EU Digital Services Act (EU DSA) Biannual VLOSE/VLOP Transparency Report

The European Union (EU) Digital Services Act (DSA) came into force on 16 November 2022. We welcome the DSA's goals of making the internet even more safe, transparent and accountable, while ensuring that everyone in the EU continues to benefit from the open web. Google is committed to promoting transparency for the users of our platforms.

Published: 27 October 2023

EU DSA Biannual VLOSE/VLOP Transparency Report

The European Union (EU) Digital Services Act (DSA) came into force on 16 November 2022. Google has long been aligned with the broad goals of the DSA and has devoted significant resources into tailoring our programs to meet its specific requirements. We welcome the DSA's goals of making the internet even safer, more transparent and more accountable, while ensuring that everyone in the EU continues to benefit from the open web.

In accordance with Articles 15, 24, and 42 of the DSA, Google is publishing biannual transparency reports for its services designated by the European Commission as a Very Large Online Search Engine (VLOSE) or a Very Large Online Platform (VLOP): Google Search, Google Maps, Google Play, Shopping and YouTube.

This report describes Google's efforts and resources to moderate content on the services listed above in the EU during the period from **28 August 2023 to 10 September 2023**[1].

Overview

Since Google was founded, our mission has been to organise the world's information and make it universally accessible and useful. When it comes to the information and content on our platforms, we take seriously our responsibility to safeguard the people and businesses using our products, and do so with clear and transparent policies and processes.

As such, our product, policy, and enforcement decisions are guided by a set of principles which enable us to preserve freedom of expression, while curbing the spread of content that is damaging to users and society.

1. **We value openness and accessibility:** We lean towards keeping content accessible by providing access to an open and diverse information ecosystem.
2. **We respect user choice:** If users search for content that is not illegal or prohibited by our policies, they should be able to find it.
3. **We build for everyone:** Our services are used around the world by users from different cultures, languages, and backgrounds, and at different stages in their lives. We take the diversity of our users into account in policy development and policy enforcement decisions.

These principles are addressed in three key ways, providing our users with access to trustworthy information and content.

First, we **protect users from harm** through built-in advanced protections, policies, and a combination of scaled technology and specially trained human reviewers. These mechanisms enable us to prevent

---

[1] Due to the short time between DSA applicability date and report deadline, and the time required to conduct data validation, it was only feasible to include a 2 week reporting period for this first report. Future reports will provide metrics collected over a longer reporting period.

distribution of harmful and illegal content before it reaches users; detect and evaluate potentially violative content; and respond to bad actors and abusive content in an appropriate way. Second, through our ranking and recommendation systems, we **deliver reliable information** to users, as well as provide tools to help users evaluate content themselves, giving them added context and confidence in what they find on our products and services, and across the internet. Third, we **partner to create a safer internet** and scale our impact, collaborating with experts, governments, and organisations to inform our tools and share our technologies.

Helpful, safe online environments do not just happen — they are designed. At Google, we aim to balance access to information with protecting users and society, while providing information and content users can trust.

In this report, we outline and provide metrics contemplated by the DSA regarding our efforts and resources to moderate potentially illegal content and policy-violative content in the EU. We are committed to improving and augmenting future iterations with further insights about our continued efforts to combat violative content on our platforms.

Section 1: Article 9 and 10 Orders from Member States' authorities

Article 15(1), point (a)

Courts and government agencies in the EU regularly request that we remove information from Google services (Removal Orders). These requests are routed to the appropriate team(s) within Google who review these requests closely to determine if information should be removed because it may violate a law or our product policies. In addition, specific Member State laws allow government agencies in the EU to request user information for civil, administrative, criminal, and national security purposes (User Data Disclosure Orders). Each request is carefully reviewed to make sure it satisfies applicable laws. No Removal Orders or User Data Disclosure Orders conforming to the requirements of Articles 9 and 10 of the DSA were received during the reporting period.

Information about other requests from government authorities around the world are published in our Government Requests for Content Removal Transparency Report and our Government Requests for User Information Transparency Report[2].

---

[2] Information in these reports is voluntarily provided and not necessarily directly comparable with information presented in this mandated DSA report, due to differences in methodologies.

Section 2: Notices received through notice and action mechanisms

Article 15(1), point (b)

Google's content and product policies apply wherever you are in the world, but we also have processes in place to remove or restrict access to content based on local laws. Users, Trusted Flaggers (as defined by Article 22) and other entities can report content that they would like to be removed from Google's services under applicable laws. Action is taken on content that is deemed to violate applicable laws or Google policies.

2.1 Number of notices submitted in accordance with Article 16, broken down by type of alleged illegal content concerned

Article 15(1), point (b)

Table 2.1.1 reflects the number of notices submitted by EU-based users and other entities in accordance with Article 16 during the reporting period, broken down by type of alleged illegal content and service.

Table 2.1.1: Number of Article 16 notices submitted, by type of alleged illegal content and service

Type of alleged illegal content	Number of Article 16 notices				
	Maps	Play	Shopping	YouTube	Multi-Services[1]
Child Sexual Abuse and Exploitation	4	0	0	235	1
Circumvention	0	0	0	6	0
Copyright	36	79	17	36,568	112
Counterfeit	0	1	6	429	7
Defamation	14,627	6	4	1,768	2
Hate and Harassment	2	0	0	767	0
Privacy	235	3	1	636	0
Trademark	0	56	1	653	171

Type of alleged illegal content	Number of Article 16 notices				
	Maps	Play	Shopping	YouTube	Multi-Services[1]
Violent Extremism	0	0	0	180	0
Other Legal	2	1	0	848	32
Total	**14,906**	**146**	**29**	**42,090**	**325**

Notes:

[1] Notices relating to advertisements that may have appeared across multiple Google services, including VLOPs, are included under Multi-Services.

2.2 Number of Article 16 notices submitted by Trusted Flaggers

Article 15(1), point (b)

In the European Union, national entities called Digital Services Coordinators may award Trusted Flagger status to entities tasked with flagging allegedly illegal content on online platforms. Trusted Flaggers are likely to have expertise in one or more fields relevant to content moderation, such as privacy or child safety. The European Commission will maintain a list of designated Trusted Flaggers in a publicly accessible database. No Trusted Flagger status has been awarded at this time, and therefore no Article 16 notices submitted by Trusted Flaggers were received during the reporting period.

2.3 Number of actions taken in response to Article 16 notices, broken down by actions based on legal grounds and actions based on policy grounds

Article 15(1), point (b)

Legal standards vary greatly by country/region. Content that violates a specific law in one country/region may be legal in others. Typically, Google removes or restricts access to content only in the country/region where it is deemed to be illegal. However, when content is found to violate Google's content or product policies or Terms of Service, Google may remove or restrict access globally.

When a legal notice is reviewed and the content violates our content policies, action may be taken on policy grounds. If the content does not violate our policies, Google may take action on legal grounds, in line with local laws (see Table 2.3.1 for breakdown by service). As a legal notice may contain one or more URLs for review, multiple actions may be taken as a result of a single notice received.

Table 2.3.1: Number of actions taken in response to Article 16 notices, by service and basis of the action[1]

Service	Actions taken on legal grounds	Actions taken on policy grounds
Maps	25,077	1,191
Play	36	57
Shopping	14	161
YouTube	32,522	5
Multi-Services[2]	107	96

Notes:

[1] More than one action can be taken on an Article 16 notice.

[2] Notices relating to advertisements that may appear across multiple Google services, including VLOPs, are included under Multi-Services.

2.4 Number of Article 16 notices processed by automated means

Article 15(1), point (b)

During the reporting period, YouTube processed 20,157 Article 16 notices by automated means (i.e., with no human involvement). Article 16 notices are not processed by automated means for any of the other VLOPs.

2.5 Median time needed to take action on content identified in Article 16 notices

Article 15(1), point (b)

Table 2.5.1 reflects the median time, in days, needed to take action on content identified in Article 16 notices for each service.

Table 2.5.1: Median time to take action on Article 16 notices, by service

Service	Median time to take action (days)
Maps	6
Play	<1
Shopping	3
YouTube	<1
Multi-Services[1]	<1

Notes:

[1] Notices relating to advertisements that may appear across multiple Google services, including VLOPs, are included under Multi-Services.

Section 3: Content moderation engaged in at Google's own initiative

3.1 Content Moderation at Google's own initiative
Article 15(1), point (c)

Across all products and services, we set clear [policies](#) for what is and is not acceptable on our platforms. These policies aim to ensure a safe and positive experience for our users and observe a high standard of quality and reliability for advertisers, publishers, and content creators alike.

Content policies establish the rules of the road for what content can be created, uploaded, sent, shared, and monetised. These policies are used to guide content moderation and enforcement actions on our products. Community Guidelines play an important role in maintaining a positive experience for everyone on our platforms no matter where they are in the world.

User data and developer policies provide rules for how developers interact with our products and services. They also describe the privacy and security requirements for handling user data to include the full spectrum of developer actions, like requesting, obtaining, using, and sharing data.

Monetised product guidelines are the policies and standards related to products Google earns revenue from and cover what can or cannot be monetised. These policies empower and protect users while promoting a thriving digital ecosystem that is safe and conducive to innovation and growth.

Content moderation actions taken at Google's "own initiative" are considered to be actions taken on content shown to or flagged by those in the EU because the content violates our policies, or where the content is illegal but action is not taken in response to an Article 9 order or Article 16 notice. These can encompass both proactive and reactive enforcement actions. Proactive enforcement takes place when Google employees, algorithms, or contractors flag potentially policy-violating content. Reactive enforcement takes place in response to external notifications, such as user policy flags or legal complaints.

To support information and content quality on our products and services, we take a wide range of enforcement actions to maintain a trusted experience for all. Enforcement actions differ from service to service.

3.1.1 Number of measures taken at Google's own initiative that affect the availability, visibility and accessibility of information provided by recipients of the service, broken down by type of illegal content or violation of terms and conditions

Article 15(1), point (c)

Google considers 'measures' as actions taken on moderated videos, URLs, listings, accounts and other content types which are of a policy-violative nature, or which are delisted as a result of applicable law. Tables 3.1.1.a through 3.1.1.j reflect the number of measures taken at Google's own initiative that affect the availability, visibility and accessibility of information provided by recipients of each service, broken down by the type of alleged illegal content or policy violation.

For each of these services except Shopping, we separately present the metrics relating to advertisements impressed on those services. The majority of the measures that Shopping takes happen before the content is shown publicly, and the actions may apply to both unpaid content (e.g., free listings) and advertisements. As such, Shopping cannot readily distinguish between unpaid content and advertisements in these metrics, therefore they are combined. In addition, for non-Shopping content, content moderation actions on advertisements that are taken before the advertisement is surfaced on a VLOSE or VLOP are not included in this report.

Google Search has a unique challenge in reporting a single level of granularity because it is a complex service that combines information from a wide range of different sources and systems, and presents information through many different formats (from web listings to dedicated Search features). Given the widely varying features and services offered in Google Search, the service's content policies and the nature of specific enforcement actions take place at varying levels of granularity. Therefore, the number of measures (actions) taken are reported alongside the following levels of granularity to reflect the scope of the actions:

- Domain Level Actions: Number of internet domains taken action on due to policy violations.
- Host Level Actions: Number of internet hostnames (or variants with common prefixes such as www) taken action on due to policy violations.
- URL Level Removals: Number of individual URLs removed due to legal or policy violations.
- URL Level Filtering: Number of times individual URLs were filtered algorithmically from Discover feeds based on Google's content policies.
- Incident Level Actions: Number of incidents originating from various reporting channels, which were actioned due to policy violations.
- Partner Feed Item Level Actions: Number of entities (URLs or images) taken action on in response to partner feeds providing "Things to Do" results that appear on Google Search.

Google services are wide-ranging and differ in their user bases, content hosted, services provided, and expectations for enforcement. Where feasible, the high-level categories identified by the European Commission for its DSA Transparency Database containing statements of reasons are used to group and report policy enforcement actions. However, some policies do not fully align with these high-level categories, and are thus reported using additional categories.

Table 3.1.1.a: Own initiative actions taken on Google Search, by type of illegal content or violation of terms and conditions[1, 2, 3, 4, 5]

Type of illegal content or violation of terms and conditions	Granularity	Number of own initiative measures taken
Data Defect[6]	Partner Feed Item Level Actions	3,199
Data Protection and Privacy Violations	URL Level Removals	7,176
Foreign Information Manipulation and Interference	Incident Level Actions	2
Healthcare and Medicine	Incident Level Actions	3
Illegal / Harmful Speech	URL Level Removals	1
	Incident Level Actions	1
Inappropriate and Unhelpful	Incident Level Actions	5
Intellectual Property Infringements	Host Level Actions	1,255
	URL Level Removals	30,456,401
Multiple Policy Violations	URL Level Filtering	4,508,249
	Partner Feed Item Level Actions	24
Non-consensual Behaviour	URL Level Removals	4,929
Online Bullying / Intimidation	Incident Level Actions	6
Pornography / Sexualised Content	URL Level Removals	30,084
	URL Level Filtering	32,887,155
	Incident Level Actions	19
Protection of Minors	URL Level Removals	1,378,633
Risk for Public Security	URL Level Removals	2
Scope of Platform Service	URL Level Filtering	4,836,895
	Incident Level Actions	98

Type of illegal content or violation of terms and conditions	Granularity	Number of own initiative measures taken
	Partner Feed Item Level Actions	167
Spam	Domain Level Actions	1,638,505
Unsafe and/or Illegal Products	Incident Level Actions	2
Violence	URL Level Filtering	19,615,160
	Incident Level Actions	14
Other	Domain Level Actions	23
	Incident Level Actions	2,020

Notes:

[1] Google Search is out-of-scope for Art 15(1)(b). However, actions taken on a legal basis in response to legal notices received about Search content are reported under Art 15(1)(c) for completeness. Policy violations may apply to some but not all Search products and features.

[2] For Google Search, only some of these actions result in complete removal from search results (e.g. "non-consensual behaviour" includes delistings under Google's policies relating to highly personal information.) Others apply only to certain Search features, such as Discover, Knowledge Graph or Featured Snippets, where prominently surfacing content might cause undue surprise to users (e.g., "violence"). Others involve the application of a ranking signal, for example, applying a demotion to domains that receive a high volume of valid copyright removal notices.

[3] Most, but not all, of policy-violating content on Google Search and its features is moderated globally. Most content delisted from Google Search on legal grounds is content subject to copyright removal notifications, which are also processed globally. However, there are classes of delistings based on local law or local court orders that affect only certain country services, based on variance in laws between countries.

[4] Google uses a variety of automated tools to provide a secure environment for users including Safe Browsing technology. This technology examines billions of URLs per day to identify malware and phishing sites and notify users and webmasters so they can protect themselves from harm. The number of URLs added to Safe Browsing block lists are not included in the metrics above.

[5] Removal actions from Google Search do not remove content from publishers' sites, but only prevent the content from being included in search results.

[6] All Data Defect items were related to "Things to Do" search results.

Table 3.1.1.b: Own initiative actions taken on advertisements presented on Google Search, by type of illegal content or violation of terms and conditions

Type of illegal content or violation of terms and conditions	Number of own initiative measures taken
Healthcare and Medicine	218,201
Intellectual Property Infringements	514,579
Pornography / Sexualised Content	57,620
Scams and/or Fraud	418,425
Scope of Platform Service	867,635
Unsafe and/or Illegal Products	69,051
Total	**2,145,511**

Table 3.1.1.c: Own initiative actions taken on Google Maps, by type of illegal content or violation of terms and conditions

Type of illegal content or violation of terms and conditions	Number of own initiative measures taken
Data Defect	2,943
Data Protection and Privacy Violations	62,656
Illegal / Harmful Speech	81,547
Inappropriate and Unhelpful	588,637
Multiple Policy Violations	24
Pornography / Sexualised Content	31,297
Risk for Public Security	3
Scams and/or Fraud	772,614
Scope of Platform Service	412,356
Unsafe and/or Illegal Products	1

Type of illegal content or violation of terms and conditions	Number of own initiative measures taken
Violence	68
Total	**1,952,146**

Table 3.1.1.d: Own initiative actions taken on advertisements presented on Google Maps, by type of illegal content or violation of terms and conditions

Type of illegal content or violation of terms and conditions	Number of own initiative measures taken
Healthcare and Medicine	4,072
Intellectual Property Infringements	10,252
Pornography / Sexualised Content	264
Scams and/or Fraud	9,214
Scope of Platform Service	44,063
Unsafe and/or Illegal Products	3,911
Total	**71,776**

Table 3.1.1.e: Own initiative actions taken on Google Play, by type of illegal content or violation of terms and conditions

Type of illegal content or violation of terms and conditions	Number of own initiative measures taken
Data Defect	638
Data Protection and Privacy Violations	12,633
Illegal / Harmful Speech	1,083
Inappropriate and Unhelpful	89,110
Negative Effects on Civic Discourse / Elections	85
Non-consensual Behaviour	18

Type of illegal content or violation of terms and conditions	Number of own initiative measures taken
Online Bullying / Intimidation	12
Pornography / Sexualised Content	340
Protection of Minors	615
Risk for Public Security	9
Scams and/or Fraud	1,229
Scope of Platform Service	4,565
Spam	352,931
Violence	5
Other	185
Total	463,458

Table 3.1.1.f: Own initiative actions taken on advertisements presented on Google Play, by type of illegal content or violation of terms and conditions

Type of illegal content or violation of terms and conditions	Number of own initiative measures taken
Healthcare and Medicine	55
Intellectual Property Infringements	4
Scams and/or Fraud	328
Scope of Platform Service	1,430
Unsafe and/or Illegal Products	331
Total	2,148

Table 3.1.1.g: Own initiative actions taken on Shopping (unpaid content and advertisements)[1], by type of illegal content or violation of terms and conditions

Type of illegal content or violation of terms and conditions	Number of own initiative measures taken
Animal Welfare	127,285
Data Defect	159,004,187
Healthcare and Medicine	146,112
Illegal / Harmful Speech	306
Intellectual Property Infringements	32,119
Negative Effects on Civic Discourse / Elections	295
Online Bullying / Intimidation	1
Pornography / Sexualised Content	6,890,349
Protection of Minors	1,438
Risk for Public Security	4,000
Scams and/or Fraud	41,725
Scope of Platform Service	14,072,178
Spam	40
Unsafe and/or Illegal Products	1,534,894
Violence	1
Other	387
Total	**181,855,317**

Notes:

[1] Shopping metrics reflect content moderation actions taken on both unpaid content (e.g., free listings) and advertisements.

Table 3.1.1.h: Own initiative actions taken on YouTube, by type of illegal content or violation of terms and conditions

Type of illegal content or violation of terms and conditions	Number of own initiative measures taken
Ad Friendly Guidelines Violation	33,459
Age Restricted	325,290
Channel-level Termination Video Removals[1]	469,889
Child Safety	40,249
Harassment / Cyberbullying	3,933
Harmful / Dangerous	13,288
Hateful / Abusive	4,522
Misinformation	2,474
Nudity / Sexual	6,320
Promotion of Violence and Violent Extremism	2,232
Violent / Graphic	2,958
Other	30,671
Total	**935,285**

Notes:

[1] This reflects the number of videos removed from the YouTube platform when the associated YouTube channel was terminated.

Table 3.1.1.i: Own initiative actions taken on advertisements presented on YouTube, by type of illegal content or violation of terms and conditions

Type of illegal content or violation of terms and conditions	Number of own initiative measures taken
Healthcare and Medicine	71,155
Intellectual Property Infringements	163,576
Pornography / Sexualised Content	2,298

Type of illegal content or violation of terms and conditions	Number of own initiative measures taken
Scams and/or Fraud	204,616
Scope of Platform Service	338,683
Unsafe and/or Illegal Products	27,685
Total	**808,013**

Table 3.1.1.j: Own initiative actions taken on Multiple Services, by type of illegal content or violation of terms and conditions

Type of illegal content or violation of terms and conditions	Number of own initiative measures taken[1]
Protection of Minors	2,079
Scams and/or Fraud	2,757
Spam[2]	115,212
Total	**120,048**

Notes:

[1] These measures reflect Google-wide account-level terminations (i.e., termination of access to all Google products and services).

[2] This metric does not refer to any possible actions taken to combat webspam in Google Search.

3.1.2 Number of measures taken at Google's own initiative that affect the availability, visibility and accessibility of information provided by recipients of the service, broken down by detection method

Article 15(1), point (c)

Table 3.1.2 reflects the number of measures taken on violative content, broken down by service and detection method, which can be:

(i) Self-detection, where Google employees, algorithms, or contractors flag potentially illegal or policy violating content;
(ii) External detection in response to an external notification (e.g., user policy flags or legal complaints); or
(iii) Unknown due to system limitations.

Table 3.1.2: Number of measures taken at Google's own initiative, by service and detection method

Service	Number of measures taken on		
	Violative content that was self-detected	Violative content that was detected externally	Violative content where detection method was unknown
Search			
Domain Level Actions	1,638,528	0	N/A[1]
Host Level Actions	1,255	0	N/A[1]
URL Level Removals	1,370,062	30,507,164	N/A[1]
URL Level Filtering	61,847,459	0	N/A[1]
Incident Level Actions	335	1,835	N/A[1]
Partner Feed Item Level Actions	3,390	0	N/A[1]
Ads on Search	1,630,713	514,798	N/A[1]
Maps	1,935,749	16,397	0
Ads on Maps	61,529	10,247	N/A[1]
Play	461,502	1,059	897
Ads on Play	1,812	336	N/A[1]

Service	Number of measures taken on		
	Violative content that was self-detected	Violative content that was detected externally	Violative content where detection method was unknown
Shopping[2]	181,823,192	32,125	N/A[1]
YouTube[3]	424,790	7,147	33,459
Ads on YouTube	642,445	165,568	N/A[1]
Multi-Services	N/A[1]	N/A[1]	120,048

Notes:

[1] N/A indicates that this is not an applicable outcome for this service.

[2] Shopping metrics reflect content moderation actions taken on both unpaid content (e.g., free listings) and advertisements.

[3] Consistent with Table 3.1.1.h, YouTube also removed 469,889 videos as the result of their associated channel's termination. There is no detection method associated with these actions, therefore these are excluded from the table above.

3.1.3 Number of measures taken at Google's own initiative that affect the availability, visibility and accessibility of information provided by the recipients of the service, broken down by type of restriction applied

Article 15(1), point (c)

Table 3.1.3 provides the number of measures taken on violative content, broken down by service and the type of restriction applied. The type of restrictions include:

(i) restrictions of visibility of content;
(ii) demonetisation;
(iii) partial service-level suspension; and
(iv) service-specific or Google-wide account-level termination.

Account level restrictions may be imposed as a result of multiple legal or policy violations across one or more services. Where possible, these restrictions are attributed to the service associated with the final violation that led to the restriction being imposed.

Table 3.1.3: Number of measures taken at Google's own initiative, by service and type of restriction applied

Service	Number of measures taken			
	Restrictions of the visibility of content	Demonetisation	Partial service level suspension	Account level termination[1]
Search				
Domain Level Actions	1,638,528	N/A[3]	N/A[4]	N/A[5]
Host Level Actions	1,255	N/A[3]	N/A[4]	N/A[5]
URL Level Removals	31,877,226	N/A[3]	N/A[4]	N/A[5]
URL Level Filtering	61,847,459	N/A[3]	N/A[4]	N/A[5]
Incident Level Actions	2,170	N/A[3]	N/A[4]	N/A[5]
Partner Feed Item Level Actions	3,390	N/A[3]	N/A[4]	N/A[5]
Ads on Search	2,137,206	N/A[3]	N/A[4]	8,305
Maps	1,950,211	N/A[3]	1,935	N/A[5]
Ads on Maps	71,688	N/A[3]	N/A[4]	88
Play	461,409	N/A[3]	572	1,477
Ads on Play	2,134	N/A[3]	N/A[4]	14
Shopping[6]	181,844,046	N/A[3]	N/A[4]	11,271
YouTube	871,465	33,973	45	29,802
Ads on YouTube	804,859	N/A[3]	N/A[4]	3,154
Multi-Services	N/A[2]	N/A[3]	N/A[4]	120,048

Notes:

[1] Service-specific account level terminations, where users are prevented from using the account for the service's main purpose, are reflected in the numbers for each service. The number of Google-wide account-level terminations, where users can no longer log into any Google products or services is reflected in Multi-Services.

[2] N/A indicates that restrictions of the visibility of content is not an applicable enforcement action for Multi-Services.

3.2 Google's use of automated tools
Article 15(1), point (e); Article 42(2), point (c)

To enforce our policies at scale, Google relies on a combination of automated and human tools to spot problematic content. While automated systems can quickly identify and take action against spam and some violative content, human judgement is needed for the many decisions that require a more nuanced determination. The context in which a piece of content is created or shared is an important factor in any assessment about its quality or its purpose. Google is attentive to educational, scientific, artistic, and documentary contexts, including journalistic intent, where the content might otherwise violate our policies. Google escalates particularly complex cases to specially-trained experts.

Additionally, Google uses the corpus of already human-reviewed and removed content to train machine learning technology to flag new content that might also violate product and service policies. Using machine learning technology trained by human decisions means the enforcement systems adapt and get smarter over time.

This section describes how Google uses automated tools, often supplemented with human review, for content moderation, along with the indicators of accuracy of any fully automated tools. While we report fully automated tools primarily on a language-agnostic basis, where applicable and feasible for this reporting period, the indicators of accuracy are broken down by language.

3.2.1 Automated tools that affect multiple services

Automated tools used to process for Legal-related Content Removal Requests
Automation plays a role in legal content moderation to help Google work at scale, and focus our efforts on actionable, authentic requests. There are a few ways that automation might be used while handling a removal request. The most common way is that Google uses automation to route a request to the right team. Google has subject matter experts in different types of content and languages, and using automation ensures the request is sent to the people best positioned to review it.

Once content removal requests are routed efficiently, Google also uses automation to manage the millions of URLs (web page addresses) that are sent to Google for review every day, and to complement and streamline human review. As an example, Google receives a significant number of Google Search removal requests for URLs that are not included in Google's search index, which is the vast and continuously updated pool of web page addresses from which all search results are drawn. We have

automated systems that detect such URLs in removal requests, enabling our teams and processes to focus on content that does appear on our services and address complex matters requiring human review.

Google also uses automation to process some legal notices. The vast majority of notices are copyright removal requests, largely from submitters with a well-established track record of submitting valid requests, allowing Google to be relatively confident in automating this processing. During the reporting period, >99.99% of all fully automated removal decisions on Web Search that impacted users based in the EU were unchanged, while <0.01% were reinstated as a result of a counter notice.

Automated Tools used to combat Child Sexual Abuse Material (CSAM)

Google takes its responsibility to fight child sexual abuse and exploitation online very seriously. We do this by combatting CSAM across Google's products and by detecting instances of abuse and enforcing robust policies. We also partner with non-governmental organisations (NGOs) and others in industry to share proprietary technology and drive the industry forward.

Built-in protections help prevent Google products from showing abusive content and deter bad actors. For example, Google deploys safety by design principles to deter users from seeking out CSAM on Google Search. It is our policy to block search results that lead to child sexual abuse imagery or material that appears to sexually victimise, endanger or otherwise exploit children. We are constantly updating our algorithms to combat these evolving threats. We apply extra protections to searches that we recognise as seeking CSAM content. We filter out explicit sexual results if the search query seems to be seeking CSAM. For queries seeking adult explicit content, Google Search won't return imagery that includes children, to break the association between children and sexual content. In many countries, users who enter queries clearly related to CSAM are shown a prominent warning that child sexual abuse imagery is illegal, with information on how to report this content to trusted organisations. When these warnings are shown, we have found that users are less likely to continue looking for this material.

In order to detect and report CSAM, we may use a combination of cutting-edge technology, including machine learning classifiers (to identify unknown CSAM) and hash-matching technology, as well as trained specialist teams. Hash-matching technology creates a "hash", or unique digital fingerprint, for an image or a video so it can be compared with hashes of known CSAM. When Google finds CSAM, our services remove it, report it to the National Center for Missing and Exploited Children (NCMEC), and take action, which may include disabling the account.

Google scales its impact by collaborating with NCMEC and partnering with NGOs and industry coalitions to help grow and contribute to a joint understanding of the evolving nature of child sexual abuse and exploitation. One of the ways Google contributes is by creating and sharing free tools to help other organisations prioritise potential CSAM images for human review. For example, Google's Child Safety Toolkit consists of two APIs. The first is CSAI Match, an API developed by YouTube that partners can use to automatically detect known videos of CSAM so they can flag for review, confirm, report, and act on it. The second is Google's Content Safety API that helps partners classify and prioritise novel potentially abusive images and videos for review. Detection of never-before-seen CSAM helps the child

safety ecosystem by identifying child victims in need of safeguarding and contributing to the list of known digital fingerprints to grow our abilities to detect known CSAM.

Google takes action not just on illegal CSAM, but also wider content that promotes the sexual abuse and exploitation of children and can put children at risk.

Automated tools that affect advertisements

Advertisements can appear across multiple VLOP and VLOSE services. To keep ads safe and appropriate for everyone, ads are reviewed to make sure they comply with Google Ads policies.

Google uses a combination of automated and human evaluation to detect and remove ads which violate our policies and are harmful to users and the overall ecosystem. Our enforcement technologies may use automated evaluation, modelled on human reviewers' decisions, to help protect our users and keep our ad platforms safe. The policy-violating content is either removed by automated means or, where a more nuanced determination is required, it is flagged for further review by trained operators and analysts who conduct content evaluations that might be difficult for algorithms to perform alone, for example because an understanding of the context of the ad is required. The results of these manual reviews are then used to help build training data to further improve our machine learning models.

When reviewing ad content or advertiser accounts to determine whether they violate our policies, Google takes various information into consideration when making a decision, including the content of the creative (e.g. ad text, keywords, and any images and video) as well as the associated ad destination. Google also considers account information (e.g., past history of policy violations) and other information provided through reporting mechanisms (where applicable) in our investigation.

During the reporting period, <2% of Google's fully automated enforcement decisions on ads placed by advertisers in the EU were overturned after subsequently undergoing human review.

3.2.2 Google Search

Google Search relies on a combination of people and technology to enforce Google Search policies. Machine learning, for example, plays a critical role in content quality on Google Search. Google Search systems are built to identify and balance signals of authoritativeness so people can find the most reliable and timely information available. Google Search algorithms look at many factors and signals to raise authoritative content and reduce low quality content. Google Search's publicly available website, How Search Works, explains the key factors that help determine which results are returned for a query. Furthermore, our systems are designed not to surface content that violates our content policies, while also aiming to provide access to an open and diverse ecosystem. Google Search works continuously to improve the effectiveness of automated systems to protect platforms and users from harmful content.

To ensure our algorithms meet high standards of relevance and quality, Google Search has a rigorous process that involves both live tests and thousands of trained external Search Quality Raters from around the world. Raters do not determine the ranking of an individual, specific page or website, but they help to benchmark the quality of Google Search's results so that Google Search can meet a high

bar for users all around the world. Under the Google Search Quality Rater Guidelines, raters are instructed to assign the lowest rating to pages that are potentially harmful to users or specified groups, misleading, untrustworthy, and spammy.

Google Search is providing precision metrics for the automated processes outlined below. Precision metrics included in this section are defined as the ratio of true positive instances (i.e., correct automated decisions) as a proportion of both true and false positives.

In addition to using automated processes related to CSAM discussed above, Google Search uses automated measures to detect racy, commercial, violent, and profane content globally on its 'Discover' feature. Precision of each automated process during the reporting period (28 August 2023 to 10 September 2023) was as follows:

- Detection of violative racy content, globally: 85%;
- Detection of violative commercial content, globally: 73%;
- Detection of violative violent content, globally: 90%; and
- Detection of violative profane content, globally: 85%.

Google Search also uses automated measures to detect webspam content. Webspam is defined as irrelevant or useless websites that exploit search engine algorithms to appear as relevant results, or pages that engage in abusive behaviour to manipulate search engine rankings, thereby inhibiting search engines from providing high quality results to users. Between 2017 and 2023, Google Search launched multiple, new automated processes that detect webspam content. The typical precision of these processes is approximately 99%.

3.2.3 Google Maps

Google's content policies for Maps user-generated content (UGC) are designed to help ensure that everyone viewing UGC has a positive experience and to keep Maps fair and honest. While most of the millions of contributions Google Maps receives each day are authentic and accurate, we sometimes receive policy-violating content.

To detect this policy-violating content, Maps' machine-learning algorithms scan contributions for signals of suspicious user activity. The policy-violating content is either removed by automated models or flagged for further review by trained operators and analysts who conduct content evaluations that might be difficult for algorithms to perform alone.

To protect users from finding inappropriate content, Maps deploys many other protections, such as suspending UGC for specific places, geographic areas and categories of places. These measures may be deployed reactively to counteract a spike in content that violates our policies, or proactively if Maps believes that these measures are necessary to prevent content that violates our policies. Maps may also restrict feature access or suspend Google accounts that violate our policies. Removing content, rejecting edits or restricting feature access may include preventing uploaded content from being displayed to other users.

Google Maps is reporting a single accuracy metric for each of the 26 official European Economic Area (EEA) Member State languages, as well as an overall accuracy metric across all automated content moderation decisions that is language-agnostic.

For each metric, accuracy is computed based on human evaluation of a random sample of all user contributions, across data types and content types (e.g., reviews, media, facts, etc.) between 1 March 2023 and 31 August 2023. The accuracy for that slice is then defined as the percentage of correct decisions made by the automated system, assuming the human evaluation is the ground truth.

The accuracy of all automated content moderation decisions affecting EEA users on Google Maps between 1 March 2023 and 31 August 2023 was 90% (95% confidence interval: 84% to 94%). Accuracy by EEA Member State language is provided in Table 3.2.3.

Table 3.2.3: Accuracy of automated measures on Google Maps, by EEA Member State language

Member State Language	% Accuracy (95% Confidence Interval)
Bulgarian	88% (81% – 92%)
Croatian	92% (86% – 96%)
Czech	96% (92% – 99%)
Danish	92% (86% – 96%)
Dutch	94% (89% – 97%)
English	88% (81% – 92%)
Estonian	91% (85% – 95%)
Finnish	95% (90% – 98%)
French	96% (91% – 98%)
German	91% (85% – 95%)
Greek	90% (84% – 94%)
Hungarian	93% (87% – 96%)
Icelandic	89% (83% – 93%)
Irish	85% (78% – 90%)
Italian	96% (91% – 98%)

Member State Language	% Accuracy (95% Confidence Interval)
Latvian	87% (80% - 92%)
Lithuanian	89% (83% - 93%)
Maltese	85% (78% - 90%)
Norwegian	87% (80% - 92%)
Polish	87% (80% - 92%)
Portuguese	91% (84% - 95%)
Romanian	84% (77% - 89%)
Slovak	91% (85% - 95%)
Slovene	90% (84% - 94%)
Spanish	95% (90% - 98%)
Swedish	91% (85% - 95%)

3.2.4 Google Play

Google Play uses a combination of human and automated evaluation to review apps and app content to detect and assess content which violates our policies and is harmful to users and the overall Google Play ecosystem. Using automated models helps us detect more violations and evaluate potential issues faster, which helps us better protect our users and developers. The policy-violating content is either removed by Google Play's automated models or by trained operators and analysts. The results of these manual reviews are then used to help build training data to further improve our machine learning models.

Developers are also able to appeal automated enforcement actions on Google Play apps. During the reporting period, <0.2% of all automated enforcement actions were reversed following a successful appeal submitted by EU developers as the original action was found to have occurred in error.

User reviews of Google Play apps also go through automated review processes to determine if the user review violates the user comment posting policies (e.g., contains hate speech, sexually explicit content, spam, etc.). This automated model's precision at the time of launch was >90%, globally, and is monitored. Should there be a performance outlier, the rule is re-evaluated and adjusted as needed.

3.2.5 Shopping

Products and merchants go through in-depth safety reviews before they can list on Google. Thanks to features such as the Shopping Graph (Shopping's data set of the world's products and sellers), Shopping's systems can quickly review whether a business is legitimate, and whether the products and other content follow Shopping's policies. This automated vetting process has helped to more efficiently and accurately review a massive amount of products.

Shopping's automated systems are always monitoring for violating activity. Some examples of automated content moderation processes used include:

- policy checks for harmful, regulated, or illegal content (e.g., weapons, recreational and prescription drugs, tobacco products);
- product image checks for policy violations such as graphic overlays or nudity;
- product data quality checks;
- landing page checks; and
- checks for recalled products such as those listed in the Rapid Exchange of Information System (RAPEX) or Organisation for Economic Co-operation and Development (OECD) public databases.

During the reporting period, <0.01% of all automated content moderation actions on Shopping were appealed by content or account owners based in the EU and consequently <0.01% of all original content moderation actions were overturned. Of the relatively few original content moderations that were appealed and subsequently closed within the reporting period, 83% were overturned.

3.2.6 YouTube

YouTube continues to invest in automated detection systems, and rely on both human evaluators and machine learning to detect and take action on problematic content at scale while simultaneously training our systems on new data. The vast majority of content reviewed and enforced on YouTube is first detected by automated systems. However, after potentially violative content has been detected by automated systems, content moderators may review the content to confirm the decision. As models continuously learn and adapt based on content moderator feedback, this collaborative approach helps improve the accuracy of these models over time. It also means that the enforcement systems can manage the scale of content that is uploaded to YouTube (over 500 hours of content every minute), while still digging into the nuances that determine whether a piece of content is violative.

Some examples of how **YouTube** uses automated processes for content moderation include:

1. **Flagging, removing, or restricting inappropriate content:** YouTube uses smart detection technology to detect content that may violate YouTube's policies and sends it for human review. In some cases, that same technology automatically takes an action, which could include removing or restricting content (e.g., age-restrict content not suitable for all audiences), limiting content's monetisation eligibility, or applying a strike to a channel.

2. **Identifying copyright-protected content:** Content ID, YouTube's automated content identification system, identifies copyright-protected content on YouTube. Videos uploaded to YouTube are scanned against a database of audio and visual reference files submitted to YouTube by copyright owners. A Content ID claim is automatically generated on behalf of a copyright owner when an uploaded video matches another video or audio reference file (in whole or in part) in YouTube's Content ID system. Depending on the copyright owner's Content ID settings, Content ID claims can:
   - Block a video from being viewed in one or more territories;
   - Enable revenue-sharing with the copyright owner based on the video's earnings;
   - Provide the video's viewership statistics to the copyright owner.

   YouTube only grants Content ID to copyright owners who meet specific criteria. More information about How Content ID works is available [here](#).

3. **Preventing re-uploads of known violative content:** YouTube utilises technology to prevent re-uploads of known violative content as quickly as possible. For example, YouTube leverages hashes (or "digital fingerprints") to detect and automatically remove child sexual abuse imagery (CSAI) videos on YouTube. YouTube has long used this technology to prevent the spread of violative content like CSAI or terrorist content. More information is available [here](#).

To improve the accuracy of our automated systems and understand what investments to make in machine learning, YouTube evaluates the amount of violative content that gets viewed before it is detected by automated technology and removed.

YouTube strives to prevent content that violates our policies from being widely viewed—or viewed at all—before it is removed. As the overwhelming majority of violative content is detected by automated systems, YouTube's Violative View Rate (VVR) is a good indication of how well our automated systems are protecting our community. VVR is an estimate of the proportion of video views that violate our Community Guidelines in a given quarter (excluding spam). In order to calculate VVR, we take a sample of the views on YouTube and send the sampled videos for review. Once we receive the decisions from reviewers about which videos in the sample are violative, we aggregate these decisions in order to arrive at our estimate. In Q2 2023, VVR was 0.09-0.10% globally. This means that out of every 10,000 views on YouTube in Q2, only 9-10 came from violative content. Additional information about the VVR methodology is available in the YouTube Community Guidelines enforcement [transparency report](#) and a third-party [statistical assessment](#) commissioned by Google.

3.3 Human Resources involved in Content Moderation

Article 42(2), points (a) and (b)

Human reviewers or content moderators play a key role in content moderation at Google. Although technology has become very helpful in identifying some kinds of problematic content (e.g., finding objects and patterns quickly and at scale in images, video, and audio), humans are able to apply a more

nuanced approach to assessing content. For example, algorithms cannot always tell the difference between terrorist propaganda and human rights footage or hate speech and provocative comedy.

In order to safeguard against content actions that could potentially contribute to or exacerbate adverse impacts due to allowing or removing content, Google utilises international human rights standards to guide policy and enforcement decision-making, considering how content could adversely impact the rights of an individual, community, or society as a whole, or further the understanding of social, political, cultural, civic, and economic affairs. As an example of public interest-informed content moderation, Google carves out exceptions to enforcement guidelines for material that is Educational, Documentary, Scientific, and/or Artistic (EDSA). Content that falls under those exceptions are crucial to understanding the world and to chronicling history, whether it is documenting wars and revolutions, or artistic expression that may include nudity. Consequently, Google takes great care in helping reviewers understand the EDSA exceptions when reviewing flagged content.

Google strives to create workplaces and economic opportunities that work for employees, as well as vendors, temporary staff, and independent contractors. While Google does not employ all of the individuals who contribute to content moderation, Google is [committed](link) to ensuring that work on Google products is conducted in environments that treat all workers with respect and dignity, ensure safe working conditions, and conduct responsible, ethical operations. For that reason, Google seeks out suppliers that embrace its [values](link), commitment to [human rights](link), and that support a safe working environment. Suppliers must operate in accordance with our [Supplier Code of Conduct](link), and comply with all applicable labour protection laws, including those related to privacy, safety, health, and wages. Google also provides [Wellness Standards](link) that promote healthy working conditions for provisioned extended workforce members performing sensitive content moderation.

Qualifications and linguistic expertise

Qualifications for Google employees who work on sensitive content may include role related knowledge in the content matter, professional experience in content moderation or sensitive workflows, linguistic expertise and computer proficiency. The linguistic expertise required varies depending on the specific workflow of a product or service, the type of content, and languages that content is available in. Some products or services require native proficiency in global supported languages, others may use translation tools if necessary, and some videos or images do not require any language proficiency in order to review. Some Google employees who work on sensitive content are also subject matter specialists skilled in specialty areas, such as child sexual abuse material or violent extremism.

Onboarding and training

Google employees that work on sensitive content teams are offered subject matter specific training on a variety of topics. Employees working in sensitive content are required to complete a training on the Psychological Impact of Sensitive Content Review at the point of onboarding, and managers are required to complete an additional training on Supporting Teams who Work with Sensitive Content. Additional optional training opportunities include those on self-compassion, emotional agility, and subject matter specific training to provide a deeper dive into the unique challenges faced by each team. The training is generally conducted via e-learning with opportunities for live facilitated training.

Wellbeing support

Google is committed to supporting the wellness of its employees that work with sensitive content through comprehensive programs and resources. Google strives for safe and healthy working conditions for all employees exposed to sensitive content and is committed to ensuring they have the highest standard of support. Google has invested significantly in these teams by:

- Providing access to on- and off-site counselling for workers who need it, dedicated wellness spaces, on-site specialist counsellor support in certain Google offices, and 24/7 phone support;
- Limiting content exposure for those focusing on sensitive content by providing guidance on daily review time limits;
- Providing peer-led peer support groups and optional listening sessions if teams experience escalations or specific events that are particularly impactful;
- Providing physical and mental wellbeing activities (e.g., gym space, workout classes, mindfulness app access, educational sessions on a variety of topics);
- Providing post-exit mental health support, including counselling services, for one year after an employee who was regularly exposed to sensitive content and situations as part of their core role exits their position at Google.

Research and technological innovation

In addition to gathering feedback directly from workers and soliciting professional input and advice, Google is committed to driving industry-leading research and technological innovation in the field of content moderation. For instance, Google published a research paper in 2019 indicating that "grayscale transformations" (i.e., where an image was converted to black and white) reduced the emotional impact of reviewing violent and extremist content. Based on these findings, Google built grayscaling into review tools, giving each reviewer an option to use this feature when performing reviews, based on their own preference.

3.3.1 Human resources evaluating content across the official EU Member State languages

Article 42(2)

Identifying the human resources who evaluate content across Google services is a highly complex process. Content moderators may review content for multiple policy violations or focus on one specific topic; they may review content that appears across one or more services; and content assigned for their review may have been posted in several different languages. In some cases and where appropriate, translation tools may be used to assist in the review process and allow us to moderate content 24/7 and at scale.

Table 3.3.1 reflects the human resources evaluating content across the official EU Member State languages, for each service. For Google Maps, the metric presented is the number of content moderators whose primary language of review is an EU language and who reviewed at least one piece of content during the period from 28 August 2023 to 10 September 2023. For YouTube, the metric reflects the number of content moderators who reviewed at least 10 videos posted in an official EU

Member State language between 1 January 2023 and 30 June 2023. The remaining metrics reflect the number of content moderators who completed at least one review of content posted in an official EU language during the period from 28 August 2023 to 10 September 2023. The metrics in Table 3.3.1 do not necessarily reflect the language in which the content was ultimately reviewed; and with the exception of Google Maps, they do not represent the number of content moderators hired to review in each official EU language and should not be aggregated as this may not reflect the total number of unique content moderators available to conduct reviews.

Table 3.3.1: Human resources evaluating content across the official EU Member State languages, by service[1, 2]

Member State Language	Human resources evaluating content				
	Maps[3]	Play[4, 5]	Shopping[4, 5]	Multi-Services[4, 5, 6]	YouTube[5, 7]
Bulgarian	0	15	16	1,478	9
Croatian	0	25	17	1,279	24
Czech	0	69	156	1,830	31
Danish	0	57	128	1,649	9
Dutch	0	131	188	2,419	24
English	2,797	2,290	632	6,402	15,142
Estonian	0	9	12	741	7
Finnish	0	38	94	1,290	15
French	2	396	236	3,222	176
German	3	302	204	3,350	231
Greek	0	24	104	1,545	28
Hungarian	0	37	123	1,398	25
Irish	0	0	0	7	0
Italian	2	184	175	2,651	91
Latvian	0	4	13	704	11
Lithuanian	0	26	14	951	11

Member State Language	Human resources evaluating content				
	Maps[3]	Play[4, 5]	Shopping[4, 5]	Multi-Services[4, 5, 6]	YouTube[5, 7]
Maltese	0	0	0	22	0
Polish	0	169	151	2,558	99
Portuguese	3	328	196	3,784	464
Romanian	0	35	117	2,067	34
Slovak	0	26	100	1,234	5
Slovene	0	11	12	904	15
Spanish	2	520	218	4,227	507
Swedish	0	68	137	1,765	16
Unknown[8]	0	2,555	861	7,112	N/A[9]
Agnostic[10]	0	0	61	213	N/A[9]

Notes:

[1] Content can be posted by users globally or reviewed by content moderators located globally. A single content moderator can be assigned content posted in several different languages for review. In some cases and where appropriate, translation tools may be used to assist in the review process. Accordingly, these metrics do not necessarily reflect the number of content moderators who speak each EU Member State language.

[2] With the exception of Google Maps, metrics relate to content posted in official EU languages. The number of content moderators who completed reviews of content posted in each language should not be aggregated as this may not reflect the total number of unique content moderators available to conduct reviews.

[3] For Google Maps, the metrics reflect the number of content moderators whose primary language of review is an EU language and who reviewed at least one piece of content during the period from 28 August 2023 to 10 September 2023.

[4] For Play, Shopping and Multi-Services, the metrics reflect the number of content moderators who completed at least one review of content posted in an official EU language during the reporting period (28 August 2023 to 10 September 2023).

[5] Metrics do not represent the number of content moderators hired to review in each official EU language.

[6] Content moderators who review content that may have appeared across multiple Google services, but not necessarily a VLOP, are included under Multi-Services.

[7] YouTube metrics cover the period from 1 January 2023 to 30 June 2023 and reflect the number of content moderators who reviewed at least 10 videos posted in an EU language.

[8] Content moderators who review content posted in a less common language (e.g., Breton, Basque, Occitan, Catalan, or Corsican) or in multiple languages, or where there are limitations in reporting the content language, are included in the 'Unknown' category.

[9] N/A indicates that this is not an applicable category for YouTube.

[10] Content moderators who review non-language content (e.g., an image) are included in the 'Agnostic' category.

Content Moderation at Google: an Illustrative Case Study

Max is a Google employee who works in sensitive content and specialises in violent extremism as a Content Specialist*. Max graduated from the University of Amsterdam with a degree in Communications. He has three years of experience working in content moderation on sensitive workflows. He is fluent in English, Spanish, Dutch and French.

Between 28 August 2023 and 10 September 2023, Max conducted 10 reviews of English-language content, 15 reviews of Spanish-language content, and 7 defamation image reviews. As Max reviews content that may have appeared across multiple Google services, including VLOPs, his reviews would be included in the "Multi-Services" category in Table 3.3.1. Max would be counted as 1 of the reviewers who conducted reviews of content in the English, Spanish and Agnostic categories.

Max is a valuable asset to the Google team, and he is dedicated to keeping the platform safe through reviewing policy escalations. The work he does is essential to keeping digital platforms safe for everyone. However, some of the content he reviews can be emotionally taxing. During Max's onboarding training, he learned about the resources available to him, such as access to counselling, limiting his exposure to content, physical well-being activities and quiet rooms, and post-exit support should he decide to leave his position at Google.

While Max is a hypothetical content moderator, his training and experience are typical of many policy escalation specialists at Google.

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

Article 15(1), point (d)

Sometimes, we make mistakes in our decisions to enforce our policies, which may result in the unwarranted removal of content from or access to our services. To address that risk, where appropriate, we make it clear to creators that we have taken action on their content and provide them the opportunity to contest that decision through designated complaint-handling systems and give us clarifications. In addition, under the DSA, EU users can submit complaints about an action that Google did not take in response to a notice/flag that they previously submitted.

4.1 Number of complaints received

Article 15(1), point (d)

Table 4.1.1 reflects the number of content moderation complaints received from creators and users located in EU Member States during the reporting period, broken down by service.

Table 4.1.1: Number of complaints received, by service

Service	Number of complaints received
Search	642
Ads on Search	179,151
Maps	130,694
Ads on Maps	3,873
Play	2,357
Ads on Play	2,932
Shopping[1]	89,749
YouTube	37,446
Ads on YouTube	75,426
Multi-Services[2]	50,535

Notes:

[1] Shopping metrics reflect complaints relating to both unpaid content (e.g., free listings) and advertisements.

4.2 Number of complaints, broken down by complaint reason

Article 15(1), point (d)

Table 4.2.1 reflects the number of complaints, broken down by service and by the complaint reason (i.e., on the basis that either action was taken against the content or account, or Google did not action the request to remove content or disable access). For complaints on the basis of action taken, the appellant is likely to be the content or account owner whereas for complaints on the basis of non-action taken, the complainant is likely to be the individual or entity who originally flagged the content as potentially violative.

Table 4.2.1: Number of complaints received, by service and complaint reason

Service	Number of complaints received on	
	Basis of action taken against the content or account	Basis that the request to remove content or disable access was not actioned
Maps	130,651	43
Ads on Maps	3,873	0
Play	2,356	1
Ads on Play	2,932	0
Shopping[1]	89,714	35
YouTube	37,216	230
Ads on YouTube	75,124	302
Multi-Services[2]	50,520	15

Notes:

[1] Shopping metrics reflect complaints about content moderation actions taken on both unpaid content (e.g., free listings) and advertisements.

[2] The Multi-Services category includes complaints about Google account-level terminations, in addition to complaints relating to actions taken in response to Article 16 notices that relate to advertisements appearing on one or more Google services, which may include a VLOSE or VLOP.

4.3 Number of complaints, broken down by outcome of complaint

Article 15(1), point (d)

Complaint outcomes include initial decision upheld, initial decision overturned, mixed result, issue resolved and other. An "initial decision" refers to the initial enforcement of Google's terms of service or product policies. These decisions may be overturned in light of additional information provided by the appellant or additional review of the content. If there are multiple items in the complaint and the decision was upheld for some items but overturned for others, this is categorised as a "mixed result". If a complaint is withdrawn, if the complaint requires no action, response or decision from Google, or if the creator resolves the issue so that their content is no longer policy-violating, this is categorised as "issue resolved". Table 4.3.1 provides the number of complaints, broken down by service and complaint outcome.

Table 4.3.1: Number of complaints, by service and complaint outcome[1]

Service	Initial decision upheld	Initial decision overturned	Mixed result[2]	Issue resolved	Other
Maps	30,670	63,610	N/A	28	0
Ads on Maps	2,521	1,296	0	0	0
Play	1,044	88	N/A	515	0
Ads on Play	2,378	226	0	0	0
Shopping[3]	15,209	61,117	0	3,468	0
YouTube	11,217	23,234[4]	N/A	1,420	472
Ads on YouTube	40,627	32,129	0	0	0
Multi-Services[5]	30,283	3,864	N/A	5,299	0

Notes:

[1] Not all complaints can be resolved during the reporting period, therefore the total number of complaint outcomes above will be less than the total number of complaints received (Table 4.1.1).

[2] The 'mixed result' outcome only applies to complaints about Ads presented on each service; 0 outcomes fell into this category for this reporting period.

[3] Shopping metrics reflect complaints about content moderation actions taken on both unpaid content (e.g., free listings) and advertisements.

[4] Within the "initial decision overturned" category, the majority of overturned decisions for YouTube are for monetary payment restrictions. This is an expected outcome intended to protect users, creators, and advertisers, enforced using YouTube's Advertiser-Friendly Content Guidelines. Monetising creators (those in the YouTube Partner Program) have easy access to a timely and user-friendly internal complaint handling system and are encouraged to use it if they believe YouTube's systems made a mistake. For information specifically about YouTube's Community Guidelines enforcement and appeals, please see YouTube's

Community Guidelines Transparency Report (note: data in YouTube's Community Guidelines Transparency Report are not directly comparable with the data presented in this DSA Transparency Report).

[5] The Multi-Services category includes complaints about Google account-level terminations, in addition to complaints relating to actions taken in response to Article 16 notices that relate to advertisements appearing on one or more Google services, which may include a VLOSE or VLOP.

4.4 Median time needed to action a complaint

Article 15(1), point (d)

Google works to provide complaint outcomes to users within a reasonable timeframe. The types of complaints vary widely, with some requiring a longer review period due to varying degrees of complexity or external factors (e.g., legally prescribed wait times). Table 4.4.1 reflects the median time, in days, needed to action a complaint for each service.

Table 4.4.1: Median time needed to action a complaint, by service

Service	Median time to action a complaint (days)
Maps	3
Ads on Maps	<1
Play	<1
Ads on Play	<1
Shopping[1]	<1
YouTube	<1
Ads on YouTube	<1
Multi-Services[2]	<1

Notes:

[1] Shopping metrics reflect content moderation actions taken on both unpaid content (e.g., free listings) and advertisements.

[2] The Multi-Services category includes complaints about Google account-level terminations, in addition to complaints relating to actions taken in response to Article 16 notices that relate to advertisements appearing on one or more Google services, which may include a VLOSE or VLOP.

Google Transparency Report

Section 5: Out-of-court dispute settlements

Article 24(1), point (a)

Out-of-court dispute settlement bodies are independent bodies with the means and expertise to consider the issues brought to them. The DSA requires that each Member State certify out-of-court settlement bodies to handle eligible disputes. No out-of-court settlement bodies have been certified at this time and thus no out-of-court disputes were submitted or resolved during the reporting period.

Google ▶ YouTube

EU DSA Report • 38

Section 6: Article 23 Suspensions imposed to protect against misuse

Article 24(1), point (b)

To protect users from significant harm and unlawful activity, Google suspends user accounts when we detect egregious content (e.g., child abuse) or repeated violations of our services' policies. Suspended user accounts are unable to access Google products and, depending on the suspension reason, may not contribute to Google platforms or engage in specific Google processes (e.g., submission of complaints through dedicated complaint channels).

6.1 Number of suspensions for Manifestly Illegal Content imposed pursuant to Article 23

Article 24(1), point (b)

Depending on the severity of the detected violation and involvement of legal enforcement authorities, users may receive a warning and/or remedial instructions to remove the violating content before their account is suspended. During the reporting period, there were 1,975 Google-wide account-level suspensions of EU users who posted manifestly illegal content across Google services, but not necessarily limited to VLOPs.

6.2 Number of suspensions for Manifestly Unfounded Notices imposed pursuant to Article 23

Article 24(1), point (b)

Users who [intentionally misuse](#) webforms and processes by repeatedly filing manifestly unfounded notices will be flagged, and their requests will be closed without assessment. Users will receive a written warning before Google takes action. If misuse continues, the user will be suspended from reporting content and their requests will be closed without assessment for a period of up to six months, and an auto reply will be issued. After a maximum of six months, new requests for content removal may be submitted.

During the reporting period, there were 0 user suspensions due to the repeated submission of manifestly unfounded legal notices. If applied, these would suspend the processing of a user's notices for any Google service – therefore they are not linked to a specific Google service. Suspended users may reach out to Google Legal via lettermail at any time to appeal a suspension.

6.3 Number of suspensions for Manifestly Unfounded Complaints imposed pursuant to Article 23

Article 24(1), point (b)

Users who intentionally misuse webforms and processes by repeatedly filing manifestly unfounded complaints will also be flagged, and their requests will be closed without assessment.

During the reporting period, there were 950 suspensions of an EU-based user's ability to submit Article 20 complaints due to the repeated submission of manifestly unfounded complaints. In addition, 5,845 appeals were suspended during the reporting period reflecting the number of appeals that were not processed.

Section 7: Average monthly active recipients of Google services in the Union

Article 24(2)

The average number of monthly active recipients of Google services in each European Union Member State is provided in the [DSA Monthly Active Recipients report](#) published on 16 August 2023.

Section 8: Additional Notes

- Metrics presented in this report generally reflect our efforts and resources to moderate potentially illegal content and policy-violative content in the EU. However, most, but not all, of policy-violating content on Google Search and its features is moderated globally. Most content delisted from Google Search on legal grounds is content subject to copyright removal notifications, which are also processed globally. However, there are classes of delistings based on local law or local court orders that affect only certain country services, based on variance in laws between countries. Finally, in some cases, EEA metrics have been voluntarily provided in this report.

- Numbers reported may fluctuate between successive reports due to various reasons, including service level changes or enhancements, changes in the number of users on a service, external events and differences in reporting periods. Therefore, report-by-report comparisons may not accurately reflect time-based improvements in our processes.

- Services designated as VLOSE and VLOPs differ in various ways, including content type on the service, underlying content moderation systems and number of users on a service, which means that in some cases, metrics may not be directly comparable.