In recent years, American Jews have faced increased threats of violence and harassment both online and offline. According to ADL’s annual Audit of Antisemitic Incidents, 2019 and 2020 were, respectively, the highest and third-
highest years on record for cases of harassment, vandalism, and assault against Jews in the United States since tracking began in 1979. The recent conflict between Israel and Hamas, after which there was an increase in antisemitic incidents reported domestically, added another layer to Jews’ concerns over surging antisemitism and safety — 60 percent of Jewish Americans witnessed behavior or comments they deemed antisemitic following the conflict between Israel and Hamas.

Against this alarming backdrop, are tech platforms doing enough to combat antisemitism? To help evaluate this, ADL analyzed how well nine platforms (Discord, Facebook, Instagram, Reddit, Roblox, TikTok, Twitch, Twitter, and YouTube) addressed submitted reports of antisemitic content.[1] This is the first time that ADL has produced a report card grading platforms on their policies and enforcement against antisemitism.

ADL’s investigation showed that the majority of platforms performed at a middling level, with most earning grades in the C range. Twitter received a B-, the highest grade given, and Roblox earned a D-, the lowest grade of all platforms studied.

To determine how platforms responded to reports of violative antisemitic content, ADL investigators first searched for a handful of examples on each site. Between three and eleven examples of anti-Jewish content were identified for each platform we investigated. Second, ADL used accounts that were not publicly affiliated with ADL to report content under the platform's hate policies to see how platforms would enforce their policies when ordinary users flagged antisemitic content. Third, for items that were not removed or otherwise actioned as a result of the initial reporting from ordinary user accounts, ADL investigators again reported the content, this time through trusted flagger programs in which ADL participates across four of the platforms included in this investigation. These programs are designed to give partners that work with tech companies a special pathway to report violative content. Fourth, ADL's Center for Technology and Society (CTS) independently researched the accessibility of data from various
platforms, because the ability for researchers to retrieve data from platforms is an essential predicate for any third-party efforts to measure the prevalence of antisemitism and hate online.

Finally, ADL created a report card with grades to reflect all these metrics.

**Online Antisemitism Report Card**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>B</td>
<td>D-</td>
</tr>
<tr>
<td>YouTube</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>C</td>
<td>D-</td>
</tr>
<tr>
<td>Reddit</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>C</td>
</tr>
<tr>
<td>Twitch</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>TikTok</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>B</td>
<td>C-</td>
</tr>
<tr>
<td>Facebook (including Instagram)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>C-</td>
<td>C-</td>
</tr>
<tr>
<td>Discord</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>Reddit</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>D</td>
<td>D-</td>
</tr>
</tbody>
</table>

**THE LANDSCAPE OF ONLINE ANTISEMITISM**

Digital spaces can be unwelcoming and unsafe for many Jewish people. Rather than being relegated to a fringe corner of the internet, antisemitic content is quickly and easily found on major technology platforms that have hundreds of millions, if not billions, of users.

In ADL’s 2021 Online Hate and Harassment survey, 36 percent of Jewish respondents said they experienced some form of online harassment. Thirty-one percent of those Jewish respondents who were exposed to online hate reported they felt they were targeted because of their religion. Thirteen percent of Jewish respondents who experienced harassment said they were physically threatened. Moreover, ADL’s 2020 survey of hate, harassment, and positive social experiences in online games found that 10 percent of American adult gamers encountered Holocaust denial while playing — a sizable figure as 65 percent of American adults play games.
Prominent figures in American life also have been specifically targeted by online antisemitism. During the weeks preceding the 2020 national elections, ADL found that Jewish members of Congress faced antisemitic abuse on Twitter. Senate Majority Leader Chuck Schumer and Representative Jerrold Nadler, both Democrats from New York, received the largest portion of abusive tweets.

Most recently, following the conflict between Israel and Hamas in May 2021, ADL documented a disturbing rise in antisemitic content on multiple social media platforms during a seven-day period, including 17,000 tweets on Twitter that were variations of “Hitler was right.” This came alongside ADL’s data on offline antisemitic incidents in May 2021 being more than double the number reported in the U.S. during the same period a year before, in May 2020.

Notably, there is a lack of knowledge about antisemitism and its history both in the U.S. and abroad, especially among younger generations. A 2020 study showed alarming levels of ignorance surrounding the most significant antisemitic event in modern Jewish history: nearly two-thirds of American adults ages 18-39 do not know that six million Jewish people were killed during the Holocaust, more than 10 percent believe Jews caused the Holocaust, and almost a quarter of respondents said they believed the Holocaust was a myth or an exaggeration. Given these findings, we cannot assume that people can effectively detect antisemitic tropes and stereotypes, whether overt or subtle.

Evidence of anti-Jewish sentiment dates back millenia; prejudice against Jewish people undergirds many conspiracy theories that blame the world’s ills on them. Bad actors use social media platforms to put a new spin on an ancient form of hatred. ADL has identified seven of the most common antisemitic myths used throughout history that still prevail today.
A lingering myth behind modern antisemitism is that Jews possess extraordinary power to harm or control people outside the Jewish community. Jewish people are depicted as megalomaniacs, as manipulating world governments, banking and financial institutions, and controlling the media and entertainment industries. Common hoaxes of Jewish people bidding for world domination include the Rothschilds and the Federal Reserve, the Zionist Occupied Government (ZOG), the Protocols of the Learned Elders of Zion, and Jewish Hollywood. Another oft-circulated myth is that the financier and philanthropist George Soros funds the Black Lives Matter and antifa movements to sow chaos. Mr. Soros, who is of Hungarian Jewish heritage, is a common bogeyman of the far right.

Disloyalty

- Jewish people are often depicted as disloyal, accused of caring more about Jewish interests — or more recently, about Israel — than non-Jewish peers or a Jewish person's home country. Relatedly, Jews are sometimes portrayed as imposters who pretend to support a particular group or issue to advance nefarious ulterior motives.

Greed

- Antisemitic content often portrays Jewish people as greedy. Accused of relentlessly pursuing wealth, Jewish people are characterized as duplicitous, and as resorting to lying, cheating, or stealing, no matter how trivial the amount of money. They are depicted as stingy misers determined not to let material wealth slip from their grasp.

Deicide

- Inaccurate historical accusations that Jewish people collectively killed Jesus Christ is an antisemitic myth that dates back centuries.

Blood

- Antisemites believe Jewish people murder non-Jews to use their blood to perform religious rituals. This allegation, which has been repudiated by the
Catholic church over the centuries, is referred to as the blood libel. Blood libels resurface during times of global health crises or plague, for instance, during the COVID pandemic.

Denial

- People who deny, doubt, or minimize the Holocaust.

Anti-Zionism

- Criticisms of Israel and Zionism are often distinct from antisemitism, but not always. ADL defines Zionism as “the movement for Jewish self-determination and statehood in their ancestral homeland in the land of Israel.” Anti-Zionism can cross the line into antisemitism when it:
  - Denies Jews the right to the same national self-determination that is afforded to other peoples.
  - Asserts Jews or Zionists are collectively responsible for the actions of Jewish or Israeli individuals, Israeli entities or the Israeli government.
  - Incorporates classic antisemitic themes or conspiracy theories into criticism of Israel.

Many users report seeing hateful cartoons drawing upon these myths, often physically caricaturing and stereotyping Jewish people. One typical antisemitic image is the “Happy Merchant,” a Jewish man with heavily stereotyped facial features, including a large nose, beard, and hunched shoulders, wearing a yarmulke, who greedily clasps his hands with a wicked grin spread across his face.
Content with the "Happy Merchant" that was reported to Instagram as part of this investigation. Instagram did not take action against this content.

Antisemitism is repackaged for younger generations using memes, emojis, and games. For example, one of the pieces of content ADL reported in this investigation was a Twitch stream of Pokemon Clover, a modification of the popular Pokemon game that shows hateful images and stereotypes, including antisemitic tropes.
Social media platforms are unmatched vectors for the rapid spread of antisemitic content; a single tweet or post can reach millions of people in mere seconds and cause lasting harm that is difficult, if not impossible, to undo. As people consume more of their news and entertainment on platforms, their ability to discern between fact and misinformation, or humor and abuse, becomes compromised.

**METHODOLOGY**

ADL evaluated the nine social media platforms across eight categories that fell within the larger areas of policy, enforcement, product, and data accessibility.

**Policy**

ADL investigators evaluated the publicly stated hate policies of nine social media platforms and their enforcement responses. In grading, ADL weighted enforcement of hate policies more heavily than existence of policies because even well-crafted policies are ineffective without meaningful and consistent enforcement at scale.

**Hate Policies and Community Guidelines**

Online platforms must have comprehensive hate policies to let users know what is or is not acceptable, and guide their own content moderation efforts. These hate policies should prohibit content or behavior targeting marginalized individuals or groups. ADL thus looked at whether technology companies had a broad policy or a set of community guidelines around hateful content on their platforms. Tech companies earned a score for the following:

- Metric #1: Does the policy explicitly mention religion, race, or ethnicity?

All nine of the platforms ADL analyzed have general hate policies[2] that explicitly mention protected groups and cover religion, race, or ethnicity.
Broad hate policies are also important because antisemitism online can take many forms. Perpetrators target Jewish people online based on a variety of actual or perceived characteristics that play to racial, ethnic, or religious bias.

For example, one post, which ADL reported to Twitter, advances a racist and antisemitic conspiracy theory about “Jews” wanting to dominate “the Blacks,” using the common antisemitic trope of a thirst for power. ADL reported this as violative of Twitter’s rules regarding hate speech. While Twitter’s hate speech policy does explicitly mention multiple protected groups, its reporting interface does not allow a user to specify the protected group being targeted.

In that post (above), ADL felt that the antisemitism on display did not necessarily target Jews based on their faith, but instead posited Jewishness as a racial or ethnic identity that exists to manipulate and dominate Black Americans for the “benefit of the Soviet Union.”
In another example, the above post — also reported to Instagram as part of this investigation — explicitly targeted Jews and Judaism in a religion-based attack, calling the practice of Judaism the “synagogue of Satan.”

These two examples illustrate some of the multi-dimensional ways that antisemitism manifests on social media. Given this complex context, platforms need policies against hate that explicitly mention religion, race and ethnicity (among other protected groups) to effectively address online antisemitism because focusing solely on the targeting of Jews based on religion might not give content moderators at platforms the flexibility, understanding and tools they need to effectively address the ways in which content expresses hatred of Jews.

**Enforcement**

**Reporting Effectiveness and Transparency**
Beyond having clear and comprehensive policies, platforms must also communicate with their users about content management decisions. Users
deserve to know that platforms will thoughtfully review their reports of hateful content. ADL established six categories in the report card to evaluate how well platforms responded to users’ reports. Platforms earned scores for the following:

Four of these six categories focused on platforms’ responses to reports from an ordinary user:

- **Metric #2:** Did the platform respond within 24 to 72 hours?
- **Metric #3:** Was the user notified whether the content they reported violated a specific platform policy (or was not violative)?
- **Metric #4:** Was the content removed or otherwise actioned as a result of the report?

Two categories focused on platforms’ responses to reports from trusted flaggers:

- **Metric #5:** Does the platform have a trusted flagger program?
- **Metric #6:** Did the platform take action on content reported through its trusted flagger program?

ADL gave the highest scores to platforms that responded within 24 hours to user reports with the results of their investigations and enforcement decisions and lower scores if they replied within 72 hours. ADL gave a score of zero if the response was received after 72 hours.

Notably, none of the “ordinary user” reports received a response that provided a specific reason as to why the platform did or did not take action. This lack of transparency on the rationale for the decisions behind content moderation is troubling. Users have reported to ADL their frustration when they are not told why a piece of content was or was not removed; they have no way of knowing whether platforms carefully reviewed their reports or how those platforms do or do not apply their policies.
Support Message

Thanks for your feedback

Thanks for your report - you did the right thing by letting us know about this. We looked over the photo, and though it doesn't go against one of our specific Community Standards, we understand that it may still be offensive to you and others. No one should have to see posts they consider hateful on Facebook, so we want to help you avoid things like this in the future.

From the list above, you can block Kevin Macaulay directly, or you may be able to unfriend or unfollow them. If you unfollow them, you'll stay friends on Facebook but you won't see their posts in your News Feed.

We know these options may not apply to every situation, so please let us know if you see something else you think we should take a look at. You may also consider using Facebook to speak out and educate the community around you. Counter-speech in the form of accurate information and alternative viewpoints can help create a safer and more respectful environment.
In the above post — reported to Facebook — ADL noted that the Jewish Star of David is called a 6-sided polygon, which supposedly contains hidden symbolism alluding to 666 (the number of the beast in the New Testament Book of Revelation), and Judaism is called “The Synagogue of Satan.” We believed that the post clearly ran afoul of Facebook’s “harmful stereotypes” policy, which the platform defines as “dehumanizing comparisons that have historically been used to attack, intimidate, or exclude specific groups, and that are often linked with offline violence.” However, Facebook declined to remove it. Furthermore, the platform failed to specify how it arrived at this decision when content was flagged by an ordinary user, which reinforces the current lack of clarity regarding what constitutes violative antisemitic content on Facebook.

**Trusted Flagger Channels**

As part of this investigation, ADL also used trusted flagger channels, if provided by a platform, to report content when companies did not take action on content reported through ordinary user channels. ADL is a member of several tech companies’ formal trusted flagger programs, which provide faster escalation paths for reports of violative content. ADL evaluated platform responses to trusted flaggers in the same way it assessed responses to ordinary users — both in terms of the speed of its response and if action was taken as a result of the report.

Each platform that includes ADL in its trusted flagger program[3] responded differently when ADL reported content through those channels. While TikTok did not take down any of the content ADL flagged as an ordinary user, the platform removed all of the content we reported as a partner organization through its trusted flagger program. Twitter and Facebook/Instagram removed none of the content reported by ordinary users, but then removed 20 percent of the content (1 out of 5 and 2 out of 10 reports, respectively) reported by us through their trusted flagger programs. Of note, YouTube removed 60 percent of the content (3 out of 5
reports) flagged by ordinary users; it removed neither of the remaining two pieces of content that we subsequently reported through the trusted flagger account.

**Product-Level Efforts Against Antisemitism**

Policy enactment and enforcement are both critical to evaluating how tech companies address antisemitism and other forms of bigotry on their platforms. But the design of the platforms is also essential to combating online hate, including antisemitism. For example, the ease by which a user can post an abusive comment directly affects both the prevalence and the impact of hateful content. For that reason, ADL has urged platforms to incorporate “friction” into their design to make users stop for a moment, hopefully encouraging reflection before posting certain types of harmful content. And, on the recipient side, the ability to effectively address online hate is based in part on how the reporting systems are designed and built. In this case, as opposed to when dealing with the poster of hateful content, it’s important to remove friction. A user should be able to report multiple instances of antisemitism to the platform with ease, instead of being required to report each piece of antisemitic content separately. The latter requires far more time and effort, decreasing the likelihood that all items will be reported for action and re-exposing targets to the harmful content.

ADL reviewed product-level changes and assessed platforms on the following:

- **Metric #7:** Does the platform have effective product-level steps to address antisemitism?

In this category, ADL gave high scores to Reddit, TikTok, Twitter, and YouTube for the following product changes:

- Both [Twitter](https://twitter.com) and [YouTube](https://www.youtube.com) implemented prompts that pop up whenever a user attempts to post potentially violative content. These prompts ask users to check themselves and confirm that they wish to continue.
In July 2020, Twitter updated its policies to remove links to off-platform content that violated its hate policies. This has also resulted in a product change involving an “interstitial” when clicking on certain links on the platform, such as the white supremacist website The Daily Stormer in the example above. In addition to warning about spam and spyware, the interstitial now includes language about content that is violent or misleading, and might lead to “real-world harm.”

- ADL has long recommended that platforms explore ways to understand the experience of users from marginalized groups and how they are impacted by hate. Among other things, this provides valuable insights to web
designers and engineers, as well as policy architects, to better understand (and measure) the harms they are seeking to mitigate. In a December 2020 update, YouTube stated that it would allow active content creators from marginalized communities to voluntarily report their identities to the platform. This would allow the platform to look more closely at how creators from specific communities are targeted by hate and better track and address hate against specific identity groups on the platform. ADL commends YouTube's measure as a step in the right direction that we also hope will influence other platforms to implement similar product changes.

- Equipping users targeted by antisemitism, hate, and other forms of online abuse with the ability to report comments in batches is another important feature that provides far more agency against trolls and hate and harassment. ADL and other organizations have long recommended that platforms provide such features to users. TikTok announced new features in May 2021 that enable users to report up to 100 comments at once and also block accounts in bulk.
Looking at how tech companies approach policy, respond to reports of antisemitism, and design their products to mitigate antisemitism is important, but even these three areas omit a key metric in evaluating a platform's performance addressing the issue: prevalence. How much antisemitism is there on any given platform? At present, no tech company provides any data on the scope of any specific form of hate on its platform. For a third party such as ADL to understand the full scope of antisemitism and other forms of hate on a social platform, tech companies must make data available to third parties, whether independent researchers, civil society organizations, or academia broadly, all while respecting user privacy.

To assess how platforms are doing on data accessibility, ADL's Center for Technology and Society evaluated the following:

- Metric #8: To what degree do platforms make it easy for researchers and civil society organizations to acquire data that makes third-party measurement and auditing achievable?

Large-scale auditing of social media platforms requires robust Application Programming Interfaces (APIs), enabling companies to share their application data with other parties. ADL used eight criteria to grade platforms' data accessibility:

- Availability of public APIs that return public posts and enable third parties to retrieve data with minimal setup
ADL recommends that every platform develop these features, but none of the nine we analyzed met all these criteria. Twitter and Reddit scored the highest because both offer strong search and streaming capabilities at high rate limits, allowing large quantities of data to be collected. However, Twitter could improve its score with fewer publication restrictions, while still finding ways to respect user privacy. Twitter allows researchers to share the IDs of tweets they analyzed, but the platform does not permit them to release the full data (text, author, likes, retweets, bios, etc.). People who ignore this restriction risk having their API keys and developer access revoked.

Reddit could do better by adding keyword search, and by providing tools to discover new subreddits and posts. ADL recommends both Reddit and Twitter also create APIs for user reporting and content moderation information.

Discord, Twitch, and YouTube scored in the middle because they could not specify search criteria or had limited collection data volume, and had no automated
discovery capabilities. TikTok lacks an API for collecting data on any type of public user post. Facebook enables data collection through the CrowdTangle API, but heavily regulates who has access to the tool. ADL has requested access to CrowdTangle, but has been denied for unspecified reasons. Roblox offers limited accessibility to researchers.

**LIMITATIONS AND CAVEATS**

ADL did not include all digital social platforms in this investigation. ADL investigators focused on a cross-section of nine mainstream technology platforms and services that have publicly expressed their commitments to addressing hate and whose platforms would be easily searchable by an ordinary user. For example, Gab, 4chan, and Telegram were excluded because they have few, if any, policies regarding hate, extremism, or antisemitism. ADL also did not include companies that provide infrastructure such as web hosts, domain registrars, or payment processors whose services are not broadly searchable, or who do not provide significant social media services.

It is also important to note that this report did not seek to measure the prevalence of antisemitism on the platforms. Rather, it evaluates policies, responses to reporting of antisemitic content, and data accessibility as components of effective and independent measurement of antisemitism online.

**RESULTS**

**Online Antisemitism Report Card**

<table>
<thead>
<tr>
<th>Platforms</th>
<th>Hate Policy That Explicitly Mentions Race, Religion, or Ethnicity</th>
<th>Response Within 24-48 Hours</th>
<th>Notification of Policy Reason for Enforcement</th>
<th>Action Taken Against User</th>
<th>Trusted Flagger Program</th>
<th>Action Taken Upon Trusted Flagger Report</th>
<th>Effective Product-Level Efforts to Address Antisemitism</th>
<th>Data Accessibility Grade</th>
<th>Total Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>C</td>
<td>B-</td>
<td></td>
</tr>
<tr>
<td>YouTube</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>C</td>
<td>B-</td>
<td></td>
</tr>
<tr>
<td>Reddit</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>C</td>
<td>B-</td>
<td></td>
</tr>
<tr>
<td>Twitch</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>C</td>
<td>D-</td>
<td></td>
</tr>
<tr>
<td>TikTok</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>C</td>
<td>C-</td>
<td></td>
</tr>
<tr>
<td>Facebook (including Instagram)</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>D</td>
<td>C-</td>
<td></td>
</tr>
<tr>
<td>Discord</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>C</td>
<td>C-</td>
<td></td>
</tr>
<tr>
<td>Roblox</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>D</td>
<td>D-</td>
<td></td>
</tr>
</tbody>
</table>
Twitter and YouTube (B-)

Twitter earned the highest grade among the platforms investigated. It has broad anti-hate policies in place and has introduced effective product-level efforts to address antisemitism, including the use of prompts and its shift last summer to scan external links for hateful content that violate its policies. However, Twitter did not remove or take other action on content reported from an ordinary user's account; it waited until ADL, using its organizational affiliation, activated the trusted flagger program. ADL weighted scores on enforcement more heavily than policy. Although Twitter responded quickly to content flagging, it did not provide information on why it took action (in fairness, no platform did). Twitter does have a trusted flagger program and took action when ADL reported content through those means. Notably, Twitter is one of the top two platforms in terms of meaningful data accessibility, which would allow for third-party measurement of antisemitism.

YouTube tied with Twitter for the highest grade due to its robust hate policies and product-level efforts to address antisemitism. Both products incorporated recommendations from ADL and other organizations by adding prompts that warn users before they post potentially violative comments and allow users who create video content on the platform to self-identify to help the platform better understand how different marginalized groups are targeted. While YouTube failed to respond to any of the reported content within 72 hours, the platform eventually took action on reports from an ordinary user. YouTube has a trusted flagger program, but it did not remove any of the content ADL subsequently flagged as part of that process. Additionally, while YouTube has some data accessibility, it is not as accessible as Reddit and Twitter. It neither has an API access tier for researchers and NGOs nor does it let third parties access data to understand its decisions behind content moderation better.

Reddit and Twitch (C)

Reddit and Twitch tied for third. Both platforms have strong policies around hate that include explicit protections for protected groups. Setting aside policy, the
reasons why each platform earned a C differ.

Reddit did not take action on any content ADL reported. ADL gave it a high score for undertaking product-level efforts to address antisemitism because its platform architecture handled the r/WallStreetBets controversy that occurred earlier this year well. While ADL saw a degree of antisemitism in r/WallStreetBets, platform and community moderation and Reddit's upvoting/downvoting features largely worked to reduce the impact of abuse. At the same time, Reddit has the highest level of data accessibility along with Twitter, which allows for third-party audits of antisemitism on the platform and earned a high score.

Twitch did significantly better than Reddit when it came to taking action on antisemitic content by responding to reported content within a 72-hour window and actioning the majority of the antisemitic content that ADL flagged. Nevertheless, its lack of a formal trusted flagger program and less than ideal data accessibility for researchers earned it the same grade as Reddit.

- **Discord, Facebook/Instagram, and TikTok (C-)**

Discord, Facebook/Instagram, and TikTok received their grades for a variety of reasons. Like the other platforms reviewed here, all four have robust hate policies yet did not provide any information behind their enforcement decisions when responding to reported content. Nor did any of these four platforms take any action against the content ADL reported through ordinary user flagging. Facebook/Instagram and TikTok have trusted flagger programs while Discord does not; only TikTok actioned content reported through the special channel. However, TikTok is at the very bottom of the platforms reviewed in terms of data access as there are no means for researchers to acquire data to understand the platform meaningfully. TikTok did not receive the lowest grade overall primarily because of its recent product change to allow mass user reporting and bulk blocking. These important tools empower users to respond to mass hate and harassment. Unlike TikTok, Facebook has an API but is opaque with data accessibility; it only provides select users with access to certain data through its...
CrowdTangle tool. Discord has superior data access compared to TikTok and Facebook; on that metric it is comparable to Twitch and YouTube. **Note:** After the release of the report, Discord notified ADL that they do have a formal trusted flagger program that is not public and about which they had not previously notified ADL. ADL plans to participate in the program going forward.

- **Roblox (D-)**

Roblox received the worst grade in this investigation. While Roblox does have policies around hate that are as substantive as more traditional social media companies, it fell short in most of the other categories. Content reported to the platform was neither responded to nor actioned. For example, ADL flagged an audio file that included in its title the phrase "6 gorillion", a popular expression used by white supremacists to dismiss the number of Jews killed in the Holocaust as hyperbole. The audio file itself consists of antisemitic stereotypes and comments making light of the Holocaust. The file is available in Roblox's library for anyone creating a Roblox game to incorporate as a sound effect.

Roblox was given a poor grade for data accessibility. The platform provides some access to data that would allow researchers to understand the platform better, but it is limited.

**CONCLUSION AND RECOMMENDATIONS**
ADL investigators found that no platform performed above a B- in addressing antisemitic content reported to it. Also, no platform provided information or a policy rationale for why it did or did not remove flagged content. Reddit and Twitter earned the highest marks for data accessibility through their APIs that can enable ADL or other researchers to study the prevalence of antisemitism on their platforms. However, neither Reddit nor Twitter took action on the content ADL reported through ordinary user channels — Twitter did so through its trusted flagger program, while Reddit did nothing in response either to ordinary user flags or trusted flagger reports.

TikTok has introduced innovative product-level efforts such as enabling users to report up to 100 comments at once, empowering users in a way rarely seen from tech companies. Still, TikTok took no action on the content ADL reported through ordinary user channels and its data accessibility is at the bottom among the nine platforms analyzed.

ADL commends YouTube for introducing prompts to users before they post potentially violative content and allowing its creators to self-identify, but the platform falls short in data accessibility and its trusted flagger program.

As the world’s biggest social media platform, Facebook’s responsibility to curb antisemitism on its platforms is even greater than that of other platforms. But as this investigation shows, Facebook’s efforts are not commensurate with its size. The company did not take action on any of the content flagged through ordinary users’ accounts and its data accessibility is heavily limited.

Roblox, robust hate policies aside, did the least to address antisemitic content even in egregious posts, clearly communicating that it does not regard user trust and safety as a priority.

Users deserve more transparency and greater protection from platforms than companies are inclined to provide. Such reluctance has consequences in the form of economic, emotional, mental, political, and physical abuses that affect many.
people's lives, as shown repeatedly in ADL's research. It is irresponsible for platforms to take, at best, piecemeal approaches that do little to address the rapidity and depth of online hate and harassment.

The challenge of enforcing platform policies consistently and at scale is not lost on ADL. Each of these platforms serves millions, if not billions, of users. ADL recognizes that even if a platform with a billion users who each post once a day enforced its policies accurately 99 percent of the time, that would still leave 10 million errors every day. Each of those 10 million posts, however, is not just a statistic, but a person exposed to harassment, a threat of violence, misinformation, or some other harmful phenomenon stemming from platforms operating at these massive scales. Though enforcement can never be perfect, ADL still believes there is significant room for improvement across all platforms in prioritizing resources toward the moderation of harmful content. This investigation captures how platforms function around antisemitism from the perspective of an affected individual user and provides constructive criticism around specific areas in terms of policy, enforcement, product and data accessibility where tech companies can improve their platforms. While some companies have made some progress, more is required to ensure that these platforms become safe, respectful and inclusive spaces for all people.

**Recommendations**

1. **Enforce policies on antisemitism and hate consistently and at scale**
   Platform policies on their own are not sufficient; they require enforcement that is consistent across a digital social platform. ADL recommends that, especially for content such as Holocaust denial, tech companies designate sufficient resources for content moderation. Appropriate training for human moderators, greater numbers of human moderators, and expanding the development of automated technologies to commit as few mistakes as possible when enforcing policies around violative content are worthwhile investments.
2. Give greater access to data for researchers
   Tech companies should use ADL’s rubric for data access to provide researchers with substantive, privacy-protecting data to aid their efforts to understand the nature of antisemitism and hate online and study whether platforms’ efforts to address hate genuinely work. Companies should make available APIs that return publicly posted content and moderation decisions, enable this content to be retroactively searched and proactively streamed, and ensure rate limits are high enough to permit data analysis at scale. When tech companies provide this meaningful data to researchers, third parties will be able to audit the prevalence of various phenomena on social media platforms. This will also allow third parties to determine whether efforts by tech companies to address different harms on their platforms, such as antisemitism, actually work.

3. Ensure expert review of content moderation training materials
   Tech companies should seek subject matter experts’ advice when drafting and revising the guidelines they use to train their human content moderators. ADL and other Jewish organizations could serve as experts who can review such materials to ensure that content moderators are adequately trained to recognize and address antisemitism.

4. Increase transparency at the user level
   Platforms must provide detailed information to users on how they make their decisions regarding content moderation. They should give the user the specific policy that guided their decisions and share information about why the content did or did not violate the policy in question.

APPENDIX

Antisemitic Content Reported But Still Active
The below content was reported to the platforms as part of this investigation in late May and June and was still active as of July 27, 2021.
Roblox

Roblox content 1

Roblox Content 2

Roblox Content 3
Roblox content 4

Discord
Discord Content 1
Discord Content 2
Discord Content 3
Discord Content 4

Judaism Intensifies [HD]
By XIZAM

- Type: Audio
- Genres: All
- Updated: Nov 18, 2017
- Description: Tags: Jew, juden, hebrew, meme, wbm, rabbi, oy vey, oy gevalt, shoah, gay, shakal, shekels, shlama, schlimo, herg, stein, gaid, man, mordecai, communism, 6 gorrillion, million
Facebook and Instagram

Instagram Content 1

![Image of Titanic meme with label "Ice(((BERG))) AHEAD"]

Instagram Content 2
Twitch

Twitch content 1

While it is unethical to help adventuring Goys, I like you

Twitch content 2
Reddit Content 10

Reddit Content 11

YouTube

YouTube Content 1

Twitter

Twitter content 1
Salamandren
@salamandrens
happy go lucky anti-Jew
⊙ the middle land ☎ Joined May 2020
98 Following 49 Followers
Not followed by anyone you're following

We Need a Natural Solution to the Jewish Menace (Salamandren's End of Days 459)

Not politics, but virtue is needed.

chute.rocks/F1hd7AVqQLj via @BitChute

VitaminD47 @vitamin_d47 · Jun 7
NEVER Trust a Jew ⚠️ Edgar J. Steele chute.rocks/zfGZGYkPZ8bq via @BitChute

Twitter content 2

Twitter content 3
Istvan Bakony: Jews want to dominate the Blacks.

"The goal of the Jewish imperialism in the USA, is using its 5th column infiltrated among the blacks, to control them to the benefit of the Jewish plans, IN THE BENEFIT OF THE SOVIET UNION."

8:27 PM · May 20, 2021 · Twitter Web App

8 Retweets 1 Quote Tweet 13 Likes
Because Facebook owns Instagram, ADL did not give separate grades for the two platforms, instead treating them as one entity.

Platforms have different names for their hate policies. For example, Facebook has “Community Standards” while Twitter has a “hateful conduct policy.” The phrase “general hate policies” encompasses all of these variations.

There are four trusted flagger programs among the nine platforms. Facebook and Instagram use the same one. The others are at TikTok, Twitter, and YouTube. ADL belongs to each of these programs.

[1] Because Facebook owns Instagram, ADL did not give separate grades for the two platforms, instead treating them as one entity.
[2] Platforms have different names for their hate policies. For example, Facebook has “Community Standards” while Twitter has a “hateful conduct policy.” The phrase “general hate policies” encompasses all of these variations.
[3] There are four trusted flagger programs among the nine platforms. Facebook and Instagram use the same one. The others are at TikTok, Twitter, and YouTube. ADL belongs to each of these programs.

DONOR RECOGNITION

ADL gratefully acknowledges The ADL Lewy Family Institute for Combating Antisemitism for its sustained support and commitment to fighting antisemitism. ADL also thanks its individual, corporate and foundation advocates and
contributors, whose vote of confidence in our work provides the resources for our research, analysis and programs fighting antisemitism and hate in the United States and around the globe.