MEDIA SAFETY INDEX SCORE:

45%
OUT OF 100%

In the 2022 SMSI Platform Scorecard, YouTube earned a score of 45. YouTube discloses a policy that protects users from hate speech, harassment, and cyberbullying based on protected attributes, including sexual orientation and gender identity. As is true of other platforms, independent researchers report that enforcement of these policies is often lacking. The company provides limited transparency on options that users have to control the company's processing of information related to their sexual orientation and gender identity. Alphabet also discloses some information on how users can control the recommended content they see on YouTube. The company also discloses that it prohibits targeted advertising based on sensitive categories, including sexual orientation and gender idenity. The company also prohibits advertising content that could be harmful and/or discriminatory to LGBTQ individuals.

However, the company falls short of providing adequate transparency in several key areas. For example, it provides only limited information on user control over the company's collection and inference of user information related to sexual orientation and gender identity. Despite advocates and LGBTQ creators raising concern over the removal and demonetization of LGBTQ-related content from ad services on YouTube, Alphabet only provides limited transparency on the state of demonetization and removal of LGBTQ creators and their content. The company does not disclose a comprehensive plan outlining concrete steps to address demonetization, filtering, and removal of LGBTQ creators. The company's transparency reports provide no data giving insights into removal and demonetization of LGBTQ creators from ad services. The company also does not disclose whether it has an LGBTQ policy lead or internal trainings in place that educate its employees to understand the needs of LGBTQ users.

KEY RECOMMENDATIONS

- **Make a policy commitment to protect transgender, non-binary, and non-conforming users from targeted deadnaming and misgendering:** The company should adopt a policy that protects users from targeted deadnaming and misgendering and removes such dehumanizing anti-trans content.

- **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. The company should also disclose comprehensive data on the wrongful removal of LGBTQ creators and their content and accounts.

- **Implement commitment to LGBTQ expression and privacy across the company:** In order to ensure that the needs of LGBTQ users are attended to across the company's products and services, Alphabet should have an LGBTQ policy lead who advises policy and product teams on the needs of LGBTQ and other vulnerable users.

GLAAD SOCIAL MEDIA SAFETY INDEX SCORE: 43% OUT OF 100%

In the 2022 SMSI Platform Scorecard, TikTok earned a score of 43. TikTok was one of only two companies evaluated with an existing policy against targeted deadnaming and misgendering, and the only company to provide comprehensive information on how it detects violations to this policy. The company also discloses a comprehensive policy protecting LGBTQ users from other attacks, threats, and violence on the platform. The company discloses a feature allowing users to add their gender pronouns to their profiles. TikTok was one of only two companies disclosing any information on trainings that educate content moderators on the needs of vulnerable users.

However, the company falls short of providing adequate transparency on other key issues. The company currently does not disclose options for users to control the company's collection of information related to their sexual orientation and gender identity. The company also provides only limited options for users to control the recommend content they see based on their disclosed or inferred sexual orientation or gender identity. Rather than an outright ban of targeted advertising based on user's sexual orientation and gender identity, limitations on ad targeting based on users' sexual orientation depend on local laws.

While the company discloses that it engages with organizations and groups that advocate on behalf of the LGBTQ community, it does not disclose that it has an LGBTQ policy lead or that it has a formal training in place that educates all employees to understand the needs of LGBTQ users. Notably, TikTok was the only company that did not disclose any information on steps it takes to diversify its workforce.

KEY RECOMMENDATIONS

- **Give users control over their own data:** TikTok should give users 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.

- **Ban targeted advertising based on sexual orientation and gender identity:** The company should ban third party advertising based on users' sexual orientation and gender identity.

- **Make a commitment to diversify TikTok's workforce:** The company should disclose a public commitment to taking proactive steps to diversify its workforce, and publish data that shows it follows up on this commitment.

MORE SUNLIGHT
The State of Transparency & Transparency Reporting

In the 2021 SMSI, GLAAD called upon 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.

According to an August 2021 Brennan Center report addressing some of these issues, Double Standards in Social Media Content Moderation, "All too often, the viewpoints of communities of color, women, LGBTQ+ communities, and religious minorities are at risk of over-enforcement, while harms targeting them often remain unaddressed."

The gravity of this problem is also emphasized in the Santa Clara Principles, the widely respected set of guidelines for transparency and accountability in content moderation:

> 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.

Rather than trying to translate each of the platform transparency reports individually, as we did in last year's SMSI, below are links to the respective documents from each company

(please also see Indicator 8 on the SMSI Platform Scorecards for that transparency-related data and notes). We once again strongly urge YouTube, Twitter, TikTok, Facebook, and Instagram to provide true transparency.

It's important to repeat here that transparency reports need to share data about all take-downs and user appeals in order to make visible the problem of the over-moderation, suppression, demonetization, or deplatforming of legitimate posts and accounts (for instance, there is extensive anecdotal research on the systemic deplatforming of sex workers and sex educators even when they are not violating terms of service — and specifically that there are double standards in this regard for non-LGBTQ vs. LGBTQ people). Again, as noted in this April 2022 piece ("Sex workers say legislation is needed to prevent censorship on online platforms") from the non-profit news outlet Prism, "The users most censored and deplatformed are people of color, LGBTQ+ folks, and people of other marginalized identities."

"The time has come for governments to act, and transparency measures are the best first step. [...] Sunlight, said renowned US Supreme Court Justice Louis Brandeis, is the best disinfectant."

– MARK MACCARTHY,
Nonresident Senior Fellow, Governance Studies, Center for Technology Innovation at the Brookings Institution

MOST RECENT TRANSPARENCY REPORTS FROM EACH PLATFORM:

f & **O** (Meta) (Q1 2022)

▶ (Q4 2021)

🐦 (Q1 & Q2 2021)

♪ (Q4 2021)

2021 ANNUAL EARNINGS BY PLATFORM:

f & **O** $117.9 BILLION

▶ $28.8 BILLION

🐦 $5.08 BILLION

♪ $4.6 BILLION

DANGEROUS & DISCREDITED:
The State of "Conversion Therapy" Policies on Social Media Platforms

Conversion therapy has been condemned by dozens of medical and psychological professional organizations, banned in numerous countries, states, and cities, and has been called 'torture' and an 'egregious violation of rights' by the UN [...]. Tech companies have failed to deplatform anti-LGBTQ+ conversion therapy disinformation and ban providers pushing the discredited practice, even though many claim to do so.

– GLOBAL PROJECT AGAINST HATE & EXTREMISM

The widely debunked practice of so-called "conversion therapy" falsely claims to change a person's sexual orientation, gender identity, or gender expression, and has been condemned by all major medical, psychiatric, and psychological organizations, including the American Medical Association and American Psychological Association. Globally, there has been a growing movement to ban conversion therapy at the national level, including bans in Greece, Canada, France, Germany, Malta, Ecuador, Brazil, Taiwan, and New Zealand. As of publication, in the United States, 20 states and the District of Columbia have bans in place.

There is continuing global consensus and momentum towards protecting LGBTQ people, and especially LGBTQ youth, from this dangerous practice. However, the spread of "conversion therapy" disinformation and extremist scare-tactic narratives (as well as profit-driven offerings of such services in content and advertising) continues to spread on social media platforms.

In February 2022, GLAAD worked with **TikTok** to add an explicit prohibition of so-called "conversion therapy" content (the company had previously grappled with the problem, as outlined in this Media Matters overview). TikTok updated their Community Guidelines to include the following language: "Adding clarity on the types of hateful ideologies prohibited on our platform. This includes [...] content that supports or promotes conversion therapy programs. Though

these ideologies have long been prohibited on TikTok, we've heard from creators and civil society organizations that it's important to be explicit in our Community Guidelines."

In 2022 GLAAD also urged both **YouTube** and **Twitter** to add an express prohibition of "conversion therapy" to their content and ad guidelines.

While **Twitter** does not currently have an express prohibition in their main hateful conduct policy, the company recently added new language to their advertising policy, to explicitly prohibit: "Content that promotes claims or services attempting to change a person's sexual orientation, gender identity or gender expression."[5]

YouTube does mitigate conversion therapy content (when it is identified by their AI systems) by showing an information pane from the Trevor Project (with the text: "Conversion therapy, sometimes referred to as 'reparative therapy,' is any of several dangeous and discredited practices aimed at changing and individuals sexual orientation or gender identity.") However, unlike TikTok and Meta, YouTube does not include an explicit prohibition of such content in their Hate Speech Policy.

Meta's **Facebook** and **Instagram** platforms currently *do* have such a prohibition, however it is listed separately from their standard three tiers of content moderation consideration as requiring, "additional information and/or context

to enforce." GLAAD has requested that it be elevated to a higher priority tier.

It is also worth noting that **Pinterest's Community Guidelines** include the following prohibition: "support for conversion therapy and related programs."

In addition to GLAAD's efforts urging platforms to add prohibitions against so-called "conversion therapy" to their community guidelines, we also urge these companies to effectively enforce these policies.

To clarify even further, we recommend that all platforms add public facing language that expressly prohibits the promotion of so-called "conversion therapy" to both their community guidelines and ad services policies. And while some platforms have described off-the-record that conversion therapy material is prohibited under the umbrella of other policies — policies prohibiting hateful ideologies, for instance — the prohibition of "conversion therapy" promotion should be explicitly stated publicly in their community guidelines.

We urge *all* social media platforms, apps, online media and technology companies to implement these same policies.

GLAAD urges all platforms to review and follow the below recommendations of two powerful January 2022 reports from the Global Project on Hate & Extremism (GPAHE).

5. Business Twitter. Inappropriate Content Policy

GPAHE RECOMMENDATIONS

To protect their users, tech companies must:

- Deplatform providers and other anti-LGBTQ+ material in accordance with community standards and hate speech policies.

- Surface more authoritative material generally on searches related to conversion therapy in all languages and countries, and de-rank disinformation.

- Incorporate the terms "same-sex attraction" and "reintegrative therapy" into the algorithms to increase authoritative results.

- Partner with experts so that the algorithms are always incorporating current terminology.

- Improve content moderation in non-English languages and devote more resources to language and cultural competencies worldwide.

- Retool the recommended content algorithms to recommend reliable information.

- Disallow payment processing platforms.

- Ensure that no conversion therapy content is monetized with ads.

Source:
New GPAHE Reports Reveal Harmful Conversion Therapy Disinformation Thriving Online (Global Project Against Hate & Extremism, Jan 2022)

TRANS & NON-BINARY PEOPLE ARE WHO THEY SAY THEY ARE:

The State of Targeted Misgendering & Deadnaming Policies on Social Media Platforms

In February 2022, following recommendations from GLAAD and gender justice advocacy group UltraViolet, **TikTok** updated their Community Guidelines to include the following language: "Adding clarity on the types of hateful ideologies prohibited on our platform. This includes deadnaming, misgendering, or misogyny [...]. Though these ideologies have long been prohibited on TikTok, we've heard from creators and civil society organizations that it's important to be explicit in our Community Guidelines. On top of this, we hope our recent feature enabling people to add their pronouns will encourage respectful and inclusive dialogue on our platform."

Twitter is the only other major platform to explicitly prohibit targeted misgendering and deadnaming in their hateful conduct policy. The 2021 GLAAD Social Media Safety Report urged all the major platforms to follow the lead of Twitter (and Pinterest) and to add an express prohibition of targeted misgendering and deadnaming into their hate and harassment policies.

In 2021, GLAAD joined Media Matters for America and 18 other organizations in calling for **YouTube** to create a policy that explicitly bans creators from intentionally misgendering and deadnaming trans people as part of YouTube's existing hate speech and harassment policies.

GLAAD has also repeatedly called on both **Facebook** and **Instagram** to initiate a policy prohibition against targeted misgendering and deadnaming.

Deadnaming and misgendering have harmful real-world effects for trans and non-binary people. Misgendering is the practice of referring to a transgender or non-binary person with the wrong gender. Revealing a transgender person's former name without their consent — often referred to as "deadnaming"— is an invasion of privacy that undermines the trans person's identity, and can put them at risk for discrimination, even violence. Studies have found that many trans people who have been misgendered face increased levels of psychological stress and depression. The Trevor Project's 2021 National Survey on LGBTQ Youth Mental Health found that transgender and non-binary youth who reported having pronouns respected by all of the people they lived with attempted suicide at half the rate of those who did not have their pronouns respected by anyone with whom they lived.

Allowing anti-trans rhetoric and misinformation, including anti-trans practices like misgendering and deadnaming, to remain unchecked on social media platforms also has implications for the protection of the trans community. In 2021, there was an unprecedented number of bills introduced in state legislatures across the country targeting transgender people, focusing primarily on restricting the rights of trans youth across sports, healthcare, and education. According to Freedom for All Americans, 2022 is already a record-setting year for state legislation targeting LGBTQ adults and children, with nearly 250 anti-LGBTQ bills proposed. More than 100 of these bills target transgender people and trans youth. Violence against the trans community also continues to increase, with 2021 being the deadliest year on record for the trans community. At least 53 transgender or gender non-conforming people were killed, with the majority of victims being Black or Latinx trans women.

Trans and non-binary people are who they say they are. Quite simply, purposefully demeaning and misidentifying them is harassment and abuse. It is important to note that even when platforms do not have explicit prohibitions against misgendering and deadnaming, such behaviors already by definition violate their existing general hate and harassment policies.

That said, GLAAD urges all platforms to immediately incorporate into their hateful conduct policies an explicit prohibition against targeted misgendering and deadnaming of transgender and non-binary people. This recommendation remains an especially high priority in our current landscape where anti-trans rhetoric and attacks are so prevalent, vicious, and harmful. We also urge these companies to effectively moderate such content and to enforce these policies.

A GROWING CHORUS:

Recent Reports on LGBTQ Social Media Safety

Since the launch of the inaugural GLAAD Social Media Safety Index report in May 2021, this past year saw the release of numerous other important reports focused specifically on the myriad of problems around LGBTQ safety and social media platforms. Below are just a few, cited with brief excerpts underscoring and echoing the call for these companies to attend to the issue of LGBTQ platform safety.

"Today, those spouting hateful anti-transgender rhetoric online are creating an ecosystem where shared ideas, themes, and language echoes. This hateful rhetoric goes from fringe to mainstream — and boomerangs between online and offline speech — in part because of social media's immense power, amplification of 'engaging' content, and sophisticated recommendation algorithms. **The impact of anti-transgender online speech is becoming concrete as lawmakers introduce and enact legislation targeting transgender people, especially youth, using language mirroring that which we see online.**"

Tracking Anti-Transgender Rhetoric Online, Offline, and In Our Legislative Chambers
Anti-Defamation League (ADL), July 2021

"**Conservative media are using Facebook to help spread anti-trans narratives around health care and student athletes** [...] part of a years-long pattern in which the right consistently earns the bulk of Facebook engagement when discussing trans issues, often spreading misinformation and bigotry."

Right-leaning Facebook pages earned nearly two-thirds ofinteractions on posts about trans issues
Media Matters for America (MMFA), November 2021

"Even though we solely interacted with transphobic content, we found that our FYP [For You page] was increasingly populated with videos promoting various far-right views and talking points. **That content did include additional transphobic videos, even though such content violates TikTok's 'hateful behavior' community guidelines, which state that the platform does not permit 'content that attacks, threatens, incites violence against, or otherwise dehumanizes an individual or group on the basis of' attributes including gender and gender identity.**"

TikTok's algorithm leads users from transphobic videos to far-right rabbit holes
Media Matters for America (MMFA), October 2021

"Prominent conservative pundits on YouTube have regularly used the platform to harass and misidentify trans people, including targeted misgendering of kids, parents, and public figures [...]. **Though YouTube's existing guidelines allegedly protect trans people, the platform has failed to create a policy that explicitly bans creators from intentionally misidentifying trans people.** If YouTube is serious about its 'ongoing commitment to support the LGBTQ+ community,' it would specifically ban content that promotes this kind of harassment of trans people."

YouTube fails to protect trans people from misgendering or deadnaming
Media Matters for America (MMFA), August 2021

"Social media platforms play a crucial role in supporting freedom of expression in today's digital societies. Platforms can empower groups that have previously been silenced. However, platforms also host hateful and illegal content, often targeted at minorities, and content is prone to being unfairly censored by algorithmically biased moderation systems. This report analyzes **the current environment of content moderation, particularly bringing to light negative effects for the LGBTIQA+ community.**"

The state of content moderation for the LGBTIQA+ community and the role of the EU Digital Services Act
The Heinrich-Böll-Stiftung, June 2021

"While social media companies dress their content moderation policies in the language of human rights, their actions are largely driven by business priorities, the threat of government regulation, and outside pres-sure from the public and the mainstream media. As a result, the veneer of a rule-based system actually conceals a cascade of discretionary decisions. Where plat-forms are looking to drive growth or facil-itate a favorable regulatory environment, content moderation policy is often either an afterthought or a tool employed to curry favor. **All too often, the viewpoints of communities of color, women, LGBTQ+ communities, and religious minorities are at risk of over-enforcement, while harms targeting them often remain unaddressed.**"

Double Standards in Social Media Content Moderation
The Brennan Center, August 2021

"The young people in our study advocated for new approaches that can sit alongside censorship mechanisms. Specifically, they advocated for **educative approaches to counter hate speech, where people — perpetrators of hate speech and bystanders — are taught about sexual and gender diversity...** queer young people's participation in debates about content moderation and censorship is critical, as platforms design effective processes and mechanisms to counter and respond to hate speech."

Social Media Insights from Sexuality and Gender Diverse Young People During Covid-19
Western Sydney University, Young & Resilient Research Centre, May 2021

Please also see the
2022 SMSI Articles & Reports Appendix
for links to numerous other reports of interest.

THE TIME HAS COME:

Industry Oversight, Regulatory Solutions, and Public Safety

The primary intent of the GLAAD Social Media Safety Index is to present recommendations to companies urging them to voluntarily undertake measures to improve their platforms. Of course, there is also a vast array of proposed legislation related to regulatory oversight of social media and tech companies. Both in the U.S. and internationally, the acronyms seem to be endless: EARN IT, NUDGE, DSOSA, PATA, GDPR, DSA. In an April 2022 overview of the EU's Digital Services Act (DSA), "As Europe Approves New Tech Laws, the U.S. Falls Further Behind," New York Times reporter Cecilia Kang points out: "In the United States, Congress has not passed a single piece of comprehensive regulation to protect internet consumers and to rein in the power of its technology giants."

While there is a clear need for some kind of industry oversight or regulation, it is also vitally important that such solutions be carefully crafted to anticipate potential harms or unintended negative impacts for LGBTQ people and other marginalized communities (and everyone). As an example of some of the important watchdog work being done in this realm, in early 2022 GLAAD joined a coalition of 60+ human rights, civil rights, & open Internet organizations — including the ACLU, the Center for Democracy & Technology, Fight for the Future, PEN America, and others — in calling on Congress to oppose the deeply-flawed EARN IT Act. Two current antitrust bills (AICOA and OAMA) hold promise to have a significant impact on Big Tech in general and social media platforms in particular, with additional significant possibilities coming from the FTC via data privacy regulations.

Other governments around the world have taken vastly different, and generally much more rigorous approaches, to prioritizing the public safety of individual citizens over the business interests of corporations — most notably, the EU's recently approved Digital Services Act (DSA). This April 2022 TechCrunch article offers a solid overview of some of the highlights of the DSA. In a May 2022 article for the Brookings Institution, tech policy expert Mark MacCarthy notes: "After years of letting them manage their own systems and content moderation practices with little or no public supervision, governments around the world are throwing a regulatory net over digital companies. Online regulation measures have been adopted or are pending in Australia, Canada, the European Union, Germany, Ireland, the United Kingdom and the United States."

Delving into the weeds of the innumerable US state and federal legislative proposals — not to mention the array of regulatory situations across the globe, including the EU's DSA — is beyond the scope of this report. Thankfully there are many organizations and entities monitoring and evaluating these proposals as well as providing leadership in policy development. One trusted source for ongoing insights is the Center for Democracy & Technology. Another excellent source is Tech Policy Press. Please also consult the SMSI's **APPENDIX OF ARTICLES & REPORTS** for additional links.

STANDING UP FOR CHANGE:

Platform Accountability Campaigns, Tools, Organizations, and Initiatives

In the field of platform accountability advocacy there are countless current projects, tools, campaigns, petitions, coalitions, organizations, and initiatives — exciting and ambitious efforts striving to create accountability, transparency, and change. Below are just a few recent and current examples; GLAAD is part of many of them.

CHANGE THE TERMS

A coalition of 60 US civil rights, human rights, tech policy, and consumer protection organizations (including GLAAD), Change the Terms has developed core recommendations to help internet companies stop hate, extremism and disinformation online and ensure that they do more to protect people of color, women, LGBTQIA people, religious minorities and other marginalized communities. The coalition first issued a set of best practice recommendations of uniform content policies for platforms in 2018, and has produced updated and revised guidance in 2022.

DISINFO DEFENSE LEAGUE

Another coalition GLAAD is proud to be a member of is the Disinfo Defense League (DDL). The DDL is a distributed network of grassroots, community-based organizations that are building a collective defense against disinformation and surveillance campaigns that deliberately target Black, Latinx, Asian Americans and Indigenous people, along with other communities of color. DDL was created by and for these communities and is supported by services and insight provided by expert partners and organizations. DDL features over 230 organizational members who work across geography, generation, and gender to equip communities with tools, training, and tactics needed to combat racialized disinformation and win.

THE FACEBOOK LOGOUT

This November 2021 campaign, led by racial-justice tech accountability organization Kairos, brought together 47 organizations from across the movement (including GLAAD) and rallied more than 55,000 people who pledged to log out of Facebook during a weekend of action which raised awareness, and called on the company with a set of demands including: prioritizing data privacy, providing transparency in content moderation, tackling disinfo and misinfo, and putting in place new company leadership.

HERE'S HOW WE STOP FACEBOOK

GLAAD is one of the 75+ sign-on organizations for this massive fall 2021 data privacy campaign launched by Fight for the Future which offers a powerful call to action pointing out that basic data privacy legislation could solve many of the fundamental problems of the toxic social media landscape. As the campaign page explains: "Whistleblower Frances Haugen has shined a light on how Big Tech companies like Facebook, Instagram, and YouTube use harmful algorithms to recommend content in order to maximize profit, and the mass surveillance and data harvesting practices that power the algorithms. Stopping these companies from amassing data by passing strong privacy laws that put people — not corporations — in control of our personal information will severely diminish these platforms' harms. Congress must pass strong data privacy legislation, and the FTC should move forward with rule making that prohibits companies from collecting, purchasing or otherwise acquiring user information beyond what is needed to provide the service requested by the user [...]." Fight for the Future founder Evan Greer (and GLAAD SMSI advisor) underscores: "It's challenging to regulate algorithms directly, but lawmakers can cut off the fuel supply for Facebook's destructive machine by enacting a real Federal data privacy law strong enough to effectively kill surveillance capitalism as a business model."

BAN SURVEILLANCE ADVERTISING

This campaign implemented by a broad coalition of 38 organizations was launched in 2021 with an open letter to Big Tech platforms. From the campaign website: "Social media giants turn massive profits by endlessly tracking and profiling us, determining how to keep us hooked, and then hyper-targeting us with ads. These

platforms manipulate each user's information flow — and boost false and divisive content — to maximize engagement, so they can show us more ads and learn more about our behavior. Big Tech is making billions off surveillance advertising; society is paying the price [...]. The surveillance advertising business model may support 'free' services, but the real costs to society are incalculable. Below is a sampling of supplemental harms: Funding the misinformation machine, a menace to public health, a vehicle for discrimination, aiding and abetting violent extremists, selling access to sensitive personal information, helping government violate 4th amendment protections, rigging the game against small businesses, stifling innovation and competition, rife with fraud and grift, gutting the journalism industry, perpetuating addiction, promoting harms to children, connecting scammers with vulnerable users."

THE REAL FACEBOOK OVERSIGHT BOARD

This unique coalition of independent researchers, activists, and academics continues to call for transparency and accountability ("a public codified system that makes transparent what Facebook's system of enforcement is") and maintains a steady presence (on Twitter) driving attention to the myriad problems of Meta's most significant product. Not to be confused with the Oversight Board (which is the actual oversight board created by Facebook, now Meta, which independently evaluates content moderation cases and issues recommendations to the company).

WORLD ECONOMIC FORUM: GLOBAL COALITION FOR DIGITAL SAFETY

From the WEF website: "The Global Coalition for Digital Safety is a public private platform for global, multi stakeholder cooperation to develop innovations and advance collaborations that tackle harmful content and conduct online. As our lives in the digital and physical world continue to intersect, harms experienced online need to be better addressed. A framework to understand the risks and set safety standards can move us forward." See the Coalition's June 2021 report: Advancing Digital Safety: A Framework to Align Global Action.

THE SANTA CLARA PRINCIPLES 2.0

First drafted in 2018, the initial Santa Clara Principles on Transparency and Accountability in Content Moderation, offered recommendations of "initial steps that companies engaged in content moderation should take to provide meaningful due process to impacted speakers and better ensure that the enforcement of their content guidelines is fair, unbiased, proportional, and respectful of users' rights." The new 2.0 iteration launched in 2021 offers additional improved guidance on "meaningful transparency and accountability." It's worth noting that those two words, "transparency" and "accountability", are the terms that appear most prominently in all of the advocacy work being done right now in this field.

MOZILLA REGRETS REPORTER

This innovative tool from Mozilla provides YouTube users with a powerful tool to protect themselves and fight back — after a quick installation of this Firefox/Chrome browser extension you can use the tool to help clean up your own user experience of YouTube and also participate in Mozilla's research. As explained on the Mozilla website: "Sometimes YouTube recommendations can be spot on. But sometimes, they're outright harmful. Research powered by Mozilla's RegretsReporter has revealed that YouTube's algorithm violates the platform's very own policies, recommending misinformation, violent content, hate speech, and scams. RegretsReporter isn't just a product — it's also part of Mozilla's ongoing research and advocacy into harmful recommendation algorithms. Mozilla's updated RegretsReporter browser extension helps you eliminate harmful YouTube recommendations — the clickbait, the outrage, and all the other content you wish you never saw. By simply clicking an icon on any video on YouTube, RegretsReporter sends a signal to YouTube asking it to stop surfacing content you find objectionable."

ALSO OF VITAL IMPORTANCE ARE THE MANY RESEARCH CENTERS AT ACADEMIC INSTITUTIONS THAT ARE DEDICATED TO PLATFORM ACCOUNTABILITY, INCLUDING SUCH NOTABLE ENTITIES AS:

UCLA CENTER FOR CRITICAL INTERNET INQUIRY

The UCLA Center for Critical Internet Inquiry (C2i2) is a critical internet studies community committed to reimagining technology, championing racial justice, and strengthening democracy through research, culture, and public policy.

BERKMAN KLEIN CENTER FOR INTERNET AND SOCIETY AT HARVARD UNIVERSITY

The Berkman Klein Center's mission is to explore and understand cyberspace; to study its development, dynamics, norms, and standards; and to assess the need or lack thereof for laws and sanctions.

TECHNOLOGY & SOCIAL CHANGE (TASC) PROJECT

Led by Dr. Joan Donovan, the Technology and Social Change Project (TaSC) at Harvard's Shorenstein Center explores media manipulation as a means to control public conversation, derail democracy, and disrupt society.

STANFORD CYBER POLICY CENTER

The Stanford Cyber Policy Center, a joint initiative of the Freeman Spogli Institute for International Studies and Stanford Law School, is Stanford University's research center for the interdisciplinary study of issues at the nexus of technology, governance and public policy focused on digital technologies impacting democracy, security, and geopolitics globally.

NYU CENTER FOR CYBERSECURITY

The NYU Center for Cybersecurity (CCS) is an interdisciplinary research institute dedicated to training the current and future generations of cybersecurity professionals and to shaping the public discourse and policy, legal, and technological landscape on issues of cybersecurity.

INFORMATION SOCIETY PROJECT

The Information Society Project (ISP) is an intellectual center at Yale Law School. It supports a community of interdisciplinary scholars who explore issues at the intersection of law, technology, and society.

These are only a handful of examples of the many tools, campaigns, petitions, coalitions, initiatives, and entities doing the vital work of demanding accountability and transparency.

IMPORTANT WORK IS BEING DONE BY COUNTLESS OTHER ORGANIZATIONS INCLUDING:

Accountable Tech

The Anti-Defamation League

AVAAZ

Center for American Progress

Center for Democracy & Technology

Color of Change

Consumer Reports

Dangerous Speech Project

Data & Society

Decode Democracy

Fight for the Future

Free Press

Global Project Against Hate & Extremism

Institute for Strategic Dialogue

International Women's Media Foundation

The Leadership Conference on Civil and Human Rights

Media Justice, Media Matters for America

MPower Change

Muslim Advocates

PEN America

Ranking Digital Rights

Southern Poverty Law Center

UltraViolet

Woodhull Freedom Foundation

World Economic Forum

And many, many others

WE KEEP US SAFE: LGBTQ DIGITAL SAFETY & ONLINE ABUSE DEFENSE

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.

Online abuse can threaten livelihoods, damage mental health, lead to self-censorship, and even migrate offline. Members of the LGBTQ community are disproportionately targeted. According to a 2022 ADL/YouGov study, 66 percent of LGBTQ people experience online hate and harassment. From hateful slurs and sexual harassment to impersonation, account hacking, and doxing, abusive tactics are intended to intimidate, discredit, and silence. LGBTQ writers, journalists, creators, and activists — whose work increasingly requires an online presence — face a double bind. But there are steps we can all take to protect ourselves and one another from online abuse.

In the fall of 2021 GLAAD partnered with PEN America, NLGJA: The Association of LGBTQ Journalists, and the Trans Journalists Association (TJA) to present a free three-part webinar series on digital safety and online abuse defense, LGBTQ Digital Safety & Online Abuse Defense. Below are links to a few of the most important resources shared in the series.

- PEN America's Online Harassment Field Manual: includes, what online abuse is and sections on how to prepare, respond, practice self-care, and offer support — also available in French and Spanish

- What to Do if You're the Target of Online Harassment (concise protocol for navigating abuse)

For even more resources and information explore PEN America's Online Abuse Defense Program.

A CONCLUDING CALL FOR ACTION

As a concluding piece of guidance, we urge everyone in positions of leadership at these companies to find ways to take meaningful action now to make these platforms safe. For LGBTQ users, and for everyone.

ACKNOWLEDGEMENTS

ON THE FIREWALL BETWEEN FINANCIAL SPONSORSHIP & 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.

glaad

The **GLAAD Media Institute** provides training, consultation, and research to develop an army of social justice ambassadors for all marginalized communities to champion acceptance and amplify media impact.

GLAAD.org

@GLAAD