



2025

EU Digital Services Act Transparency Report

Report Date: Feb 27, 2026

Reporting Period: Jan 1, 2025 - Dec 31, 2025

Roblox is an immersive gaming and creation platform where every day, millions of users create, play, learn, and connect with each other through online games and other virtual 3D experiences built by a global community of creators.

Our guiding vision is to create the safest and most civil online community in the world. As our platform evolves and scales, forging a new future for communication and connection, our investment in preventative measures remains fundamental. With each passing year, we implement new strategies and technology to improve the speed and effectiveness of our safety and moderation systems.

In 2025, we released dozens of safety enhancements, and we have more to come. These enhancements include new tools for [age-based communications](#), [Roblox Sentinel](#)—a system designed to proactively identify content and behaviors, including communications, that may violate our policies—and continued improvements to our [voice and text](#) moderation. We've rolled out [new technology](#) designed to detect servers where users are breaking our rules in experiences that are otherwise innocuous, and take them down. We've also [joined a global content rating organization](#) in order to provide parents globally with more clarity and confidence regarding age-appropriate content, and [expanded our experience guidelines](#) to restrict social hangout experiences depicting private spaces, make unrated experiences unplayable, and [increase the age limit for restricted experiences](#) from 17 to 18.

Transparency reports such as this one demonstrate our deep commitment to keeping all Roblox users safe and complying with our obligations under the EU's Digital Services Act ("DSA"). This document provides additional information about our approach to moderation on Roblox.

On our [Transparency Site](#), we have published a DSA XLSX file with moderation data localized to the European Union, in the machine-readable format specified by Annex I to the Commission Implementing Regulation (EU) 2024/2835.

Roblox Terms of Use and Community Standards

Safety on Roblox is at the core of everything we do. This starts with our robust set of [Community Standards](#) that govern what is acceptable content and behavior, and are incorporated into our Terms of Use. If any users are found to be violating these Standards, we take a range of actions, as described further below.

During the reporting period*, our Community Standards were broken down into four main categories and cover the following topics:

1. Safety
 - a. Child Exploitation
 - b. Terrorism and Violent Extremism
 - c. Threats, Bullying, and Harassment
 - d. Suicide, Self Injury, and Harmful Behavior
 - e. Discrimination, Slurs, and Hate Speech
 - f. Harmful Off-Platform Speech or Behavior
2. Civility
 - a. Real-World Sensitive Events
 - b. Violent Content and Gore
 - c. Romantic and Sexual Content
 - d. Illegal and Regulated Goods and Activities
 - e. Profanity
 - f. Political Figures and Entities
3. Integrity
 - a. Cheating and Scams
 - b. Spam
 - c. Intellectual Property Violations
 - d. Advertising
 - e. Roblox Economy
4. Security
 - a. Sharing Personal Information
 - b. Directing Users Off-Platform
 - c. Misusing Roblox Systems

** In February of 2026, Roblox updated its Community Standards and combined "Threats, Bullying, and Harassment" and "Discrimination, Slurs, and Hate Speech" into a single Community Standard called "Harassment and Discrimination."*

Content moderation practices

To efficiently moderate the increasing volume of content produced on Roblox, we are continuously innovating along three dimensions: scale, speed, and quality. A combination of automated and human content moderation systems plays a crucial role in enforcing our policies by proactively identifying and removing violative content.

Machine Learning (ML) can make decisions in milliseconds, repeatedly, consistently, and 24 hours a day. However, we still need, and employ, humans to address less common cases where a deeper, nuanced human judgement is required. Our system combines robust, innovative safety and moderation tools with thousands of human moderators around the world who provide oversight and continuous training of our systems to address new and evolving challenges. When experiences are published or updated on the Roblox Platform, they are evaluated by a suite of AI driven tools that identify problematic language, potential bypasses to our safety systems, and content that falls outside our policies, and a human review team is continuously operating to evaluate flagged content. Additionally, content not flagged for removal by ML can still be subject to human review, including when (1) reported by users, (2) reviewed for quality assurance purposes, or (3) determinations require human judgment to assess context or nuance.

Moderation with advanced AI models

The magnitude of content produced daily on Roblox demands scalable infrastructure, ML models, and purpose-built tools. Underpinning our entire moderation system are large, transformer-based models, with knowledge across various modalities. [Advanced AI models](#) monitor the billions of pieces of content added to Roblox every day and block or remove content that violates our Community Standards.

These models are trained on Roblox-specific language and abbreviations, and are able to consider the context of potential violations that would likely be missed by other, non-Roblox-specific ML models. For visual assets, including avatars and avatar accessories, we use computer vision. One technique involves photographing 3D assets from multiple angles. The system then reviews these photographs to determine what the next step should be. Only items that meet our exacting criteria are cleared for the platform, ensuring a consistently safe experience for all users. For text, we use techniques such as hash matching and keyword lists to identify violating content; this helps ensure that once a piece of content is removed, it won't appear on the platform again.

For voice chat, a feature available for age-checked over 13 users, we use Automatic Speech Recognition to transcribe voice into text, then use an in-house AI model to classify and detect policy-violating language. This is done in real-time, allowing us to take action such as temporarily suspending users violating policies in voice chats on Roblox. Our [voice safety classifier](#) moderates chat across multiple languages - including English, Spanish, German, French, Portuguese and Italian. We've also [open-sourced](#) this model as part of our broader commitment to [sharing safety innovations](#) with the industry.

In addition, we back up these models with thousands of safety professionals because we know AI is just one layer in this process. We also partner closely with parental advocacy and mental health organizations, law enforcement, and government officials on best practices, the latest policies and guidelines, and new research about what parents and users expect from us.

Responding to user reports

Our goal is to catch any problematic content before users see or interact with it. We know no system is perfect, so we also make it easy for users to report anything they may spot through our [“Report Abuse”](#) functionality, as well as our DSA illegal content report form. These reports not only help us remove inappropriate content, but also help us train and strengthen our system.

In addition, we’ve improved our reporting tool so users have the option to capture an entire scene, including avatar and object IDs, and highlight the part they want to report. This additional context helps us proactively identify problematic experiences where users frequently report concerns.

Enforcement actions

Once a policy violation has been determined, we act in accordance with our policies. The actions taken can vary based on the severity and impact of the violation, and they can include:

- Warnings;
- Removing content; or
- Roblox account-level and feature-level restrictions.

In addition to considering the specific violation, we also consider 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 remove content or take action against a user, we notify users of our decision and provide them with an opportunity to appeal. Appeal requests consider the severity of the violation, the user’s reason for appealing, and their behavior on the platform.

DSA obligations

Below, we set out some additional contextual information relating to our DSA obligations and some of the data points included in our comprehensive, machine-readable DSA data XLSX file (located on our [Transparency Site](#)). Further information about how Roblox is complying with its DSA obligations can be found [here](#).

Member states orders

In accordance with Articles 9 and 10 of the DSA, we have established a designated electronic point of contact and streamlined operational workflows to receive and process orders from EU member state judicial and administrative authorities. We maintain a dedicated channel that allows authorities to submit legally binding orders securely. Upon receipt, we review orders to ensure they contain the necessary elements required by the DSA, such as a reasoned justification and precise identification of the content or user, and act expeditiously to either remove the illegal content or provide the requested specific information. **In the reporting period, Roblox did not receive any orders to act against illegal content under DSA Article 9.**

In accordance with the instructions in Annex II to the Commission Implementing Regulation (EU) 2024/2835, we have reported all median time indicators in hours, rounded to the nearest whole number. Thus, a median time of "0" indicates that the median time was under 1 hour.

EU Member State	Statement Category	Number of orders to provide information	Median time to inform the authority of receipt	Median time to give effect to the order
CZ	Risk for public security	1	0	3
DE	Cyber violence	3	0	150
	Illegal or harmful speech	3	0	18
	Protection of minors	16	0	2
	Risk for public security	2	0	17
	Scams and fraud	14	0	1
ES	Protection of minors	1	0	2
FR	Scams and fraud	2	0	47
MT	Scams and fraud	1	0	2
PL	Scams and fraud	1	0	2
Grand total		44	0	2

Illegal content

EU users may report content that they believe violates the law of EU member states using our dedicated [illegal content reporting form](#). We review all illegal content reports against our [Community Standards](#) first. Virtually all content reported via illegal content reports is removed for violations of our Community Standards. **In the reporting period, only 1 illegal content report resulted in a removal under EU local law.**

In accordance with the instructions in Annex II to the Commission Implementing Regulation (EU) 2024/2835, we have reported all median time indicators in hours, rounded to the nearest whole number. Thus, a median time of "0" indicates that the median time was under 1 hour.

Statement Category	Article 16 Notices	Median time
Illegal or harmful speech	48,905	0
Intellectual property infringements	46	0
Protection of minors	77,011	0
Risk for public security	42,417	0
Scams and fraud	24,147	0
Unsafe and prohibited products	81,196	0
Statement category not specified in notice	136,130	0
Grand total	409,852	0

Trusted flaggers

Trusted flaggers [accredited](#) in accordance with the Digital Services Act can submit notices of illegal content via our dedicated illegal content reporting form.

In accordance with the instructions in Annex II to the Commission Implementing Regulation (EU) 2024/2835, we have reported all median time indicators in hours, rounded to the nearest whole number. Thus, a median time of "0" indicates that the median time was under 1 hour.

Statement Category	Article 16 Notices	Median time
Illegal or harmful speech	2	0
Protection of minors	2	26
Risk for public security	6	1
Statement category not specified in notice	52	0
Grand total	62	0

Own-initiative content moderation

In accordance with the DSA, Roblox publishes data on content moderation actions performed at our own initiative in the European Union. All moderation is anchored in Roblox's [Community Standards](#) that set out what is and isn't permitted on the platform. These standards are designed to be extremely robust, prohibit illegal activity and are organized around pillars such as Safety, Civility, Integrity, and Security. We do not proactively identify or remove content based on EU law at our own initiative. Consequently, **all measures taken at our own initiative are for content incompatible with our terms and conditions**, inclusive of Community Standards.

In the context of the DSA and this report, "own-initiative" refers to Roblox's proactive moderation actions not motivated by any external notice or report. Therefore, abuse reports received from users are not included in the data.

DSA Statement Category	Number of measures	Detection with solely automated means	Content removal	Account suspension	Account termination
Consumer Information	112	83	83	29	0
Cyber violence	12,761	10,942	10,942	1,817	2
Data protection and privacy violations	659,420	628,869	628,869	29,496	1,055
Illegal or harmful speech	217,500	157,493	157,493	59,886	121
Intellectual property infringements	8,241	7,124	7,124	912	205
Negative effects on civic discourse or elections	18,436	16,312	16,312	1,755	369
Protection of minors	3,417	2,765	2,765	651	1

DSA Statement Category	Number of measures	Detection with solely automated means	Content removal	Account suspension	Account termination
Risk for public security	11,941	5,986	5,986	5,953	2
Scams and fraud	318,274	170,612	170,612	84,101	63,561
Self harm	34	20	20	12	2
Unsafe and prohibited products	81,325	75,029	75,029	4,694	1,602
Violence	81,892	68,828	68,828	6,417	6,647
Other violation of terms and conditions	1,150,178	820,220	820,220	224,592	105,366
Grand total	2,563,531	1,964,283	1,964,283	420,315	178,933

Appeals

Whenever we make a decision to remove content on the platform, we notify users of our decision and provide them with an opportunity to appeal. Appeal requests consider the severity of the violation, the user's reason for appealing, and their behavior on the platform.

EU users can appeal an initial moderation action within a 6-month period. EU reporters who file DSA illegal content reports also have the right to appeal Roblox's decision on their report.

The median time needed for taking decisions under Art 15(1)(d) was calculated based on the time from receipt of submission to issuance of decision. In accordance with the instructions in Annex II to the Commission Implementing Regulation (EU) 2024/2835, we have reported all median time indicators in hours, rounded to the nearest whole number. Thus, a median time of "0" indicates that the median time was under 1 hour.

Metric		Value
Complaints submitted to the internal-complaints mechanism (i.e., appeals)	Total number	3,130,004
	Decisions upheld	2,713,701
	Decisions partially reversed	25,446
	Decisions reversed	390,857
	Median time	0
Complaint regarding a decision to remove information	Total number	374,241
	Decisions upheld	253,568
	Decisions partially reversed	25,407
	Decisions reversed	95,266
	Median time	0
Complaint regarding a decision to suspend or terminate an account	Total number	2,755,763
	Decisions upheld	2,460,133
	Decisions partially reversed	39
	Decisions reversed	295,591
	Median time	0

Out-of-court dispute settlement

For further disputes related to content moderation, in accordance with Article 21 of the DSA, EU users can also select a certified out-of-court dispute settlement body. The EU Commission has a list of accredited out-of-court dispute settlement bodies for this purpose, which can be found [here](#).

In accordance with the instructions in Annex II to the Commission Implementing Regulation (EU) 2024/2835, we have reported all median time indicators in hours, rounded to the nearest whole number.

Metric	Value	Contextual information
Total number	247	This figure reflects the number of complaints Roblox received from ODS bodies during the reporting period.
Decisions upheld	62	
Decisions partially reversed	1	
Decisions reversed	26	
Median time	1,032	This metric was calculated using a start time of when Roblox received a complaint and an end time of when Roblox received a final decision from the ODS body. Roblox maintains no control over the procedural timelines of the eventual issuance of a final decision by these independent bodies. Consequently, this figure represents the median time for all ODS decisions finalized during the reporting period, including cases initiated in 2024 but where a final resolution was issued in 2025.
Decision omitted	72	This figure is inclusive of decisions where Roblox reversed its moderation action prior to the ODS body rendering a decision, and decisions where the ODS body deemed the complaint inadmissible.
Percentage of outcomes implemented	51.9%	This figure reflects decisions during the reporting period, including decisions for complaints originating in 2024 where a final decision was not rendered until 2025.

Use of automated means for content moderation

We have integrated specific disclosures into both public reporting and user-facing notifications. We explicitly outline in our transparency reports and Community Standards how Roblox utilizes a hybrid model of AI-driven proactive detection and human review to enforce policies, detailing the volume of content removed automatically versus that which requires human intervention. Users are also informed of this when a moderation decision is taken against a user in the EU.

Metric	Value
Number of measures solely taken by automated means	1,813,995
Number of measures not taken by automated means	749,536
Number of notices solely processed by automated means	0
Number of notices not processed by automated means	409,852

Closing remarks

At Roblox, our commitment to safety and civility is an ongoing journey, not a final destination. As this 2025 EU Digital Services Act Transparency Report demonstrates, we are dedicated to maintaining a secure, welcoming, and trusting environment for our global community through rigorous policy enforcement, innovative technological solutions, and transparent reporting. The data and practices detailed in this report highlight our comprehensive, hybrid approach to content moderation—combining the speed and scale of advanced AI models with the nuanced judgment of thousands of human safety professionals.

Looking ahead, we recognize that the digital landscape and the challenges it presents will continue to evolve. In 2026 and beyond, we will continue to invest heavily in our safety infrastructure, refining our automated detection systems, and expanding our partnerships with industry experts, regulators, and safety organizations. We remain steadfast in our commitment to empowering our users, protecting our community, and upholding our obligations under the DSA.

The Roblox Team

Feb 27, 2026