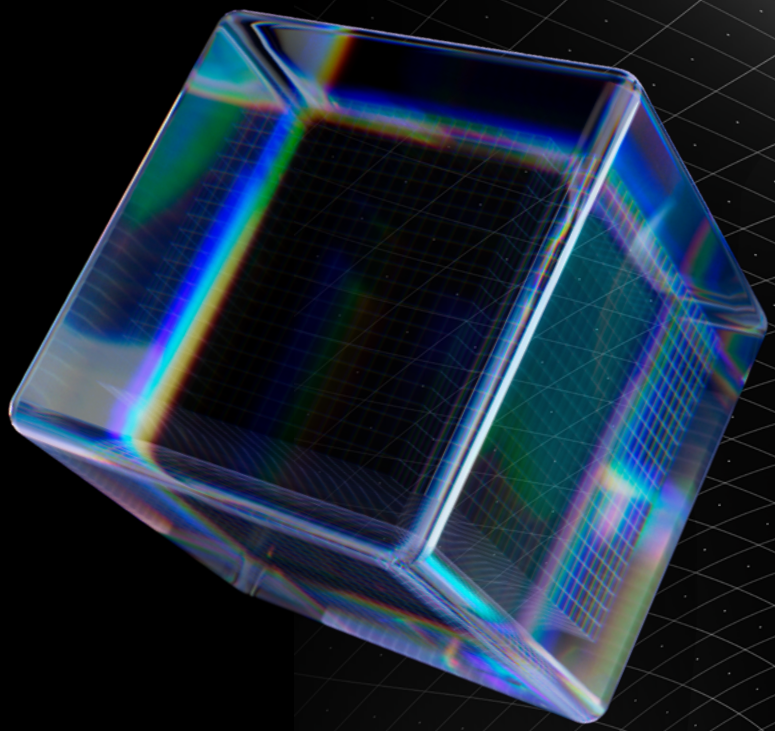
ROBLOX



California AB 587
Transparency Report

Q3 2023

Roblox is an immersive platform for connection and communication. Every day, millions of people come to Roblox to create, play, work, learn, and connect with each other in experiences built by our global community of creators.

Our vision is to reimagine the way people come together – in a world that's safe, civil, and optimistic. To achieve this vision, we are building an innovative company that, together with the Roblox community, has the ability to strengthen our social fabric and support economic growth for people around the world.

Transparency reports demonstrate our deep commitment to keeping all Roblox users safe.

Roblox Terms and Community Standards

Roblox's Community Standards, which are incorporated into our Terms of Use, govern user conduct on the platform. Roblox makes the Terms of Use available in all languages where we offer product features.

The current version of the terms of service (last updated on September 13, 2023) can be found here. Our full community standards can be found here. The specific Community Standards for the violation categories included in this report can be found in the appendix to this report.

Content Moderation Practices

Automated and Human Content Moderation

A combination of automated and human content moderation systems play a crucial role in enforcing our terms of service by proactively identifying and addressing content that violates our policies. Content submitted by creators, including assets and communication such as text and voice, undergoes a comprehensive review process before it is published to the site.

Content not flagged for removal by automated systems are subject to human review, both proactively and when reported for policy violations. Our automated systems also prevent the re-upload of content that we've previously removed or previously rejected, ensuring that content we've taken down isn't reintroduced to the platform.

Responding to User Reports

Roblox has a robust reporting process in place to address user reports of Terms of Use and Community Standards violations. Users have the ability to easily mute or block other community members and report inappropriate content or behavior on the Roblox application and our website. Roblox also provides users with a direct channel to our Customer Support team to report concerns or other issues.

Enforcement Actions

Once content has been flagged, our moderation systems assess the validity of the flag, and take action in accordance with our Terms of Use and Community Standards. The actions taken can vary based on the severity and impact of the violation, and they can include:

- Warnings
- Removing content
- Roblox account-level and feature-level restrictions
- Reporting users to relevant authorities in cases presenting an imminent risk of harm

The duration of these enforcement actions and the type of action taken can be manual, automated, or a combination of both. In addition to considering the specific violation, we also take into account a user's historical use of the platform and whether they have repeatedly violated our policies. Repeated violations of our policies may increase the severity of the enforcement actions.

Whenever we make a decision to restrict access to content on the platform, we notify users of our decision and provide them with an opportunity to appeal by contacting the Roblox Appeals team. Moderators holistically evaluate appeal requests, considering the severity of the violation, the user's reason for appealing, and their behavior on the platform.

Users also have the option to discontinue their use of the service at any time by requesting an account deletion, which can be done in-app or via the Support Form, utilizing the Right to Be Forgotten request feature.

Q3 2023 Data

Roblox provides the information below in response to the CA Assembly Bill 587 in order to comply with the requirements of a terms of service report. The following data covers the time period from July 1, 2023 through September 30th, 2023, referred to as the reporting period.

To provide greater context on the charts below, in the reporting period, Roblox users generated and uploaded approximately 282 Billion (281,946,911,600) total pieces of content to Roblox's platform, of which a significant amount is text chat, as well as other content types including audio, voice, and images. Every piece of content generated and uploaded is reviewed by our content moderation tools. Thus, this number represents the total pieces of content that potentially could have been flagged via our detection and reporting systems. There were 15,233,444 total flagged items of content (0.0054% of total).

The following tables detail the actual amount of flagged content on our platform, during the reporting period, broken down by the type of policy that the content has been actioned under. Per legislative requirements, we report on entity and content type. In providing this report, we have endeavored to present the required information in a way that avoids unnecessary repetition and most accurately reflects Roblox's content moderation practices.

1. The total number of flagged items of content.

1.1: Category of content

VIOLATION CATEGORY	TOTAL FLAGGED
Threats, Bullying, and Harassment	9,846,952
Discrimination, Slurs, and Hate Speech	5,332,666
Terrorism and Violent Extremism	53,826
GRAND TOTAL	**15,233,444**

1.2: Type of content and type of media of the content

ENTITY TYPE	CONTENT TYPE	TOTAL FLAGGED
Account	Chat	13,565,015
	Voice	1,154,919
	Other	98,425
	Account	201
ACCOUNT TOTAL		**14,818,560**
Asset	Image	358,023
	Audio	19,145
	Asset	13,969
	Model	673
	Mesh	158
	3D Accessory	29
ASSET TOTAL		**391,997**
Group	Group	17,370
GROUP TOTAL		**17,370**
Place	Place	5,517
PLACE TOTAL		**5,517**
GRAND TOTAL		**15,233,444**

1.3: How the content was flagged

GRANULAR IDENTIFICATION SOURCE	TOTAL FLAGGED
Moderator Review	8,086,044
Abuse Report	6,751,968
Machine Detection	394,997
Customer Support	435
GRAND TOTAL	**15,233,444**

2. The total number of actioned items of content.

There were approximately 15 million (15,233,444) total pieces of content flagged and available to be actioned on in the reporting period. The following table details the amount of actioned content on our platform during the reporting period, broken down by the type of policy that the content has been actioned under.

2.1: Category of content

VIOLATION CATEGORY	TOTAL ACTIONS
Threats, Bullying, and Harassment	9,846,952
Discrimination, Slurs, and Hate Speech	5,332,666
Terrorism and Violent Extremism	53,826
GRAND TOTAL	**15,233,444**

2.2: Type of content and type of media of the content

ENTITY TYPE	CONTENT TYPE	TOTAL ACTIONS
Account	Chat	13,565,015
	Voice	1,154,919
	Other	98,425
	Account	201
ACCOUNT TOTAL		**14,818,560**
Asset	Image	358,023
	Audio	19,145
	Asset	13,969
	Model	673
	Mesh	158
	3D Accessory	29
ASSET TOTAL		**391,997**
Group	Group	17,370
GROUP TOTAL		**17,370**
Place	Place	5,517
PLACE TOTAL		**5,517**
GRAND TOTAL		**15,233,444**

2.3: How the content was actioned

GRANULAR IDENTIFICATION SOURCE	TOTAL ACTIONS
Moderator Review	8,086,044
Abuse Report	6,751,968
Machine Detection	394,997
Customer Support	435
GRAND TOTAL	**15,233,444**

3. The total number of actioned items of content that resulted in action against the user or group of users responsible for the content.

There were 15,233,415 actioned items that resulted in 15,037,520 enforcements. There is a discrepancy between total actions and actions with enforcements because content may be actioned for violations against multiple violation categories; however, it will only receive an enforcement for the most severe violation category.

Whenever content on Roblox is actioned on, it is removed from the platform. For some media types where content is ephemeral in nature (e.g. chat, voice), the content cannot be removed as it is no longer available.

The following table details the number of actioned items of content on our platform that resulted in user enforcement during the reporting period, broken down by the type of policy that the content has been actioned under.

TOTAL ACTIONS	ACTIONS WITH ENFORCEMENT
15,233,444	**15,037,520**

3.1: Category of content

VIOLATION CATEGORY	TOTAL ACTIONS	ACTIONS WITH ENFORCEMENT
Threats, Bullying, and Harassment	9,846,952	9,840,231
Discrimination, Slurs, and Hate Speech	5,332,666	5,166,542
Terrorism and Violent Extremism	53,826	30,747
GRAND TOTAL	**15,233,444**	**15,037,520**

3.2: Type of content and type of media of the content

ENTITY TYPE	CONTENT TYPE	TOTAL ACTIONS	ACTIONS WITH ENFORCEMENT
Account	Chat	13,565,015	13,563,640
	Voice	1,154,919	1,154,886
	Other	98,425	98,425
	Account	201	201
ACCOUNT TOTAL		**14,818,560**	**14,817,152**
Asset	Image	358,023	172,003
	Audio	19,145	11,476
	Asset	13,969	13,969
	Model	673	669
	Mesh	158	0
	3D Accessory	29	0
ASSET TOTAL		**391,997**	**198,117**
Group	Group	17,370	17,370
GROUP TOTAL		**17,370**	**17,370**
Place	Place	5,517	4,881
PLACE TOTAL		**5,517**	**4,881**
GRAND TOTAL		**15,233,444**	**15,037,520**

4. The total number of actioned items of content that were removed, demonetized, or deprioritized.

There were 513,510 actioned items of content that were removed, demonetized, or deprioritized. Whenever content on Roblox is actioned on, it is removed from the platform. For some media types where content is ephemeral in nature (e.g. chat, voice), the content cannot be removed as it is no longer available.

4.1: Category of content

VIOLATION CATEGORY	TOTAL CONTENT REMOVED
Threats, Bullying, and Harassment	92,973
Discrimination, Slurs, and Hate Speech	372,100
Terrorism and Violent Extremism	48,437
GRAND TOTAL	**513,510**

4.2: Type of content and type of media of the content

ENTITY TYPE	CONTENT TYPE	TOTAL CONTENT REMOVED
Account	Other	98,425
	Account	201
ACCOUNT TOTAL		**98,626**
Asset	Image	358,023
	Audio	19,145
	Asset	13,969
	Model	673
	Mesh	158
	3D Accessory	29
ASSET TOTAL		**391,997**
Group	Group	17,370
GROUP TOTAL		**17,370**
Place	Place	5,517
PLACE TOTAL		**5,517**
GRAND TOTAL		**513,510**

5. The number of times actioned items of content were viewed by users.

During the reporting period, Roblox averaged 70 million daily active users (DAU) and those users spent 16 billion hours engaged with over 5 million unique experiences created by our community. And, between assets that comprise those experiences and communication between users, there were 282 billion pieces of content viewed during the reporting period. We estimate that during the reporting period, there were 13 million unique views of content that was actioned.

Total user views on moderated content was primarily driven by chats. As we moderate chat messages at the time of upload to the platform, the user creating the chat may be the only viewer of the violating content.

The following table details the number of times the actioned content was viewed by users on our platform during the reporting period, broken down by the type of policy that the content has been actioned under.

5.1: Category of content

VIOLATION CATEGORY	TOTAL USER VIEWS
Threats, Bullying, and Harassment	8,817,546
Discrimination, Slurs, and Hate Speech	4,713,884
Terrorism and Violent Extremism	12,395
GRAND TOTAL	**13,543,825**

5.2: Type of content and type of media of the content

ENTITY TYPE	CONTENT TYPE	TOTAL USER VIEWS
Account	Chat	12,330,147
	Voice	1,042,934
	Username/Profile	68,416
ACCOUNT TOTAL		**13,441,497**
Place	Place	87,157
PLACE TOTAL		**87,157**
Group	Group	15,171
GROUP TOTAL		**15,171**
GRAND TOTAL		**13,543,825**

6. The number of times actioned items of content were shared, and the number of users that viewed the content before it was actioned.

Roblox does not currently offer the ability for users to share/re-share content via the platform as of Q3 2023.

7. The number of times users appealed actions taken on the platform and the number of reversals of actions on appeal disaggregated by each type of action.

During the reporting period, there were 15,233,444 moderation actions on individual pieces of content that were available to be appealed. The total appeals were 533,050. The following table details the number of appeals and reversals on our platform during the reporting period, broken down by the type of policy that the content has been actioned under.

TOTAL APPEALS	TOTAL APPROVED APPEALS	% SUCCESSFUL APPEALS
533,050	**29,348**	**5.51%**

7.1: Category of content

VIOLATION CATEGORY	TOTAL APPEALS	TOTAL APPROVED APPEALS
Threats, Bullying, and Harassment	346,024	19,589
Discrimination, Slurs, and Hate Speech	183,348	9,605
Terrorism and Violent Extremism	3,678	154
GRAND TOTAL	**533,050**	**29,348**

7.2: Type of content and type of media of the content

ENTITY TYPE	CONTENT TYPE	TOTAL APPEALS	TOTAL APPROVED APPEALS
Account	Chat	397,553	9,190
	Voice	129,425	19,398
	Username/Profile	1,596	550
	Unclassified	559	7
ACCOUNT TOTAL		**529,133**	**29,145**
Asset	Image	2,884	169
	Mesh	165	3
	Audio	76	1
	Unclassified	57	0
	Video	9	8
ASSET TOTAL		**3,191**	**181**
Group	Group	261	4
GROUP TOTAL		**261**	**4**
Unclassified	Unclassified	36	3
UNCLASSIFIED TOTAL		**36**	**3**
Other	Other	10	1
OTHER TOTAL		**10**	**1**
Place	Place	419	14
PLACE TOTAL		**419**	**14**
GRAND TOTAL		**533,050**	**29,348**

Appendix

Appendix 1.1: Current Version of our Community Standards

The current version of our Community Standards as of November 2023 describes certain categories of violations as follows:

Harassment

Roblox's "Threats, Bullying, and Harassment" Community Standard reads as follows:

"Threatening others with real-world or online harm, inciting violence against people or property, bullying, stalking, trolling, harassment, intimidation, extortion, and blackmail are not permitted on Roblox. We also do not allow any content that depicts, glorifies, or promotes such behavior, including:

- Threatening physical or sexual assault or violence
- Threatening to harm someone in real life
- Revealing or threatening to reveal others' personal information
- Threatening to take over another's Roblox account or to file false abuse reports against another user
- Singling out a user or group for ridicule or abuse, either publicly or privately
- Sexual harassment
- Impersonating individuals, groups, or entities, in ways that may damage their reputation or cause others to harm them, either online or in real life"

Extremism or Radicalization

Roblox's "Terrorism and Violent Extremism" Community Standard reads as follows:

"Roblox has a zero tolerance policy for content or behavior that incites, condones, supports, glorifies, or promotes any terrorist or extremist organization or individual (foreign or domestic), or their ideology, or actions, including:

- Depictions of or support for terrorist or extremist attacks
- Depictions of or support for the leaders or representatives of terrorist or extremist organizations
- Sharing the slogans, images, flags, manifestos, or icons of terrorist or extremist organizations, either in whole or in readily identifiable part
- References to the ideologies, messages, or strategies of terrorist and extremist organizations
- Expressing support, condoning, or glorifying terrorist extremist ideologies or actions
- Recruiting membership for a terrorist or extremist organization, or encouraging others to leave Roblox to find such information
- Fundraising for terrorist or extremist organizations, people, or supporting groups
- Expressing support, condoning, or glorifying mass shootings and other acts of domestic terrorism or violent extremism"

Hate Speech or Racism

Roblox's "Discrimination, Slurs, and Hate Speech" Community Standard reads as follows:

"Roblox honors and welcomes users of all ages, backgrounds, and identities. We do not allow content or behavior that supports, glorifies, or promotes hate groups, their ideologies, or actions. You also may not discriminate, mock, or promote hatred against individuals or groups, or encourage others to do so directly or indirectly, on the basis of their:

- Age
- Race, perceived race, or ethnicity
- National origin
- Sexual orientation
- Gender, gender identity, or gender expression
- Religion or religious affiliation or beliefs
- Disability status including diseases, bodily conditions, disfigurement, mobility issues, and mental impairment
- Physical or mental disability status
- Veteran status
- Caste
- Familial status"

Disinformation or Misinformation, and Foreign Political Interference

Roblox's Community Standards do not specifically define "disinformation or misinformation" or "foreign political interference" and Roblox does not flag or moderate content as belonging to those specific categories.

Our Community Standards do broadly prohibit certain political content on the platform. Our "Political Content" Community Standard reads as follows:

"We value friendly debate about issues and topics that matter to Robloxians. However, to maintain a civil and respectful environment, we prohibit the discussion or depiction of certain political content, including:

- Current candidates running for public office, including their slogans, campaign material, rallies, or events
- Political parties, including official party-affiliated organizations
- Specific races for elected office
- Sitting real-world elected officials
- Recent, previously-elected, officials in their official capacity
- Individuals who have previously run for political office in their capacity as candidate
- Desecration of political entity symbols, including flag burning
- Inflammatory content related to real-world border, territorial, or jurisdictional relationships"

Controlled Substance Distribution

Roblox does not specifically flag or moderate content as belonging to the category of controlled substance distribution. However, our Community Standards more broadly prohibit users from discussing or engaging in illegal activities or encouraging others to do so on Roblox. We also broadly prohibit the discussion, depiction, or promotion of some illegal and regulated goods or activities. Our "Illegal and Regulated Goods and Activities" Community Standard reads as follows:

"We do not allow users to discuss or engage in illegal activities or encourage others to do so on Roblox. We also prohibit the discussion, depiction, or promotion of some illegal and regulated goods or activities. This includes:

- Controlled substances such pharmaceutical and recreational drugs, as well as as alcohol, tobacco, vaping, and their associated paraphernalia
- Dietary supplements and enhancers such as weight loss pills and steroids
- Depictions of intoxicated behavior associated with consuming alcohol or drugs
- Purveyors of illegal and regulated substances
- Bomb/explosive and weapon-making instructions or schematics in the real world
- Realistic modern firearms outside of in-experience items
- On-platform contests where Robux or anything else of value is offered as a prize
- Sweepstakes-style games

Except where prohibited by local law or regulation, we allow the portrayal of gambling in experiences. However, no real money, Robux, or in-experience items of value may be exchanged in connection with any gambling activities. We also require that the odds of winning be fair and clearly disclosed to the user prior to playing."

Appendix 1.2: Glossary of Terms

Audio: An audio clip.

Group: For purposes of the Roblox Terms of Use, a "group" exists where creators have joined together and, via a single email address, registered on the services as a single unit for the purpose of releasing an experience or other virtual content through the services. Groups may also be used to connect socially.

Asset: An asset in Roblox refers to any item or object that can be used in a game or experience. Assets can include models, meshes, audio files, images, fonts, videos, and more. These assets can be either objects within the data model or applied as properties to other objects. Overall, assets are essential building blocks that allow creators to enhance their games and experiences with various elements, such as 3D models, sounds, visuals, and more.

Place: A place in Roblox refers to an individual 3D world or environment within a Roblox experience. Each place contains all the components necessary for that specific portion of the experience, including its environment, parts, meshes, scripts, and user interface. Places are the building blocks of Roblox experiences and can be created and managed using Roblox Studio, the all-in-one IDE for Roblox development. A place can be thought of as a level or a specific area within a game or experience. While an experience can consist of multiple places, each experience can only have one starting place that all users load into when they join. From within any place, users can teleport to another place.

Account: A Roblox account, which a user must create in order to access certain elements and functionality of the Services.

Mesh: A mesh is a collection of vertices, edges, and faces that make up a 3D object. Meshes allow for more detailed and complex 3D objects compared to what can be created using the built-in modeling tools in Studio. They can include textures, rigs, and animations, making it possible to create lifelike and customizable objects. Meshes are commonly used for creating characters, accessories, and other detailed objects in Roblox games. They can be imported and used in games to enhance the visual quality and realism of the creations.

Experience: Interactive content published on the Services by developers for the engagement and enjoyment of users.

Content Type: The type of content, including, but not limited to, posts, comments, messages, profiles of users, or groups of users..

Media Type: The type of media of the content, including, but not limited to, text, images, and videos.

Entity Type: The entity with the violation, including but not limited to an account, asset, group, place, etc.

Flagged: Content labeled by Roblox moderators or models as Terrorism and Violent Extremism, Threats, Bullying, and Harassment, or Discrimination, Slurs, and Hate Speech for the purposes of CA AB 587 report.

Moderator Review: Enforcements applied to users when violations are found that were not initially reported in the abuse report, which can involve both machine and human detection methods.

Total User Views on Moderated Content: The total views from users on a moderated asset.