Transparency report

Parent entity: **Discord Inc.**

Content-sharing service: **Discord**

https://discord.com/

Meets OECD baseline transparency standard

Publication date: Nov 4, 2022, 07:55 AM UTC

Select your transparency reporting period

1 January – 30 June 2021

Section 1

Which of the following best describes the primary type(s) of online service(s) that you provide?

- Messaging
- Video chat

If you would like to add a comment or other information, please do so here.

Discord is a voice, video, and text chat app that's used by tens of millions of people ages 13+ to talk and hang out with their communities and friends. The purpose of Discord is to give everyone the power to create belonging in their lives. Users communicate through direct messages, group messages, or in communities which we call servers. While messaging and video-chats are the primary services we provide, Discord currently also offers the possibility to share files in servers, play games together through “activities”, as well as stream content in servers and group direct messages through the “sharing” option.

Section 2

Do you prohibit terrorist and/or violent extremist content on your service?

Yes

Do you use one or more specific, publicly available definitions or understandings of terrorist and/or violent extremist content?

Yes

Please provide the publicly available definition(s) or understanding(s) that you use, along with your relevant terms of service or policies.

We use a dedicated category of "violent extremism". In addition, we have two additional categories called Violent and Graphic Content and Hate Speech which are distinct from our Violent Extremism category.

The relevant community guideline is: "Do not use Discord for the organization, promotion, or support of violent extremism. This also includes glorifying violent events, the perpetrators of violent acts, or similar behaviors". You can find further information on our community guidelines at https://discord.com/guidelines

Categorizing violent extremism itself is difficult because not all extremists have the same motives or believe in the same ideas. Some individuals who adopt violent ideologies act on their beliefs by joining organized hate, terrorist, or violent extremist groups.

Others don't want to officially identify themselves as belonging to a particular movement, and may instead form looser connections with others who have adopted the same worldview. Different cultural contexts also influence belief systems and behaviors, so violent extremist ideologies in one country will naturally be different from those on the other side of the world.

Violent extremism is nuanced and the ideologies and tactics behind them evolve fast. We don't try to apply our own labels or identify a certain "type" of extremism.

Instead, we evaluate user accounts, servers, and content that is flagged to us based on common characteristics and patterns of behavior, such as:

- Individual accounts, servers, or organized hate groups promote or embrace radical and dangerous ideas that are intended to cause or lead to real-world violence
- These accounts, servers, or groups target other groups or individuals who they perceive as enemies of their community, usually based on a sensitive attribute.
- They don't allow opinions or ideas opposing their ideologies to be expressed or accepted.
- They express a desire to recruit others who are like them or believe in the same things to their communities and cause.

It's important to note that the presence of one or two of these signals doesn't automatically mean that we would classify a server as "violent extremist." While we might use these signs to help us determine a user or space's intent or purpose, we always want to understand the context in which user content is posted before taking any action.

Section 3

Do you use any of the following methods to detect terrorist and/or violent extremist content on your platform or service?

- Flagging by individual users or entities
- Trusted notifiers
- Internal flagging done by human review
- Internal flagging done by automated technologies
- Hybrid system

Can you determine the total amount of content that was flagged or reported as terrorist and/or violent extremist content on your service during the reporting period?

No

If you would like to add a comment or other information, please do so here.

In H1 2021, Discord received 13,102 reports for violent extremism.

Section 5

Please select all interim or final actioning methods that you use on terrorist and/or violent extremist content.

- Content removal
- Warnings to account holder
- Suspension/removal of account
- Limitations to account functions and/or tooling
- Other

You have selected "Other" on the previous question. Please describe:

In addition to the actions mentioned in response to the previous question, we also have the ability to remove entire Discord's servers, for specific violations of our community guidelines.

Section 6

Can you determine the total amount of terrorist and/or violent extremist content on which you took action during the reporting period?

No

Can you determine the total number of accounts on which you took action during the reporting period for violations of your policies against the use of your service for terrorist and/or violent extremist purposes as a percentage of the average number of monthly active accounts during the reporting period?

Yes

Are you willing and able to disclose it?

Yes

Please provide that percentage, along with any breakdowns that are available.

We cannot determine the amount of content we took action on. We can determine the number of accounts disabled and servers removed.

H1 2021:
Accounts Disabled: 30,645
Servers Removed: 1,834
Proactive: 1,351
Reactive: 483

If you would like to add other comments or information, please do so here.

In addition to disabling accounts, we remove servers on Discord. A total of 1,834 servers were removed for Violent Extremism in the six months between January and June 2021. We are proud that upwards of 73% of these high-harm spaces were removed before being reported to Trust & Safety. We are continuing to invest in proactive tooling and resources to ensure that violent and hateful groups do not find a home on Discord.

Section 7

If your service includes livestreaming functionality (even if it is not among what you consider to be the primary functionalities), then given the potential for terrorists and violent extremists to exploit livestreaming in ways that could promote, cause, or publicize imminent violence or physical harm, do you implement controls or proactive risk parameters on livestreaming to reduce misuse?

No livestreaming functionality

If you would like to add other comments or information, please do so here.

Our livestreaming services are available to users only through private direct messages to other users, or through Discord communities (called servers) which are mostly only available through private invitation. Content livestreamed on Discord does not automatically broadcast to a publicly available website and is consequently limited in its reach to viewers.

When we can confirm that users have used this service to livestream terrorist or violent extremist content, we have the ability to disable the account livestreaming the content and/or remove the server where the content is livestreamed.

Section 8

Please provide details on how you balance the need to action terrorist and/or violent extremist content with the risk that such content can be mislabelled and may actually be denouncing and documenting human rights abuses, or that it does not otherwise violate your terms of service.

Our trained agents from the trust and safety team regularly evaluate whether to take action on specific content violating our guidelines, or whether that content should stay on the service. Our internal policies provide guidelines on how to assess content, to ensure we implement them uniformly across our service. For instance, we have carved out exceptions for content that is newsworthy or educational and is very clearly identified as such in the context.

Section 9

Do you have an appeal or redress process for content and/or account actioning decisions made under your terms of service on terrorist and/or violent extremist content?

Yes

Please provide a detailed overview of those processes.

Discord allows users to appeal actions taken on their accounts if they feel that the enforcement was incorrect. In this section, we'll discuss our approach to how users can have actions taken on their accounts reviewed.

We welcome appeals and take our users seriously when they make the effort to raise context or information we may not have known of when we made our decision. We review appeal requests and reinstate accounts if we determine that a mistake was made, or if we have good faith in the user's appeal that they have recognized the violation made for a lower-harm issue and will abide by our Community Guidelines once back on Discord.

Is your appeal or redress process available to the user who posted the content or owns the account in question?

Yes

Is the outcome of your appeal and redress process available to the user who posted the content or owns the account in question?

Yes

Is your appeal or redress process available to the person or entity who requested actioning?

No

If you would like to add an explanation or other comments, please do so here.

To preserve the privacy of our users, we do not reveal the appeal or redress process, nor the outcome of this process to reporters for any of our policies, including violent extremism.

What is the total number of appeals received from all sources, during the reporting period, following content or account actioning decisions under your policies against terrorist and/or violent extremist content?

7,248

How many such appeals were decided during this reporting period (regardless of when those appeals were received)?

We reviewed 7,248 appeals.

Of those, how many were granted?

103

If you can break these numbers (appeals received, decided and granted) down with any more detail, please do so here.

N/A

Section 10

How, and how often, do you measure and evaluate the efficacy and/or room for improvement of your policies in each of the following areas?

We plan to audit and update all of our platform policies at least once a year, or when real-world events or new research or data necessitates an immediate review. We are currently working with a third-party consultant to review and update our violent extremism policy.

We also regularly reassess the efficacy of our automated and manual detection strategies based on transparency reports and plan adjustments accordingly.

Section 11

Do you have a point of contact (such as a dedicated email alias, desk or department) that can be contacted during a real-world or viral event with direct online implications and which works to address harmful content on your service?

Yes

Provide any additional details that you would like to add about this point of contact.

We have an email address, leading to our specialized trust and safety teams, which we share with our partners to ensure swift communication during real world events. Please see the answer to the following question for further details.

Section 12

Are you a member of an international crisis protocol aimed at reducing the volume and impact of terrorist and/or violent extremist content online during a crisis?

Yes

Please identify the protocol.

We are members of the EU Internet Forum's EU Crisis Protocol, as well as of GIFCT's Content Incident Protocol (CIP).

Did your company participate in or benefit from the activation of a crisis protocol aimed at reducing the volume and impact of terrorist and/or violent extremist content during the reporting period?

No such crisis protocol was activated during the reporting period.

If you would like to add other comments or information, please do so here.

We accept emergency law enforcement requests. Please see this page for further information: https://discord.com/safety/360044157931-Working-with-law-enforcement#Emergency-Requests