

DSA Transparency Report - October 2024

Introduction

This report covers the content moderation activities of X's international entity Twitter International Unlimited Company (TIUC) under the Digital Services Act (DSA), during the date range 1 April, 2024 to 30 September, 2024.

We may refer to "notices" as defined in the DSA as "user reports" and "reports".

Description of our Content Moderation Practices

Our content moderation systems are designed and tailored to mitigate systematic risks without unnecessarily restricting the use of our service and fundamental rights, especially freedom of expression. Content moderation activities are implemented and anchored on principled policies and leverage a diverse set of interventions to ensure that our actions are reasonable, proportionate and effective. Our content moderation systems blend automated and human review paired with a robust appeals system that enables our users to quickly raise potential moderation anomalies or mistakes.

Policies

X's purpose is to serve the public conversation. Violence, harassment, and other similar types of behaviour discourage people from expressing themselves, and ultimately diminish the value of global public conversation. Our Rules are designed to ensure all people can participate in the public conversation freely and safely.

X has policies protecting user safety as well as platform and account integrity. The [X Rules and policies](#) are publicly accessible on our Help Center, and we are making sure that they are written in an easily understandable way. We also keep our Help Center regularly updated anytime we modify our Rules.

For the purposes of the summary tables below, the X policy titles in use at the start of the reporting period have been retained, even if they changed throughout the period.

Enforcement

When determining whether to take enforcement action, we may consider a number of factors, including (but not limited to) whether:

- The behaviour is directed at an individual, group, or protected category of people;
- The report has been filed by the target of the abuse or a bystander;
- The user has a history of violating our policies;
- The severity of the violation; and
- The content may be a topic of legitimate public interest.

When we take [enforcement actions](#), we may do so either on a specific piece of content (e.g., an individual post or Direct Message) or on an account. We may employ a combination of these options. In most cases, this is because the behaviour violates the X Rules.

To enforce our Rules, we use a combination of machine learning and human review. Our systems are able to surface content to human moderators who use important context to make decisions about potential violations. This work is led by an international, cross-functional team with 24-hour coverage and the ability to cover multiple languages. We also have a complaints process for any potential errors that may occur.

To ensure that our human reviewers are prepared to perform their duties we provide them with a robust support system. Each human reviewer goes through extensive training and refreshers, they are provided with a suite of tools that enable them to do their jobs effectively, and they have a suite of wellness initiatives available to them. For further information on our human review resources, see the section titled "Human resources dedicated to Content Moderation".

Reporting violations

X strives to provide an environment where people can feel free to express themselves. If abusive behaviour happens, we want to make it easy for people to report it to us. EU users can also report any violation of our Rules or their local laws, no matter where such violations appear.

Transparency

We always aim to exercise moderation with transparency. Where our systems or teams take action against content or an account as a result of violating our Rules or in response to a valid and properly scoped request from an authorised entity in a given country, we strive to provide context to users. Our [Help Center article](#) explains notices that users may encounter following actions taken. We promptly notify affected users about legal requests to withhold content, including a copy of the original request, unless we are legally prohibited from doing so. We have also updated our global [transparency centre](#) covering a broader array of our transparency efforts.

Content Moderation Governance Structure

Own Initiative Content Moderation Activities

X employs a combination of heuristics and machine learning algorithms to automatically detect content that we believe violates the [X Rules and policies](#) enforced on our platform. We use combinations of natural language processing models, image processing models and other sophisticated machine learning methods to detect potentially violative content. These models vary in complexity and in the outputs they produce. For example, the model used to detect abuse on the platform is trained on abuse violations detected in the past. Content flagged by these machine learning models are either reviewed by human content reviewers before an action is taken or, in some cases, automatically actioned, based on the historical accuracy of the model's output. Heuristics are typically utilised to enable X to react quickly to new forms of violations that emerge on the platform. Heuristics are common patterns of behaviours, text, or keywords that may be typical of a certain category of violations. Pieces of content detected by heuristics may also get reviewed by human content reviewers before an action is taken on the content. These heuristics are used to flag content for review by human agents proactively.

TESTING, EVALUATION, AND ITERATION

Automated enforcements under the [X Rules and policies](#) undergo rigorous testing before being applied to the live product. Both machine learning and heuristic models are trained and/or validated on thousands of data points and labels (e.g., violative or non-violative) including those that are generated by trained human content moderators. For example, inputs to content-related models can include the text within the post itself, the images attached to the post, and other characteristics. Training data for the models comes from both the cases reviewed by our content moderators, random samples, and various other samples of pieces of content from the platform.

USE OF HUMAN MODERATION

Before any given algorithm is launched to the platform, we verify its detection of policy violating content or behaviour by drawing a statistically significant test sample and performing item-by-item human review. Reviewers have expertise in the applicable policies and are trained by our Policy teams to ensure the reliability of their decisions. Human review helps us to confirm that these automations achieve a level of precision, and sizing helps us understand what to expect once the automations are launched.

In addition, humans proactively conduct manual content reviews for potential policy violations. We conduct proactive sweeps for certain high-priority categories of potentially violative content both periodically and during major events, such as elections. Content moderators also proactively review content flagged by heuristic and machine learning models for potential violations of other policies, including our [sensitive media](#), [child sexual exploitation](#) (CSE) and [violent and hateful entities](#) policies.

Once reviewers have confirmed that the detection meets an acceptable standard of accuracy, we consider the automation to be ready for launch. Once launched, automations are monitored dynamically for ongoing performance and health. If we detect anomalies in performance (for instance, significant spikes or dips against the volume we established during sizing, or significant changes in user complaint/overturn rates), our Engineering (including Data Science) teams - with support from other functions - revisit the automation to diagnose any potential problems and adjust the automations as appropriate.

AUTOMATED MODERATION ACTIVITY EXAMPLES

A vast majority of all accounts that are suspended for the promotion of terrorism and CSE are proactively flagged by a combination of technology and other purpose-built internal proprietary tools. When we remove CSE content with these automated systems, we immediately report it to the National Center for Missing and Exploited Children (NCMEC). NCMEC makes reports available to the appropriate law enforcement agencies around the world to facilitate investigations and prosecutions.

Our current methods deploy a range of internal tools and third party solutions that utilises industry standard hash libraries (e.g., [PhotoDNA](#)) to ensure known CSAM is caught prior to any user reports being filed. We leverage the hashes provided by NCMEC and industry partners. We scan media uploaded to X for matches to hashes of known CSAM sourced from NGOs, law enforcement and other platforms. We also have the ability to block keywords and phrases from Trending and block search results for certain terms that are known to be associated with CSAM.

We [commit](#) to continuing to invest in technology that improves our capability to detect and remove, for instance, terrorist and violent extremist content online before it can cause user harms, including the extension or development of digital fingerprinting and AI-based technology solutions. Our participation in multi-stakeholder communities, such as the [Christchurch Call to Action](#), [Global Internet Forum to Counter Terrorism](#) and EU Internet Forum (EUIF), helps to identify emerging trends in how terrorists and violent extremists are using the internet to promote their content and exploit online platforms.

You can learn more about our commitment to eradicating CSE and terrorist content, and the actions we've taken [here](#). Our continued investment in proprietary technology is steadily reducing the burden on people to report this content to us.

SCALED INVESTIGATIONS

These moderation activities are supplemented by scaled human investigations into the tactics, techniques and procedures that bad actors use to circumvent our rules and policies. These investigations may leverage signals and behaviours identifiable on our platform, as well as off-platform information, to identify large-scale and/or technically sophisticated evasions of our detection and enforcement activities. For example, through these investigations, we are able to detect coordinated activity intended to manipulate our platform and artificially amplify the reach of certain accounts or their content.

CLOSING STATEMENT ON CONTENT MODERATION ACTIVITIES

Our content moderation systems are designed and tailored to mitigate systematic risks without unnecessarily restricting the use of our service and fundamental rights, especially freedom of expression. Content moderation activities are implemented and anchored on principled policies and leverage a diverse set of interventions to ensure that our actions are reasonable, proportionate and effective. Our content moderation systems blend automated and human review paired with a robust appeals system that enables our users to quickly raise potential moderation anomalies or mistakes.

INDICATORS OF ACCURACY FOR CONTENT MODERATION

The possible rate of error of the automated and human means used in enforcing [X Rules and policies](#) is represented by the number of Content Removal Complaints (appeals) received and the number of Content Removal Complaints that resulted in reversal of our enforcement decision (successful appeals) by remediation type and by country.

Human resources dedicated to Content Moderation

Today, we have 1275 people working in content moderation. Our teams work on both initial reports as well as on complaints of initial decisions across the world (and are not specifically designated to only work on EU matters).

Linguistics Expertise of our Content Moderation Team

X's scaled operations team possesses a variety of skills, experiences, and tools that allow them to effectively review and take action on reports across all of our Rules and policies. X has analysed which languages are most common in reports reviewed by our content moderators and has hired content moderation specialists who have professional proficiency in the commonly spoken languages. The following table is a summary of the the number of people in our content moderation team who possess professional proficiency in the most commonly spoken languages in the EU on our platform:

Primary Language	People
Bulgarian	1
English	1,117
French	67
German	69
Italian	1
Portuguese	5
Spanish	15
Total	1,275

In addition to the primary language support, we have also have people supporting additional languages. The following is the list of secondary EU language support:

Secondary Language	People
Bulgarian	1
Croatian	1
French	74
German	71
Greek	1
Irish	1
Italian	2
Latvian	1
Polish	1
Portuguese	22
Spanish	41
Total	216

Please note that the numbers included in the secondary language support are not separate or distinct from the numbers included in the primary language support data.

Qualifications of our Content Moderation Team

Content Moderation Team Qualifications	
Years in Current Role	Headcount
0 to 1	422
1 to 2	188
2 to 3	245
3 to 4	161
4 to 5	77
5 to 6	127
6 to 7	62
7 or more	47

The above table includes all moderators who support EU member state languages as of September 2024. The content moderation team collectively provides linguistic capacity in multiple languages. In situations where we need additional language support, we use translation services and/or machine translation tools, to investigate and address challenges in additional languages. Additionally, content moderators will leverage playbooks that contain colloquial terms and phrases that are consistently being updated to reflect various EU languages spoken within the region.

Moderators are recruited using a standard job description that includes a language requirement which states that the candidate should be able to demonstrate written and spoken fluency in the language and have at least one year of work experience for entry-level positions. In the interview and application process, each agent candidate must meet certain linguistic standards to be considered "language qualified". This determination is made through multiple tests (i.e. written, oral, etc.) of the candidate's respective language, to determine their respective proficiency level. Candidates must also meet the educational and background requirements in order to be considered, as well as demonstrate an understanding of current events for the country or region of content moderation they will support.

Organisation, Team Resources, Expertise, Training and Support of our Team that Reviews and Responds to Reports of Illegal Content

Description of the team

X has built a specialised team made up of individuals who have received specific training in order to assess and take action on illegal content that X becomes aware of via reports or other processes on our own initiative. This team consists of different tier groups, with higher tiers consisting of more senior, or more specialised, individuals.

When handling a report of illegal content or a complaint against a previous decision, content and senior content reviewers first assess the content under X's Rules and policies. If no violation of X's Rules and policies is determined warranting a global removal of the content, the content moderators will assess the content for potential illegality under Local Laws. If the content is not manifestly illegal, it can be escalated for second or third opinions. If more detailed investigation is required, content moderators can escalate reports to experienced policy and/or legal request specialists who have also undergone in-depth training and/or have language expertise in the respective cases language. These individuals take appropriate action after carefully reviewing the report and/or complaint in close detail. In cases where this specialist team still cannot determine a decision regarding the potential illegality of the reported content, the report can be discussed with in-house legal counsel. Everyone involved in this process works closely together with daily exchanges through meetings and other channels to ensure the timely and accurate handling of reports. Additionally, in the instance that a case warrants in-house legal counsel, the lessons learned and actions made on this case will be disseminated to all relevant content moderator parties to ensure consistency in review and an understanding of best practices made by the agent, if a similar case is encountered in the future.

Furthermore, all teams involved in solving these reports closely collaborate with a variety of other policy teams at X who focus on safety, privacy, authenticity rules and policies. This cross-team effort is particularly important in the aftermath of tragic events, such as violent attacks, to ensure alignment, swift consistency in review, and the same potential remediation actions if the content is found violative.

Content moderators are supported by team leads, subject matter experts, quality auditors and trainers. We hire people with diverse backgrounds in fields such as law, political science, psychology, communications, sociology and cultural studies, and languages.

Training and support of persons processing legal requests

All team members, i.e. all employees hired by X as well as vendor partners working on these reports, are trained and retrained regularly on our tools, processes, Rules and policies, including special sessions on cultural and historical context. Initially when joining the team at X, each individual follows an onboarding program and receives individual mentoring during this period, as well as thereafter through our Quality Assurance (QA) program (for external employees), in house and external counsels (for internal employees).

All team members have direct access to robust training and workflow documentation for the entirety of their employment, and are able to seek guidance at any time from trainers, leads, and internal specialist legal and policy teams as outlined above as well as managerial support.

Updates about significant current events or Rules and policy changes are shared with all content reviewers in real time, to give guidance and facilitate balanced and informed decision making. In the case of Rules and policy changes, all training materials and related documentation is updated. Calibration sessions are carried out frequently during the reporting period. These sessions aim to increase collective understanding and focus on the needs of the content reviewers in their day-to-day work, by allowing content moderators to ask questions and discuss aspects of recently reviewed cases, X's Rules and policies, and/or local laws.

The entire team also participates in obligatory X Rules and policies refresher training as the need arises or whenever Rules and policies are updated. These trainings are delivered by the relevant policy specialists who were directly involved in the development of the rules and policy change. For these sessions we also employ the "train the trainer" method to ensure timely training delivery to the whole team across all of the shifts. All team members use the same training materials to ensure consistency.

QA is a critical measure to the business to help ensure that we are delivering a consistent service at the desired level of quality to our key stakeholders, both externally and internally as it pertains to our case work. We have a dedicated QA Team within our vendor team to help us identify areas of opportunity for training and potential defect detection in our workflow or Rules and policies. The QA specialists perform quality checks of reports to ensure that content is actioned appropriately.

The standards and procedures within the QA team ensure the team's QA is assessed equally, objectively, efficiently and transparently. In case of any mis-alignments, additional training is scheduled, to ensure the team understands the issues and can handle reports accurately.

In addition, given the nature and sensitivity of their work, the entire team has access to online resources and regular onsite group and individual sessions related to resilience and well-being. These are provided by mental health professionals. Content reviewers also participate in resilience, self-care, and vicarious trauma sessions as part of our mandatory wellness plan during the reporting period.

Training and Support provided to those Persons performing Content Moderation Activities for our TIUC Terms of Service and Rules

Training is a critical component of how X maintains the health and safety of the public conversation through enabling content moderators to accurately and efficiently moderate content posted on our platform. Training at X aims to improve the content moderators' enforcement performance and quality scores by enhancing content moderators' understanding and application of X Rules through robust training and quality programs and a continuous monitoring of quality scores.

TRAINING PROCESS

There is a robust training program and system in place for every workflow to provide content moderators with the adequate work skills and job knowledge required for processing user cases. All content moderators must be trained in their assigned workflows. These focus areas ensure that content moderators are set up for success before and during the content moderation lifecycle, which includes:

- Training analysis/design focused on agent and learning needs;
- Classroom training with expert trainers;
- Nesting period to apply new skills;
- Cross-skilling opportunities;
- Upskilling opportunities;
- Refresher programs;
- New launch/update roll-outs process; and
- Remediation plans.

TRAINING ANALYSIS & DESIGN

Before commencing design work on any content moderators program or resource, a rigorous learner analysis is conducted in close collaboration with training specialists and quality analysts to identify performance gaps and learning needs. Each program is designed with key stakeholder engagement and alignment. The design objective is to adhere to visual and learning design principles to maximise learning outcomes and ensure that agents can perform their tasks with accuracy and efficiency. This is achieved by making sure that the content is:

- Easy to experience
- Easy to understand; and
- Easy to apply.

X's training programs and resources are designed based on needs, and a variety of modalities are employed to diversify the content moderators learning experience, including:

- Self-led learning: microlearning, scenario-based learning, e-learning modules, and gamification (where appropriate);
- Virtual live instructor-led trainings;
- Face-to-face classroom training; and
- Videos.

CLASSROOM TRAINING

Classroom training is delivered either virtually or face-to-face by expert trainers. Classroom training activities can include:

- Instructor-led policy training;
- Interactive e-learning;
- Scenario-based learning sets;
- Shadowing sessions with seasoned agents;
- Guided casework sessions with trainers; and
- Knowledge checks, quizzes and assessments.

NESTING (ON-THE-JOB TRAINING)

When content moderators successfully complete their classroom training program, they undergo an onboarding period. The onboarding phase includes case study by observation, demonstration and hands-on training on live cases. Onboarding activities include content moderator shadowing, guided case work, Question and Answer sessions with their trainer, coaching, feedback sessions, etc. Quality audits are conducted for each onboarding content moderator and content moderators must be coached for any mis-action spotted in their quality scores the same day that the case was reviewed. Trainers conduct needs assessment for each onboarding content moderator and prepare refresher training accordingly. After the onboarding period, content is evaluated on an ongoing basis with the QA team to identify gaps and address potential problem

areas. There is a continuous feedback loop with quality analysts across the different workflows to identify challenges and opportunities to improve materials and address performance gaps.

UP-SKILLING

When a content moderator needs to be upskilled they receive training of a specific workflow within the same pillar that the content moderator is currently working. The training includes a classroom training phase and onboarding phase which is specified above.

REFRESHER SESSIONS

Refresher sessions take place when a content moderator has previously been trained, has access to all the necessary tools, but would need a review of some or all topics. This may happen for content moderators who have been on prolonged leave, transferred temporarily to another content moderation policy workflow, or ones who have recurring errors in the quality scores. After a needs assessment, trainers are able to pinpoint what the content moderator needs and prepare a session targeting their needs and gaps.

NEW LAUNCH / UPDATE ROLL-OUTS

There are also processes that require new and/or specific product training and certification. These new launches and updates are identified by X and the knowledge is transferred to the content moderators.

REMEDATION PLANS

There are remediation plans in place to support content moderators who do not pass the training or onboarding phase, or are not meeting quality requirements.

Member States Orders to Act Against Illegal Content

Removal Orders Received - Apr 1 to Sep 30						
Illegal Content Category	France	Germany	Ireland	Italy	Slovak Republic	Spain
Illegal or harmful speech			1	1	1	2
Risk for public security		1				
Unsafe and illegal products	8					

Removal Orders Median Handle Time (Hours) - Apr 1 to Sep 30						
Illegal Content Category	France	Germany	Ireland	Italy	Slovak Republic	Spain
Illegal or harmful speech			14.1	45.4	5.7	46.8
Risk for public security		30.9				
Unsafe and illegal products	6.2					

Removal Orders Median Time To Acknowledge Receipt

X provides an automated acknowledgement of receipt of removal orders submitted by law enforcement through our [Legal Request submission portal](#). As a consequence of this immediate acknowledgement of receipt, the median time was zero hours.

Important Notes about Removal Orders:

- To improve clarity, we've omitted countries and violation types with no legal requests from the tables above;
- "Removal Orders Median Handle Time" shows the category that we considered to fit best and under which we handled the order. This category might deviate from the information provided by the authority when submitting the order via the X online submission platform; and
- In the cases from France, Germany, Ireland, Italy, Slovak Republic and Spain, we asked the submitting authority to fulfil Article 9 information requirements but did not receive responses in the reporting period.

Information Requests Received - Apr 1 to Sep 30																
Illegal Content Category	Austria	Belgium	Czech Republic	Denmark	Estonia	Finland	France	Germany	Greece	Hungary	Ireland	Italy	Malta	Netherlands	Poland	Port
Animal welfare								1								
Data protection &		1					10	12	1			1	1		1	

privacy violations																
Illegal or harmful speech	7	6		1		6	107	4556	11		6	10	1		8	63
Intellectual property infringements							1	13				1				
Issue Unknown							1	2								1
Negative effects on civic discourse or elections							5	27							1	
Non-consensual behaviour							7	22	1							1
Pornography or sexualized content							8	47								1
Protection of minors		3		1		1	6	76			3					1
Risk for public security	24	75					1847	196			1	1			10	2
Scams and fraud					1	1	54	67	3	1		2	1		1	10
Self-harm							1		2							
Unsafe and illegal products							2	5			2					
Violence	2	10	3	2		9	264	160	1		14	12			20	35

Information Request Median Handle Time (Hours) - Apr 1 to Sep 30

Illegal Content Category	Austria	Belgium	Czech Republic	Denmark	Estonia	Finland	France	Germany	Greece	Hungary	Ireland	Italy	Malta	Netherlands	Poland	Port
Animal welfare								17								
Data protection & privacy violations		550					494	581	510			627	604		590	.1
Illegal or harmful speech	600	660		647		603	603	572	652		41	580	575		729	528
Intellectual property infringements							769	623				509				
Issue Unknown							1	275								484
Negative effects on civic discourse or elections							678	589							550	
Non-consensual behaviour							195	645	744							549
Pornography or sexualized content							637	624								584
Protection of minors		18		4		38	9	5			24					2
Risk for public security	622	314					147	604			143	530			557	339
Scams and fraud					601	621	590	582	595	550		623	602		550	532
Self-harm							25		10							
Unsafe and illegal products							599	627			1					
Violence	677	504	550	497		605	576	604	745		46	590			690	525

Information Request Median Time To Acknowledge Receipt

X provides an automated acknowledgement of receipt of information requests submitted by law enforcement through our [Legal Request submission portal](#). As a consequence of this immediate acknowledgement of receipt, the median time is zero.

Important Notes about Information Requests:

- To improve clarity, we have omitted countries and violation types with no legal requests from the tables above;
- The content category for each request is determined by the information law enforcement provides when submitting requests through the X online submission platform. If law enforcement does not provide sufficient information during form submission, the category is determined based on the allegations provided in the legal process. Where multiple illegal content categories were provided, only the gravamen offence was included;
- The median handling time is the time between receiving the order and either: 1) disclosing information to law enforcement if the order is valid; or 2) pushing back due to legal issues. The median handling time does not include additional time where X pushes back due to legal issues, receives a valid order or additional information later, and disclosure is eventually made; and
- The "Issue Unknown" category shows cases where the illegal content category could not be determined based on the information law enforcement provided during the submission process and/or in the legal process.

Illegal Content Notices

Illegal Content Notices Received - Apr 1 to Sep 30																
Reason Code	Austria	Belgium	Bulgaria	Croatia	Cyprus	Czechia	Denmark	EU	Estonia	Finland	France	Germany	Greece	Hungary	Ireland	Italy
Animal welfare	103	33	14	3	2	18	21	426	3	14	372	236	14	7	26	57
Data protection & privacy violations	136	202	49	25	38	104	93	2501	18	55	2883	2620	140	44	310	503
Illegal or harmful speech	2316	917	104	93	112	1032	577	37389	111	683	25044	33861	332	148	2020	3350
Intellectual Property Infringements	47	28	20	34	9	77	34	N/A*	9	135	1987	5667	112	35	1150	614
Negative effects on civic discourse or elections	189	102	21	7	14	95	61	1461	14	66	4445	3387	29	19	169	451
Non-consensual behaviour	32	44	7	5	17	24	23	1008	4	30	1244	977	13	2	42	117
Pornography or sexualized content	190	235	44	29	16	195	180	2508	83	65	4060	2329	121	146	130	581
Protection of minors	54	143	28	10	13	43	101	2019	46	340	2096	9021	30	33	121	240
Risk for public security	109	67	22	9	14	125	151	1092	104	77	1869	2689	60	18	133	197
Scams and fraud	344	501	74	56	60	335	192	3655	47	221	5837	2897	121	301	970	1197
Scope of platform service	5	4	9	2	1	5		316	4	1	117	160	5	1	14	63
Self-harm	17	9	21	6	3	12	4	653	2	10	287	270	5	3	17	55
Unsafe and illegal products	22	27	1	3	12	99	68	397	22	56	1365	616	5	21	46	64
Violence	194	170	49	16	26	150	92	5824	23	96	3865	4465	83	101	276	706

*This category is not applicable since such a field option does not exist in the Intellectual Property infringement reporting form.

Actions Taken on Illegal Content Notices - Apr 1 to Sep 30															
Closure Type	Action Type	Grounds for Action	Reason Code	Austria	Belgium	Bulgaria	Croatia	Cyprus	Czechia	Denmark	EU	Estonia	Finland	France	Germany
Automated Means	Global content deletion based on a violation of TIUC Terms of	Terms of Service and/or X's Rules or Policies	Violence								3				

	Service and Rules													
	Country withheld Content	Basis of Law and/or Local Laws												
No Violation Found	Terms of Service and/or X's Rules or Policies	Violence								1				
		Animal welfare								1				
		Data protection & privacy violations								1				
		Illegal or harmful speech	1							4				
		Negative effects on civic discourse or elections				1						1		
		Non-consensual behaviour								1				
		Pornography or sexualized content	8	3				4		22		1		
		Protection of minors	12	13	6			4	12	326	17	136		
		Risk for public security		3					1	2				
		Scams and fraud	12	13	3				2	45		5		
		Unsafe and illegal products								1				
		Violence								3	1			
Manual Closure	Global content deletion based on TIUC Terms of Service and Rules	Animal welfare	7	4				2	2	119	1	2	32	
		Data protection & privacy violations	12	10	4	1		16	18	204	1		138	4
		Illegal or harmful speech	53	27	1	5	6	69	35	738	11	17	463	9
		Negative effects on civic discourse or elections						1	2	14			6	
		Non-consensual behaviour		3					2	45		3	37	
		Pornography or sexualized content	22	27	4		1	12	69	281	40	21	314	5
		Protection of minors	9	36	9	4	1	8	19	640	2	144	665	65
		Risk for public security	10	7		1	2	8	83	64	49	23	134	6
		Scams and fraud			1			2	1	14			7	
		Scope of platform service								29			1	
		Self-harm	1			1		2		62			13	
		Unsafe and illegal products	1	5		1		20	39	25	13	4	318	1
		Violence	22	29	6	4	2	33	12	858	4	19	373	8
	Offer of help in case of self-harm and suicide concern based on TIUC Terms of	Illegal or harmful speech								2				
		Protection of minors								1				
		Self-harm				1				40		1	2	

Service and Rules		Violence								3				
Country withheld Content	Basis of Law and/or Local Laws	Animal welfare	4	1				1	1	18		1	3	
		Data protection & privacy violations	12	21		7	4	14	10	322		20	210	3
		Illegal or harmful speech	1221	348	28	29	19	431	226	14012	32	207	5252	171
		Negative effects on civic discourse or elections	27	5	1		1	6	5	170	1	1	190	6
		Non-consensual behaviour	4	9				2	6	248		4	93	3
		Pornography or sexualized content	24	40	18	9	3	63	20	1128	12	15	827	8
		Protection of minors	1	7	3	2		11	6	244		5	107	8
		Risk for public security	24	7	2	1	1	14	10	215	3	7	153	5
		Scams and fraud	37	29	2	3	5	53	31	463	25	65	359	2
		Scope of platform service	1					2		58			4	
		Self-harm	2					2		54		1	19	
		Unsafe and illegal products	7	2		1	6	44	9	69	4	14	556	1
		Violence	34	34	7	2	6	29	12	1442	4	10	390	9
Content removed globally following illegal content notice	Terms of Service and/or X's Rules or Policies	Intellectual Property Infringements	21	20	10	2	4	41	16		6	88	986	1
Account Suspension	Terms of Service and/or X's Rules or Policies	Intellectual Property Infringements	1	1	2	16	2	3	0		0	1	355	19
Content removed globally following illegal content notice	Terms of Service and/or X's Rules or Policies	Data protection & privacy violations											4	
		Illegal or harmful speech											7	
		Non-consensual behaviour												
		Pornography or sexualized content	4	4						10			4	
		Protection of minors					2	1	22	69	14	20		
		Risk for public security												
		Scams and fraud								2				
		Unsafe and illegal products								1				
		Violence											3	
No Violation Found	Basis of Law and/or Local Laws	Animal welfare	92	27	14	3	2	12	17	279	2	10	235	1
		Data protection & privacy violations	110	168	45	17	32	73	64	1924	17	35	1866	17
		Illegal or harmful	1012	531	75	56	87	506	315	21767	67	450	15464	152

			speech														
			Negative effects on civic discourse or elections	159	95	19	7	13	88	54	1261	13	64	3873	26		
			Non-consensual behaviour	28	31	7	5	16	22	15	703	4	21	649	5		
			Pornography or sexualized content	130	156	20	20	12	112	90	1031	31	27	850	7		
			Protection of minors	31	84	9	4	10	19	39	694	11	34	625	10		
			Risk for public security	74	49	20	7	10	100	57	792	51	46	1333	14		
			Scams and fraud	286	458	68	53	47	274	157	3085	18	149	1414	25		
			Scope of platform service	4	4	9	2		3		225	4	1	71	1		
			Self-harm	14	9	21	4	3	8	4	485	1	7	181	1		
			Unsafe and illegal products	13	19	1	1	6	34	19	299	5	37	279	2		
			Violence	133	104	35	10	18	84	66	3399	14	66	2527	26		

Reports of Illegal Content Median Handle Time (Hours) - Apr 1 to Sep 30

Reason Code	Austria	Belgium	Bulgaria	Croatia	Cyprus	Czechia	Denmark	EU	Estonia	Finland	France	Germany	Greece	Hungary	Ireland	Italy	Li
Animal welfare	5.8	8.3	1.1	0.4	0.6	3.2	2.2	6.1	8.7	0.3	14.2	2.1	3	1.5	8.2	2.1	
Data protection & privacy violations	1.5	3.7	5.8	4.7	2	6.7	7.8	3.6	21.5	6.6	8.2	2.7	2.1	2.4	8.2	3.9	
Illegal or harmful speech	2	2.2	2	3	3	2.1	1.6	2.8	2.7	2.7	5.6	1.1	2.5	5.6	3.1	2.2	
Intellectual Property Infringements	32.9	4.3	4.3	2.6	5.1	20.4	7.5	N/A*	8.1	8.9	4.6	10.4	5.2	1.7	12.4	6.3	
Negative effects on civic discourse or elections	2.2	1.2	1	12	1.8	1.8	1	1.7	1.6	2.5	6.6	1	3.3	2.4	3.1	2	
Non-consensual behaviour	1.8	10.5	19.2	7.5	0.4	1.4	1.5	9.5	9.9	3.2	9.3	1	3.5	74.9	1.2	3.8	
Pornography or sexualized content	6.2	8.4	3.7	2.2	6.1	2.9	1.8	4	1.6	2.9	5.5	1.5	9.3	4.3	9.4	4.1	
Protection of minors	4.8	5.3	3.8	4.4	3.1	2.4	5.4	2.9	3.2	2.4	4.5	1.1	6.1	8.5	5.4	5.6	
Risk for public security	4.9	2.4	3.6	2.7	2.3	1.2	2.2	3.1	1.6	2.4	6.1	1.2	2.1	1	9.1	1.9	
Scams and fraud	6.7	9	7	4.8	9.8	2.4	4	5.4	9.7	4.6	14	8.9	4.4	5.2	10.4	5.2	
Scope of platform service	9.8	0.6	11	1.9		0.1		1.8	3	10.3	9	1	0.9	0.4	0.3	5.7	
Self-harm	1.8	9.8	11.3	1	0.3	0.8	3.4	2.4	10.8	0.9	2.8	3.4	1.9	1.2	2.6	3.4	
Unsafe and illegal products	2	1.8	0.1	1	0.6	1.6	2.3	1.6	1.5	2.9	1.1	1.3	3.7	14.1	8.4	1.9	
Violence	2.4	4.2	1.1	0.4	1.7	1.9	9.7	2.9	8.8	2.4	6.1	0.9	1.4	2.2	2.2	2	

*This category is not applicable since such a field option does not exist in the Intellectual Property infringement reporting form.

Own Initiative Enforcements

RESTRICTED REACH LABELS DATA

Restricted Reach Labels - Apr 1 to Sep 30

Detection Method	Enforcement	Policy	Austria	Belgium	Bulgaria	Croatia	Cyprus	Czechia	Denmark	Estonia	Finland	France	Germany	Greece	Hungary	Ireland	Italy
Own Initiative	Automated Means	Hateful Conduct	4385	8572	3008	3415	932	5198	5856	1270	5638	39,454	38,046	4441	3356	16,268	11
User Report	Manual Review	Abuse & Harassment	130	344	81	69	46	184	217	28	135	1,541	1,588	187	93	305	
	Manual Review	Hateful Conduct	1,219	3,023	618	704	187	1499	1,718	264	1671	15,251	12,291	1,695	705	4,130	7
	Manual Review	Violent Speech	640	886	157	158	116	425	463	55	607	3,488	8,276	465	139	999	2
	Manual Review	Abuse & Harassment	11	25	5	4	4	9	37	7	31	36	54	11	11	62	
Own Initiative	Manual Review	Hateful Conduct	380	1007	221	363	173	212	829	70	357	1701	2507	421	251	1661	1
	Manual Review	Violent Speech	13	42	25	20	4	17	65	18	44	79	155	32	21	84	

ACTIONS TAKEN ON CONTENT FOR TIUC TERMS OF SERVICE AND RULES VIOLATIONS

TIUC Terms of Service and Rules Content Removal Actions - Apr 1 to Sep 30																
Detection Method	Enforcement	Policy	Austria	Belgium	Bulgaria	Croatia	Cyprus	Czechia	Denmark	Estonia	Finland	France	Germany	Greece		
Own Initiative	Automated Means	Abuse & Harassment	5	6				3	2	1	3	15	47	3		
		Child Sexual Exploitation		6	1	1	1		3		3	38	19			
		Deceased Individuals														
		Hateful Conduct	2	7			2	4	3	1	5	17	13	2		
		Illegal or certain regulated goods and services						2				1	2	1		
		Non-Consensual Nudity	21	26	2	10		13	22	2	7	248	556	35		
		Other	18	56	2	40	20	64	60		34	328	380	88		
		Perpetrators of Violent Attacks	1								1	7	5			
		Private Information & media	3	19	1			1	4	1	3	81	39	1		
		Sensitive Media	1039	1422	797	768	236	1981	638	282	848	15076	12121	1147		
		Suicide & Self Harm											1			
		Violent Speech	1238	3484	792	852	260	1196	1435	354	1467	39291	11505	1417		
User Report	Manual Review	Abuse & Harassment	962	1276	3158	2771	214	3496	696	308	776	21227	12048	1115		
		Child Sexual Exploitation	3	33	2	2		4	5	1	9	50	46	9		
		Deceased Individuals	4	4	1	1		3	3	3	5	31	34	1		
		Hateful Conduct	48	82	22	13	8	42	37	12	25	942	428	70		
		Illegal or certain regulated goods and services	245	127	969	418	117	808	137	110	144	15631	3280	186		
		Intellectual property infringements											1			
		Misleading & Deceptive Identities		1								2				
		Non-Consensual Nudity	112	194	162	17	26	127	93	55	59	2087	1610	147		
		Perpetrators of Violent Attacks	3					1	2		1	2	29			
		Private Information & media	43	108	56	6	10	57	69	14	40	1042	732	48		
		Sensitive Media	325	699	211	138	78	379	268	30	202	5286	4086	304		
		Suicide & Self Harm	197	209	72	74	21	199	269	34	169	1375	3177	200		
		Synthetic & Manipulated Media														

Own Initiative	Manual Review	Violent & Hateful Entities					1		1		3	1	3	
		Violent Speech	1838	2747	728	682	223	1795	1435	312	1564	22442	22298	1329
		Abuse & Harassment		1				1				1	1	1
		Deceased Individuals												
		Hateful Conduct			1				1					
		Illegal or certain regulated goods and services												
		Non-Consensual Nudity												
		Private Information & media												
		Sensitive Media		1		1						1		
		Suicide & Self Harm	2	2				1		1	1		3	
		Violent Speech	1	6	4	3		5	9	1	7	8	43	7

ACTIONS TAKEN ON ACCOUNTS FOR TIUC TERMS OF SERVICE AND RULES VIOLATIONS

TIUC Terms of Service and Rules Account Suspensions - Apr 1 to Sep 30														
Detection Method	Enforcement	Policy	Austria	Belgium	Bulgaria	Croatia	Cyprus	Czechia	Denmark	Estonia	Finland	France	Germany	
Own Initiative	Automated Means	Abuse & Harassment										4	3	
		Ban Evasion	6	3	6			2	1	1	4	62	101	
		CWC for various countries for illegal activity										1		
		Child Sexual Exploitation	1164	2302	1860	559	317	1803	1056	379	1125	32782	16501	
		Financial Scam	15	29	12	8	2	21	8	1	13	244	369	
		Illegal or certain regulated goods and services	11	27	17	5	5	16	11	1	31	398	276	
		Misleading & Deceptive Identities	1342	869	798	509	391	1443	889	92	414	10352	8686	
		Non-Consensual Nudity	2	5				9	1	5	1	84	27	
		Other	391	629	4205	97	72	3030	84	95	125	10430	10888	
		Perpetrators of Violent Attacks	10	6	9	4		9	12	4	25	64	126	
		Platform Manipulation & Spam	730104	1657499	2176791	660640	170772	1256894	563526	605145	1033063	16170433	15223104	
		Sensitive Media	1	1			1			1	1	18	19	
		Suicide & Self Harm												
		Violent & Hateful Entities	88	163	40	13	32	89	91	18	103	1040	1265	
User Report	Manual Review	Abuse & Harassment	401	435	2013	1616	122	2013	298	196	339	12759	5803	
		Ban Evasion	2	6		2	1	2	3			71	24	
		CWC for various countries for illegal activity										13	3	
		Child Sexual Exploitation	10	13	16	3	6	15	15	6	3	211	137	
		Deceased Individuals						1				1	1	
		Financial Scam										4	2	

			Hateful Conduct	7	21	7	5	3	12	8		6	247	73
			Illegal or certain regulated goods and services	150	212	718	410	42	575	102	86	142	9670	3212
			Intellectual property infringements	9	18	6		1	7	6	2	7	335	106
			Misleading & Deceptive Identities	204	289	216	92	42	311	160	1850	109	2760	2443
			Non-Consensual Nudity	31	52	55	5	5	41	19	23	19	661	512
			Other	8	16	8	2	2	21	13	1	1	184	82
			Perpetrators of Violent Attacks	4	1		1		1			42	11	18
			Platform Manipulation & Spam	142	211	226	99	32	289	133	45	91	6564	2403
			Private Information & media	1	8	1	1		5	4		5	57	28
			Sensitive Media	1	3	1	1		1	2	1		29	22
			Suicide & Self Harm	4	4	3			4	6			27	40
			Username Squatting						1				1	1
			Violent & Hateful Entities	5	16	1	3	2	5	7		7	99	126
			Violent Speech	207	474	129	138	24	291	237	52	205	4397	2692
	Own Initiative	Manual Review	Child Sexual Exploitation*	90	74	80	25	19	90	59	26	41	1064	717

*This data was previously included in the 'user report' section, but with this iteration, we were able to better categorise and clarify that the detection method included a proactive element.

Overall Figures

COMPLAINTS OF ACTIONS TAKEN FOR ILLEGAL CONTENT RECEIVED - Apr 1 to Sep 30																		
	Austria	Belgium	Bulgaria	Croatia	Cyprus	Czechia	Denmark	Estonia	EU	Finland	France	Germany	Greece	Hungary	Ireland	Italy	Latvia	Lithu:
Complaints Received	18	21	0	1	2	14	11	3	0	2	52	237	5	7	56	133	5	
Overturned Appeals	11	2	0	9	1	1	1	3	0	9	2	3	4	9	1	0	4	
Median Time to Respond (Hours)	1	1		9	1	1	1	3		3	1	1	1	1	1	0	4	

COMPLAINTS OF ACTIONS TAKEN FOR TIUC TERMS OF SERVICE AND RULES VIOLATIONS RECEIVED - Apr 1 to Sep 30																	
Appeal Category	Metric	Austria	Belgium	Bulgaria	Croatia	Cyprus	Czechia	Denmark	Estonia	Finland	France	Germany	Greece	Hungary	Ireland	Italy	Latvia
Account Suspension Complaints	Complaints Received	2628	4461	1508	936	540	2366	2125	733	4004	46222	55216	2487	2236	3562	11080	4
	Overturned Appeals	632	799	297	188	111	498	538	164	1564	12047	16381	410	407	849	2016	
	Median Time to Respond (Hours)	0.3	0.2	0.2	0.6	0.6	0.3	0.2	0.3	0.6	0.4	0.1	0.1	0.2	0.5	0.3	
Content Action Complaints	Complaints Received	470	748	199	154	87	313	341	90	454	6772	7006	323	186	978	1618	
	Overturned Appeals	41	60	21	10	6	30	33	7	12	793	454	14	10	107	111	
	Median Time to Respond (Hours)	2.8	2.3	1.6	5.5	244.8	7.4	3.7	49.0	357.7	4.6	186.9	11.8	342.2	0.6	349.3	3
Live Feature Action Complaints	Complaints Received	22	44	20	34	0	19	18	2	33	388	393	22	22	37	115	
	Overturned Appeals	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
	Median Time to Respond (Hours)	50.3	102.3	63.7	80.5		349.3	63.2	398.1	8.4	58.0	38.8	72.2	74.1	53.5	57.6	1
Sensitive Media Action Complaints	Complaints Received	49	46	33	11	10	43	33	5	50	284	622	57	53	71	128	
	Overturned Appeals	31	37	18	9	8	30	33	3	39	192	480	44	40	32	84	
	Median Time to Respond (Hours)	0.2	1.1	2.0	0.5	4.3	0.8	2.5	37.6	0.9	0.4	0.5	0.9	0.1	0.6	0.7	
Restricted Reach Complaints	Complaints Received	307	503	140	219	52	339	313	73	299	2046	2668	281	146	1161	865	
	Overturned Appeals	135	233	51	80	24	153	148	33	141	942	1224	113	57	577	399	

	Median Time to Respond (Hours)	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.1	0.2	0.2	0.2	0.2	0.3	0.2	0.3
--	--------------------------------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

INDICATORS OF ACCURACY FOR CONTENT MODERATION

VISIBILITY FILTERING INDICATORS

Metric	Enforcement	Policy	Bulgarian	Croatian	Czech	Danish	Dutch	English	Finnish	French	German	Greek	Hungarian	Irish	Italian	Latvian	Polish	Po
Appeal Rate	Automated Means	Hateful Conduct	1.0%	1.9%	5.5%	4.7%	6.3%	3.1%	3.2%	4.3%	6.0%	2.7%	2.0%	15.4%	4.4%	0.0%	2.6%	
	Manual Closure	Abuse & Harassment	11.1%	2.5%	23.1%	3.7%	5.8%	4.2%	0.0%	8.5%	13.8%	8.5%	0.0%	0.0%	12.5%		5.0%	
		Hateful Conduct	1.3%	2.7%	2.8%	2.6%	2.1%	1.9%	1.0%	1.5%	3.2%	0.9%	2.9%	0.0%	1.9%	0.0%	1.4%	
		Violent Speech	0.0%	0.0%	2.7%	3.2%	0.8%	1.3%	0.4%	1.1%	2.5%	0.4%	0.0%	0.0%	1.8%	0.0%	1.5%	

Note: Cells that are blank mean that there was no enforcement. For cells containing '0.0%' value, there were no cases of successful appeals or overturns.

Metric	Enforcement	Policy	Bulgarian	Croatian	Czech	Danish	Dutch	English	Finnish	French	German	Greek	Hungarian	Irish	Italian	Latvian	Polish	Po
Overturn Rate	Automated Means	Hateful Conduct	33.3%	41.2%	55.8%	71.4%	61.6%	56.8%	63.5%	68.9%	60.8%	34.6%	38.1%	50.0%	62.4%		48.9%	
	Manual Closure	Abuse & Harassment	100.0%	100.0%	88.9%	100.0%	82.6%	71.6%		77.3%	90.2%	100.0%			80.0%		83.3%	
		Hateful Conduct	0.0%	30.0%	23.1%	45.5%	45.8%	45.2%	42.9%	52.9%	44.9%	33.3%	16.7%		44.4%		33.3%	
		Violent Speech			20.0%	0.0%	18.8%	24.8%	0.0%	26.3%	31.6%	0.0%			38.5%		0.0%	

Note: Cells that are blank mean that there was no enforcement. For cells containing '0.0%' value, there were no cases of successful appeals or overturns.

INDICATORS OF ACCURACY FOR CONTENT REMOVAL

Metric	Enforcement	Policy	Bulgarian	Croatian	Czech	Danish	Dutch	English	Finnish	French	German	Greek	Hungarian	Irish	Italian	Latvian	Polish	Po
Appeal Rate	Automated Means	Abuse & Harassment			0.0%	0.0%	0.0%	8.1%	0.0%	11.1%	5.9%				16.7%		14.3%	
		Child Sexual Exploitation					0.0%	0.0%		0.0%	0.0%				0.0%		0.0%	
		Hateful Conduct				0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%			0.0%		0.0%	
		Illegal or certain regulated goods and services						0.0%					0.0%		0.0%			
		Non-Consensual Nudity	0.0%	0.0%	0.0%	0.0%	3.3%	2.8%	0.0%	4.2%	11.0%	0.0%	0.0%		10.5%		2.0%	
		Other		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%		0.0%		0.0%	
		Perpetrators of Violent Attacks						0.0%		0.0%	0.0%						0.0%	
		Private Information & media				0.0%	0.0%	3.2%	0.0%	0.0%	0.0%				0.0%		0.0%	
		Sensitive Media					0.0%	0.0%	0.0%	0.0%	0.0%	0.0%			0.0%		0.0%	
		Suicide & Self Harm						0.0%										0.0%
		Violent Speech	2.1%	0.0%	1.6%	2.5%	1.7%	4.8%	0.7%	5.7%	5.8%	1.1%	0.9%	0.0%	2.3%	0.0%	0.6%	
	Manual Closure	Abuse & Harassment	3.6%	3.6%	4.3%	5.7%	2.4%	1.3%	0.8%	5.6%	8.7%	1.4%	3.0%	0.0%	2.0%	0.0%	2.0%	
		Child Sexual Exploitation			0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%		0.0%	0.0%	0.0%	
		Deceased Individuals					0.0%	6.6%		4.5%	20.0%		0.0%		0.0%		0.0%	
		Hateful Conduct			0.0%		0.0%	11.5%	0.0%	0.0%	16.7%	0.0%			0.0%		0.0%	
		Illegal or certain regulated goods and services	0.0%	0.0%	0.0%	0.0%	2.4%	0.0%	0.0%	0.7%	2.6%	0.0%	0.0%		0.0%		0.0%	
		Intellectual property infringements									0.0%							
		Non-Consensual Nudity	0.0%	0.0%	0.0%	4.8%	0.0%	1.0%	0.0%	2.8%	1.5%	2.3%	0.0%		0.6%	0.0%	0.8%	
		Perpetrators of Violent Attacks				0.0%		5.9%		0.0%	40.0%						0.0%	
		Private Information & media	22.2%	0.0%	0.0%	0.0%	1.0%	5.4%	0.0%	5.6%	13.9%	0.0%	0.0%		3.5%	0.0%	4.7%	

		Sensitive Media	0.0%	0.0%	0.0%	0.0%	5.7%	5.7%	0.0%	7.6%	9.7%	0.0%	0.0%	12.2%	0.0%	0.3%
		Suicide & Self Harm	0.0%	0.0%	0.0%	5.2%	1.7%	6.4%	0.0%	7.4%	14.2%	7.8%	0.0%	4.8%	0.0%	1.6%
		Violent & Hateful Entities						0.0%								
		Violent Speech	2.9%	0.0%	3.3%	1.5%	2.2%	4.7%	0.7%	5.2%	7.9%	0.2%	0.3%	2.2%	12.5%	1.9%

Note: Cells that are blank mean that there was no enforcement. For cells containing '0.0%' value, there were no cases of successful appeals or overturns.

Metric	Enforcement	Policy	Bulgarian	Croatian	Czech	Danish	Dutch	English	Finnish	French	German	Greek	Hungarian	Irish	Italian	Latvian	Polish
Overturn Rate	Automated Means	Abuse & Harassment						33.3%		0.0%	0.0%				0.0%		100.0%
		Child Sexual Exploitation															
		Hateful Conduct															
		Illegal or certain regulated goods and services															
		Non-Consensual Nudity					100.0%	50.0%		0.0%	44.4%				75.0%		50.0%
		Other															
		Perpetrators of Violent Attacks															
		Private Information & media						28.6%									
		Sensitive Media															
		Suicide & Self Harm															
		Violent Speech	0.0%		0.0%	25.0%	18.8%	22.6%	0.0%	28.2%	20.5%	33.3%	0.0%		50.0%		25.0%
	Manual Closure	Abuse & Harassment	0.0%	0.0%	25.0%	0.0%	5.9%	7.9%	0.0%	12.8%	11.6%	0.0%	0.0%		7.1%		3.2%
		Child Sexual Exploitation															
		Deceased Individuals						12.5%		0.0%	0.0%						
		Hateful Conduct						42.3%			50.0%						
		Illegal or certain regulated goods and services					0.0%	0.0%		100.0%	0.0%						
		Intellectual property infringements															
		Non-Consensual Nudity				0.0%		6.1%		10.0%	14.3%	0.0%			0.0%		0.0%
		Perpetrators of Violent Attacks						0.0%			30.0%						
		Private Information & media	66.7%				0.0%	9.0%		5.9%	10.4%				50.0%		16.7%
		Sensitive Media					0.0%	2.6%		3.8%	2.7%				1.6%		0.0%
		Suicide & Self Harm				0.0%	0.0%	12.1%		11.5%	7.5%	0.0%			35.7%		22.2%
		Violent & Hateful Entities															
		Violent Speech	50.0%		22.7%	0.0%	12.5%	10.0%	0.0%	16.0%	14.3%	0.0%	0.0%		13.3%	33.3%	5.7%

Note: Cells that are blank mean that there was no enforcement. For cells containing '0.0%' value, there were no cases of successful appeals or overturns.

INDICATORS OF ACCURACY FOR SUSPENSIONS

Metric	Enforcement	Policy	Bulgarian	Croatian	Czech	Danish	Dutch	English	Estonian	Finnish	French	German	Greek	Hungarian	Irish	Italian
Appeal Rate	Automated Means	Abuse & Harassment						20.0%			0.0%					
		Ban Evasion			0.0%		25.0%	34.3%			20.0%	44.1%	0.0%	0.0%		33.3%
		Child Sexual Exploitation	45.7%	30.9%	69.9%	58.8%	57.1%	26.1%	0.0%	43.6%	66.8%	70.1%	78.3%	51.8%		76.7%
		CWC for various						266.7%								0.0%

	countries for illegal activity																
	Financial Scam			0.0%		20.0%	6.9%			114.2%	124.1%	0.0%	0.0%			45.2%	
	Illegal or certain regulated goods and services					0.0%	4.3%			0.0%	0.0%	0.0%	0.0%			28.6%	
	Misleading & Deceptive Identities	22.7%	41.7%	9.3%	2.3%	17.1%	8.0%		15.0%	38.7%	18.0%	34.6%	9.4%			13.9%	
	Non-Consensual Nudity					100.0%	4.4%			0.0%	0.0%		0.0%			0.0%	
	Other	180.0%	0.0%	64.7%	128.6%	100.0%	0.9%		0.0%	103.8%	41.7%	60.0%	55.6%			53.7%	
	Perpetrators of Violent Attacks		0.0%	0.0%		0.0%	11.6%		0.0%	12.5%	6.7%		0.0%	0.0%		11.1%	
	Platform Manipulation & Spam	9.5%	28.4%	4.6%	3.0%	5.0%	0.3%	0.6%	5.7%	7.0%	4.5%	9.5%	4.8%	0.0%		3.9%	
	Sensitive Media					0.0%	4.2%			50.0%	0.0%						
	Suicide & Self Harm						50.0%										
	Username Squatting	0.0%	100.0%	200.0%	50.0%	111.8%	61.8%		100.0%	70.6%	95.8%	0.0%	0.0%			84.2%	
	Violent & Hateful Entities		0.0%	0.0%	0.0%	7.5%	6.4%		0.0%	6.4%	9.6%	0.0%	0.0%			18.9%	
	Manual Closure	Abuse & Harassment	37.5%	21.6%	27.5%	45.7%	35.4%	2.1%		23.5%	35.9%	43.3%	25.0%	23.4%			31.6%
		Ban Evasion				100.0%	0.0%	170.1%		33.3%	98.8%	135.7%	272.7%				150.0%
Child Sexual Exploitation		14.3%	3.4%	9.9%	9.1%	18.2%	2.9%		27.8%	18.8%	20.3%	11.8%	14.9%			22.2%	
Civic Integrity						0.0%	0.0%										
CWC for various countries for illegal activity						0.0%	49.2%			53.3%	300.0%					50.0%	
Deceased Individuals					0.0%		14.3%				0.0%		0.0%				
Financial Scam		100.0%	0.0%	0.0%	0.0%	11.5%	7.9%		0.0%	124.6%	9.3%	200.0%	0.0%			33.9%	
Hateful Conduct		200.0%	0.0%	33.3%	50.0%	29.4%	42.2%		36.4%	56.0%	83.8%	52.6%	0.0%			65.0%	
Illegal or certain regulated goods and services		100.0%	33.3%	0.0%	0.0%	71.2%	3.5%		14.3%	81.1%	74.7%	33.3%	35.3%			45.9%	
Intellectual property infringements		850.0%	600.0%	175.0%	87.5%	94.1%	233.1%		366.7%	1304.6%	271.8%	725.0%	125.0%			574.6%	
Misleading & Deceptive Identities		50.0%	100.0%	265.2%	10.9%	32.7%	21.8%		107.1%	64.4%	61.1%	136.0%	28.8%			37.6%	
Non-Consensual Nudity		20.0%	12.5%	34.6%	0.0%	42.1%	23.3%		42.9%	41.8%	41.3%	48.9%	24.2%			44.9%	
Other		57.1%	111.1%	140.0%	72.7%	119.8%	8.4%		90.0%	567.7%	219.2%	172.7%	363.3%			302.2%	
Perpetrators of Violent Attacks			0.0%	18.2%	600.0%	0.0%	33.7%		4.7%	334.8%	38.8%	0.0%	0.0%	0.0%		41.2%	
Platform Manipulation & Spam		23.9%	21.1%	21.3%	17.9%	33.4%	2.9%	14.3%	24.4%	37.2%	25.9%	39.1%	30.7%	0.0%		28.1%	
Private Information & media			0.0%	0.0%	0.0%	33.3%	19.8%		50.0%	41.0%	81.8%	0.0%	0.0%			150.0%	
Sensitive Media		33.3%	50.0%	66.7%	100.0%	84.6%	30.9%		0.0%	58.0%	54.1%	50.0%	157.1%			76.5%	
Suicide & Self Harm		0.0%		33.3%	0.0%	162.5%	79.0%		0.0%	50.0%	44.4%		220.0%			45.8%	
Username Squatting						0.0%	184.8%			20.0%	50.0%					66.7%	
Violent & Hateful Entities		0.0%	33.3%	0.0%	4.8%	11.8%	8.3%		6.7%	5.8%	15.6%	17.7%				30.7%	
Violent Speech		43.1%	43.1%	40.6%	26.0%	39.2%	46.1%		37.9%	63.0%	89.2%	54.0%	29.7%			63.7%	

Note: Cells that are blank mean that there was no enforcement. For cells containing '0.0%' value, there were no cases of successful appeals or overturns.

Metric	Enforcement	Policy	Bulgarian	Croatian	Czech	Danish	Dutch	English	Estonian	Finnish	French	German	Greek	Hungarian	Irish	Italian	Latvi
Overturn Rate	Automated Means	Abuse & Harassment						0.0%									

		Ban Evasion					0.0%	3.6%			0.0%	0.0%			0.0%	
		Child Sexual Exploitation	0.0%	0.6%	0.3%	0.3%	0.4%	0.5%		0.3%	0.6%	0.4%	0.7%	0.3%	0.4%	0.0%
		CWC for various countries for illegal activity						0.0%								
		Financial Scam					0.0%	0.0%			0.0%	0.0%			0.0%	
		Illegal or certain regulated goods and services						0.0%							0.0%	
		Misleading & Deceptive Identities	40.0%	10.0%	26.3%	0.0%	26.2%	15.9%		17.6%	23.5%	24.7%	23.4%	16.0%	16.4%	
		Non-Consensual Nudity					0.0%	0.0%								
		Other	0.0%		0.0%	11.1%	12.2%	9.4%			16.3%	7.2%	16.7%	20.0%	15.9%	
		Perpetrators of Violent Attacks						11.5%			0.0%	0.0%			0.0%	
		Platform Manipulation & Spam	9.0%	5.6%	8.2%	8.2%	8.0%	16.4%	0.0%	12.4%	10.3%	10.9%	8.1%	8.0%	6.9%	9.6%
		Sensitive Media						0.0%			0.0%					
		Suicide & Self Harm						0.0%								
		Username Squatting		0.0%	25.0%	0.0%	5.3%	7.9%		0.0%	25.0%	34.8%			12.5%	0.0%
		Violent & Hateful Entities					40.0%	10.0%			5.9%	0.0%			28.8%	
	Manual Closure	Abuse & Harassment	0.0%	0.0%	0.0%	0.0%	4.4%	4.1%		0.0%	4.0%	3.5%	0.0%	0.0%	3.7%	0.0%
		Ban Evasion				0.0%		1.1%		0.0%	2.4%	0.0%	0.0%		0.0%	
		Child Sexual Exploitation	0.0%	0.0%	5.6%	0.0%	10.7%	3.5%		3.1%	6.2%	4.2%	9.1%	0.0%	4.3%	
		Civic Integrity														
		CWC for various countries for illegal activity						0.0%			12.5%	33.3%			0.0%	
		Deceased Individuals						0.0%								
		Financial Scam	0.0%				0.0%	3.9%			7.9%	0.0%	0.0%		0.0%	
		Hateful Conduct	0.0%		100.0%	40.0%	60.0%	30.9%		25.0%	29.1%	32.8%	50.0%		42.3%	
		Illegal or certain regulated goods and services	0.0%	0.0%			0.0%	0.1%		0.0%	0.4%	0.0%	0.0%	0.0%	0.0%	
		Intellectual property infringements	5.9%	0.0%	14.3%	0.0%	6.3%	2.5%		0.0%	1.1%	5.2%	0.0%	0.0%	1.7%	0.0%
		Misleading & Deceptive Identities	16.7%	20.0%	4.1%	40.0%	5.2%	4.7%		13.3%	6.3%	8.6%	4.4%	0.0%	4.7%	
		Non-Consensual Nudity	0.0%	0.0%	0.0%		0.0%	1.1%		0.0%	2.6%	1.7%	0.0%	0.0%	0.0%	
		Other	0.0%	0.0%	0.0%	0.0%	0.8%	1.2%		0.0%	0.8%	0.4%	0.0%	0.9%	0.7%	0.0%
		Perpetrators of Violent Attacks			0.0%	0.0%		0.9%		0.0%	0.0%	0.0%			0.0%	
		Platform Manipulation & Spam	2.7%	0.0%	0.2%	0.7%	0.8%	1.0%	0.0%	0.5%	0.8%	1.1%	1.5%	0.6%	0.8%	6.7%
		Private Information & media					0.0%	27.8%		0.0%	12.5%	0.0%			11.1%	
		Sensitive Media	0.0%	0.0%	0.0%	100.0%	18.2%	17.3%			31.0%	25.0%	0.0%	9.1%	0.0%	
		Suicide & Self Harm			0.0%		0.0%	1.0%			0.0%	4.2%		0.0%	0.0%	
		Username Squatting						0.0%			0.0%	0.0%			0.0%	
		Violent & Hateful Entities		0.0%		0.0%	0.0%	1.1%		100.0%	4.8%	3.7%	0.0%		3.7%	
		Violent Speech	3.6%	10.0%	4.3%	12.0%	5.4%	4.8%		5.5%	3.9%	3.3%	1.6%	5.2%	3.3%	0.0%

Note: Cells that are blank mean that there was no enforcement. For cells containing '0%' value, there were no cases of successful appeals or overturns.

Art. 24.2: Average Monthly Active Recipients - Apr 1 to Sep 30			
Country Name	Logged In Users	Logged Out Users	Total
Austria	810,875	596,764	1,407,639
Belgium	1,506,657	1,048,225	2,554,882
Bulgaria	448,180	256,928	705,107
Croatia	328,973	439,587	768,561
Cyprus	170,672	104,181	274,853
Czechia	1,029,161	1,029,173	2,058,333
Denmark	772,227	400,955	1,173,182
Estonia	179,326	115,516	294,843
Finland	1,488,241	829,364	2,317,605
France	13,039,693	7,084,102	20,123,795
Germany	11,272,823	5,683,320	16,956,143
Greece	1,001,335	900,135	1,901,470
Hungary	717,275	521,725	1,239,000
Ireland	1,465,243	862,010	2,327,252
Italy	5,455,121	2,743,750	8,198,871
Latvia	274,666	165,222	439,888
Lithuania	409,848	146,978	556,826
Luxembourg	153,726	78,721	232,447
Malta	82,705	41,075	123,780
Netherlands	5,109,779	3,262,163	8,371,941
Poland	5,575,832	3,537,284	9,113,116
Portugal	1,667,325	806,391	2,473,716
Romania	1,372,621	533,260	1,905,881
Slovakia	281,635	251,873	533,508
Slovenia	193,824	252,193	446,017
Spain	10,073,378	6,038,881	16,112,258
Sweden	1,792,584	867,529	2,660,113

Further Information on Suspensions

During the applicable reporting period 1 April, 2024 to 30 September, 2024, there were zero actions taken for: provision of manifestly unfounded reports or complaints; or manifestly illegal content. While manifestly illegal content is not a category that we have taken action on during the reporting period, we suspended 159,011 accounts for violating our Child Sexual Exploitation policy and 7,321 for violating our Violent and Hateful Entity policy.

Disputes submitted to out-of-court dispute settlement bodies.

To date, zero disputes have been submitted to the out-of-court settlement bodies.

Reports received by trusted flaggers.

To date, we have received 6 reports from Article 22 DSA approved trusted flaggers. Once Article 22 DSA awarded trusted flaggers information is published, we immediately enrol them in our trusted flaggers program, which ensures prioritisation of human review, via their email, username, and account.