

Twitter's India Transparency Report: User Grievances & Proactive Monitoring March 2022



Overview

Twitter was founded on a commitment to transparency. This commitment is part of our effort to serve the public conversation and to increase its collective health, openness, and civility around the world.

Twitter is reflective of real conversations happening in the world and that sometimes includes perspectives that may be offensive, controversial, and/or bigoted to others. While we welcome everyone to express themselves on our platform, we do not tolerate behavior that harasses, threatens, dehumanizes, or uses fear to silence the voices of others. We have the [Twitter Rules](#) in place to help ensure everyone feels safe expressing their beliefs and we strive to enforce them with uniform consistency. We are continually working to update, refine, and improve both our enforcement and our policies, informed by in-depth research around trends in online behavior both on and off Twitter, feedback from the people who use Twitter, and input from a number of external entities, including members of our Trust & Safety Council.¹

When it comes to enforcing these rules, we have a global team that manages this with 24/7 coverage in every language that Twitter supports and endeavors to take into account regional and cultural context and complexities of the various markets. Twitter is committed to being fair, informative, responsive, and accountable and our goal is to apply the [Twitter Rules](#) objectively and consistently. Enforcement actions are taken on content that is determined to violate the [Twitter Rules](#). Read more about our approach to policy development and enforcement philosophy in the [Twitter Help Center](#).²

In compliance with Rule 4(1)(d) of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, Twitter is required to publish a monthly compliance report which includes the details of complaints from users via the India grievance mechanism and action taken on them, as well as information related to Twitter's proactive monitoring efforts under the IT Rules. This report covers the reporting period from January 26, 2022 to February 25, 2022. We will continue to publish this report on a monthly basis, and will make improvements over time, based on feedback received from the government, or in accordance with internal changes that allow for us to provide more granular data.

As part of our work to increase the collective health of the platform, Twitter continues to launch initiatives, projects, and updates. Here are some recent highlights:

- [Replies people do not want to see](#)³: Earlier in July 2021, we rolled out a test where some users on iOS may have seen different options to up or down vote on replies. Through this test, we learned a lot about the types of replies people don't find relevant and this month, we

¹ https://blog.twitter.com/en_us/topics/company/2019/rules-refresh.html

² <https://help.twitter.com/de/rules-and-policies/enforcement-philosophy>

³ <https://twitter.com/TwitterSupport/status/1489351648245956613>

expanded this test, so more people on web and soon iOS and Android will have the option to use reply downvoting. Downvotes aren't public.

- [Reply directly with a DM icon on Tweets](#)⁴: We understand that sometimes replying publicly can be daunting. Therefore, we have started testing a DM icon on Tweets with select people on iOS. For when people want to reply directly to a Tweet's author, we're making it easier to DM them from your timeline to start a conversation. However, people can limit who can DM them.
- [Sensitive media Tweet warnings](#)⁵: Everyone across Android, iOS, and web now has the option to add one-time sensitive content warnings to photos and videos they Tweet out, if it contains sensitive content.

How Twitter Processes Grievances from Users in India

In India, Twitter users can report grievances via the Grievance mechanism by using the contact details available on the [Grievance Officer - India](#) page. Twitter does not require such reporters to be registered with Twitter, or have a Twitter user ID in order to file a grievance. Separately, global users can report directly from the Tweet or account in question while logged into Twitter, or reports can be made via [Twitter's Help Center](#).

Each user complaint we receive via the India Grievance channel is assessed under Twitter's Terms of Service ([TOS](#)) and the [Twitter Rules](#) and any content that is determined to be in violation is actioned in line with our [range of enforcement options](#). Where appropriate, we also assess whether the grievance meets local legal requirements for Twitter to action the reported content, based on a court order. Users generally have the option to appeal an enforcement action via our [appeals process](#). NOTE: Law enforcement requests should, ideally, be submitted at Twitter's Legal Request Submission site, available at t.co/lr or legalrequests.twitter.com, to facilitate efficient processing.

To process these grievances, we have a dedicated team of trained individuals. These individuals take appropriate action after carefully reviewing the grievance, and any available context, in close detail. All of these team members are fluent in one or more of the languages spoken in India and participate in regular training on our tools and policies. Updates about significant current events or internal rules and policy changes are shared with all agents in real time, to give guidance and facilitate balanced and informed decision-making.

Grievance Officer - India Data

Below is the data we received via our Grievance Officer - India channel between **January 26, 2022 through to February 25, 2022** that included content on Twitter. This includes complaints received from individual users with accompanying court orders.

NOTE: We receive complaints in our Grievance Officer - India channel that relate to account verification, account access, or seeking assistance or information regarding an account or Twitter's

⁴ <https://twitter.com/TwitterSupport/status/1489328529133211651>

⁵ <https://twitter.com/TwitterSupport/status/1497312745557049345>

enforcement actions that are not included in the data below. The majority of complaints received in this channel during this reporting period fell into these categories.

NOTE: The data in these reports is as accurate and comprehensive as possible at the time of reporting.

| Issue Type | Total Number of Grievances | Total Number of URLs Actioned |
|--|----------------------------|-------------------------------|
| Abuse / Harassment | 161 | 471 |
| Child Sexual Exploitation | 0 | 0 |
| Defamation | 11 | 2 |
| Hateful Conduct | 18 | 42 |
| Illegal Activities ⁶ | 0 | 0 |
| Impersonation | 40 | 15 |
| IP-related Infringement ⁷ | 7 | 35 |
| Misinformation / Synthetic and Manipulated Media | 14 | 23 |
| Promoting Suicide or Self-Harm | 0 | 0 |
| Terrorism / Violent Extremism | 1 | 0 |
| Sensitive Adult Content | 5 | 50 |
| Privacy Infringement | 1 | 0 |

In addition to the above, we processed **80** grievances which were appealing Twitter account suspensions. These were all resolved and the appropriate responses were sent. We overturned **6** of the account suspensions based on the specifics of the situation, but the other accounts remain suspended. We received **11** requests related to general questions about Twitter accounts during this reporting period.

Twitter’s Proactive Monitoring Efforts

There are many measures that have been in place for a long time on Twitter that relate to the mitigation of child sexual exploitation and terrorism activity. Twitter does not tolerate any material that features or promotes [child sexual exploitation](#) — whether in Direct Messages or elsewhere throughout the service. This includes media, text, illustrations, or computer-generated images. When we remove content, we immediately report it to the National Center for Missing and Exploited Children (NCMEC). NCMEC makes reports available to the appropriate law enforcement agencies around the world to facilitate investigations and prosecutions.

⁶ This includes requests related to money laundering, gambling, hacking, doxxing, or similar issues.

⁷ This includes Copyright, Trademark, or other IP-related issues.

Twitter was one of the [founding members](#) of, and continues to participate in, the [Global Internet Forum to Counter Terrorism](#).

A vast majority of all accounts that are suspended for the [promotion of terrorism](#) and [child sexual exploitation](#) are proactively flagged by a combination of technology and other purpose-built internal proprietary tools. You can learn more about our commitment to eradicating child sexual exploitation and terrorist content and the actions we've taken [here](#).

Proactive Monitoring Data⁸

| Issue | Total Accounts Suspended |
|---|--------------------------|
| Child Sexual Exploitation, Non-Consensual Nudity, and similar content | 30,992 |
| Promotion of Terrorism | 3,924 |

Footnotes

- “**Proactive Monitoring**” refers to content proactively identified by employing internal proprietary tools and industry hash sharing initiatives.
- Each grievance may identify multiple items. For example, a single grievance may ask us to remove individual Tweets, an entire account, or both.
- We may not remove content or all aspects of a grievance for a variety of reasons. For example, the grievance may fail to identify content on Twitter, the content may have been removed by the user that posted it, or we find no violation of our Terms of Service ([TOS](#)) or the [Twitter Rules](#). Our analysis, and any reasons for not taking action, are specific to each grievance.
- Data on government requests can be found in our existing [Twitter Transparency Center](#).
- We have removed all duplicate requests from the data provided in this report, so that only unique requests are counted.

⁸ This data represents global actions taken, and not just actions related to content from India