HateAid

Rights without reach

The DSA put to the test

Funded by

STIFTUNG MERCATOR

Trigger warning:
This publication contains depictions of discriminatory, racist language.

Contents

1 Executive Summary

The Digital Services Act (DSA) entered into force in February 2024 with an ambitious goal: to "create a safe digital environment in which the fundamental rights of all users of digital services are protected."[1] Expectations for the new regulatory framework were accordingly high. It was hoped that users would be better protected from digital violence and harmful content. To achieve this, the DSA introduced EU-wide standards for reporting and moderating illegal content. While the definition of "illegal content" remains within the competence of the Member States, the DSA obliges all online platforms to provide user-friendly reporting channels, access to internal complaint systems, and out-of-court dispute settlement mechanisms.

Between April 2024 and July 2025, HateAid examined how very large online platforms are implementing these new obligations on content moderation. To this end, HateAid submitted reports of illegal content and pursued internal complaints as well as dispute settlement options. The aim was to assess the effectiveness of the user rights established by the DSA in practice and to document the experiences of users systematically.

The findings reveal serious **deficiencies in the implementation of the new rules** and a **clear need for improvement**. Very large online platforms, dispute settlement bodies, and supervisory authorities are currently failing to meet their obligations adequately. Only about **half of the illegal content reported by HateAid was removed**, even though internal complaints and, in many cases, additional out-of-court dispute settlement proceedings were pursued after the initial reports. Consequently, users and people affected by digital violence are often unable to exercise their rights effectively.

There are significant differences between platforms. On Facebook and Instagram, **removal rates of illegal content declined** significantly in early 2025. YouTube consistently removed only around one-third of reported content, while TikTok and X showed modest improvements. **Access to redress mechanisms** remains difficult: incomplete feedback, the absence of reference numbers, and in some cases the complete lack of internal complaint procedures makes it nearly impossible for affected users to follow up on decisions, preserve evidence, or effectively assert their rights. Likewise, **out-of-court dispute settlement** under Article 21 DSA, intended as a swift and low-cost alternative to judicial proceedings, often proves ineffective in practice. While some providers issue prompt decisions, others fail to respond for months.

The **quality of content moderation** raises further concern. Highly standardised processing times on TikTok and X suggest a large-scale **use of automated decision-making**, yet the platforms fail to label such decisions as required by law. Illegal content expressed in so-called **"algospeak"**, code words or symbols easily understood by humans but bypassing automated filters, often remains unprocessed. In addition, there are **shortcomings in supervision**. Complaints submitted to the Digital Services Coordinator (DSC) are subject to long delays, possibly due in part to limited staffing. The certification process for organisations seeking recognition as Trusted Flaggers under Article 22 DSA has likewise proven lengthy and burdensome, with unexpectedly high procedural demands on civil society actors reporting content in this capacity.

Overall, it is clear that although the DSA has laid the theoretical foundations for stronger protection, its potential remains largely untapped in practice. Without clear standards, sufficient resources for supervision and consistent monitoring of platforms, the DSA is at risk of failing to achieve its central objective of effectively protecting users from digital violence.

[1] https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package.

2 Methodology

Between April 2024 and July 2025, HateAid examined how core user rights under the DSA, in particular Articles 16, 20, and 21, are implemented on five very large online platforms: Facebook, Instagram, TikTok, YouTube, and X. The focus lay on the perspective of people who are confronted with illegal content or affected by it themselves. The basis was a total of 301 reported illegal contents, 134 internal complaints and 50 proceedings before four EU-certified out-of-court dispute settlement bodies (User Rights GmbH, Appeals Centre Europe, ADROIT, RTR).[2]

The assessment relied on legal expertise and a strict four-eyes principle. All reported content was reviewed by qualified lawyers to ensure that only material highly likely to constitute “illegal content” under the DSA was included. This was documented extensively, for example through formal legal assessments.

The content was selected through manual screening and notices from the HateAid counselling service for people affected by digital violence. All cases were legally reviewed and classified as illegal incitement to hatred, anti-constitutional symbols, incitement to or approval of criminal offences, insults, defamation or slander. Each report was comprehensively documented. The results are not representative but provide qualitative insights into structural deficits in the implementation of DSA obligations. The findings presented here refer to the period up to mid-August 2025, the end of the project’s test phase. The project is funded by the Mercator Foundation.

[2] The aforementioned bodies were selected because they were already certified in October 2024 and either German or English are permitted as procedural languages. However, HateAid only submitted two cases to the Appeals Centre Europe and refrained from submitting further cases, as the scope of review there is limited to violations of platform terms and conditions and does not include an examination of the illegality of content. Therefore, the findings focus on User Rights GmbH, ADROIT and RTR.

3 Practical findings: reports, complaints and out-of-court dispute settlement

3.1 Illegal content remains online – despite the DSA

One of the DSA’s key objectives is to ensure that illegal content is promptly and thoroughly reviewed and removed after being reported. However, the analysis shows that this promise remains largely unfulfilled in practice: in 2024, only 44% of the total 301 illegal pieces of content were removed from all platforms included in the investigation after a simple report to the platforms. In 2025, this rate rose slightly to 47%. While 24% of internal complaints about a negative decision were successful in 2024, this figure fell to just 7.2% in 2025. This represents a decline of more than two-thirds (71%). Referral to an out-of-court dispute settlement body led to the removal of the content in 34% of cases. In many cases, however, the bodies did not deliver a decision or the platforms did not implement the communicated decision (35% of decisions). After exhausting all available legal remedies, only 57% of reported illegal content was removed during the project period.

3.1.1 Meta platforms: decline in the removal of illegal content

The Meta-operated platforms **Facebook and Instagram removed significantly less illegal content in 2025 than in 2024**, see figure 1. The change in the handling of internal complaints is particularly striking. Although these were decided much more quickly in 2025 than in 2024, they were generally rejected.

In 2024, **Facebook removed a total of 65% of reported content** after a simple report. At the same time, 37.5% of internal complaints about unfavourable decisions were successful. Overall, Facebook removed 78% of reported content following reports and complaints. In 2025, however, this figure was only 44%. In 2025, Facebook removed only 41% of content following a simple report, and only 6% of complaints about negative decisions on illegal content were successful.[3]

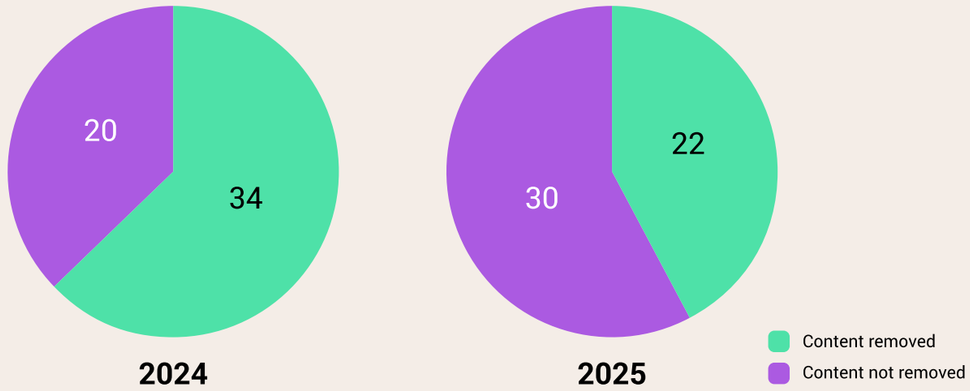
Instagram also saw a **reduction in the removal rate of illegal content**. In 2024, the platform removed 52% of reported illegal content after a report or complaint, while in 2025 this figure was only 40%. The main reason for this is a decline in the success rate of internal complaints against negative decisions. Instagram removed only 29% of content in 2024 and 40% in 2025 following a simple report. However, while 33% of internal complaints against negative decisions were successful in 2024, not a single complaint was successful in 2025.

[3] HateAid has not yet been able to obtain any findings from Facebook through out-of-court dispute settlement. No decision has yet been received from a dispute settlement body in the proceedings initiated against Facebook. HateAid has initiated dispute settlement proceedings against Facebook starting in March 2025.

This development coincides with a change to Meta's internal guidelines made in January 2025, which was announced by CEO Mark Zuckerberg in February 2025. The reformed guidelines stipulate that certain politically or religiously motivated content will no longer be considered a violation of the platform's terms and conditions.[4] However, the policy change must not alter the handling of illegal content. The investigation revealed, however, that comparable illegal content that was removed in 2024 following a report under the DSA remained online in 2025.[5]

The **processing time for reports** is less than 24 hours for 80% of reports on Facebook and for 63% of reports on Instagram. Most decisions can therefore be assessed as 'immediate' within the meaning of Article 16(5) DSA.[6] Facebook in particular showed significant outliers: in 2025, two reports took more than two months to be resolved. When comparing platforms, it is striking that meta platforms process few cases in less than an hour (Facebook: 14%, Instagram: 5%). We consider this to be an indication that both platforms, as stated in their transparency reports, have reports assessed manually.

Figure 1: Meta platforms: Removal rate of illegal content after report and complaint



[4] https://transparency.meta.com/de-de/policies/community-standards/hateful-conduct/.

[5] Example 1 (original content in German): Content removed from Facebook in 2024 after a report: "My greatest wish is that every Green Party member will soon die or be killed." Left online on Facebook in 2025 after a report and complaint: „Now the Green bitch is completely losing it. Throw the guy in the locked ward and leave him there to rot……..🤬🤬🤬‼️"; Example 2 (original content in German): 2024 removed on Facebook after report: "Stasi whore no thanks". Left online on Facebook in 2025 after report and only removed after complaint "Then Fuck yourself bitch" (both in reference to female politicians); Example 3 (original content in German): Content removed from Instagram in 2024 after being reported: "Scum…🤮", "Dirt belongs in the loony bin". Content left online on Instagram in 2025 after report and complaint: "There's still a few of those scumbags missing. People say and do that every day. Waste separation, waste disposal. They're hazardous waste and have to be disposed of SEPARATELY." .

[6] The Code of Conduct on Countering Illegal Hate Speech Online+ can be used as a benchmark for the 24-hour timeframe (see 2.3 of the Code of Conduct), which has been signed by all platforms examined in the report.

The **processing time for internal complaints** has been significantly reduced during the project period: in 2024, Facebook needed more than seven days to process 50% of complaints submitted, while Instagram needed more than seven days for 32%. Such lengthy processing suggests a violation of Article 20(4) of the DSA, which requires complaints to be processed promptly. In 2025, both platforms were significantly faster and generally processed complaints (Facebook: 94%, Instagram: 93%) in less than three days. In two cases, the platforms took more than a month to reach a decision. As explained, all complaints in 2025 were unsuccessful.

3.1.2 YouTube: lowest removal rates

The Google-operated platform **YouTube has the lowest removal rate of illegal content** of all providers examined. Overall, **only 32% of reported content was removed** during the period under review.[7] While 38% of content was removed in 2024, this figure fell to 28% in 2025. It is striking that YouTube did **not communicate any decision** on a significant part of reports even though the platform acknowledged the receipt of all reports: in 2024, this affected almost half (48%) of cases, and in 2025, around a third (34%). Both platform notifications in the profile and communications by email were examined. The lack of feedback raises significant questions about compliance with Article 16(5) and (6) of the DSA, which requires platforms to process reports and communicate their decision to the reporting party without delay.

In cases where feedback was provided, **YouTube usually processed reports within 24 hours** (2024: 92%, 2025: 80%). However, YouTube often sent an email to the reporting person after a report was submitted, requesting further information. In many cases, the questions, for example about the violated law, had already been answered in the report. If the person reporting does not respond immediately, the processing time can be significantly delayed: for one reported piece of content in 2025, YouTube only responded 23 days after receiving such an email request.

3.1.3 TikTok: slight improvement with ratings remaining low

During the project period, **TikTok removed a total of 38% of reported illegal content**, with this rate increasing in 2025 compared with the previous year: in 2024, a total of 33% of content was removed, while in 2025 the rate was 46%. In 2024, only 26% of content was removed following a simple report, while internal complaints were successful in 19% of cases. In 2025, 46% of content was removed following a simple report, but internal complaints against negative decisions were never successful. TikTok also frequently failed to communicate a decision, namely in 21% of cases.

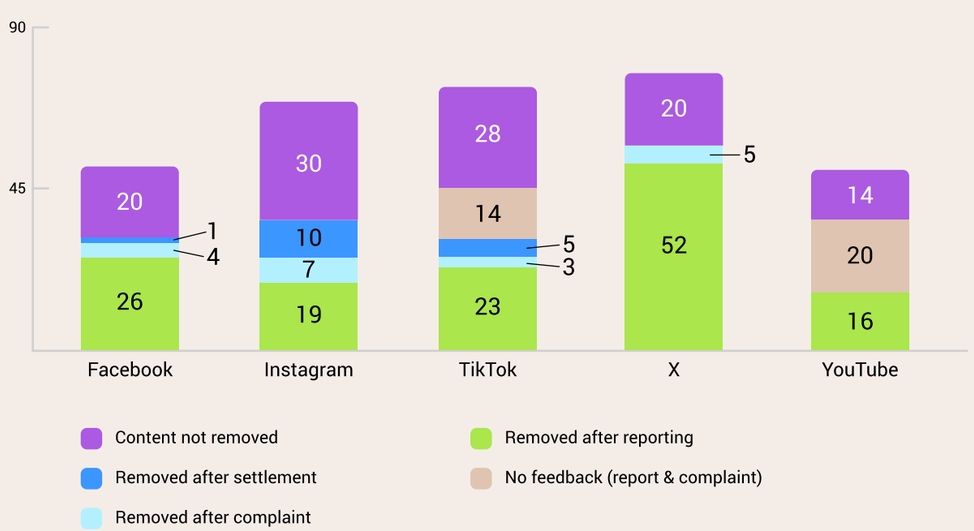
TikTok generally **responded to reports and complaints within 24 hours** (2024: approx. 87% of reports and complaints, 2025: 64% of reports, 100% of complaints).

[7] The removal rate is based solely on content removed following a report. Following a complaint or out-of-court dispute resolution proceedings, no further content was removed from YouTube in our analysis, which is due to the fact that these legal remedies are hardly accessible, see 3.3.1.

3.1.4 X (formerly Twitter): high removal rates

X has the highest removal rate of all platforms examined: during the project years, a total of 74% of reported content was removed after reporting and, where applicable, internal complaints.[8] This figure remained consistent: throughout the entire project period, around 67% of content was removed following a simple report. An internal complaint led to success in approximately 20% of cases (2024: 18%, 2025: 21%). X generally processes reports and complaints very quickly: 99% of reports and 100% of all complaints submitted were processed in less than 24 hours, with 70% of reports and 88% of complaints being processed in less than an hour.

Figure 2: Total: reports, complaints and dispute settlements sorted by platform during the project period from april 2024 to july 2025



[8] HateAid has not yet been able to achieve the removal of content from X through out-of-court dispute resolution. No decision has yet been received from a dispute resolution body in the proceedings initiated against X. HateAid has initiated dispute resolution proceedings against X starting in February 2025.

3.2 Access to reporting channels is made difficult for users on all platforms

The ability to report illegal content easily is a central pillar of the DSA's system for dealing with illegal content. According to Article 16 of the DSA, all providers of intermediary platforms are obliged to provide easy to access and user-friendly reporting channels. Our analysis shows that none of the platforms examined adequately meets these legal requirements. On the contrary, users on all platforms are systematically discouraged from reporting illegal content due to excessive complexity, untransparent structuring, deterrent wording and targeted design elements.

A key **shortcoming here is the unclear separation of reporting channels**: all platforms strictly differentiate between reports of violations of the terms and conditions of the platforms and those of possible legal violations under Article 16 DSA. However, this differentiation is not clear to users. The two channels are not presented in a way that distinguishes them either visually or in terms of content, and there is no assistance for users. The option 'Report as illegal' is often only found at the end of a long list of options. This is because it is preceded by categories such as 'bullying or harassment' or 'violence, hate or exploitation'. These are legal-sounding terms that mislead users into reporting legal violations under these categories. However, these categories conceal reporting channels exclusively for violations of terms and conditions. This is advantageous for the platforms, as reports based on the terms and conditions that are not explicitly marked as reports of illegal content under the DSA are not subject to the strict obligations of Article 16. There is a high potential for users to unwittingly waive their rights in this way, including the right to confirmation of receipt (4), an immediate decision with information on legal remedies (5), and the right to prompt and non-arbitrary processing as well as the labelling of automated decisions (6).

In addition, the **reporting channels for illegal content under the DSA are generally much more complicated, time-consuming and convoluted than those for violations of terms and conditions**. Users, especially on mobile devices, often must click through a multitude of confusing categories to complete the reporting process. The actual reporting process also often asks for complex information that requires in-depth legal knowledge. This is also confirmed by an evaluation of the reporting channels conducted by the organisation das NETTZ in 2025 through a user survey: the test subjects considered reporting illegal content under the DSA to be significantly more difficult and complicated than reporting under the terms and conditions. For example, users of X must select a 'territory' whose laws the content violates, with both individual EU member states and the option 'In the EU' available for selection. Facebook, Instagram and YouTube also require the full name of the reporting person to be provided, although this is not a mandatory requirement under Art. 16 DSA. Platforms must simply give users the option to provide their name.[9]

These structural hurdles can lead to what is known as **'click fatigue'**,[10] a psychological fatigue effect, whereby users become exasperated and abandon the reporting process or unconsciously select the wrong category. Research by das NETTZ[11] also confirms that reporting channels can overwhelm users: more than a quarter (27%) of reports submitted via the DSA's illegal content reporting channel were abandoned. The reasons frequently given for abandoning the report were that the person reporting felt overwhelmed by the choices or did not know how to proceed.

[9] See: edpb_guidelines_202503_interplay-dsa-gdpr_v1_en.pdf, para. 25 ff.

[10] See: Martini/Kramme/Kamke: Dark Patterns im Scheinwerferlicht des Digital Services Act, MMR 2023, 323, 325.

[11] https://www.das-nettz.de/publikationen/zwischen-klick-und-konsequenz-eine-evaluation-der-meldeverfahren-von-plattformen-nach, p. 34.

On YouTube, for example, people who select the 'legal issue' category, must confirm several times that they actually want to report the content for a legal violation, whereas no such double confirmation is required for reports under the terms and conditions. Requesting information again by email, even though it was already included in the report, creates additional hurdles and tires users. Another example is the length of the click paths on Facebook: if users want to report an illegal threat, for example, five clicks are required to report it as a violation of the terms and conditions, but at least 15 clicks are required to report it as a legal violation. **This form of deliberate complexity can be considered a so-called dark pattern**, which is a prohibited practice under Article 25 of the DSA.

A dark pattern is the deliberately manipulative design of digital processes that prevents users from exercising their legal rights. Repeatedly asking users to make a decision they have already made is even cited in Article 25(3) as an example of a dark pattern.

Technical limitations and intimidating wording make matters worse: many forms contain explicit warnings about possible sanctions for providing false information. In some cases, users are advised to submit a report 'with guidance from a qualified lawyer.' Technical restrictions also have a hindering effect: TikTok, for example, limits the number of characters in the explanation to 400, which makes it difficult to provide a comprehensible description of the violation.

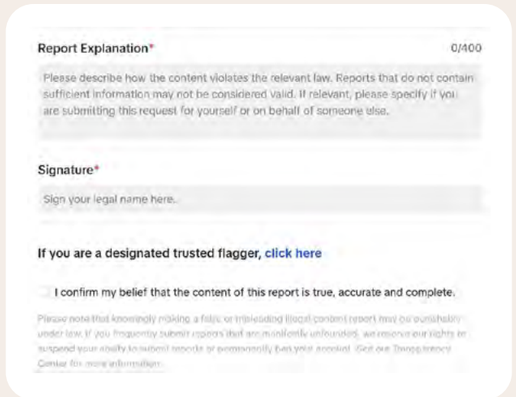


Figure 3: Excerpt from TikTok's reporting form, in which various factors can make it difficult or discourage users from submitting a report: the character limit for the report statement, the requirement to provide the full name of the person submitting the report, and the claim that reports could be punishable by law.

Figure 4: Excerpt from YouTube's reporting form. Here, too, a full name must be provided, and reporting persons must swear 'under penalty of perjury' that the information given is accurate. The wording 'I swear [...] that I am authorised to report this alleged violation' is also highly misleading: according to Article 16(1) of the DSA, any person has the right to report content to a platform that they consider to be illegal. No specific authorisation is required. It is not clear how a blanket criminal liability for false statements could arise, nor are individuals liable if the reported content is not punishable by law.

3.3 Shortcomings in the redress system: internal complaint procedures and out-of-court dispute settlement

The **DSA provides for a tiered system of redress**: after reporting a violation, users should be able to use the platforms' internal complaint procedures in accordance with Article 20 DSA. They can also initiate out-of-court dispute settlement proceedings in accordance with Article 21 DSA. This interaction is intended to ensure low-threshold, effective and cost-efficient legal protection. However, the analysis shows that this promise has so far only been fulfilled to a limited extent. Significant shortcomings are apparent at both levels: platforms make access to internal complaint procedures difficult or effectively impossible, while not all dispute resolution bodies comply with deadlines, communication and responsibilities.

3.3.1 Platforms make it difficult to access legal remedies

The analysis revealed that **platforms often fail to provide clear information about legal remedies or make them difficult to find**. In doing so, platforms violate the right to an effective remedy under Article 16(5) and the right to an easily accessible and effective complaint procedure under Article 20(1) and (3) of the DSA.

A key problem for users is **incomplete and unclear feedback on reports**. Under Article 16(5) of the DSA, platforms are required to provide reporters with a decision 'in relation to' the reported information. Such a reference is necessary to be able to assign the feedback to a specific report. If this is missing, effective enforcement is significantly impeded. On **TikTok**, for example, acknowledgements of receipt are sent exclusively via notifications in the app, not by email, and do not contain a reference number. If several reports are submitted in parallel, it is practically impossible to assign such feedback to the corresponding report. Only when a complaint is submitted via the 'Request further review' button does **TikTok** display the originally reported content, thus enabling assignment.

The **practice at X is similarly problematic**: although the platform sends a confirmation of receipt immediately after a report is submitted, this is purely generic and contains neither a reference number nor any details about the report. Users who submit multiple reports cannot distinguish between the different cases. The notification of the moderation decision by X does contain a link to the reported content and a reference number. As with TikTok, however, users face a problem when no decision is made. They cannot use the purely generic confirmation of receipt without a reference to prove that a report has been made to courts or supervisory authorities. However, this is necessary to establish liability in accordance with Art. 6(1)(a) DSA. **Facebook and Instagram** are in stark contrast to this: the platforms send acknowledgements of receipt and decision feedback by email to the address provided in the reporting process, assign each report an individual reference number and enclose a copy of all information.

In addition to insufficient feedback, **some platforms also lack an internal complaint procedure altogether or conceal it**. Article 20(1) of the DSA obliges platforms to offer an internal complaint procedure for all types of decisions following reports. This shortcoming is particularly evident once again on **YouTube**: while an internal complaint procedure is available for reported videos, it is completely absent for reported comments. Accordingly, there is no information about the possibility of an internal complaint. This constitutes a violation of Article 16(5) (information on legal remedies) and Article 20(1) DSA. X has a similar problem when it comes to reporting violations of terms and conditions: contrary to Article 20(1) DSA, X does not allow internal complaints to be lodged against negative decisions, even though Recital 58 makes it clear that this should be possible.

Another obstacle is the **lack of or limited possibilities for preserving evidence**. For an effective appeal process, it is crucial that users can document the reported content and the feedback received in a legally secure manner. TikTok in particular does not offer sufficient functions for this, neither in the desktop view or in the mobile version: the date and time of comments are not displayed precisely. Instead, only relative time information such as “two weeks ago” is displayed, rather than a specific date with an exact time. YouTube and Instagram also only display relative time information in both versions; absolute time information can only be obtained indirectly. This makes legal proceedings particularly difficult, as they regularly require complete documentation. The situation is different in the desktop version of Facebook and both versions of the X platform: here, the date and time are available in such a way that legally secure screenshots can be taken and, in case of doubt, events can be documented.

3.3.2 Shortcomings in out-of-court dispute settlement bodies

Out-of-court dispute settlement under Article 21 DSA is designed as a key instrument to offer users a fast, cost-effective and technically sound alternative to court proceedings. Between October 2024 and July 2025, HateAid initiated a total of 50 proceedings with four EU-certified dispute resolution bodies: **User Rights GmbH** (Germany), **Appeals Centre Europe** (Ireland), **ADROIT** (Malta) and **RTR** (Austria).[12] The aim was to test the practical effectiveness of the rights enshrined in Article 21 DSA, in particular processing time, quality of decisions, communication with parties involved in the proceedings, and implementation by the platforms.

The **competencies of the dispute resolution bodies are regulated differently** and are listed on the EU Commission’s website: **User Rights GmbH** handles cases involving illegal content on LinkedIn, Instagram, TikTok and, since July 2025, Facebook. The **Appeals Centre Europe**, on the other hand, is limited to cases involving alleged violations of the platforms’ terms and conditions and is responsible for Facebook, YouTube and TikTok; cases involving illegal content are not reviewed there. **ADROIT** in Malta is platform-neutral and accepts cases relating to all online platforms, including X, Snapchat, Skype and Telegram. The same applies to the Austrian **RTR**, which has no restrictions on the platforms it covers.

The results:
In half of the cases submitted (50%), the dispute resolution body called upon has made a decision. One of these decisions rejects a request,[13] four decisions (16%) find that the content has already been removed during the ongoing proceedings and that the matter has therefore been resolved, and the remaining 80% confirm the illegality of the content and request the platform to remove the content in question. To date, this decision has led to the removal of the content by the platform in approximately two-thirds (65%) of cases.[14] This is unexpected, as the removal is voluntary and the effectiveness of dispute resolution is therefore unclear at this stage.

In practice, however, there are **significant differences in the quality and effectiveness of the procedures**: while some settlement bodies, above all User Rights GmbH, demonstrate that Article 21 DSA can be effectively implemented, **long processing times, limited review standards and a lack of communication** at other bodies prevent enforcement.

[12] The aforementioned bodies were selected because they were already certified in October 2024 and either German or English are permitted as procedural languages. However, HateAid only submitted two cases to the Appeals Centre Europe and refrained from submitting further cases, as the standard of review there is limited to violations of platform terms and conditions and does not examine the illegality of content. Therefore, the findings focus on User Rights GmbH, ADROIT and RTR.

[13] This dismissive decision by the Appeals Centre Europe was not further justified. However, it stands to reason that the application was rejected because the Appeals Centre Europe only accepts cases relating to platform terms and conditions, and our application concerned the illegality of content.

[14] It cannot be established beyond doubt in every case that the content was removed specifically as a result of the dispute resolution procedure. In some cases, there is no confirmation from the platform that it removed the content because of the decision in the dispute resolution, but the content can no longer be found online.

The legal framework stipulates that **decisions must generally be made within 90 days**, or within 180 days in exceptional cases. In 50% of cases (25 out of 50), no decision has been delivered at least 90 days after the start of proceedings, and in seven cases, proceedings have been ongoing for over 180 days without a decision being communicated. At ADROIT alone, this applies to all proceedings filed (11), and at the Austrian dispute resolution body RTR, the proceedings also remained largely undecided (13 out of 14 proceedings undecided for at least 90 days). This indicates potential violations of Article 21(4) DSA. Particularly serious is that in many cases, no response was received for months. In one-fifth of the cases decided (20%), a decision was only delivered more than 90 days after the proceedings were filed. The removal of content by the platforms was repeatedly not communicated and had to be checked manually; in at least two cases, it took the platforms more than a month after the decision was delivered to remove the content. The outcome of the proceedings is therefore often incomprehensible or very time-consuming for users.

User Rights GmbH confirmed the illegality of the content in all closed cases, **provided detailed and transparent reasons for its decisions**, and generally achieved implementation by the platforms within three weeks of the decision. A total of 73% of the content that was subject to proceedings at User Rights was removed. A good quarter (27%) of the content remained online, even though the dispute resolution body requested the platform to remove it. **ADROIT and RTR show considerable shortcomings**: in some cases, the dispute resolution bodies provided updates on the proceedings, such as the delivery of the application to the platform. However, decisions were generally not made despite inquiries, in several cases at ADROIT for over ten months. RTR only decided in one of 14 cases submitted, while ADROIT did not decide on any of the eleven cases submitted. The Appeals Centre Europe is also of limited use, as this body only checks compliance with the platforms’ terms and conditions when illegal content is reported. Although removal was achieved in one case, TikTok only implemented this after more than two and a half months.

It is particularly problematic that, in the case of platforms X and YouTube, it remains unclear to this day which dispute settlement body can effectively conduct proceedings concerning illegal content. This lack of clarity is also reflected in the transparency reports: X states that no proceedings were received by dispute settlement bodies in the last reporting period,[15] **although two proceedings were brought against X** during this period as part of this investigation, **one with ADROIT and the other with RTR**. While ADROIT reported that it had forwarded the case to X during the reporting period, RTR did not forward it until more than five months later, well beyond the 90-day deadline. Google states that although cases were received during the reporting period, no decision was made regarding **YouTube**.[16] Here, too, the effectiveness of the proceedings remains unclear. In contrast, **Instagram and TikTok** provide specific information on pending and already concluded dispute resolution procedures. In the investigation underlying this report, the two platforms voluntarily implemented the decision of the dispute resolution body in each of the cases decided in 2024 by User Rights GmbH.[17]

[15] X DSA Transparency Report, October 2024 – March 2025, S. 29, https://transparency.x.com/assets/dsa/transparency-report/dsa-transparency-report-april-2025.pdf.

[16] Google EU Digital Services Act (EU DSA) Biannual VLOSE/VLOP Transparency Report, Juli – Dezember 2024, S. 41 f., https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-27_2024-7-1_2024-12-31_en_v1.pdf.

[17] Facebook also states in its latest transparency report that the platform did not receive any cases from out-of-court dispute resolution bodies during the reporting period. However, as User Rights GmbH after the reporting period has expanded its scope of review to cover Facebook, this is likely to change in the future.

3.4 Moderation systems

A functioning moderation system is central to curbing the spread of illegal content. Current regulation is based on liability privileges: platforms are not liable for illegal content, but only if they fail to act after receiving a notification, Art. 6(1) DSA. The legislator has decided to rely on reports from users to deal with illegal content. This makes it even more crucial that reporting and moderation processes function reliably and are accessible.

Article 16(6) of the DSA requires platforms to indicate when a **decision has been made using automated means**. When processing complaints, purely automated decision-making is even completely prohibited under Article 20(6) of the DSA.

The analysis raises doubts about the implementation of these rules: TikTok shows recurring and strikingly symmetrical processing times, which strongly indicate the use of automated processes without the platform informing users about this, as required by Article 16(6) of the DSA. In 2024, 62% of processed reports and 76% of submitted complaints were processed in exactly 30 minutes. In 2025, this pattern occurred less frequently but was still evident in 36% of all reports. Even when the processing time varied, the pattern remained recognisable: many decisions were made at exact intervals of 30 minutes, for example after 1 hour 30 minutes or 1 day 7 hours 30 minutes. Nevertheless, **in no case was there any indication of the use of automated systems, as required by Article 16(6) DSA**. In its most recently published transparency report[18] TikTok states that reports of illegal content are first automatically checked for possible violations of the community guidelines (terms and conditions). Only if no violation is found is a human moderator supposed to check whether a legal violation has occurred. However, this cannot be considered as justification.

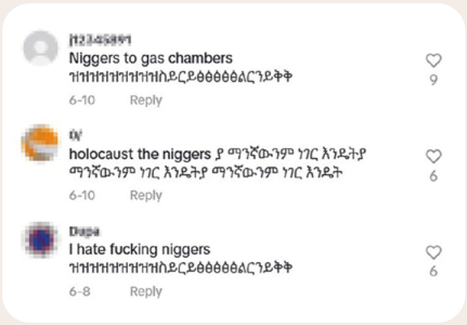
X also decided on reports with conspicuous speed, which also suggests that **automated means are being used**: Overall, X processed 30% of the reports submitted in less than 10 minutes, 9% in less than five minutes, and one report within one minute in each of the two project years. X does not provide any information about the potential use of automated means. In its transparency report, the platform also states that all reports and complaints are reviewed by human oversight.[19] It remains unclear how this process can be completed within a few minutes.

A particularly serious shortcoming is also evident in the **moderation of content on TikTok in so-called 'algospeak'**. This refers to deliberate linguistic concealment, such as the use of non-Roman characters or emojis, to circumvent platform filters. Although the content is clearly understandable to human readers and often clearly illegal, the investigation found that TikTok often did not respond at all to reports of such content. In 2024, no decision was made on 71% of all cases of clearly illegal content in algospeak over a period of at least three months. Even in the few cases where TikTok did decide on such reports, there were repeated instances of incomprehensible decisions: even obvious formal insults or inflammatory content were not classified as illegal.

The consequences extend far beyond individual cases: when illegal content in algospeak is systematically overlooked by moderators, it creates a 'safe space' for perpetrators. They can spread illegal content, such as incitement to hatred, anti-Semitic insults or sexualised threats of violence, undisturbed. Algospeak is often used specifically in coordinated hate campaigns, when identical or similar coded comments appear en masse in comment sections.

Trigger warning: Depictions using discriminatory, racist language.

Figure 5:[20] Examples of illegal 'algospeak' comments on TikTok.



[18] TikTok DSA Transparency Report, july – december 2024, p. 5, https://sf16-va.tiktokcdn.com/obj/eden-va2/zayvwlY_fjulyhwzuhy[/ljhwZthlaukjlkulzlp/DSA_H2_2024/TikTok-DSA-Transparency-Report-July-to-Dec-2024.pdf.

[19] X DSA Transparency Report, october 2024 – march 2025, p. 7, https://transparency.x.com/assets/dsa/transparency-report/dsa-transparency-report-april-2025.pdf.

[20] The comments were published on the platform TikTok and anonymised for data protection reasons. They are no longer online.

4 Other findings on DSA implementation

Under Article 53, users and, explicitly, organisations and associations can lodge complaints about platforms for breaches of the DSA with the national Digital Services Coordinators (DSC). This includes the right to be heard and to be informed about the status of the complaint. If the national DSC considers itself to be incompetent, it forwards the complaint to the DSC of the country where the platform in question is based and, if necessary, adds its own opinion on the matter. However, experience with filing complaints with the Federal Network Agency (BNetzA) as the Digital Services Coordinator (DSC) shows that civil society is currently not sufficiently empowered to fulfil its role as complainant in the supervisory regime.

During the project period, HateAid submitted a total of seven complaints, mostly to the German DSC, the Federal Network Agency. The results: **complaints took a very long time to process**, often several months. Furthermore, it was not clear what stage a complaint was at and whether it had led to the initiation of supervisory proceedings. In several cases, HateAid only received information after six to seven months that the complaint had been forwarded to the responsible Irish DSC. Repeatedly, months after a complaint was filed, the German DSC informed us that it assumed the proceedings had been completed and offered to continue processing only if we explicitly requested this. This may also reflect the limited staffing resources of the German DSC, which, even after an increase in staff, still has to make do with only 24 employees across all departments. In the summer of 2025, there was a first positive sign in terms of transparency: in mid-August, the Irish DSC informed HateAid that the complaint against Meta, which had been submitted to the German DSC in June and forwarded, had been incorporated into ongoing proceedings by the European Commission under Article 16 of the DSA.

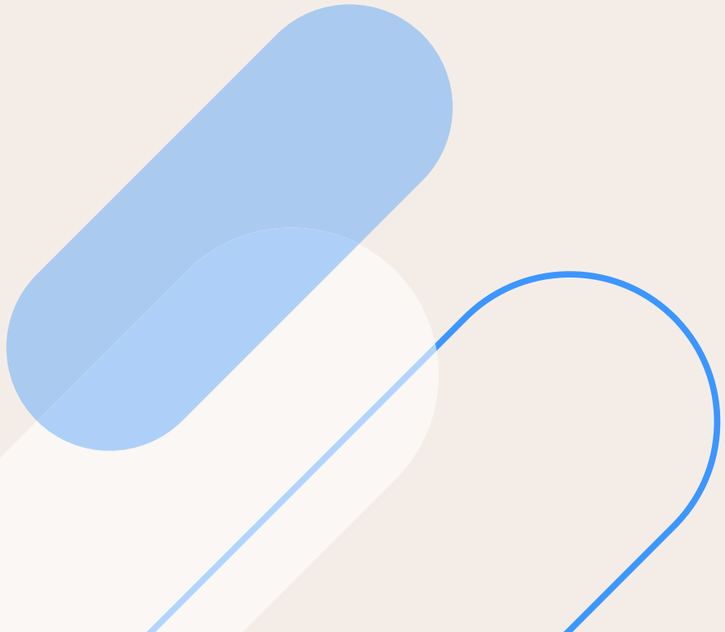
Challenges in the application process as a Trusted Flagger became even more apparent. In June 2025, HateAid was certified as a Trusted Flagger in accordance with Article 22 of the DSA. The process proved to be lengthy, taking over eight months, and was characterised by considerable uncertainty regarding the content requirements. The scope and the occasional lack of transparency of the procedure were particularly challenging: for long periods, it remained unclear which of the information requested was a mandatory requirement for approval and which was merely supplementary. HateAid had to submit extensive documentation on its structure, working methods and internal documentation without knowing which information was actually relevant to the decision. The evidence requested included examination certificates of employees. Although it is entirely justified and even desirable to set high quality requirements for Trusted Flaggers, some of these went beyond the official guidelines.[21] This resulted in the unnecessary expenditure of resources in some cases and prolonged the process. At the same time, it should be acknowledged that the approval process was still largely untested at the time of application.

Another bottleneck arises from the unexpectedly restrictive interpretation of the expertise requirement: Article 22(2) requires ‘special expertise and competence in relation to the identification, detection and reporting of illegal content’. The BNetzA, on the other hand, required that each report be either made directly by a person who has passed at least the first state law exam or approved by such a person. It is unclear whether this standard is set so high across Europe. This narrow interpretation not only limits the pool of potential Trusted Flaggers but also makes it more difficult to use organisations specialising in specific topics, such as anti-Semitism or stalking counselling centres. The discrepancy in comparison with platforms is particularly problematic: there, moderation teams with no legal training decide on the removal of content, some of which has a global impact. Trusted Flaggers, who cannot remove content themselves but can only report it, are thus subject to higher requirements than the platforms that carry out deletions. In addition, during the application period, there was noticeable public and media sensitivity, which was partly linked to concerns about attacks.

[21] https://www.bundesnetzagentur.de/DSC/DE/4TrustedF/leitfaden.pdf?__blob=publicationFile&v=3.

5 Political assessment & recommendations: what needs to be done

For the DSA to achieve real impact in practice, supervisory authorities and politicians must take decisive action. Experience with the implementation of the DSA to date clearly shows that without **decisive supervision** and **structural improvements**, the law is at risk of falling short of its own ambitions in key areas. Although the DSA has created important rights for users, practice reveals serious gaps in transparency, enforcement and access to justice. For users, and especially those affected by digital violence, it is often a matter of chance whether they receive support.



5.1 Decisive supervision and enforcement

Despite **dozens of ongoing supervisory proceedings, neither the national Digital Services Coordinators (DSCs) nor the European Commission have so far issued any formal orders or sanctions** against platforms. Yet consistent enforcement is essential if the law is to be effective. Delays or political considerations must not influence enforcement. Most recently, the imposition of a fine on X, which had been in preparation since January 2025, was postponed awaiting the outcome of negotiations in the tariff dispute with the US.[22] It is vital to warn against making DSA enforcement dependent on such extraneous considerations. Digital sovereignty requires that the EU's democratic values be upheld and enforced in the digital sphere. While the imposition of fines is not the goal of regulation, it remains an indispensable means to ensure compliance and credibility.

5.2 Clarifications and amendments to the legal text

The forthcoming evaluation of the DSA must address existing ambiguities and loopholes within the legislation. Platforms actively exploit interpretative uncertainties and design flexibilities in the DSA to undermine the exercise of users' rights. In practice, users are discouraged from reporting illegal content through the design of reporting channels and forms. These shortcomings stem from ambiguities within the DSA itself, which must be clarified in the legislative text.

In particular, anonymous reporting without the disclosure of a real name must be explicitly permitted to effectively protect users, especially those affected by severe digital violence. This requires clarification in Article 16(2) and in Recitals 50 and 53, which are currently open to divergent interpretations. The Guidelines of the European Data Protection Board (EDPB) from 11 September 2025[23] (EDPB) already state that while platforms may offer users the option to provide a name and email address, they must not make this a mandatory condition for submitting a report under Article 16. Nonetheless, several platforms still require users to provide their full name and appear to regard this as compliant with the law. This practice deters users from filing reports — an outcome that ultimately benefits the platforms.

[22] https://www.nytimes.com/2025/09/02/business/dealbook/elon-musk-x-europe-trump.html, https://www.handelsblatt.com/politik/international/zollverhandlungen-eu-pausiert-offenbar-ermittlungen-gegen-x/100142302.html, last accessed on October 13, 2025, at 3 pm.

[23] See: edpb_guidelines_202503_interplay-dsa-gdpr_v1_en.pdf.

5.3 Ensuring access to redress

Internal complaint procedures and out-of-court dispute resolution must be available for all decisions concerning reported content, regardless of whether they are based on terms-of-service violations or breaches of law. Articles 20(1) DSA and Recital 58 already provide for this explicitly. However, for users to access these procedures, they must be made aware of them. A redress information should therefore be made a mandatory part of every decision notification, including for reports submitted under the platforms' terms of service. This obligation is currently not clearly established in law. As a result, users who, misled by the confusing design of reporting interfaces, report content as a TOS violation rather than as illegal, or those who report disinformation, remain unaware of their right to appeal, even though the complaints procedure could provide an effective remedy.

5.4 Improving quality and consistency in dispute settlement

There is also considerable **need for reform in out-of-court dispute settlement**. Surprisingly, this mechanism has proven to be a potentially useful tool for resolving content moderation disputes quickly and at low cost, even though the decisions of dispute settlement bodies are not binding on platforms. However, the differences between existing bodies are so significant that the outcome for users often depends on chance — whether their case happens to be handled in line with the legal requirements.

To remedy this, two complementary instruments are needed: First, a **Code of Conduct for dispute settlement bodies under Article 45 DSA** should be developed to establish standards on independence, neutrality, user communication, and transparency of decisions. Such a code could be elaborated in a practical manner involving dispute bodies, platforms, civil society, and the European Commission. However, it would only apply to those bodies that voluntarily sign it and would therefore not ensure uniform binding standards. For this reason, the European Commission should additionally be obligated to issue **guidelines on Article 21 DSA**. Such guidelines would serve as interpretative tools for all dispute settlement bodies and establish EU-wide standards, for instance regarding time limits, obligations to state reasons for decisions made, and the duty to inform users about whether platforms have implemented decisions. Unlike codes of conduct, guidelines create legal clarity across the board and apply independently of voluntary adherence.

Supervisory authorities must also prioritise user rights in their work and sanction violations. This includes both explicit procedural obligations, such as labelling automated decisions and the broader duty to interpret the law in a user-friendly manner. Moreover, platforms must be urged to assign reference numbers to reports to make decisions traceable and verifiable.

5.5 Reforming the trusted flagger certification process

EU-wide minimum standards are also needed for the certification of trusted flaggers, particularly concerning expertise requirements, which should not be set disproportionately high. Uniform, practical criteria would provide clarity and facilitate access for qualified civil society organisations. Instead of rigid legal qualification thresholds, specialised and trained non-legal organisations with proven experience in identifying and reporting illegal content should likewise be eligible for certification.

5.6 Ensuring transparency and traceability

Another key shortcoming is the **absence of a unified EU-wide system for tracking complaints** submitted to supervisory authorities under Article 53 DSA. Currently, neither affected individuals nor civil society organisations have a reliable, low threshold means to monitor the progress of their reports and complaints. This lack of transparency hampers not only oversight of procedural progress but also civil society support for authorities and the independent evaluation of platform practices.

Civil society organisations — whose expertise supervisory bodies crucially rely upon — must be able to demonstrate to funders and the public the tangible impact of their work. Without reliable feedback from authorities or traceability of complaint procedures, such evidence is lacking, reducing both accountability and incentives to invest scarce resources in time-consuming reporting and complaint processes.

Only through transparent, verifiable, and adequately resourced cooperation between civil society, supervisory authorities and platforms can the DSA live up to its promise of protecting fundamental rights effectively in the digital sphere.

6 Annex

Table 1: Removal of content by platform and stage of proceedings (as of October 16th 2025)

	Number of reports	Content removed (total)	Content not removed (total, incl. cases without decision)	Content removed after report	>1 month no decision on report	Number of complaints	Content removed after complaint	> 1 month no decision on c omplaint	Number of out-of-court-dispute settlement cases[24]	Content removed after out-of-court dispute settlement	> 90 days no decision on out-of-court dispute settlement
Facebook total	**50**	**30**	**20**	**26**	**0**	**24**	**4**	**1**	**10**	**1**	**6**
2024	23	18	5	15	0	8	3	0	0	0	0
2025	27	12	15	11	0	16	1	1	10	1	6
Instagram total	**56**	**35**	**21**	**19**	**0**	**36**	**7**	**1**	**16**	**10**	**4**
2024	31	22	9	9	0	21	7	0	7	6	1
2025	25	13	12	10	0	15	0	1	9	4	3
TikTok total	**68**	**31**	**37**	**23**	**14**	**22**	**3**	**0**	**11**	**5**	**3**
2024	42	18	24	11	10	16	3	0	7	4	3
2025	26	13	13	12	4	6	0	0	4	2	1
X total	**77**	**57**	**20**	**52**	**0**	**25**	**5**	**0**	**11**	**0**	**11**
2024	34	25	9	23	0	11	2	0	0	0	0
2025	43	32	11	29	0	14	3	0	11	0	11
YouTube total	**50**	**17**	**33**	**17**	**18**	**1**[25]	**0**	**1**	**2**[26]	**0**	**0**
2024	21	8	13	8	9	1	0	1	1	0	0
2025	29	9	20	9	9	0	0	0	1	0	0

[24] In 2024, we only submitted dispute settlement cases against Instagram and TikTok (and a test case against YouTube), because User Rights, the only platform from which we received feedback, only handled cases against Instagram and TikTok. Regarding the other arbitration bodies, we therefore also only submitted cases concerning these two platforms, in order to be able to compare the work of the dispute resolution bodies with each other. In 2025, we expanded the investigation to the remaining platforms to find out whether effective dispute resolution with them was possible at all.

[25] The low number of complaints against YouTube is due to the fact that the platform does not provide a complaints procedure for a comment being reported as a legal violation. One complaint was possible because in this case a video was reported and not a comment.

[26] The low number of dispute settlement cases against YouTube can also be explained by the fact that a complaints procedure must first be completed before a dispute settlement procedure can be initiated. Despite the lack of decisions in the complaints procedures, we therefore called on a dispute settlement body twice only for test purposes – in one case, the application was rejected, and in the other, the dispute settlement body RTR confirmed the illegality and requested YouTube to remove the content. The platform did not comply with this decision within 28 days.

Table 2: Processing times for reports and complaints by platform (as of October 16th 2025)

	Number of reports / complaints	<1 hour (of which <10 min)	1–24 hours	1–2 days	3–6 days	7–14 days	>14 days	≥1 month no decision
Facebook reports total	**50**	**7 (1)**	**33**	**7**	**1**	**0**	**2**	**0**
Reports 2024	23	7 (1)	13	2	1	0	0	0
Reports 2025	27	0	20	5	0	0	2	0
Facebook complaints total	**24**	**1 (1)**	**12**	**3**	**3**	**3**	**2**	**0**
Complaints 2024	8	0	1	0	3	3	1	0
Complaints 2025	16	1 (1)	11	3	0	0	1	0
Instagram reports total	**56**	**3 (0)**	**32**	**19**	**2**	**0**	**0**	**0**
Reports 2024	31	1 (0)	24	5	1	0	0	0
Reports 2025	25	2 (0)	8	14	1	0	0	0
Instagram complaints total	**36**	**6 (2)**	**14**	**4**	**5**	**3**	**4**	**0**
Complaints 2024	21	4 (2)	6	0	5	3	3	0
Complaints 2025	15	2 (0)	8	4	0	0	1	0
TikTok reports total	**68**	**28 (0)**	**14**	**9**	**1**	**2**		**14**
Reports 2024	42	20 (0)	8	1	1	2	0	10
Reports 2025	26	8 (0)	6	8	0	0	0	4
TikTok complaints total	**31**	**22 (0)**	**5**	**3**	**1**	**0**	**0**	**0**
Complaints 2024	21	16 (0)	1	3	1	0	0	0
Complaints 2025	10	6 (0)	4	0	0	0	0	0
X reports total	**77**	**54 (23)**	**21**	**1**	**0**	**1**	**0**	**0**
Reports 2024	34	21 (11)	11	1	0	1	0	0
Reports 2025	43	33 (12)	10	0	0	0	0	0
X complaints total	**25**	**22 (12)**	**3**	**0**	**0**	**0**	**0**	**0**
Complaints 2024	11	9 (7)	2	0	0	0	0	0
Complaints 2025	14	13 (5)	1	0	0	0	0	0
YouTube reports total	**50**	**15 (3)**	**12**	**2**	**1**	**0**	**2**	**18**
Reports 2024	21	4 (0)	7	1	0	0	0	9
Reports 2025	29	11 (3)	5	1	1	0	2	9
YouTube complaints total	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**1**
Complaints 2024	1	0	0	0	0	0	0	1
Complaints 2025	0	0	0	0	0	0	0	0

Imprint

HateAid gGmbH
Greifswalder Straße 4
10405 Berlin

Telephone: +49 (0) 30 25208802
Email: kontakt@hateaid.org
hateaid.org

Registered office of the company: Berlin
Register court: Amtsgericht Charlottenburg
Commercial register number: HRB 203883 B
VAT ID No: DE322705305

CEOs: Anna-Lena von Hodenberg and
Josephine Ballon