Meta

California Assembly Bill No. 587-22677

Terms of Service Report

December 29, 2023

Current Version of the Terms of Service

Facebook's Terms of Service are found [here](#). Instagram's Terms of Use are found [here](#).

In addition to our Terms of Service and Terms of Use, we maintain a set of globally applicable [Community Standards](#) and [Community Guidelines](#) that define what is and isn't allowed on Facebook and Instagram.

Content Moderation Practices

Meta builds technologies that help people connect, find communities, and grow businesses. Facebook and Instagram help users connect with friends, family, and communities of people who share their interests. In doing so, Meta is committed to giving people a voice and keeping them safe.

Our approach to content moderation is detailed in the [Meta Transparency Center](#), where you can find more information on the following topics:

Policies and Tools:

We maintain a set of globally applicable [Community Standards](#) and [Community Guidelines](#) that define what is and isn't allowed on Facebook and Instagram. In addition, we have [Advertising Standards](#) in place for advertising and [Commerce Policies](#) applicable to the sale of products on Facebook, Instagram, and WhatsApp. We collaborate with global experts in technology, public safety, and human rights in developing these policies. We also build and enhance features for safety, so people can, for example, report content and block, hide, or unfollow accounts.

The following policies may address, in whole or in part, hate speech or racism, extremism or radicalization, disinformation or misinformation, harassment, foreign political interference, and controlled substance distribution:

a) [Hate Speech Community Standard](#) and [Advertising Standard](#); [Misleading, Violent, or Hateful Commerce Policy](#) (hate speech or racism)

b) [Dangerous Organizations and Individuals Community Standard](#) and [Advertising Standard](#) (extremism or radicalization)

c) [Misinformation Community Standard](#) and [Advertising Standard](#) (disinformation or misinformation)

d) [Bullying and Harassment Community Standard](#) and [Advertising Standard](#) (harassment)

e) [Inauthentic Behavior Community Standard](#); [Social Issues, Elections, or Politics Advertising Standard](#) (foreign political interference)

f) [Restricted Goods and Services Community Standard](#); [Unsafe Substances](#), [Prescription Drugs](#), [Online Pharmacies](#), and [CBD and Related Products](#) Advertising Standards; [Prescription Products, Drugs, and Drug Paraphernalia Commerce Policy](#) (controlled substance distribution)

We provide the relevant terms and policies in languages in addition to English if we offer products in those languages.

Enforcement:

Meta uses technology and human review teams to detect, review, and take action on millions of pieces of content and accounts every day on Facebook and Instagram. In most cases, this happens automatically, with technology to detect, restrict, and remove content and accounts that may go against our policies. In other cases, our technology selects content for human review. Our teams review a blend of user reports and content surfaced by our technology.

Removals:

We remove harmful content that goes against our policies, reduce the distribution of problematic content that doesn't violate our policies, and inform people with additional context so they can decide what to click, read, or share. We also remove accounts that are used to egregiously or persistently violate our policies across our products, including those that compromise the security of other accounts and our services.

Automation and Human Review:

Most removals of content and accounts happen automatically, with technology working behind the scenes to remove violating content—often before anyone sees it. Other times, our technology will detect potentially violating content but send it to review teams to check and take action on it.

Our technology also supports the review teams by prioritizing the most critical content to be reviewed, based on severity, virality, and likelihood of a violation. In order to prevent as much harm as possible, our review systems use technology to prioritize high-severity content with the potential for offline harm (e.g., posts related to terrorism and suicide) and viral content that is spreading quickly and has the potential to reach a large audience.

These technologies run on accounts, posts, comments, photos, and other pieces of content uploaded to Facebook and Instagram. They determine how probable or likely it is that this content violates a certain policy, based on those signals or patterns, and if the content should be automatically removed.

Additionally, our human reviewers are provided with various tools and resources when undertaking content review. For example, human reviewers receive in-depth training and often specialize in certain policy areas.

User Reports:

When an individual reports a piece of content (e.g., a photo, comment, post) to us, we'll take action to remove the content if we find that it goes against our policies. Where we've reviewed the content, we'll let the reporting person know whether it did or did not go against our policies. At that time, if the reporting person disagrees with our decision, we generally offer an opportunity to request another review.

Metrics

Below, we provide metrics related to actions, automation volumes, appeals, and appeal overturns for the United States for the time period between July 1, 2023 and September 30, 2023, to the extent such data was available and complete for that time period. These metrics are categorized by type of violation of our policies.

Actions

"Actions" refers to the amount of content removed.

Policy	Actions
Bullying and Harassment Community Standard and Advertising Standard	3.7 M
Dangerous Organizations and Individuals Community Standard and Advertising Standard	682.2 K
Hate Speech Community Standard and Advertising Standard	2.1 M
Restricted Goods and Services Community Standard: Drugs	1.8 M
Misinformation Community Standard and Advertising Standard	49.0 K
Social Issues, Elections, or Politics Advertising Standard	221.8 K

Actioned Automation Volume

"Actioned automation volume" refers to pieces of content removed by the use of automation.

Policy	Actioned automation volume
Bullying and Harassment Community Standard	3.5 M
Dangerous Organizations and Individuals Community Standard	651.6 K
Hate Speech Community Standard	2.0 M
Restricted Goods and Services Community Standard: Drugs	1.7 M
Misinformation Community Standard	13.0 K

Appeals Volume

"Appeals volume" refers to the number of appeals received via appeals-handling systems.

Policy	Appeals volume
Bullying and Harassment Community Standard and Advertising Standard	742.3 K
Dangerous Organizations and Individuals Community Standard and Advertising Standard	31.0 K
Hate Speech Community Standard and Advertising Standard	344.4 K
Restricted Goods and Services Community Standard: Drugs	259.7 K
Misinformation Community Standard and Advertising Standard	2.4 K
Social Issues, Elections, or Politics Advertising Standard	44.3 K

Restored Content After Appeal

"Restored content after appeal" refers to the number of instances where content was reinstated following an appeal.

Policy	Restored content after appeal
Bullying and Harassment Community Standard and Advertising Standard	79.0 K
Dangerous Organizations and Individuals Community Standard and Advertising Standard	6.7 K
Hate Speech Community Standard and Advertising Standard	38.8 K
Restricted Goods and Services Community Standard: Drugs	37.1 K
Misinformation Community Standard and Advertising Standard	0.8 K
Social Issues, Elections, or Politics Advertising Standard	21.2 K

Additional information on actioned content and appeals is available in Meta's Q3 2023 Community Standards Enforcement Report.

Actions Against Users, Accounts, or Other Entities

"Actions against users, accounts, or other entities" refers to the removal of a user, account, or other on-platform entity (such as a Page or Group) from Facebook or Instagram. "Appeals volume" refers to the number of appeals received through our appeals-handling system, and "restored after appeal" refers to users, accounts, or other entities who were restored after appeal.

Users	Actions	Appeals volume	Restored after appeal
Facebook Users	133.8 M	23.8 M	1.8 M
Instagram Users	8.1 M	2.5 M	1.3 M
Facebook Pages	140.1 K	0.7 K	0.3 K
Facebook Groups	43.6 K	1.6 K	0.6 K
Advertising Accounts	4.1 M	183.4 K	71.4 K

Additional information regarding actions against users, accounts, and other entities is available in Meta's [Adversarial Threat Report, for third quarter 2023](#) (regarding enforcement of the Inauthentic Behavior Community Standard).