Regulation (EU) 2022/2065
Digital Services Act
Transparency Report for WhatsApp

14 February 2025

1. Introduction

Name of service provider: This Report is published by WhatsApp Ireland Limited ('WhatsApp' or 'WA') in accordance with the transparency reporting requirements under Articles 15 and 24 of the European Union's Digital Services Act (Regulation (EU) 2022/2065) ('DSA'). This report compiles responses for WhatsApp Messenger App ("WhatsApp Messenger"), WhatsApp Channels ("Channels"), and the WhatsApp Business App ("Business App").

Date of the publication of the report: 14 February 2025

Starting and ending date of reporting period: The Report contains information for a reporting period from 17 February 2024 to 31 December 2024.

2. Orders received from Member States' Authorities

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

WhatsApp 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, first, we review the reported account in line with the relevant policies, which may include the WhatsApp Channels Guidelines, the WhatsApp Messaging Guidelines, and/or the WhatsApp Business Messaging Policy. If we determine that the account has violated our policies,

we take appropriate enforcement action. If the account has not violated 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 hosted content in the jurisdiction where it is alleged to be unlawful (for Channels content) or take appropriate enforcement action against the violating account (for WhatsApp Messenger or Business App).

Similarly, we have processes to handle orders from a Member State authority that request disclosure of information about individual recipients of the service. In connection with official investigations, Member State authorities sometimes make requests for data about WhatsApp users. WhatsApp reviews and scrutinises every Member State authority order, regardless of which authority issues the order, to make sure it is legally valid. WhatsApp requires authorities that send orders to comply with applicable laws and our policies, including our Privacy Policy. Upon receipt of a legally valid order, if we have a good faith belief that we are required by law in that jurisdiction to respond, and that providing the requested information is consistent with internationally recognized standards, WhatsApp will produce narrowly tailored responsive user information to comply with such orders. 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 Channels, Business App and WhatsApp Messenger services

The breakdown below refers to Member States' Authorities' Orders to act against illegal content, including under Article 9 DSA, which cover orders relating to users of WhatsApp Channels, the WhatsApp Business App and WhatsApp Messenger for the reporting period of 17 February 2024 to 31 December 2024.

Member State	Number of Authority Orders to act against illegal content (including Article 9 orders)		
	WA Channels	WA Business App	WA Messenger
Austria	0	0	0
Belgium	0	0	0
Bulgaria	0	0	0

Croatia	0	0	0
Cyprus	0	0	0
Czech Republic	0	0	0
Denmark	1	0	1
Estonia	0	0	0
Finland	1	0	1
France	0	0	7
Germany	2	2	18
Greece	0	0	0
Hungary	0	0	0
Ireland	0	0	2
Italy	0	1	0
Latvia	0	0	0
Lithuania	0	0	0
Luxembourg	0	0	0
Malta	0	0	0
Netherlands	0	0	0
Poland	0	0	0
Portugal	0	0	0
Romania	0	0	0
Slovakia	0	0	0

Slovenia	0	0	0
Spain	0	0	0
Sweden	0	0	0
Total	**4**	**3**	**29**

Note: The table above does not include 93 submissions under Article 9 DSA which were not related to any WhatsApp service.

Table 15.1.a.(2) - Number of Authority Orders from Member States to act against illegal content by type of reported illegality for WhatsApp Channels, Business App and Messenger services

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 Channels, Business App and Messenger services for the reporting period of 17 February 2024 to 31 December 2024.

Type of reported illegality	**Number of Authority Orders to act against illegal content (including Article 9 orders)**		
	WhatsApp Channels	**WhatsApp Business App**	**WhatsApp Messenger**
Account Access	4	0	7
Account Impersonation	0	0	0
Adult Intimate Imagery	0	0	1
Bullying and Harassment	0	0	0
Child Exploitation Imagery	0	2	7
Criminal Organisations	0	0	1
Fraud and Scams	0	0	0

Graphic Content	0	0	0
Hate Speech	0	0	0
Human Exploitation	0	0	3
Misinformation	0	0	0
Other	0	1	9
Payer and Beneficiary Information	0	0	0
Suicide and Self-Injury	0	0	1
Violence and Incitement	0	0	0
Total	**4**	**3**	**29**

Note: The table above does not include 93 submissions under Article 9 DSA which were not linked to any WhatsApp service.

Table 15.1.a.(3) - Number of Authority Orders from Member States to provide information under Article 10 DSA categorised by Member States

The breakdowns provided below refer to the number of Authority orders to provide information under Article 10 DSA, categorised by Member State, and include the following:

- **Table 15.1.a.(3a)** – for the period of 17 February 2024 to 31 December 2024: combined metrics for the number of Article 10 orders relating to Channels, Business App and Messenger services; and

- **Table 15.1.a.(3b)** – for the period of 1 October 2024 to 31 December 2024: the number of Article 10 orders relating to Channels, Business App and Messenger Services, categorised by the service.

Please note that WhatsApp is continuously improving its reporting processes to align with the transparency obligations of DSA and in anticipation of the future entry into force of the associated Implementing Regulation. As of 1 October 2024, WhatsApp introduced the technical capability to report on specific metrics by service, as detailed in Table 15.1.a.(3b) below. Future reports will provide this service specific breakdown for the entire reporting period(s).

Table 15.1.a.(3a) - Number of Authority orders from Member States to provide information for the period of 17 February 2024 to 31 December 2024 for WhatsApp Channels, Business App and Messenger combined

Member State	Number of Authority Orders to provide information (Article 10 DSA)
Austria	0
Belgium	0
Bulgaria	0
Croatia	0
Cyprus	0
Czech Republic	0
Denmark	0
Estonia	0
Finland	0
France	87
Germany	129
Greece	0
Hungary	5
Ireland	12
Italy	43
Latvia	0
Lithuania	0

Luxembourg	0
Malta	10
Netherlands	0
Poland	29
Portugal	36
Romania	7
Slovakia	0
Slovenia	0
Spain	7
Sweden	1
Total	**366**

Note: The above Table 15.1.a.(3a) solely concerns Article 10 orders typically self-selected by Member State Authorities at the time of submission via Article 11 DSA Point of Contact. WhatsApp 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 WhatsApp receives from Governments and, accordingly, the associated metrics may not be representative of the nature and extent of all requests WhatsApp receives.

Table 15.1.a.(3b) - Number of Authority orders from Member States to provide information for the period 1 October 2024 to 31 December 2024 (categorised by service)

Member State	WA Channels	WA Business App	WA Messenger
Austria	0	0	0
Belgium	0	0	0
Bulgaria	0	0	0
Croatia	0	0	0
Cyprus	0	0	0
Czech Republic	0	0	0

Denmark	0	0	0
Estonia	0	0	0
Finland	0	0	0
France	14	15	39
Germany	13	17	46
Greece	0	0	0
Hungary	0	0	2
Ireland	3	4	6
Italy	6	7	16
Latvia	0	0	0
Lithuania	0	0	0
Luxembourg	0	0	0
Malta	0	0	2
Netherlands	0	0	0
Poland	4	5	8
Portugal	5	5	8
Romania	1	1	1
Slovakia	0	0	0
Slovenia	0	0	0
Spain	1	1	3
Sweden	0	0	0
Total	**47**	**55**	**131**

Note: The above Table 15.1.a.(3b) solely concerns Article 10 Orders categorised by the specific WhatsApp service(s), typically self-selected by Member State Authorities at the time of submission via Article 11 DSA Point of Contact. WhatsApp 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 WhatsApp receives from Governments and, accordingly, the associated metrics may not be representative of the nature and extent of all requests WhatsApp receives.

Table 15.1.a.(4) - Number of Authority Orders from Member States to provide information by type of reported illegality for WhatsApp Channels, Business App and Messenger services

The breakdowns provided below refer to the number of Authority orders to provide information under Article 10 DSA, categorised by the type of reported illegality, and include the following:

- **Table 15.1.a.(4a)** – for the period of 17 February 2024 to 31 December 2024: combined metrics for the number of Article 10 orders relating to Channels, Business App and Messenger services; and

- **Table 15.1.a.(4b)** – for the period of 1 October 2024 to 31 December 2024: the number of orders relating to Channels, Business App and Messenger services, categorised by the service.

Please note that WhatsApp is continuously improving its reporting processes to align with the transparency obligations of DSA and in anticipation of the future entry into force of the associated Implementing Regulation. As of 1 October 2024, WhatsApp introduced the technical capability to report on specific metrics by service, as detailed in Table 15.1.a.(4b) below. Future reports will provide this service specific breakdown for the entire reporting period(s).

Table 15.1.a.(4a) - Number of Authority orders for the period of 17 February 2024 to 31 December 2024 (combined across services)

Type of reported illegality*	Number of Authority Orders to provide information under Article 10 DSA for WhatsApp Channels, Business App and Messenger Services

Bullying/Harassment	6
Child Safety	29
Defamation	1
Drugs/Narcotics	31
Fake/Impersonation Account	2
Financial Fraud/Scam	163
Firearms/Weapons	2
Fugitive	3
Gang Activity	6
Hacked Account	14
Hate Speech	1
Homicide/Murder	15
Human Smuggling	17
Human Trafficking	5
Missing/Kidnapped Person	7
Other	12
Physical Assault	5
Robbery/Theft	15
Sex Crime/Sexual Assault	6
Sexual Extortion	11
Terrorist Activity	5
Threats of Violence	10
Total	**366**

*__Note:__ The above Table 15.1.a.(4a) solely concerns Article 10 orders (typically 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 WhatsApp receives from Governments and, accordingly, the associated metrics may not be representative of the nature and extent of all requests WhatsApp receives.

The orders are categorised by the type of reported illegality under investigation or prosecution, which is also typically self-selected by Member State Authorities at the time of submission. WhatsApp 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.

Table 15.1.a.(4b) - Number of Authority orders for the period 1 October 2024 to 31 December 2024 (categorised by service)

Type of reported illegality*	WA Channels	WA Business App	WA Messenger
Bullying/Harassment	0	0	3
Child Safety	2	3	9
Defamation	1	0	1
Drugs/Narcotics	4	6	9
Fake/Impersonation Account	0	0	1
Financial Fraud/Scam	20	23	52
Firearms/Weapons	0	0	1
Gang Activity	3	4	4
Hacked Account	2	1	7
Hate Speech	1	1	1
Homicide/Murder	4	4	9
Human Smuggling	0	2	4
Human Trafficking	1	0	3
Missing/Kidnapped Person	1	1	1
Other	3	2	6
Robbery/Theft	0	0	5
Sex Crime/Sexual Assault	0	0	2
Sexual Extortion	2	5	7
Terrorist Activity	1	1	1
Threats of Violence	2	2	5

Total	**47**	**55**	**131**

***Note:** The above Table 15.1.a.(4b) solely concerns Article 10 orders (typically 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 WhatsApp receives from Governments and, accordingly, the associated metrics may not be representative of the nature and extent of all requests WhatsApp receives.

The orders are categorised by the specific WhatsApp service(s) and by the type of reported illegality under investigation or prosecution, which are also typically self-selected by Member State Authorities at the time of submission. WhatsApp 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 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 an Article 9 order to act against illegal content

The median time taken to give effect to the Member States' Authorities' Orders to act against alleged illegal content for WhatsApp Channels, Business App and Messenger combined is 22 hours.

Stat 15.1.a.(7) - Median time to give effect to the Authority Order

The median time taken to give effect to the Member States' Authorities' Orders to provide information for WhatsApp Channels, Business App and Messenger combined is 10.1 days.

***Note:** The information refers to Member States' Authorities' Orders to provide information under Article 10 DSA, which covers requests relating to WhatsApp Channels, Business App and Messenger combined. This information solely concerns Article 10 orders (typically 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 WhatsApp receives from Governments and, accordingly, the associated information provided here may not be representative of all requests WhatsApp receives (including emergency requests).

Please note that the time to give effect to an Authority Order is calculated based on the interval between valid receipt of an Article 10 order and WhatsApp giving effect to it. This metric excludes time passed – where applicable – for WhatsApp to respond to the requesting authority to seek clarification, further context or resolution of formal defects with respect to the order.

3. Notices

> Information about notices submitted in accordance with Article 16 (Article 15(1)(b) DSA).

WhatsApp has in place notice mechanisms in accordance with Article 16 DSA allowing users, individuals, and entities to notify WhatsApp of the presence of specific items of information on the Channels service that they allege to be illegal content. This mechanism is available directly from the Channel and is easily accessible. The reporting form is also available from the Help Center. Once we receive such a notice, we review the reported content in line with our WhatsApp Channels Guidelines and other applicable policies, and take appropriate enforcement action 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.

For Trusted flaggers (as designated by the Digital Services Coordinator of the Member State in which the applicant is established), we have established a dedicated reporting channel to appropriately prioritise notices submitted by designated trusted flaggers.

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 Channels

Type of alleged illegal content	Total number of notices submitted	Total number of notices from Trusted flaggers	Total number of notices resulting in enforcement for policy violations	Total number of notices resulting in restriction of access to content due to alleged illegality

Intellectual Property (IP)	190	0	19	0
Defamation	10	0	0	0
Privacy	0	0	0	0
Other illegal content	21	0	6	0

As of 31 December 2024, the number of notices submitted by trusted flaggers through the dedicated trusted flagger reporting form for Channels is 0.

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

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 Channels

- Median time needed to take action on reported content after receiving Article 16 notices: **12.3 days**

4. WhatsApp Integrity: Measures taken on Our Own Initiative

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

WhatsApp maintains sets of globally applicable WhatsApp Channels Guidelines, WhatsApp Messaging Guidelines, and the WhatsApp Business Messaging Policy, and the applicable Terms of Service, that define what is and isn't allowed on our services. We employ a combination of product design, automated data processing and human review, as well as a range of enforcement actions to implement these policies. This Section 4 of the Report focuses on the actions taken by WhatsApp for Channels, Business App and Messenger on its own initiative.

WhatsApp Messenger and the Business Messaging App

WhatsApp has built its Messenger and small business apps to be simple, reliable, and private. Product features (including the absence of in-app user discovery and the lack of an algorithmic feed) are designed to ensure users are able to communicate reliably with their most important contacts. Personal messages and calls on WhatsApp are always protected by end-to-end encryption.

WhatsApp uses a range of automated tools to enforce our messaging policies. These include tools designed to prevent of scripted account creation or bulk messaging. In addition, WhatsApp uses automated data processing on non-encrypted account, group, and community profile information, as well as on reported messages, to help detect potential violations of our guidelines. These tools automate decisions for certain areas where account behavior or reported message content is highly likely to be in violation of WhatsApp's policies. Automation also helps us prioritize and expedite reviews by routing potentially violating accounts, groups, or communities to human reviewers, so our teams can focus on the most important cases first.

When a messaging account, group or community requires further review, our automated systems send it to a human review team to make the final decision. Our human review teams - who are located across the globe, receive in-depth training, and often specialize in certain policy areas and regions - are able to review account information and reported messages. Because WhatsApp does not have access to encrypted message content, in most cases a violation of our messaging policies results in account termination. We may, in certain cases, also take actions to suspend groups or communities. We do not have the ability to restrict access to messaging content in a particular jurisdiction.

WhatsApp Channels

WhatsApp Channels is an optional, one-way broadcasting feature within WhatsApp, separate from WhatsApp Messenger. WhatsApp Channel updates are not encrypted. To enforce our policies on Channels, WhatsApp runs automated data processing over channels information, including picture, description, and updates. In addition, WhatsApp provides users multiple entry points to report potentially violating channels or specific channel updates. WhatsApp employs automated decisions for certain areas where channels content is highly likely to violate Channels Guidelines.

In addition, automation also helps us prioritize review by routing potentially violating channels to human reviewers, so our teams can focus on the most important cases first. Either as a result of automated enforcement or human review, WhatsApp can take a number of measures to enforce our policies and restrict access to certain Channels content. WhatsApp may choose to warn a channel admin, apply strikes, remove a violating channel picture, or prevent a channel from being accessed in certain jurisdictions or discovered by non-followers. WhatsApp will suspend a channel if its admin repeatedly posts content that violates our terms and policies, including illegal content. The decision to suspend a channel will depend on the amount, nature, and severity of the violating content, and, if identifiable, the intent of the user. WhatsApp may also terminate a user's access to the entire service for violating our Channels policies.

Persons in charge of WhatsApp Integrity

Human reviewers are provided with various tools and resources when undertaking account review. For example, human reviewers receive in-depth training and often specialise in certain policy areas. On-screen tooling such as tooltips, highlighters and training protocols are available to explain definitions and inform decisions. An example would be highlighting lists of policy-violating drugs, common precursors and related slang terms.

Metrics

Our metrics in the below tables provide an overview of the number and type of measures taken that affect the availability, visibility, or 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.

Table 15.1.c.(1) - Number of account termination measures in the European Union for Business App and WhatsApp Messenger

Account Restriction: Termination	WhatsApp Business Termination volume	WhatsApp Business Termination Automation Volume	WhatsApp Messenger Termination Volume	WhatsApp Messenger Termination Automation Volume
Adversarial Data Use	56,869	56,869	214,476	214,476
Child Sexual Exploitation	3,771	2,255	35,333	26,332
Fake Accounts	351,467	351,467	819,711	819,711
Scams	260,372	251,445	433,382	382,337
Spam	1,318,547	1,311,341	4,885,456	4,813,218
Total (including other violations)	**2,043,164**	**2,019,007**	**6,406,067**	**6,261,062**

Note: The above Table 15.1.c.(1) highlights the type of violations through WhatsApp's content moderation systems between 17 February 2024 and 31 December 2024 on Channels, Business App, and Messenger.

The WhatsApp Messenger refers to the application that allows users access to both the private messaging and channels services. Violations against the terms of service for either service can result in account termination for both services.

Table 15.1.c.(2) - Number of WhatsApp Channel Icon Pictures Removed

Reason	Removal volume	Removal automation volume
Nudity and Pornography	876	1
Child Sexual Exploitation	482	50
Total (including other violations)	**1,359**	**52**

Table 15.1.c.(3) - Number of WA Channel Suspensions

Reason	Suspension volume	Suspension automation volume
Child Sexual Exploitation	1,756	265
Nudity and Pornopraphy	678	36
Hate Orgs	516	36
Sexual Solicitation And Prostitution	214	2
Total (including other violations)	**3,468**	**374**

Table 15.1.c.(4): Number of WA Channel Messages Causing a Channel to be Hidden

Reason	Hidden volume	Hidden automation volume
Nudity and Pornography	19,558	58
Hate Orgs	11,829	91
Sexual Solicitation And Prostitution	4,802	60
Child Sexual Exploitation Indirect	1,941	1
Total (including other violations)	**47,714**	**310**

Note: Table 15.1.c.(4) refers to the number of messages resulting in a WhatsApp Channel being hidden where some messages are from the same channel. This table does not refer to unique channels hidden.

Table 15.1.c.(5): Number of WA Channels Geographically Hidden

Reason	Hidden volume	Hidden automation volume

Account Security	6	6
Local Law Violation	25	25
Privacy Violation	1	1
Sexually Explicit Language	1	1
Total (including other violations)	**36**	**36**

5. Complaints received through WhatsApp's Internal Complaint-Handling Systems

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

Appeals

If a Channel, group and/or account is restricted on the basis of local law or actioned for going against our WhatsApp terms or policies, the affected users can request a review of that decision. The specific review mechanisms differ based on the product, but in all cases we inform the parties that we have received their appeal and respond accordingly. 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: Complaints are categorized below based on the reason the content was originally actioned.

Table 15.1.d.(1) - Number of complaints for Business App and Messenger

	Business App	Messenger
Total complaints volume	1,888,781	5,614,105

Table 15.1.d.(2) - Number of Channels complaints volume and total resulting restores after complaint for Channels

Channel complaint reasons	Total complaints volume	Total restored Channels after complaint
Child Sexual Exploitation Indirect	229	5
Hate Orgs	507	139
Nudity & Pornography	483	79
Sexual Solicitation And Prostitution	278	95
Total (including other violations)	**2364**	**706**

Table 15.1.d.(3) - Number of channel complaints from reporters and resulting suspended channels for Channels

Appealed by reporter volume	Removed after reporter appeal volume
87,138	574

Stat 15.1.d.(4) - Median time needed for decision or action on complaints for for Channels

- The median time taken for decisions on all complaints from Channel admins is **0.25 hours**;

- The median time taken for channels to be actioned after receiving reporter appeals is **72 hours**.

6. Use of Automated Means to Promote Integrity

> Any use made of automated means for the purpose of content moderation (Article 15(1)(e).

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 detect content on our services that might be harmful and violate our Whatsapp Channels Guidelines, WhatsApp Messaging Guidelines, and/or WhatsApp Business Messaging Policy, so we can take faster action. We described in Section 4 our primary content moderation automated tools.

These technologies assess account behavior and run on non-encrypted account, group and community profile information, channel profile information and channels content, as well as messages reported by users. These tools (as described in Section 4) determine how probable or likely it is that the messaging account or channel admin has violated WhatsApps terms or policies. The result of this determination can either be an enforcement action, where account behavior or reported message content is highly likely to be in violation, or the referral of the account or channel for further human review.

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 the WhatsApp Channels Guidelines, there are ongoing quality evaluation processes in place. WhatsApp uses overlapping techniques and systems for maintaining a high overall accuracy for our automation.

Prior to fully launching any new technologies, 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 account 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 account or behaviour violates. This helps to improve the quality of our artificial intelligence algorithms and our lists of known policy-violating accounts 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 channels actioned using automated means that are later restored. The automation overturn rate captures channel enforcements via automation that were later restored. While not all restores are errors and not all errors are restored, the metric still is a directionally approximate indicator of accuracy.

Service	Automation Overturn Rate
WhatsApp Channels	3.18%

7. 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 31 December 2024, we did not receive any disputes from certified out-of-court settlement bodies pursuant to Article 21 DSA.

8. Measures and protection against misuse

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

On our WhatsApp services, suspension decisions are made on the basis of WhatsApp policies, which are not necessarily coterminous with manifestly illegal content.

Note: WhatsApp 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 WhatsApp in its core products or services. Thus, while WhatsApp has employed rigorous practices to provide the most accurate information required by applicable law, it is possible for inaccuracies to persist.