reddit

Reddit's EU Digital Services Act Transparency Report

This report is published by Reddit Netherlands B.V. in accordance with the transparency reporting requirements under Article 15 of the European Union's Digital Services Act ("DSA"). This report contains information for a reporting period from **17 February 2024** to **31 December 2024**.

Published: February 14, 2025

tl;dr

Reddit's mission is to empower communities and make their knowledge accessible to all, which requires ongoing efforts to keep Reddit safe, healthy, and real. As part of this, we regularly assess our platform and share insights and findings about the health of our communities, our efforts to foster safe interactions, and how we enforce our policies.

As a provider of intermediary services and in accordance with Article 15's transparency reporting requirements, we publish this report regarding our content moderation efforts in the EU that are regulated by the DSA. The information contained in this report is for the reporting period from 17 February 2024 through 31 December 2024.

In addition to this report, we share findings about how we enforce our policies in our biannual Transparency Reports, which are available in our [Transparency Center](#). These reports provide global insights and metrics about content that was removed from Reddit and accounts that were suspended, as well as information about legal requests we receive from governments, law enforcement agencies, and third parties.

Table of contents

Section 1: Member State authority requests

Article 15.1(a)

From time to time, Reddit may receive legal requests from courts and government agencies in the EU to remove content or disclose user data. The tables below reflect the number of orders received from Member States' authorities in accordance with DSA Articles 9 and 10. Information about other requests from government authorities around the world are published in Reddit's global, biannual Transparency Reports, which are available in our [Transparency Center](#).

a) Member State removal orders

DSA Article 9 refers to orders to act against illegal content, issued by relevant national judicial or administrative authorities, on the basis of the applicable Union law or national law in compliance with Union law (removal orders). All reported content, including removal orders, are first reviewed against our sitewide [Reddit Rules](#).[1] Reddit Rules serve as a set of principles-based rules that apply to the entire Reddit platform, including our users and content on Reddit. Reddit Rules are applied universally, regardless of how we became aware of the violating content or who submits the report. When there is no underlying rule violation, we evaluate the removal order under local law. Reddit scrutinizes each removal order to determine its legal sufficiency, and may push back, seek more context, or deny the request in whole or in part for any number of reasons, including that the request is overbroad. In cases where Reddit determines that the reported content appears to violate local law, Reddit may [restrict access](#) to (aka "geoblock") the content in the relevant EU country.

During the reporting period, Reddit received zero removal orders from EU Member States' authorities issued in accordance with Article 9.

b) Member State disclosure orders

DSA Article 10 refers to orders to provide specific information about one or more specific individual recipients of the service, issued by the relevant national judicial or administrative authorities on the basis of the applicable Union law or national law in compliance with Union law (disclosure orders). Reddit may disclose specific account information in response to a valid disclosure order when required by law and in certain emergency situations. A specialized Reddit

---

[1] The name of our sitewide rules was recently changed from "Reddit Content Policy" to "Reddit Rules."

team carefully reviews every disclosure order for procedural validity and legal sufficiency, and objects when appropriate.

Table 1.1 - Disclosure orders received from EU Member States' authorities

Reflects the total number of user data disclosure orders received from EU Member States' authorities issued in accordance with Article 10 during the reporting period.

Member State	Disclosure orders
Austria	0
Belgium	0
Bulgaria	0
Croatia	0
Cyprus	0
Czechia	0
Denmark	0
Estonia	0
Finland	0
France	1
Germany	16
Greece	0
Hungary	0
Ireland	0
Italy	0
Latvia	0
Lithuania	0
Luxembourg	0
Malta	0
Netherlands	0
Poland	0
Portugal	0
Romania	0

Slovakia	0
Slovenia	0
Spain	0
Sweden	0

Table 1.2 - User data disclosure orders by investigation type

Reflects the user data disclosure orders received during the reporting period issued in accordance with Article 10 broken out by the type of investigation.

Type of Investigation	Disclosure orders
Child safety (i.e., sexual exploitation of children)	1
Cybercrime (i.e., fraud, doxxing, illegal gambling)	1
Non-consensual intimate media (NCIM)	3
National security related investigations (i.e., terrorism)	0
Non-violent crimes (i.e., burglary, drugs, defamation)	0
Other types of crimes	11
Self-harm	0
Violent crimes (i.e., murder, armed robbery, sexual assault)	1

When Reddit receives a disclosure order from a Member State authority, confirmation of receipt is sent back to the authority. During the reporting period, confirmation of receipt to the Member State authority occurred within 1 day. The median time needed to take action (or non-action) in response to disclosure orders received from Member States' authorities during the reporting period was 3 days.

Section 2: Article 16 notices
Article 15.1(b)

In accordance with Article 16, EU users can notify Reddit of content that is potentially illegal in the EU via the EU Illegal Content Report Form (Article 16 notices). The form is used by individuals claiming legal rights in the EU, as well as individuals or entities designated as trusted flaggers pursuant to the DSA. Trusted flagger notices are prioritized for review in accordance with DSA Article 22. As indicated above, any reported content is first assessed under Reddit

Rules. When there is no underlying rule violation, we evaluate the removal order under local law. In cases where Reddit determines that the reported content appears to violate local law, Reddit may restrict access to (aka "geoblock") the content in the relevant EU country.

Table 2.1 - Article 16 notices

Reflects the number of Article 16 notices, including trusted flagger notices, during the reporting period broken down by the type of alleged illegal content. This data includes notices that we may later deem to be invalid during processing, including spam, invalid URLs, and duplicate notices.

Type of alleged illegal content	User notices	Trusted flagger notices
Copyright infringement	123	0
Counterfeit goods	6	0
Defamation	603	0
Harassment	272	0
Hateful content	992	0
Illegal transactions	363	0
Minor sexual exploitation	1,391	0
Non-consensual intimate media	382	0
Other legal issue	655	0
Private or personal information	463	0
Terrorist content	311	0
Trademark infringement	96	0
Violence	389	1

Table 2.2 - Removal actions in response to Article 16 notices

Reflects the number of posts, comments, and private messages that were removed in response to Article 16 notices broken out by Reddit Rule violation.

Reddit Rule violation	Removed posts, comments, and PMs
Harassment	113
Hate	292

Impersonation	5
Intellectual property policies	56
Minor abuse or sexualization	260
Non-consensual intimate media	659
Personal information	244
Prohibited transactions	246
Violence	315

Note: We review each notice carefully. It's possible that content was reported via an Article 16 notice under one illegal content category and removed under a different Reddit Rule category. Content removals might also result in account-level actions, including warnings, temporary suspensions, and permanent suspensions. Table 3.3 below reflects temporary and permanent suspensions of EU accounts broken out by Reddit Rule violation during the reporting period.

Table 2.3 - Restriction actions in response to Article 16 notices

Reflects the number of content restriction actions in response to Article 16 notices during the reporting period broken down by the type of alleged illegal content. Local law restrictions result in a "geoblock" wherein the content is no longer visible for users in a given jurisdiction.

Type of alleged illegal content	Restriction actions
Defamation	1
Hateful content	23
Other legal issue	14

The median time for processing notices submitted via Reddit's EU illegal content report form is 2 days. Reddit processed zero notices by automated means.

Section 3: Reddit's content moderation

Article 15.1(c)

Reddit has a layered, community-driven approach to content moderation. Reddit Rules are enforced globally by Reddit safety teams. Individual communities on Reddit (known as "subreddits") also have community-specific rules, tailored to their unique purpose and needs.

These rules are set and enforced by the users who create and maintain these communities, known as moderators.

Reddit's volunteer-based, self-moderation efforts at the community level are supported at the sitewide level by Reddit employees (known as "admins"). These employees enforce Reddit's site-wide rules in cases that are beyond the scope of volunteer moderators. When content violates our rules, admins step in to remove the content globally. In addition, depending upon the circumstances, admins may also sanction the user who submitted the content by temporarily or permanently suspending that user's account. When a subreddit repeatedly hosts content that violates Reddit's rules, either through the neglect or encouragement of its moderators, Reddit admins can take actions against the moderators (including suspension of moderator privileges or account suspension) or against the subreddit itself (including removing the whole subreddit from the platform).

Reddit admins use a variety of signals and enforcement tools to identify and remove rule-violating content from the platform, including automated detection, manual review, and user reports.

a) Use of automated tools
Article 15.1(c); Article 15.1(e)

Reddit uses automated measures across the site to help identify potentially rule-violating content. Reddit's automated systems use a number of signals to assess the likelihood of violating content, including heavily downvoted, mod removed, abusive language, etc. We then apply machine learning to help us prioritize the high-likelihood-violating content to ensure that we get to the most actionable content first, along with the content that is most likely to have real-world consequences. This automated process then sends the high-priority content for human review by our teams.

We also use automated tooling to proactively remove content. This includes the use of industry standard automated moderation capabilities, including hashing, to detect and prevent the dissemination of harmful content such as child sexual abuse material, terrorist content, and non-consensual intimate imagery. Reddit also recently launched use of large language models (LLMs) trained in-house to automatically review certain categories of content that have been flagged for review via other automated content moderation systems to make certain policy decisions.

We recognize that it's important to maintain a high degree of accuracy with our use of these automated systems. Prior to deploying new automated tools, we assess the accuracy, latency, and stability of the tools. For example, we typically set up internal testing datasets that we can use for purposes of tool evaluation. Results are assessed by various safety teams (including engineering, operations, and policy teams) before we move forward with a deployment. We also test tools to ensure they meet performance requirements under real-world conditions and can consistently perform over time and under varying loads or external conditions (e.g., short-term traffic spikes or long-term operational demands).

Following deployment, we evaluate the performance of our automated content moderation tools by monitoring proactive detection rates vs. user reports, reversal rate following user appeals, and actionability rates. Specific to automated decisions impacting EU users, we estimate an error rate by calculating the number of automated decisions that were appealed and overturned, divided by the total number of automated decisions. During the reporting period, less than 2% of automated decisions were appealed and subsequently overturned.

b) Humans involved in Reddit's content moderation
Article 15.1(c)

Admins involved in content moderation and enforcement include our Safety team that handles Reddit Rules violations, a Community team that handles moderation violations, and a Trust & Safety Policy team that specializes in legal content removals, including the review and processing of Article 16 notices.

Coverage, Training and Quality
We rely on dedicated, globally distributed admins who provide continuous coverage and support for multiple languages. Admins bring diverse backgrounds and perspectives, giving them a nuanced understanding of cultural context. We recognize that some content requires an even higher level of scrutiny and expertise. For our most complex and sensitive rules, particularly those concerning child safety, violence, privacy, and illegal content removals, we rely on our most experienced admins.

All admins receive an in-depth, multi-week initial training, as well as consistent refreshers and testing to ensure they maintain a thorough understanding of our evolving rules and policies. We regularly monitor trends in admin enforcement to ensure accuracy and provide real-time feedback. Furthermore, we conduct weekly operational proficiency reviews of our performance metrics to identify areas of improvement and ensure the highest standards of moderation.

Support and Wellness

The health and wellbeing of our content moderation teams is a company priority. Every admin has access to confidential counseling sessions with psychologists specializing in the unique challenges of trust and safety work, as well as dedicated mental health resources. Furthermore, we foster a supportive environment through:

- **Regular check-ins:** admins have a dedicated space to discuss their experiences, share concerns, and receive support from their colleagues and supervisors.
- **Wellness days:** designed to encourage team members to disconnect from work, recharge, and prioritize their personal wellbeing.
- **Industry-focused discussions and presentations:** led by experts in the trust and safety field, these sessions provide valuable insights into industry best practices, emerging trends, and strategies for managing the emotional impact of content moderation. They also offer a platform for admins to connect with peers and share their experiences.
- **Non-work activities and resources:** we encourage our admins to engage in non-work activities and offer flexible benefits that can be used for a wide range of purposes, including gym memberships, mindfulness apps, and classes.

By supporting our admins in these ways, we empower them to make the best possible decisions, ensuring a safer and more positive experience for our community.

c) Removals and account actions imposed by Reddit
Article 15.1(c)

Tables 3.1, 3.2, and 3.3 below provide an overview of the number and type of measures taken that affect the availability, visibility, and accessibility of content posted by or sent from EU users and actions against EU-based Reddit accounts.

reddit

Table 3.1 - Removals of posts, comments, and private messages from EU users

Reflects admin-removed posts, comments, and private messages, broken out by Reddit Rule violation and the number of posts, comments, and private messages that were detected by automation or detected by a user (i.e., user reports). Content "detected by automation" may include content that was also reported by a user.

Reddit Rule violation	Admin-removed posts, comments, and PMs	Admin-removed posts, comments, and PMs detected by automation	Admin-removed posts, comments, and PMs detected by users
Harassment	57,768	56,665	1,103
Hate	43,068	28,807	14,261
Impersonation	23	11	12
Intellectual property policies	77,209	0	77,209
Minor abuse or sexualization	142,847	108,821	34,026
Non-consensual intimate media	359,098	357,153	1,945
Personal information	2,489	1,865	624
Prohibited transactions	5,367	5,239	128
Violence	32,592	30,152	2,440

Table 3.2 - Removals of chat messages from EU users

Reflects admin-removed chat messages broken out by Reddit Rule violation and the number of chat messages that were detected by automation or detected by a user (i.e., user reports). Content "detected by automation" may include content that was also reported by a user.

Reddit Rule violation	Admin-removed chat messages	Admin-removed chat messages detected by automation	Admin-removed chat messages detected by users
Harassment	28,138	13,061	15,077
Hate	9,271	7,922	1,349
Impersonation	79	1	78

Intellectual property policies	0	0	0
Minor abuse or sexualization	106,559	99,990	6,569
Non-consensual intimate media	57,660	54,865	2,795
Personal information	158	108	50
Prohibited transactions	1,343	625	718
Violence	4,176	3,343	833

Table 3.3 - EU account-level actions

Reflects temporary and permanent suspensions of EU accounts broken out by Reddit Rule violation.

Reddit Rule violation	Temporary suspensions	Permanent suspensions
Harassment	24,484	3,213
Hate	20,311	4,576
Impersonation	1	86
Intellectual property policies	0	415
Minor abuse or sexualization	4,564	90,276
Non-consensual intimate media	8,757	138,732
Personal information	361	589
Prohibited transactions	2,040	255
Violence	13,391	3,203

Note: Multiple temporary suspensions can be applied to a single account over time.

Section 4: Complaints (i.e., appeals) received through internal complaint handling systems

Article 15.1(d)

When we remove a piece of content for violating Reddit's rules, or take an associated account-level enforcement action, the account that posted the content is notified of the removal reason and provided instructions for how to appeal. In our notifications, we aim to educate users about how and why they violated our rules, to foster and encourage positive contributions to the platform going forward. Users may submit an appeal within six months of receiving the

notification. We process appeals in a timely, non-discriminatory, diligent, and non-arbitrary manner, and we reverse our original decision if we determine that our initial assessment was incorrect.

Table 4.1 - Appeals related to EU Member State users

Reflects appeals from EU accounts of account- and content-related actions during the reporting period broken out by the Reddit Rule violation and the number of appeals granted.

Reddit Rule violation	Content and account-level appeals	Appeals granted
Harassment	10,406	4,143
Hate	9,265	2,438
Impersonation	8	0
Intellectual property policies	0	0
Minor abuse or sexualization	3,262	461
Non-consensual intimate media	5,453	1,630
Personal information	287	98
Prohibited transactions	302	68
Violence	6,666	1,498

The median time for processing appeals from EU accounts of account- and content-related actions during the reporting period was less than 24 hours.

When we action a piece of content reported via the EU Illegal Content Report Form—an Article 16 report—the account that posted the content is notified of the removal reason and provided instructions for how to appeal. Under the DSA, EU users can also submit appeals about an action that Reddit did not take in response to an Article 16 report.

reddit

Table 4.2 - Article 16 notice appeals

Reflects appeals by Article 16 reporters during the reporting period and the number of appeals granted broken out by the type of alleged illegal content.

Type of alleged illegal content	Article 16 reporter appeals	Article 16 reporter appeals granted
Copyright infringement	0	0
Counterfeit goods	0	0
Defamation	56	8
Harassment	15	1
Hateful content	29	4
Illegal transactions	5	1
Minor sexual exploitation	21	0
Non-consensual intimate media	7	1
Other legal issue	25	1
Private or personal information	23	7
Terrorist content	12	0
Trademark infringement	0	0
Violence	15	6

Note: During the reporting period, Reddit had zero appeals of content that was restricted (aka "geoblocked") as a result of an Article 16 notice.

The median time for processing appeals submitted via our EU illegal content report form appeals process during the reporting period was 11 days.

Section 5: Out-of-court settlement body disputes
Article 24.1(a)

Reddit did not receive notice of any disputes submitted to out-of-court dispute settlement bodies in accordance with Article 21 during the reporting period.