



# Regulation (EU) 2022/2065

## Digital Services Act

# Transparency Report for Facebook

---

25 October 2024

*Update: The values in Table 42.3 of the report were provided on 31 October 2024.*

## 1. Introduction

### Background

This Report is published by Meta Platforms Ireland Limited ('Meta') in relation to the Facebook Online Platform ('Facebook') in accordance with the transparency reporting requirements under Articles 15, 24, and 42 of the European Union's Digital Services Act (Regulation (EU) 2022/2065) ('DSA').

This Report is limited to Facebook, the Very Large Online Platform as designated by the European Commission under Article 33 DSA, and does not cover any other Meta provided services that are outside the scope of that designation. In some situations, due to process and/or product limitations, it was not possible to provide Facebook-only data. Where applicable, this is noted in the relevant sections below.

The Report contains information for a reporting period from 1 April 2024 to 30 September 2024.

Below we provide information per the respective sections of Articles 15, 24 and 42 DSA. More information on our approach to content moderation can be found in the [Meta Transparency Center](#).

## 2. Orders received from Member States' Authorities

Information about orders received from Member States' authorities (Article 15(1)(a) DSA).

Meta may receive orders from Member States' authorities, including orders issued in accordance with Articles 9 and 10 DSA (collectively, 'Authority Orders'). Article 9 DSA 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. Article 10 DSA 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.

In the event we receive orders from a Member State authority to act against specific items of alleged illegal content on our platform, first, we review the reported content in line with our Community Standards and other relevant policies, for example Advertising Standards. If we determine that the content goes against our policies, we remove it. If content does not go against our policies, in line with our commitments as a member of the [Global Network Initiative](#) and our [Corporate Human Rights Policy](#), we conduct a review to confirm whether the order is valid. We may then restrict access to the content in the jurisdiction where it is alleged to be unlawful.

Similarly, we have a process to handle orders from a Member State authority that requests the disclosure of information about individual recipients of the service. As explained in more detail in our [Transparency Center](#), Member States' authorities sometimes make requests for data about people who use Facebook as part of official investigations. Meta scrutinises every Member State authority order we receive, regardless of which authority issues the order, to make sure it is legally valid. Meta requires authorities that send orders to comply with applicable laws and our policies. We only produce narrowly tailored user information in response to such orders, and only when we have a good faith belief that

the response is required by law in that jurisdiction, affects users in that jurisdiction, and is consistent with internationally recognised standards. In certain scenarios, we may also require such Member State authorities to use the Mutual Legal Assistance Treaty process.

**Table 15.1.a.(1) - Number of Authority Orders to act against illegal content by Member State for Facebook**

The breakdown below refers to Member States' Authorities' Orders to act against illegal content, including under Article 9 DSA, which cover orders relating to the Facebook service.

<b>Member State</b>	<b>Number of Authority Orders to act against illegal content (including Article 9 orders) addressed to Meta</b>
Austria	1
Belgium	6
Bulgaria	1
Croatia	0
Cyprus	8
Czech Republic	1
Denmark	5
Estonia	8
Finland	1
France	136
Germany	369
Greece	2
Hungary	4
Ireland	4
Italy	52
Latvia	1
Lithuania	5
Luxembourg	0

Malta	2
Netherlands (the)	8
Poland	25
Portugal	1
Romania	2
Slovakia	6
Slovenia	0
Spain	30
Sweden	1
<b>Total</b>	<b>679</b>

**Table 15.1.a.(2) - Number of Authority Orders from Member States to act against illegal content by type of reported illegality for Facebook**

The breakdown below refers to Member States’ Authorities’ Orders to act against illegal content, including under Article 9 DSA, which cover orders relating to the Facebook service.

<b>Type of reported illegality</b>	<b>Number of Authority Orders to act against illegal content (including Article 9 orders) addressed to Meta</b>
Account Access	30
Account Impersonation	7
Adult Intimate Imagery	4
Bullying and Harassment	3
Child Exploitation Imagery	9
Criminal Organisations	35
Fraud and Scams	19
Graphic Content	5
Hate Speech	125
Human Exploitation	2
Misinformation	2
Other	434

Payer and Beneficiary Information	1
Suicide and Self-Injury	0
Violence and Incitement	3
<b>Total</b>	<b>679</b>

**Table 15.1.a.(3) - Number of Authority Orders from Member States to provide information for Facebook and other Meta services combined**

The breakdown below refers to Member States' Authorities' Orders to provide information under Article 10 DSA, which may cover requests relating to the Facebook service as well as other Meta services.

<b>Member State</b>	<b>Number of Authority Orders to provide information (Article 10 orders) addressed to Meta</b>
Austria	2
Belgium	6
Bulgaria	0
Croatia	11
Cyprus	0
Czech Republic	0
Denmark	0
Estonia	4
Finland	0
France	462
Germany	1500
Greece	29
Hungary	213
Ireland	115
Italy	294
Latvia	0

Lithuania	0
Luxembourg	1
Malta	5
Netherlands (the)	0
Poland	550
Portugal	105
Romania	57
Slovakia	2
Slovenia	4
Spain	54
Sweden	0
<b>Total</b>	<b>3414</b>

**Note:** The above Table 15.1.a.(3) solely concerns Article 10 orders as self-selected by Member State Authorities at the time of submission via Article 11 DSA Point of Contact. Meta does not take responsibility for any misleading, inaccurate, or incomplete reporting by the Member States’ Authorities.

Article 10 orders are a small subset of the user data requests Meta receives from Governments and, accordingly, the associated metrics may not be representative of the nature and extent of all requests Meta receives. For a more comprehensive record of government requests for user data, we recommend referring to Meta’s [Global Transparency Reports](#).

**Table 15.1.a.(4) - Number of Authority Orders from Member States to provide information by type of reported illegality for Facebook and other Meta services combined**

The breakdown below refers to Member States’ Authorities’ Orders to provide information under Article 10 DSA, which may cover requests relating to the Facebook service as well as other Meta services.

<b>Type of reported illegality*</b>	<b>Number of Authority Orders to provide information (Article 10 orders) addressed to Meta</b>
Advertising	1

Bullying/Harassment	211
Child Safety	386
Defamation	191
Drugs/Narcotics	68
Fake/Impersonation Account	123
Financial Fraud/Scam	634
Firearms/Weapons	7
Fugitive	43
Gang Activity	18
Hacked Account	231
Hate Speech	383
Homicide/Murder	85
Human Smuggling	52
Human Trafficking	41
Missing/Kidnapped Person	13
Other	78
Physical Assault	87
Possibly Counterfeiting	15
Robbery/Theft	98
Sex Crime/Sexual Assault	58
Sexual Extortion	211
Terrorist Activity	112
Threats of Violence	268
<b>Total</b>	<b>3414</b>

**\*Note:** The above Table 15.1.a.(4) solely concerns Article 10 orders (as self-selected by Member State Authorities at the time of submission via Article 11 DSA Point of Contact) – these are a small subset of the user data requests Meta receives from Governments and, accordingly, the associated metrics may not be representative of the nature and extent of all requests Meta receives. The orders are categorised by the type of reported illegality under investigation or prosecution, which is also self-selected by Member State Authorities at the time of submission. Meta does not take responsibility for any misleading, inaccurate, or incomplete

reporting by the Member States' Authorities. Furthermore, the submission of Orders does not of itself reflect the existence of illegality.

### **Stat 15.1.a.(5) - Time to inform the authority of receipt of an Authority Order**

Automated instant responses are sent to inform the authority of the receipt of Authority Orders to act against allegedly illegal content as well as Authority Orders for data requests.

### **Stat 15.1.a.(6) - Median time to give effect to the Authority Order**

- Median time taken to give effect to the Member States' Authorities' Orders to act against alleged illegal content: **25.8 hours**
- Median time taken to give effect to the Member States' Authorities' Orders to provide information addressed to Meta\*: **14.5 days**

**\*Note:** The information refers to Member States' Authorities' Orders to provide information under Article 10 DSA, which may cover requests relating to the Facebook service as well as other Meta services. This information solely concerns Article 10 orders (as self-selected by Member State Authorities at the time of submission via Article 11 DSA Point of Contact) – these are a small subset of the user data requests Meta receives from Governments and, accordingly, the associated information provided here may not be representative of all requests Meta receives (including emergency requests).

## **3. Notices**

[Information about notices submitted in accordance with Article 16 \(Article 15\(1\)\(b\) DSA\).](#)

Facebook has in place notice mechanisms in accordance with Article 16 DSA allowing users, individuals, and entities to notify Facebook of information on the service that they allege to be illegal content. This mechanism is available directly from the piece of content and is easily accessible. It is also available from the Help Center. Once we receive such a notice, we review the reported content in line with our Community Standards and other relevant policies, for example Advertising Standards, and action the content for violation of our policies as outlined in Section 2. If the reported content does not violate our policies, we review it for legality based on the information provided in the report and may restrict access to it in the jurisdiction where it is alleged to be unlawful.



Trusted flaggers (as designated by the Digital Services Coordinator of the Member State in which the applicant is established) can file a notice within that notice mechanism in line with Article 22 DSA. In order to appropriately prioritise these reports, Meta onboards trusted flaggers to our dedicated reporting channel once they have been designated by the relevant Digital Services Coordinators.

**Table 15.1.b.(1) - Number of notices submitted in accordance with Article 16 DSA, by type of alleged illegal content and actions taken for Facebook**

Type of alleged illegal content	Total number of notices submitted	Total number of notices resulting in content removal for policy violations	Total number of notices resulting in restriction of access to content due to alleged illegality
Intellectual Property (IP)	113,839	45,643	0
Defamation	107,584	11,620	390
Privacy	41,862	8,477	42
Other illegal content	248,156	33,159	4,740

As of 30 September 2024, the number of notices submitted by trusted flaggers from the dedicated trusted flagger reporting channel on Facebook content is 0.

**Stat 15.1.b.(2) - Notices processed by using automated means for Facebook**

All Article 16 DSA notices are processed using manual review. Instances of duplicate submissions are handled by applying the original manual decision, to avoid conflicting decisions.

**Stat 15.1.b.(3) - Median time needed for taking action for Facebook**

- Median time needed to take action on reported content after receiving Article 16 notices: **15.3 hours**

The time periods refer to the time between when the notice was submitted and the first action we took in response to the notice. In instances where there are multiple pieces of content reported, we calculate turnaround time as the time between when the notice was submitted and the first action we take. As an example, if a notice contained two pieces of content and we actioned one piece within 24 hours and the other within 7 days, 24 hours was used for the median calculation. Some decisions can require different time frames due to specific nuances. More complex decisions may require additional guidance from specialised staff and therefore more time.

## 4. Content Moderation Engaged in Meta's Own Initiative

Information about the content moderation engaged in at Meta's own initiative, including the use of automated tools, the measures taken to provide training and assistance to persons in charge of content moderation, and restrictions of the service (Article 15(1)(c) DSA).

Facebook maintains a set of globally applicable Community Standards that define what is and isn't allowed on Facebook. In addition, we have Advertising Standards and Commerce Policies in place for advertising and commerce content, respectively. As part of our content moderation efforts, we employ a combination of human review and technology. This Section 4 of the Report focuses on the actions taken by Meta for Facebook on its own initiative.

### Use of automated tools

Every day, we remove millions of violating pieces of content and accounts on Facebook. In most cases, this happens automatically, with technology to detect, restrict, and remove content and accounts that may go against our Community Standards, Advertising Standards, and Commerce Policies. In other cases, our technology selects content for human review. Our review teams review a blend of user reports and content surfaced by our technology. Our technology also supports the review teams by prioritising the most critical content to be reviewed, based on severity, virality, and likelihood of a violation. Our review systems use technology to prioritise 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, in order to prevent as much harm as possible.

Our technology is trained to identify violations of our Community Standards, Advertising Standards, and Commerce Policies. There are three primary forms of technology used to detect possible policy violations. First, we employ rate limits (speed limits) on how rapidly accounts can perform multiple actions on our platforms, including making posts, to prevent the usage of bots. Next, we have matching technology that identifies identical or near identical copies of URLs, text, images, audio, and videos that we have previously identified as violating our policies. When we match the content exactly or we determine it is near identical, we will typically remove the content. Finally, we also use artificial intelligence (AI) to augment and scale our human review capacity with appropriate oversight: like with the matching technology, when confident enough that a post violates one of our Community Standards, the artificial intelligence will typically remove the content or demote it. We also use artificial intelligence to select the content and account for human review on the basis of severity, virality, and likelihood of a violation. As with matching technology, artificial intelligence operates on URLs, text, images, audio, and videos. Unlike technologies that can only match violations they've seen before, artificial intelligence has the potential to identify certain violations it has never seen before.

In the context of advertisements, when advertisers place an order, each ad is reviewed against our policies. Our Advertising Standards provide policy detail and guidance on the types of ad content we allow and the types of ad content we prohibit. Our Advertising Standards also provide guidance on advertiser behaviour that may result in advertising restrictions being placed on a business account or its assets (an ad account, Page, or user account). Our ad review system relies primarily on automated tools to check ads and business assets against our policies.

### **Persons in charge of content moderation**

Human reviewers are provided with various tools and resources when undertaking content review. For example, human reviewers receive in-depth training and often specialise in certain policy areas and regions. Please see Section 7 below on human resources dedicated to content moderation, which includes details on the measures taken to provide training and assistance to persons in charge of content moderation. Reviewers may be able to use a highlighting tool for slurs and dangerous organisations based on the region where the content is reviewed and tooltips that explain the definitions of certain words and how they should be used to inform decisions.

## Metrics

Our metrics in the below tables provide an overview of the number and type of measures taken that affect the availability, visibility, and accessibility of information provided by the recipients of the service and the recipients' ability to provide information through the service, and other related restrictions of the service, categorised by the type of violation of the terms and conditions, by the use of automation, and by the type of restriction applied.

**Note:** The data provided in Section 4 does not cover content moderation outcomes for French Guiana and Aland Islands due to a technical limitation.

**Table 15.1.c.(1) - Number of organic content removal measures in the European Union for Facebook**

<b>Organic content removals</b>	<b>Removed volume</b>	<b>Removed automation volume</b>
Adult Nudity and Sexual Activity	2,246,208	2,125,014
Bullying and Harassment	915,187	508,435
Child Endangerment - Child Nudity and Physical Abuse	168,641	26,476
Child Endangerment - Child Sexual Exploitation	312,717	287,710
Dangerous Organisation - Hate Orgs	191,002	148,544
Dangerous Organisation - Terrorism	292,393	269,003
Hate Speech	1,151,054	949,336
(Restricted Goods and Services) Drugs	83,109	21,524
(Restricted Goods and Services) Firearms	170,542	153,205
Spam	7,382,238	7,250,783
Suicide and Self-Injury	129,153	114,297
Violence and Incitement	565,616	425,712
Violent and Graphic Content	67,295	57,762

<b>Total (including other violations)</b>	<b>49,361,368</b>	<b>46,758,906</b>
---	-------------------	-------------------

**Note:** The above Table 15.1.c.(1) highlights the type of violations and the use of automation through Meta's content moderation systems between 1 April 2024 and 30 September 2024 on Facebook.

**Table 15.1.c.(2) - Number of organic content demotion measures in the European Union for Facebook**

“Demotion” refers to an enforcement action that we may take to reduce the distribution of content that may be problematic, or low quality. Additional information regarding our strategy and the types of content we demote can be found in our [Transparency Center](#).

<b>Organic content demotions</b>	<b>Demoted volume</b>	<b>Demoted automation volume</b>
Adult Nudity and Sexual Activity	1,088,521	1,088,518
Bullying and Harassment	311,158	311,157
Fact-Checked Misinformation	18,989,014	18,977,541
Hate Speech	205,577	205,576
(Restricted Goods and Services) Drugs	384,783	384,783
Suicide and Self-Injury	43,955	43,955
Violence and Incitement	398,226	398,226
Violent and Graphic Content	5,600,461	5,600,449

**Note:** Demoted content does not indicate a Community Standards violation, therefore we have categorised demotion actions based on our [Content Distribution Guidelines](#), with an additional category for [Fact-Checked Misinformation demotions](#).

**Table 15.1.c.(3) - Number of Marketplace content removals in the European Union for Facebook**

	<b>Removal volume</b>	<b>Removal automation volume</b>
<b>Products removed from Marketplace listing</b>	8,959,521	8,437,045

**Table 15.1.c.(4) - Number of business content removals in the European Union for Facebook and Instagram combined**

	<b>Removal volume</b>	<b>Removal automation volume</b>
<b>Advertising + Commerce Content Removed</b>	16,967,962	14,118,109

**Note:** The data in “Table 15.1.c.(4)” above contains actions for Facebook and Instagram combined due to technical and product limitations.

**Table 15.1.c.(5) - Number of account termination measures in the European Union for Facebook**

	<b>Termination volume</b>	<b>Termination Automation Volume</b>
<b>Account Restriction: Termination</b>	87,825,945	79,195,203

**Note:** Account restriction terminations restrict access to a user’s account in its entirety.

**Table 15.1.c.(6) - Number of provision of service termination measures in the European Union for Facebook and Instagram combined**

	<b>Termination volume</b>	<b>Termination Automation Volume</b>
<b>Provision of the Service: Termination</b>	2,783,273	2,480,724

**Note:**

1. Provision of the service terminations restrict access to a subset of a user’s account or accounts they manage. For example, a user may lose access to their advertising account while retaining access to the remainder of Facebook.
2. The data in Table “15.1.c.(6)” above contains actions for Facebook and Instagram combined due to technical and product limitations.

## 5. Complaints received through Meta’s Internal Complaint-Handling Systems

Information about complaints received through the internal complaint-handling systems (Article 15(1)(d) DSA).

### Reporter complaint

When an individual reports a piece of content (e.g., a photo, comment, post) to us, we'll take action to restrict access to the content or remove the content from Facebook if we find that it goes against our Community Standards or other applicable policies. Where we've reviewed the content, we'll let the individual or entity know whether it did or did not go against our policies.

If the individual or entity reports content but we find that the content does not go against our Community Standards or other applicable policies, we will let the individual or entity know. At that time, if the individual or entity disagrees with our decision to leave the content up, we offer the individual or entity the opportunity to request another review.

In the following, we refer to our internal complaints handling systems as “complaints” or “appeals”.

As outlined above, any user can report content for going against our Community Standards using the in-app product feature. They also have the option to report content as illegal content in line with Article 16 DSA. We offer the possibility to appeal the decisions we take on reports about alleged illegal content or about content going against our Community Standards. If content is not found to go against local law or our Community Standards, the individual or entity that submitted the report can request a review of that decision.

If content is restricted on the basis of local law or actioned for going against our Community Standards, Advertising Standards, Commerce Policies, or Facebook’s Terms of Service, the affected user responsible for that piece of content can request a review of that decision. If a user’s content distribution has been demoted in feed, the user will be able to request another review as described in DSA Article 20. Any affected user whose account is

restricted on the basis of local law or actioned for going against our Community Standards, Advertising Standards, Commerce Policies, or Facebook’s Terms of Service can also request a review of that decision.

We inform the parties that we have received their appeal and respond accordingly.

After that, if we uphold our original decision, there may still be an opportunity for the user to appeal to the [Oversight Board](#). As out-of-court dispute settlement bodies become established under Article 21 DSA, we will also take steps to engage in this process.

We set out below our metrics on the number of complaints received through our internal complaints-handling systems described above, the basis for those complaints, decisions taken with respect to those complaints, the median time needed by us for taking those decisions, and the number of instances where those decisions were reversed.

**Note:**

1. The data provided in Section 5 does not cover content moderation outcomes for French Guiana and Aland Islands due to a technical limitation.
2. Complaints are categorized below based on the reason the content was originally moderated.

**Table 15.1.d.(1) - Number of organic content removal complaints and resulting restored content for Facebook**

<b>Organic content removals</b>	<b>Total removal complaints volume</b>	<b>Total restored content after removal complaint</b>
Adult Nudity and Sexual Activity	302,679	79,693
Bullying and Harassment	257,018	46,803
Child Endangerment - Child Nudity and Physical Abuse	19,625	5,455
Child Endangerment - Child Sexual Exploitation	92,163	43,720
Dangerous Organisation - Hate Orgs	50,692	18,265
Dangerous Organisation - Terrorism	39,008	9,902
Hate Speech	299,485	38,171



(Restricted Goods and Services) Drugs	12,464	1,354
(Restricted Goods and Services) Firearms	35,810	13,084
Spam	144,048	15,279
Suicide and Self-Injury	14,937	3,946
Violence and Incitement	128,382	19,620
Violent and Graphic Content	16,104	3,199
<b>Total (including other violations)</b>	<b>3,509,362</b>	<b>871,759</b>

**Table 15.1.d.(2) - Number of organic content demotion complaints and resulting restored content for Facebook**

<b>Organic content demotions</b>	<b>Total demoted complaints volume</b>	<b>Total demotion lifted after complaint</b>
Adult Nudity and Sexual Activity	139,545	110,195
Bullying and Harassment	46,683	42,977
Fact-Checked Misinformation	172,550	5,440
Hate Speech	26,489	16,601
(Restricted Goods and Services) Drugs	41,687	30,059
Suicide and Self-Injury	6,964	5,987
Violence and Incitement	53,348	46,178
Violent and Graphic Content	999,390	885,603

**Note:** Demoted content does not indicate a Community Standards violation, therefore we have categorised demotion actions based on our [Content Distribution Guidelines](#), with an additional category for [Fact-Checked Misinformation demotions](#).

**Table 15.1.d.(3) - Number of complaints and restores for removed content on Marketplace for Facebook**

	<b>Total complaints volume</b>	<b>Total restored content after complaint</b>
<b>Products removed from Marketplace listing</b>	1,128,780	367,892

**Table 15.1.d.(4) - Number of business content removal complaints and restores for removed business content, for Facebook and Instagram combined**

	<b>Total complaints volume</b>	<b>Total restored content after complaint</b>
<b>Advertising and Commerce Content Removed</b>	4,253,714	968,420

**Note:** The data in “Table 15.1.d.(4)” above contains complaints and restores for Facebook and Instagram combined due to technical and product limitations.

**Table 15.1.d.(5) - Number of complaints and restores by type of restriction for Facebook**

	<b>Total complaints volume</b>	<b>Total restored entities after complaint</b>
<b>Account Restriction: Termination</b>	5,131,032	1,464,993

**Note:** Account restriction terminations restrict access to a user’s account in its entirety. The volumes provided in Table 15.1.d.(5) above describe the number of appeals of this decision and the appeals resulting in account restoration. Users lose access to all Facebook services during this process.

**Table 15.1.d.(6) - Number of complaints and restores by type of restriction, for Facebook and Instagram combined**

	<b>Total complaints volume</b>	<b>Total restored entities after complaint</b>
--	--------------------------------	--

<b>Provision of the Service: Termination</b>	192,194	123,286
--	---------	---------

**Note:**

1. Provision of the service terminations restrict access to a subset of a user's account or accounts they manage. The volumes provided in "Table 15.1.d.(6)" above describe the number of appeals of this decision and the appeals resulting in restoration of the applicable service. The user retains access to all other Facebook services regardless of the outcome of the appeal.
2. The data in Table 15.1.d.(6) above contains complaints and restores for Facebook and Instagram combined due to technical and product limitations.

**Table 15.1.d.(7) - Number of organic content complaints from reporters and resulting removed content for Facebook**

	<b>Appealed by reporter volume</b>	<b>Removed after reporter appeal volume</b>
<b>Organic Content</b>	1,707,932	177,484

**Table 15.1.d.(8) - Other complaints on legal basis (Intellectual Property, Defamation, Privacy, and other illegal content) for Facebook in the European Union**

	<b>Total complaints volume</b>	<b>Total restored content after complaint</b>
<b>Actor appeal</b>	49,552	41,901

	<b>Total complaints volume</b>	<b>Total removed content after complaint</b>
<b>Reporter appeal</b>	4,524	1,797

**Stat 15.1.d.(9) - Median time needed for decision or action on complaints for Facebook and Instagram combined**

- The median time taken for decisions on all complaints from content creators or account/entity owners on Facebook and Instagram to be made or overturned is **0.6 hours**;
- The median time taken for entities to be actioned after receiving reporter appeals on Facebook and Instagram is **11.8 hours**.

**Note:** The above contains turnaround times for Facebook and Instagram combined due to technical and product limitations.

## 6. Automated Means for Content Moderation

Any use made of automated means for the purpose of content moderation (Article 15(1)(e), Article 42(2)(c) DSA).

### Use of automated means for the purpose of own initiative and other content moderation, and purpose of those tools

As described in Section 4, we use technology to help us proactively detect content on our services that might be harmful and violate our Community Standards, so we can remove it faster. We described in Section 4 our primary content moderation automated tools.

These technologies run on accounts, posts, comments, photos, and other pieces of content uploaded to Facebook. 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.

### Indicators of accuracy, error rates, safeguards

Our technology learns and improves from each human decision. Over time – after learning from thousands of human decisions – the technology gets better. When reviewing violating content, review teams manually label the policy guiding their decision, which means that they mark or “label” the relevant policy that the content, account, or behaviour violates. This labelling of data helps us improve the quality of our algorithms that proactively detect and remove harmful content, accounts, and behaviour.

To ensure and improve the quality, i.e., how accurate the technologies mentioned above are in enforcing Community Standards and other policy violations, there are ongoing

quality evaluation processes in place. Meta uses overlapping techniques and systems for maintaining a high overall accuracy for our automated content moderation.

Prior to fully launching any new rate limit (speed limit), matching technology or artificial intelligence (AI), we use the technology to only log how the technology would have behaved instead of immediately acting. We then use human reviewers to assess the accuracy against current content, behaviour, or accounts, rather than just historical ones, as we did during the technology's training. After launching rate limits, matching technologies, or artificial intelligence, we monitor the volumes of actions and appeals by the user who posted the content as well as the rate at which appeals are granted. If any of the metrics we monitor are abnormal, our engineering teams may investigate.

For each primary form of automation technology, the investigation of abnormal metrics can vary. With rate limits, engineers typically reevaluate if the limit is preventing bot behaviour. For our matching technologies, if an entry in our list of previously identified instances of policy violations has abnormal signals, we will re-review the entry to confirm it continues to go against our policies. Similarly, if one of our artificial intelligence tools has abnormal signals, we will either send a sample of the artificial intelligence tool's recent results to human labelling to confirm the accuracy rate or deprecate the artificial intelligence tool if abnormal signals indicate a clear breakage.

In addition, many of our machine learning classifiers are automatically reassessed for accuracy after each human review. This classifier reassessment is an example of the general feedback loop between human review and technology. The content labelling decisions taken by human reviewers are used to train and refine our technology. As a part of this process, the review teams manually label the policy guiding their decision, i.e., they mark the policy that the content, account, or behaviour violates. This helps to improve the quality of our artificial intelligence algorithms and our lists of known policy-violating content used by our matching technology. To maintain quality control in all of these decisions, we regularly audit random samples of decisions taken by the algorithm and our content reviewers and measure them against our expectations for policy enforcement. In the context of automation relating to language, some automation is developed to support specific languages whilst others are language agnostic.

While various types of automation necessitates different and overlapping techniques for assessing accuracy, an indicator of accuracy across all automation techniques is the automation overturn rate: the percentage of content actioned using automated means that are later restored. The automation overturn rate only captures content removed via automation that was later restored. While not all restores are errors and not all errors are restored, the metric still is a directionally approximate indicator of accuracy.

**Table 15.1.e.(1) - Indicator of Accuracy for Facebook and Instagram combined**

<b>Automation Overturn Rate</b>	7.47%
---------------------------------	-------

## 7. Human Resources dedicated to Content Moderation

The human resources that Meta dedicates to content moderation in respect of the service offered in the Union, their qualifications and linguistic expertise and training and support given to such staff (Article 42(2) (a) and (b) DSA).

Meta uses human resources dedicated to content moderation, including for compliance with the obligations set out in Articles 16 and 22 DSA, as well as for compliance with the obligations set out in Article 20 DSA.

### Qualifications and training

Human reviewers come from different backgrounds, reflect our diverse community, and include experts in enforcement in policy areas such as child safety, hate speech, and counterterrorism. Human reviewers undergo extensive training when they join and are regularly trained and tested beyond this initial training, with specific examples, such as on how to uphold the Community Standards and take the correct action on a report. We also do our own proactive audits, where we conduct re-reviews that help us figure out if we are getting it right.

Human reviewers who review content alleged to be illegal receive distinct training based on the nature of their respective work. Every member receives several weeks of training focused heavily on operational proficiency and in preparation for processing such content. For example, the reviewers who review content for defamation receive training specifically on assessing defamation.

### Support

We recognise that reviewing content can be challenging work. Keeping people safe online sometimes means review teams have to look at content that may be objectionable or graphic. We respect the difficulty of this work and work with industry leading vendors to ensure reviewers have access to the resources they need to do their job and support their

health. There is a robust and diverse program to support human reviewers. Our vendor contracts mandate high quality support in a variety of areas, including pay, benefits, work environment, and wellbeing and psychological support. The assistance model depends on what type of content reviewers work on. Such assistance, for example, can take the form of psychological support, including individual and group sessions, and a 24/7 independent support program which includes a range of offerings that include clinical services.

**Volume of human resources dedicated to content moderation and linguistic capabilities**

The team working on safety and security is made up of around 40,000 people across full-time employees and content reviewers. We partner with companies to help with content review, which allows us to scale globally with coverage across time zones, languages, and markets.

For content that requires specific language review in the EU, there are dedicated teams of reviewers that perform content moderation activities specifically for that content. For EU languages that are widely spoken outside the EU, like Portuguese and Spanish, we have content moderation teams that provide global coverage. Our global scale allows us to quickly redeploy capacity if there is a surge in demand in a specific country, in times of crisis, or when unpredictable events occur. When these instances transpire, we are able to adapt our resources appropriately. This means that we are also able to rely temporarily on our teams that provide global coverage to mitigate EU risks.

This content review team is global and reviews content 24/7 in over 80 languages. The team includes reviewers with language expertise to enforce our policies in cases where certain words or content require additional contextual understanding. Not all content requires language expertise; for example, nudity and sexual activity is language agnostic. For this type of content, Meta has a global pool of content moderators who review these types of reports. Currently there are around 5,400 language agnostic reviewers globally. These reviewers also review EU content that falls under this language agnostic category. Content in this category generally does not contain language.

**Table 42.2.a - Content Moderators by Official EU Language for Facebook and Instagram combined**

EU Language	Number of Reviewers
-------------	---------------------

Bulgarian	39
Croatian	43
Czech	45
Danish	30
Dutch	96
English*	183
Estonian	3
Finnish	20
French*	540
German	406
Greek	30
Hungarian	30
Irish	0
Italian	362
Latvian	3
Lithuanian	11
Maltese	1
Polish	103
Portuguese*	1,122
Romanian	64
Slovak	39
Slovenian	8
Spanish*	2,325
Swedish	45

**Note:**

1. These numbers in Table 42.2.a apply to Facebook and Instagram. The numbers are reflective of reviewers, as outlined above, who work on EU-specific content across EU official languages.
2. For EU content that is in a language other than the 24 official languages of the EU, there are additional language-based content reviewers.



3. \*For languages that are widely spoken outside of the EU, e.g., French, English, Spanish, Portuguese, there are additional content reviewers that review reports from non-EU countries in these languages.

## 8. Out-of-court dispute settlement submissions

Information about disputes submitted to the out-of-court dispute settlement bodies referred to in Article 21 (Article 24(1)(a) DSA)

We inform users, individuals, and entities that if they do not agree with relevant enforcement decisions, they may have the right to challenge the decision in a relevant court and that they may also be able to refer the decision to a certified dispute settlement body. As of 30 September 2024, we did not receive any disputes from certified out-of-court settlement bodies pursuant to Article 21 DSA.

## 9. Measures and protection against misuse

The number of suspensions imposed pursuant to Article 23 (Article 24(1)(b) DSA).

If users post content that goes against Community Standards, we remove it and may then apply a strike to their Facebook account. Whether we apply a strike depends on the severity of the content, the context in which it was shared, and when it was posted. More information can be found in the [Meta Transparency Center](#). When users persistently or severely breach our policies, we may suspend or disable their account. More information can be found in the [Meta Transparency Center](#). We have also developed a Misuse Policy in line with Article 23(1) DSA for users who frequently upload manifestly illegal content.

We may suspend, for a limited period of time, users who, after being warned, repeatedly post manifestly illegal content. This does not preclude us from terminating their use of the service under Article 4.2 of the Facebook Terms of Service.

In line with Article 23(2) DSA, we may suspend the processing of notices and complaints submitted through our notice and complaints mechanisms, for a limited period of time, where individuals and entities have, after being warned, frequently submitted notices and complaints that are manifestly unfounded. We do not have any volumes to report for this

reporting period relating to manifestly unfounded notices or manifestly unfounded complaints.

We set out below the number of suspensions imposed pursuant to Article 23(1) DSA for the provision of manifestly illegal content.

**Table 24.1.b - Number of suspensions by basis for Facebook**

Suspension basis	Number of suspensions
Provision of manifestly illegal content	0

**Note:** The data provided is currently representative of the number of suspensions enacted for the provision of manifestly illegal content.

## 10. Average Monthly Active Users for each Member State

Information about the average monthly active recipients of the service for each Member State (Article 42(3) DSA).

Meta is required under Article 42(3) to publish a breakdown of the average monthly users for each Member State in this Report. The breakdown per Member State for Facebook below is calculated as an average over a 6-month period, ending 30 September 2024 (i.e., 1 April 2024 - 30 September 2024), rounded to one decimal points and then approximated. For this period, there were a total of approximately **260.6 million** average monthly active users on Facebook in the EU.

For Facebook, we define a monthly active user as a registered and logged-in Facebook user who visited Facebook through our website or a mobile device in the last 30 days as of the date of measurement.

We note that Article 24(2) DSA also mandates the publication of information on the average monthly active users on Facebook in the EU. This data must be made available on a publicly accessible section of our online platform at least every six months. To avoid duplication in such publications on our website, we reference this information from our transparency reports which are also required to be published at least every six months. The Article 24(2) publications are available in the 'Digital Services Act - Information on

Average Monthly Active Recipients in the European Union' section of our [Transparency Center](#).

**Table 42.3 - Average Monthly Active Users for each Member State for Facebook**

<b>Member State</b>	<b>Average Monthly Active Users</b>
Austria	4.2M
Belgium	8.0M
Bulgaria	4.4M
Croatia	2.2M
Cyprus	1.0M
Czech Republic	6.5M
Denmark	4.3M
Estonia	Less than 1M
Finland	3.2M
France	42.3M
Germany	32.9M
Greece	6.4M
Hungary	7.0M
Ireland	3.3M
Italy	35.9M
Latvia	1.1M
Lithuania	2.1M
Luxembourg	Less than 1M
Malta	Less than 1M
Netherlands (the)	10.3M
Poland	24.0M

Portugal	7.7M
Romania	12.3M
Slovakia	3.3M
Slovenia	1.2M
Spain	26.9M
Sweden	7.1M

**Note:**

1. This information on the use of Facebook in the EU has been prepared for Articles 24(2) and 42(3) DSA. This information may differ from user metrics reported in other contexts in certain key respects, including, for example, periodic reports filed with other regulatory authorities, and should not be used for other purposes.
2. Where applicable, the Member State breakdown above includes any monthly active user metrics available for outermost regions or other territories associated with such Member States.
3. There are inherent challenges in measuring usage of our services across large online and mobile populations across the world. Many people in our community have user accounts on more than one of our services, and some people have multiple user accounts within an individual service. The above monthly active user estimates by online platform do not represent estimates of the number of unique people using these services.

---

Note: Meta works diligently and utilises a variety of quality assurance measures to strive for accuracy and reliability of the data and metrics it releases. With respect to the data and metrics provided here, they are novel, voluminous, and generally not of the type operationalised by Meta in its core products or services. Thus, while Meta has employed rigorous practices to provide the most accurate information required by applicable law, it is possible for inaccuracies to persist.