



Hate Aid

Recht ohne

Reichweite

Der DSA im Praxistest

Inhalt

Gefördert durch

STIFTUNG
MERCATOR



Triggerwarnung:

Diese Veröffentlichung enthält Darstellungen mit diskriminierender, rassistischer Sprache.

© HateAid gGmbH

1 Zusammenfassung	4
2 Methodik	6
3 Erkenntnisse aus der Praxis: Meldungen, Beschwerden, außergerichtliche Streitbeilegung	7
3.1 Illegale Inhalte bleiben sichtbar – trotz DSA	7
3.1.1 Meta-Plattformen: Rückgang bei der Entfernung illegaler Inhalte	7
3.1.2 YouTube: Niedrigste Entfernungsquoten	9
3.1.3 TikTok: Leichte Verbesserung bei weiterhin niedrigen Quoten	9
3.1.4 X (ehemals Twitter): Hohe Entfernungsquoten	10
3.2 Der Zugang zu Meldewegen wird Nutzenden auf allen Plattformen erschwert	11
3.3 Defizite im Rechtsbehelfssystem: interne Beschwerdeverfahren und außergerichtliche Streitbeilegung	13
3.3.1 Plattformen erschweren Zugang zu Rechtsbehelfsmöglichkeiten	13
3.3.2 Defizite bei außergerichtlichen Streitbeilegungsstellen	14
3.4 Moderationssysteme	16
4 Sonstige Erkenntnisse zur DSA-Implementierung	18
5 Politische Bewertung & Forderungen: Was jetzt zu tun ist	20
5.1 Entschlossene Aufsicht und Durchsetzung	20
5.2 Klarstellungen und Nachbesserungen im Gesetzestext	21
5.3 Zugang zu Rechtsbehelfen sicherstellen	21
5.4 Qualität und Einheitlichkeit der Streitbeilegung stärken	22
5.5 Trusted-Flagger-Zertifizierungsverfahren reformieren	22
5.6 Transparenz und Nachvollziehbarkeit schaffen	23
6 Anhang	24

1 Zusammenfassung

Der Digital Services Act (DSA) trat im Februar 2024 in Kraft. Sein ambitioniertes Ziel: Die „Schaffung eines sicheren digitalen Raums, in dem die Grundrechte aller Nutzenden digitaler Dienste geschützt werden“.¹ Die Erwartungen an das neue Regelwerk waren entsprechend hoch. Es bestand die Hoffnung, dass Nutzende künftig besser vor digitaler Gewalt geschützt würden. Dazu setzte man im Regelwerk auf europaweit einheitliche Systeme zur Moderation illegaler Inhalte. Obwohl der DSA die Definition von illegalen Inhalten den Mitgliedsstaaten überlässt, schreibt er für alle Online-Plattformen verbindlich die Einrichtung nutzungsfreundlicher Meldewege, Zugang zu internen Beschwerdesystemen und außergerichtlichen Streitbeilegungsstellen vor.

Zwischen April 2024 und Juli 2025 untersuchte HateAid, wie sehr große Online-Plattformen die neuen Vorschriften zur Moderation illegaler Inhalte umsetzen. Dafür wurden illegale Inhalte gemeldet und alle Möglichkeiten der Moderation ausgeschöpft. Ziel war es, die Wirksamkeit der durch den Digital Services Act geschaffenen, Nutzendenrechte zu überprüfen und die Erfahrungen aus Sicht der Nutzenden systematisch auszuwerten.

Im Ergebnis zeigt die Untersuchung gravierende **Mängel** bei der Umsetzung der neuen Regeln und eindeutige Nachbesserungsbedarfe. Sehr große Online-Plattformen, Streitbeilegungsstellen und Aufsichtsbehörden kommen ihren Pflichten bisher nur unzureichend nach. Nur etwa **die Hälfte der von HateAid gemeldeten illegalen Inhalte wurde entfernt**, obwohl nach jeder erfolglosen Meldung eine interne Beschwerde eingelegt und oftmals zusätzlich ein außergerichtliches Streitbeilegungsverfahren angestoßen wurde. Nutzende und Betroffene digitaler Gewalt haben so zwar formal Rechte, der Zugang bleibt ihnen jedoch in vielen Fällen verwehrt.

Zwischen den Plattformen selbst zeigen sich dabei große Unterschiede: Auffällig war, dass bei Facebook und Instagram die **Entfernungsquoten illegaler Inhalte** seit Anfang 2025 deutlich zurückgingen. Bei YouTube wurde hingegen durchgehend nur rund ein Drittel der gemeldeten illegalen Inhalte entfernt. Bei TikTok und X waren sogar moderate Verbesserungen bei der Entfernung illegaler Inhalte erkennbar. Auch beim **Zugang zu Rechtsbehelfen** bestehen erhebliche Hürden: Unzureichende Rückmeldungen, fehlende Referenznummern oder

gänzlich fehlende interne Beschwerdeverfahren erschweren es Betroffenen, Entscheidungen nachzuvollziehen, Beweise zu sichern und ihre Rechte wirksam geltend zu machen. Ebenso erweist sich die **außergerichtliche Streitbeilegung** nach Artikel 21 DSA, die als schnelle und kostengünstige Alternative zu Gerichtsverfahren konzipiert ist, in der Praxis vielfach als unwirksam. Während einige Anbieter zügige Entscheidungen hervorbringen, versäumen es andere, den Nutzenden über Monate hinweg eine Entscheidung mitzuteilen.

Zudem wirft die **Qualität der Moderation** Fragen auf: Auffällig standardisierte Bearbeitungszeiten bei TikTok und X deuten auf einen großflächigen **Einsatz automatisierter Entscheidungen** hin. Diese werden jedoch, entgegen der gesetzlichen Vorschriften, von den Plattformen nicht als solche gekennzeichnet. Illegale Inhalte in sogenanntem „**Algospeak**“, deren Inhalt für Menschen problemlos zu verstehen sind, bleiben in der Moderation oft unbearbeitet. Hinzu kommen **Defizite in der Aufsicht**. Beschwerden beim Digital Services Coordinator (DSC) haben lange Bearbeitungszeiten, was wohl auch auf die personelle Ausstattung der Behörde zurückzuführen ist. Das Bewerbungsverfahren für Organisationen als Trusted Flagger erwies sich als langwierig und stellte unerwartet hohe Anforderungen an zivilgesellschaftliche Akteur*innen, die in dieser Funktion Inhalte melden.

Insgesamt zeigt sich: Der DSA hat zwar die theoretische Grundlage für einen stärkeren Schutz geschaffen, doch in der Praxis bleibt sein Potenzial bislang weitgehend ungenutzt. Ohne klare Standards, ausreichend Ressourcen für die Aufsicht und eine konsequente Kontrolle der Plattformen droht der DSA, sein zentrales Ziel zu verfehlen, Nutzende wirksam vor digitaler Gewalt zu schützen.

¹ <https://digital-strategy.ec.europa.eu/de/policies/digital-services-act-package>.

2 Methodik

HateAid hat für diese Untersuchung zwischen April 2024 und Juli 2025 die Umsetzung zentraler Rechte des DSA, insbesondere Artikel 16, 20 und 21, auf fünf sehr großen Online-Plattformen, Facebook, Instagram, TikTok, YouTube und X, untersucht. Im Mittelpunkt stand dabei die Perspektive von Menschen, die mit illegalen Inhalten konfrontiert oder selbst davon betroffen sind. Grundlage waren insgesamt 301 gemeldete strafbare Inhalte, 134 interne Beschwerden sowie 50 Verfahren vor vier, von der EU zertifizierten, Streitbeilegungsstellen (User Rights GmbH, Appeals Centre Europe, ADROIT, RTR).²

Die Untersuchung basiert auf juristischer Expertise in der Bewertung illegaler Inhalte sowie einem Vier-Augen-Prinzip. Alle gemeldeten Inhalte wurden von Jurist*innen bewertet, um sicherzustellen, dass nur Inhalte gemeldet werden, die mit hoher Wahrscheinlichkeit als „illegal“ im Sinne des DSA einzustufen sind. Dies wurde umfangreich, zum Beispiel durch Rechtsgutachten, dokumentiert.

Die Auswahl der Inhalte erfolgte über manuelles Screening und Hinweisen aus der Betroffenenberatung von HateAid. Alle Fälle wurden juristisch geprüft und als illegale Volksverhetzung, verfassungsfeindliche Symbole, Aufforderung zu oder Billigung von Straftaten, Beleidigung, üble Nachrede oder Verleumdung eingestuft. Jede Meldung wurde umfassend dokumentiert. Die Ergebnisse sind nicht repräsentativ, sondern liefern qualitative Einblicke in strukturelle Defizite bei der Umsetzung der DSA-Pflichten. Der hier dargestellte Erkenntnisstand bezieht sich auf den Zeitraum bis Mitte August 2025, dem Ende der Testphase im Projekt. Das Projekt wird durch die Stiftung Mercator gefördert.

² Die genannten Stellen wurden ausgewählt, weil sie bereits im Oktober 2024 zertifiziert waren und entweder Deutsch oder Englisch als Verfahrenssprachen zugelassen sind. Allerdings legte HateAid beim Appeals Centre Europe nur zwei Fälle ein und verzichtete auf weitere Einlegungen, da der Prüfungsmaßstab dort auf Verletzungen von Plattform-AGB begrenzt ist und keine Illegalität von Inhalten überprüft wird. Daher konzentrieren sich die Erkenntnisse auf die Stellen User Rights GmbH, ADROIT und RTR.

3 Erkenntnisse aus der Praxis: Meldungen, Beschwerden, außergerichtliche Streitbeilegung

3.1 Illegale Inhalte bleiben sichtbar – trotz DSA

Ein zentrales Ziel des DSA ist es, dass illegale Inhalte nach einer Meldung zeitnah und sorgfältig geprüft und entfernt werden. Die Untersuchung zeigt jedoch, dass dieses Versprechen in der Praxis weitgehend unerfüllt bleibt: Im Jahr 2024 wurden auf allen in der Untersuchung enthaltenen Plattformen nur 44 % der insgesamt 301 illegalen Inhalte nach einer einfachen Meldung bei den Plattformen entfernt. Im Jahr 2025 stieg diese Quote leicht auf 47 % an. Während im Jahr 2024 noch 24 % der internen Beschwerden über eine abschlägige Entscheidung Erfolg hatten, waren es im Jahr 2025 nur noch 7,2 %. Dies entspricht einem Rückgang von mehr als zwei Drittel (71 %). Die Anrufung einer außergerichtlichen Streitbeilegungsstelle führte in 34 % der Fälle zur Entfernung des Inhalts. Vielfach blieben die angerufenen Stellen jedoch untätig oder die Plattformen setzten eine getroffene Entscheidung nicht um (35 % der Entscheidungen). Nach Ausschöpfung aller verfügbaren Rechtsbehelfe wurden insgesamt im Projektzeitraum nur 57 % der gemeldeten illegalen Inhalte entfernt.

3.1.1 Meta-Plattformen: Rückgang bei der Entfernung illegaler Inhalte

Die von Meta betriebenen Plattformen **Facebook und Instagram entfernten im Jahr 2025 deutlich weniger illegale Inhalte als noch im Jahr 2024**, siehe Abbildung 1. Auffällig ist vor allem die Veränderung der Behandlung interner Beschwerden. Diese wurden zwar im Jahr 2025 deutlich schneller entschieden als noch im Jahr 2025, dafür jedoch in der Regel abgewiesen.

So entfernte Facebook 2024 nach einer einfachen Meldung insgesamt 65 % der gemeldeten Inhalte. Gleichzeitig waren 37,5 % der internen Beschwerden über abschlägige Entscheidungen erfolgreich. Insgesamt entfernte Facebook so auf **Meldung und Beschwerde** 78 % der gemeldeten Inhalte. 2025 waren es hingegen nur noch 44 %. Im Jahr 2025 entfernte Facebook im Jahr 2025 nur noch 41 % der Inhalte nach einer einfachen Meldung und nur noch 6 % der Beschwerden über abschlägige Entscheidungen bei illegalen Inhalten waren erfolgreich.³

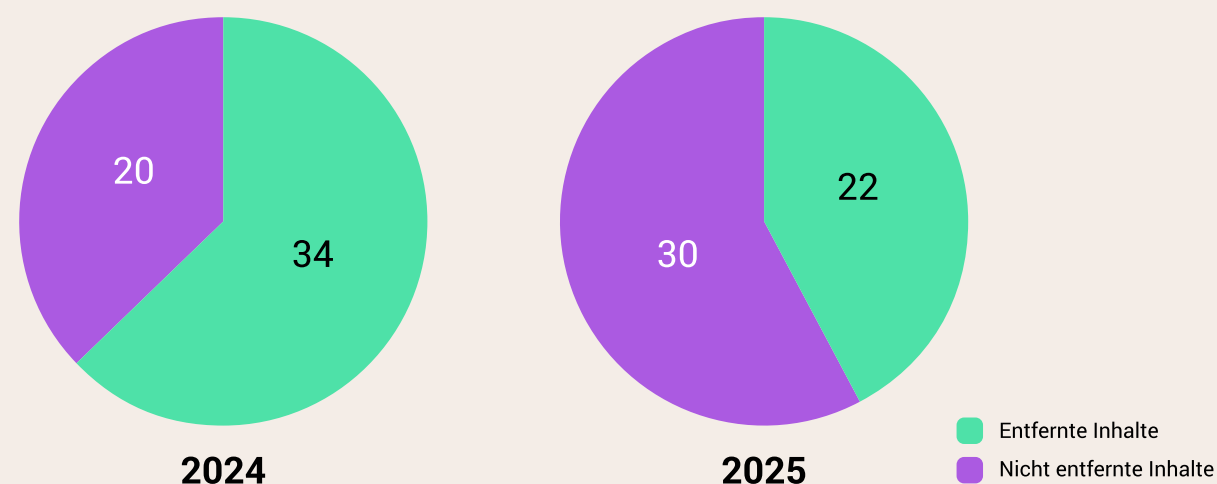
Auch bei **Instagram** war eine Verringerung der Entfernungsquote illegaler Inhalte erkennbar. Im Jahr 2024 entfernte die Plattform 52 % der gemeldeten illegalen Inhalte nach Meldung oder Beschwerde, während es im Jahr 2025 nur noch 40 % waren. Grund hierfür ist vor allem ein **Rückgang der Erfolgsquote von internen Beschwerden gegen abschlägige Entscheidungen**. Instagram entfernte im Jahr 2024 nur 29 % und im Jahr 2025 sogar 40 % der Inhalte auf eine einfache Meldung hin. Während jedoch im Jahr 2024 33 % der internen Beschwerden gegen abschlägige Entscheidungen Erfolg hatte, hatte 2025 keine einzige Beschwerde Erfolg.

³ Durch die außergerichtliche Streitbeilegung konnte HateAid bei Facebook noch keine Erkenntnisse erzielen. In den zu Facebook angestoßenen Verfahren ist derzeit noch keine Entscheidung einer Streitbeilegungsstelle eingegangen. HateAid hat Streitbeilegungsverfahren zu Facebook ab März 2025 eingelegt.

Diese Entwicklung fällt zeitlich mit einer im Januar 2025 vorgenommenen Änderung der internen Richtlinien von Meta zusammen, die im Februar 2025 vom CEO Mark Zuckerberg verkündet wurde. Diese sieht vor, dass bestimmte, politisch oder religiös motivierte Inhalte nicht mehr als Verstoß gegen die Plattform-AGB gelten sollen.⁴ Die Richtlinienänderung darf jedoch den Umgang mit illegalen Inhalten nicht verändern. Bei der Untersuchung zeichnete sich allerdings ab, dass illegale Inhalte, die 2024 nach einer Meldung nach dem DSA noch entfernt wurden, 2025 trotz vergleichbarer Sachverhalte online blieben.⁵

Unter 24 Stunden liegt die **Bearbeitungsdauer von Meldungen** auf Facebook bei 80 % der Meldungen und auf Instagram bei 63 % der Meldungen. Der Großteil der Entscheidungen kann daher als „unverzüglich“ im Sinne des Art. 16 Abs. 5 DSA bewertet werden.⁶ Ausreißer wurden vor allem bei Facebook deutlich: 2025 erging bei zwei Meldungen erst nach über zwei Monaten eine Entscheidung. Im Plattformvergleich fällt auf, dass die Meta-Plattformen wenige Verfahren in unter einer Stunde bearbeiten (Facebook: 14 %, Instagram: 5 %). Wir bewerten dies als Indiz dafür, dass beide Plattformen, wie in ihrem Transparenzbericht ausgewiesen, Meldungen manuell bewerten lassen.

Abbildung 1: Meta-Plattformen: Gegenüberstellung der entfernten Inhalte insgesamt nach Meldung oder Beschwerde



⁴ <https://transparency.meta.com/de-de/policies/community-standards/hateful-conduct/>.

⁵ Beispiel 1: 2024 auf Facebook nach einer Meldung entfernter Inhalt: „Mein größter Wunsch ist das jeder Grüne bald verreckt oder erschlagen wird“, 2025 auf Facebook nach Meldung und Beschwerde online gelassen: „Jetzt dreht das grüne Miststück völlig durch. Schmeißt den Kerl in die Geschlossene und lasst ihn dort verr.....“; Beispiel 2: 2024 auf Facebook nach Meldung entfernt: „Stasi Hure nein danke“, 2025 auf Facebook erst nach Beschwerde entfernt: „Dan Fick dich selber Schlampe“ (beides in Bezug auf Politikerinnen); Beispiel 3: 2024 auf Instagram nach Meldung entfernte Inhalte: „Abschaum...“, „Schmutz gehört in de Klappe“, 2025 auf Instagram nach Meldung und Beschwerde online gelassen: „Da fehlen noch ein paar des Drecks.Sagt und macht man täglich. Muelltrennung Müllentsorgung.Die sind Sondermüll müssen EXTRA entsorgt werden“.

⁶ Als Maßstab für den Zeitrahmen von 24 Stunden kann der Code of Conduct on Countering Illegal Hate Speech Online + herangezogen werden (siehe 2.3 des Codes of Conduct), welcher von allen im Bericht untersuchten Plattformen unterzeichnet wurde.

Die **Bearbeitungszeit von internen Beschwerden** hat sich im Projektzeitraum deutlich verkürzt: Noch 2024 benötigte Facebook bei 50 %, Instagram bei 32 % der eingereichten Beschwerden über sieben Tage für eine Bearbeitung. Eine so langsame Bearbeitung legt eine Verletzung von Art. 20 Abs. 4 DSA, der Verpflichtung zur zeitnahen Bearbeitung von Beschwerden, nahe. 2025 zeigten sich beide Plattformen deutlich schneller und bearbeiteten Beschwerden in der Regel (Facebook: 94 %, Instagram: 93 %) in unter drei Tagen. In zwei Fällen brauchten die Plattformen mehr als einen Monat für eine Entscheidung. Wie schon dargelegt, waren außerdem alle Beschwerden im Jahr 2025 erfolglos.

3.1.2 YouTube: Niedrigste Entfernungsquoten

Die von Google betriebene Plattform **YouTube weist die niedrigste Entfernungsquote illegaler Inhalte aller untersuchten Anbieter auf**. Insgesamt wurden im Untersuchungszeitraum lediglich **32 % der gemeldeten Inhalte entfernt**.⁷ Während 2024 noch 38 % der Inhalte entfernt wurden, sank der Wert 2025 auf 28 %. Auffällig ist, dass YouTube zu einem erheblichen Teil der Meldungen zwar eine Eingangsbestätigung abgab, aber **keine inhaltliche Entscheidung** mitteilte: 2024 betraf dies fast die Hälfte (48 %) und 2025 noch rund ein Drittel (34 %) der Fälle. Geprüft wurden dabei sowohl interne Benachrichtigungen im Profil als auch ergänzende Mitteilungen per E-Mail. Das Ausbleiben dieser Rückmeldungen wirft erhebliche Fragen zur Vereinbarkeit mit Artikel 16 Abs. 5 und 6 DSA auf, wonach Plattformen verpflichtet sind, Meldungen zu bearbeiten und die Entscheidung den Meldenden unverzüglich mitzuteilen.

In den Fällen, in denen eine Rückmeldung erfolgte, **bearbeitete YouTube Meldungen in der Regel in unter 24 Stunden** (2024: 92 %, 2025: 80 %). Allerdings versandte YouTube nach einer Meldung häufig eine E-Mail an die meldende Person mit der Anfrage nach weiteren Informationen. Oftmals waren die Nachfragen, zum Beispiel nach dem verletzten Gesetz, zuvor bereits in der Meldung beantwortet. Wird die meldende Person nicht sofort tätig, kann sich die Bearbeitungsdauer erheblich verzögern: Bei einem 2025 gemeldeten Inhalt reagierte YouTube erst 23 Tage nach Beantwortung einer solchen E-Mail-Nachfrage.

3.1.3 TikTok: Leichte Verbesserung bei weiterhin niedrigen Quoten

TikTok entfernte im Projektzeitraum insgesamt **38 % der gemeldeten illegalen Inhalte**, wobei sich diese Quote im Jahr 2025 im Vergleich zum Vorjahr erhöhte: 2024 wurden insgesamt 33 % der Inhalte entfernt, 2025 lag die Quote bei 46 %. Auf eine einfache Meldung hin wurden 2024 nur 26 % der Inhalte entfernt, interne Beschwerden waren in 19 % der Fälle erfolgreich. 2025 wurden dann zwar 46 % der Inhalte auf eine einfache Meldung hin entfernt, interne Beschwerden gegen abschlägige Entscheidungen hingegen waren niemals erfolgreich. Auch TikTok teilte häufig, nämlich in 21 % der Fälle, keine Entscheidung mit.

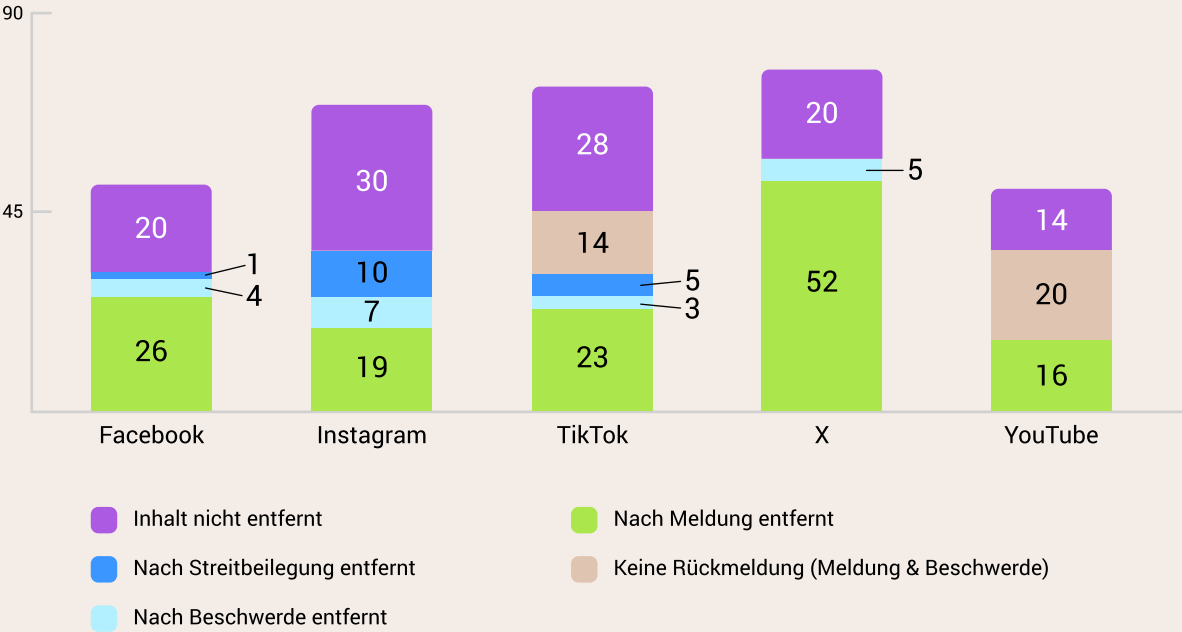
TikTok reagierte sowohl auf Meldungen als auch auf Beschwerden in der Regel in **unter 24 Stunden** (2024: ca. 87 % bei Meldungen und Beschwerden, 2025: 64 % der Meldungen, 100 % der Beschwerden).

⁷ Die Entfernungsquote ergibt sich ausschließlich aus den nach einer Meldung entfernten Inhalten. Nach einer Beschwerde oder im außergerichtlichen Streitbeilegungsverfahren wurden auf YouTube in unserer Analyse keine weiteren Inhalte entfernt, was sich daraus ergibt, dass diese Rechtsbehelfsmöglichkeiten kaum zugänglich sind, siehe 3.3.1.

3.1.4 X (ehemals Twitter):
Hohe Entfernungsquoten

X weist die höchste Entfernungsquote aller untersuchten Plattformen auf: In den Projektjahren wurden insgesamt 74 % der gemeldeten Inhalte nach Meldung und ggf. interner Beschwerde entfernt.⁸ Dieser Wert war gleichbleibend: Im gesamten Projektzeitraum wurden auf eine einfache Meldung hin rund 67 % der Inhalte entfernt. Eine interne Beschwerde führte in ca. 20 % der Fälle zum Erfolg (2024: 18 %, 2025: 21 %). X bearbeitet Meldungen und Beschwerden in aller Regel sehr schnell: 99 % der Meldungen und 100 % aller eingereichten Beschwerden wurden in unter 24 Stunden bearbeitet, 70 % der Meldungen und 88 % der Beschwerden sogar in unter einer Stunde.

Abbildung 2: Meldungen, Beschwerden und Streitbeilegung nach Plattform im Projektzeitraum April 2024 – Juli 2025



⁸ Durch die außergerichtliche Streitbeilegung konnte HateAid bei X noch keine Entfernungen von Inhalten erzielen. In den zu X ange-
stoßenen Verfahren ist derzeit noch keine Entscheidung einer Streitbeilegungsstelle eingegangen. HateAid hat Streitbeilegungsver-
fahren zu X ab Februar 2025 eingelegt.

3.2 Der Zugang zu Meldewegen
wird Nutzenden auf allen Plattformen
erschwert

Die Möglichkeit, illegale Inhalte niedrigschwellig zu melden, stellt eine zentrale Säule des Systems zum Umgang mit illegalen Inhalten im DSA dar. Gemäß Artikel 16 DSA sind alle Anbieter von Vermittlungs-
plattformen verpflichtet, leicht zugängliche und
nutzungsfreundliche Meldewege bereitzustellen. Unsere Analyse zeigt, dass keine der untersuchten
Plattformen diese gesetzlichen Vorgaben in an-
gemessener Weise erfüllt. Im Gegenteil: Nutzenden wird auf sämtlichen Plattformen die Meldung illegal-
er Inhalte durch übermäßige Komplexität, intrans-
parente Strukturierung, abschreckende Formulie-
rungen und gezielte Designelemente systematisch erschwert.

Dabei liegt ein zentrales Defizit in der **undurchsich-
tigen Trennung der Meldewege**: Sämtliche Platt-
formen unterscheiden streng zwischen Meldungen
wegen Verstößen gegen die Nutzungsbedingungen
(AGB) und solchen wegen möglicher Rechtsverstö-
ße nach Art. 16 DSA. Diese Differenzierung ist für
Nutzende jedoch nicht nachvollziehbar ausgestal-
tet. Beide Wege werden weder optisch noch inhalt-
lich unterscheidbar präsentiert, es gibt keinerlei
Hilfestellung für die Nutzenden. Die Option „Als
rechtswidrig melden“ findet sich häufig erst am
Ende einer langen Liste an Auswahlmöglichkeiten.
Denn vorangestellt sind ihre Kategorien wie „Mob-
bing oder Belästigung“ oder „Gewalt, Hass oder
Ausbeutung“. Hierbei handelt es sich um juristisch
anmutende Begriffe, die Nutzende dazu verleiten,
hierunter auch rechtliche Verstöße zu melden. Hin-
ter diesen Kategorien verstecken sich allerdings die
Meldewege ausschließlich für AGB-Verletzungen.
Für die Plattformen ist dies vorteilhaft, da Meldun-
gen auf Grundlage der AGB, die nicht explizit als
Meldung illegaler Inhalte nach dem DSA gekenn-
zeichnet sind, unter anderem nicht den strengen
Pflichten des Art. 16 unterliegen.

Das Potenzial, dass Nutzende so oft unwissentlich
auf ihre Rechte verzichten, ist groß: Darunter zählt
das Recht auf eine Empfangsbestätigung (Abs.
4), eine unverzügliche Entscheidung mit Rechts-
behelfsbelehrung (Abs. 5) sowie das Recht auf eine
zeitnahe und willkürfreie Bearbeitung und die Kenn-
zeichnung automatisierter Entscheidungen (Abs. 6).

Hinzu kommt, dass die Meldewege für illegale
Inhalte nach dem DSA in der Regel deutlich **kom-
plizierter, aufwendiger und verschachtelter** sind
als jene für AGB-Verstöße. Nutzende müssen sich
insbesondere auf mobilen Geräten häufig durch
eine Vielzahl unübersichtlicher Kategorien klicken,
um den Meldepfad zu beenden. Der eigentliche
Meldeprozess erfordert zudem häufig komplexe
Angaben, die tiefgehendes juristisches Wissen
voraussetzen. Dies bestätigt auch eine Evaluation
der Meldewege durch Befragung von Nutzenden
durch die Organisation das NETTZ aus dem Jahr
2025: Die Meldung illegaler Inhalte nach dem DSA
wurde von den Testpersonen als deutlich schwerer
und komplizierter angesehen als eine Meldung
nach AGB. So müssen Nutzende bei X etwa ein
„Territorium“ auswählen, dessen Gesetze der Inhalt
verletzt, wobei sowohl einzelne EU-Mitgliedsstaa-
ten als auch die Option „In der EU“ zur Auswahl
stehen. Facebook und Instagram sowie YouTube
verlangen zusätzlich die Angabe des vollständigen
Namens der meldenden Person, obwohl dies nicht
zwingende Voraussetzung für eine Meldung nach
Art. 16 DSA ist. Plattformen müssen Nutzenden
allenfalls die Möglichkeit einräumen.⁹

Diese strukturellen Hürden können zu sogenannter
„**Click Fatigue**“ führen:¹⁰ einem psychologischen
Ermüdungseffekt, bei dem Nutzende den Melde-
prozess entnervt abbrechen oder unbewusst die
falsche Kategorie wählen. Auch die Untersuchung
von das NETTZ¹¹ bestätigt, dass die Meldewege
Nutzende überfordern können: Mehr als ein Viertel
(27 %) der Meldungen über den Meldeweg für ille-
gale Inhalte nach dem DSA wurde abgebrochen.

⁹ Dazu siehe: edpb_guidelines_202503_interplay-dsa-gdpr_v1_en.pdf, Rn. 25 ff.

¹⁰ Vgl. 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-meldevorgaenge-von-plattformen-nach>, S. 34.

Als Gründe für den Abbruch wurde häufig angegeben, dass sich meldende Personen „mit der Auswahl überfordert“ fühlten oder nicht wussten, „wie ich weitermachen kann“. Bei YouTube etwa müssen Personen, die die Kategorie „Rechtliches Problem“ wählen, mehrfach bestätigen, dass sie den Inhalt tatsächlich wegen einer Rechtsverletzung melden möchten, während bei AGB-Meldungen eine solche doppelte Bestätigung nicht erforderlich ist. Auch das erneute Abfragen von Angaben per E-Mail, obwohl diese bereits Bestandteil der Meldung waren, schafft zusätzlich Hürden und ermüdet Nutzende. Ein anderes Beispiel ist die Länge der Klickwege bei Facebook: Möchten Nutzende beispielsweise eine illegale Drohung melden, sind für eine Meldung als AGB-Verletzung fünf Klicks erforderlich, für eine Meldung als Rechtsverletzung dagegen mindestens 15 Klicks. Diese Form der gezielten Komplexitätssteigerung kann als sogenanntes Dark Pattern bewertet werden, was nach Art. 25 DSA

eine verbotene Praktik darstellt. Ein Dark Pattern ist die bewusst manipulative Gestaltung digitaler Prozesse, die Nutzende davon abhält, ihre gesetzlich vorgesehenen Rechte wahrzunehmen. Das wiederholte Auffordern von Nutzenden, eine Entscheidung zu treffen, die sie bereits getroffen haben, benennt Artikel 25 Absatz 3 sogar exemplarisch als Beispiel eines Dark Pattern.

Erschwerend kommen technische Einschränkungen und abschreckende Formulierungen hinzu: Viele Formulare enthalten explizite Warnungen über mögliche Sanktionen bei falschen Angaben. Zum Teil wird Nutzenden geraten, eine Meldung „am besten mit anwaltlicher Unterstützung“ vorzunehmen. Auch technische Einschränkungen wirken behindernd: So begrenzt etwa TikTok die Zeichenzahl bei der Begründung auf 400 Zeichen, was eine nachvollziehbare Beschreibung des Verstoßes erschwert.

Abbildung 3: Ausschnitt des Meldeformulars von TikTok, bei welchem verschiedene Faktoren das Einreichen einer Meldung erschweren oder davon abschrecken können: Die Zeichenzahlbegrenzung bei der Meldeerklärung, die Verpflichtung der Angabe des vollen Namens der meldenden Person sowie die Behauptung, Meldungen könnten strafbar sein.

Abbildung 4: Ausschnitt des Meldeformulars bei YouTube. Auch hier muss ein vollständiger Name angegeben werden, außerdem müssen meldende Personen bestätigen, dass sie „in Kenntnis der Strafbarkeit einer falschen Angabe“ handeln. Auch die Formulierung „Ich versichere [...], dass ich zur Meldung dieser mutmaßlichen Rechtsverletzung autorisiert bin“, ist stark irreführend: Nach Art. 16 Abs. 1 DSA hat jede Person das Recht, einer Plattform Inhalte zu melden, die sie für illegal hält. Eine bestimmte Autorisierung ist nicht notwendig. Woraus sich eine pauschale strafrechtliche Haftung für falsche Angaben ergeben würde, ist nicht ersichtlich.

3.3 Defizite im Rechtsbehelfssystem: interne Beschwerdeverfahren und außergerichtliche Streitbeilegung

Der DSA sieht ein **gestuftes System von Rechtsbehelfen** vor: Nach einer Meldung sollen Nutzende zunächst interne Beschwerdeverfahren der Plattformen nach Artikel 20 DSA nutzen können. Außerdem können sie ein außergerichtliches Streitbeilegungsverfahren nach Artikel 21 DSA einleiten. Dieses Zusammenspiel soll einen niedrigschwelligen, wirksamen und kostengünstigen Rechtsschutz gewährleisten. Die Analyse zeigt jedoch, dass dieses Versprechen bislang nur eingeschränkt eingelöst wird. Auf beiden Ebenen treten erhebliche Defizite zutage: Plattformen erschweren den Zugang zu internen Beschwerdeverfahren oder machen ihn faktisch unmöglich, während nicht alle Streitbeilegungsstellen Fristen, Kommunikation und Zuständigkeiten einhalten.

3.3.1 Plattformen erschweren Zugang zu Rechtsbehelfsmöglichkeiten

Die Untersuchung ergab, dass Plattformen es vielfach versäumen, verständlich auf Rechtsbehelfsmöglichkeiten hinzuweisen oder diese schwer aufzufinden sind. Damit verstoßen Plattformen gegen das Recht auf einen wirksamen Rechtsbehelf nach Art. 16 Abs. 5 und ein leicht zugängliches und wirksames Beschwerdeverfahren aus Art. 20 Abs. 1, Abs. 3 DSA.

Ein zentrales Problem für Nutzende liegt in **unvollständigen und unübersichtlichen Rückmeldungen auf Meldungen**. Nach Art. 16 Abs. 5 DSA sind Plattformen verpflichtet, den Meldenden eine Entscheidung „in Bezug auf“ die gemeldete Information mitzuteilen. Eine solche Bezugnahme ist erforderlich, um die Rückmeldung einer konkreten Meldung zuordnen zu können. Fehlt diese, wird die effektive Rechtsdurchsetzung erheblich behindert. Bei **TikTok** etwa erfolgen Empfangsbestätigungen ausschließlich über Benachrichtigungen in der App, nicht per E-Mail, und enthalten keine Referenznummer. Bei mehreren parallel eingereichten Meldungen ist die Zuordnung einer solchen Rückmeldung

zu der entsprechenden Meldung praktisch unmöglich. Erst wenn eine Beschwerde über den Button „Weitere Überprüfung anfordern“ eingereicht wird, blendet TikTok den ursprünglich gemeldeten Inhalt ein und ermöglicht so eine Zuordnung.

Ähnlich problematisch ist die Praxis bei **X**: Zwar versendet die Plattform unmittelbar nach einer Meldung eine Empfangsbestätigung, diese ist jedoch rein generisch und enthält weder eine Referenznummer noch Angaben zur Meldung. Nutzende, die mehrere Meldungen einreichen, können die Vorgänge nicht auseinanderhalten. Zwar enthält die Mitteilung der Moderationsentscheidung durch X einen Link zum gemeldeten Inhalt und eine Referenznummer. Wie bei TikTok stehen Nutzende aber vor einem Problem, wenn eine Entscheidung ausbleibt: Sie können die rein generische Empfangsbestätigung ohne Referenz nicht verwenden, um gegenüber Gerichten oder Aufsichtsbehörden eine Meldung nachzuweisen. Diese ist jedoch erforderlich, um gemäß Art. 6 Abs. 1 a) DSA eine Haftung zu begründen. Im deutlichen Gegensatz dazu stehen **Facebook und Instagram**: Die Plattformen versenden Empfangsbestätigungen und Entscheidungsrückmeldungen per E-Mail an die im Meldeprozess angegebene Adresse, versehen jede Meldung mit einer individuellen Referenznummer und legen eine Kopie aller Angaben bei.

Neben unzureichenden Rückmeldungen **fehlt bei einigen Plattformen auch ein internes Beschwerdeverfahren vollständig** oder es wird verschleiert. Art. 20 Abs. 1 DSA verpflichtet Plattformen dazu, für alle Arten von Entscheidungen nach Meldungen ein internes Beschwerdeverfahren anzubieten. Besonders deutlich zeigt sich das Defizit erneut bei **YouTube**: Während für gemeldete Videos ein internes Beschwerdeverfahren zur Verfügung steht, fehlt es bei gemeldeten Kommentaren vollständig. Dementsprechend erfolgt auch keine Belehrung über die Möglichkeit einer internen Beschwerde. Dies stellt zugleich einen Verstoß gegen Art. 16 Abs. 5 (Rechtsbehelfshinweis) und Art. 20 Abs. 1 DSA dar. Bei X zeigt sich ein ähnliches Problem **bei der Meldung von AGB-Verstößen**: X ermöglicht es entgegen Art. 20 Abs. 1 DSA nicht, eine interne Beschwerde gegen abschlägige Entscheidungen einzureichen, obwohl Erwägungsgrund 58 klarstellt, dass dies möglich sein soll.

Ein weiteres Hindernis besteht in **fehlenden oder eingeschränkten Möglichkeiten zur Beweissicherung**. Für ein effektives Rechtsbehelfsverfahren ist es entscheidend, dass Nutzende den gemeldeten Inhalt und die erfolgte Rückmeldung rechtssicher dokumentieren können. Gerade **TikTok** bietet hierfür weder in der Desktopansicht noch in der mobilen Version ausreichenden Funktionen: Datum und Uhrzeit von Kommentaren werden nicht exakt angezeigt. Stattdessen werden lediglich relative Zeitangaben wie ‚vor zwei Wochen‘ angezeigt, statt eines konkreten Datums mit exakter Uhrzeit. Auch bei **YouTube und Instagram** sind in beiden Versionen bloß relative Zeitangaben sichtbar, absolute Zeitangaben können nur über Umwege erfasst werden. Dies erschwert vor allem gerichtliche Verfahren, für die eine lückenlose Dokumentation regelmäßig erforderlich ist. Anders verhält es sich in der Desktopversion bei **Facebook** und beiden Versionen der Plattform **X**: Hier sind Datum und Uhrzeit so verfügbar, dass rechtssichere Screenshots erstellt und Vorgänge im Zweifel belegt werden können.

3.3.2 Defizite bei außergerichtlichen Streitbeilegungsstellen

Die außergerichtliche Streitbeilegung nach Artikel 21 DSA ist als zentrales Instrument konzipiert, um Nutzenden eine schnelle, kostengünstige und fachlich fundierte Alternative zum Gerichtsverfahren zu bieten. Zwischen Oktober 2024 und Juli 2025 hat HateAid insgesamt 50 Verfahren bei vier von der EU zertifizierten Streitbeilegungsstellen eingeleitet: **User Rights GmbH** (Deutschland), **Appeals Centre Europe** (Irland), **ADROIT** (Malta) und **RTR** (Österreich).¹² Ziel war es, die praktische Wirksamkeit der in Artikel 21 DSA verankerten Rechte zu erproben, insbesondere Bearbeitungsdauer, Qualität der Ent-

scheidungen, Kommunikation mit den Verfahrensbeteiligten und Umsetzung durch die Plattformen.

Die **Zuständigkeiten der Streitbeilegungsstellen** sind dabei unterschiedlich geregelt und jeweils auf der Webseite der EU-Kommission ausgewiesen: Die **User Rights GmbH** bearbeitet Verfahren zu illegalen Inhalten auf LinkedIn, Instagram, TikTok und seit Juli 2025 auch Facebook. Das **Appeals Centre Europe** ist dagegen auf Verfahren zu mutmaßlichen Verstößen gegen die Plattform-AGB beschränkt und zuständig für Facebook, YouTube und TikTok; Fälle zu illegalen Inhalten werden dort nicht geprüft. **ADROIT** in Malta ist plattformoffen und nimmt Verfahren zu sämtlichen Online-Plattformen an, darunter auch X, Snapchat, Skype und Telegram. Gleiches gilt für die österreichische **RTR**, die keine Einschränkungen hinsichtlich der abgedeckten Plattformen vorsieht.

Die Ergebnisse:

In der Hälfte der eingereichten Fälle (50 %) hat die jeweils angerufene Streitbeilegungsstelle eine Entscheidung gefällt. Eine dieser Entscheidungen weist einen Antrag ab,¹³ vier Entscheidungen (16 %) stellen fest, dass der Inhalt bereits während des laufenden Verfahrens entfernt wurde und sich das Anliegen damit erledigt habe und die übrigen 80 % bestätigen die Illegalität des Inhaltes und fordern die Plattform dazu auf, den entsprechenden Inhalt zu entfernen. Diese Entscheidung führte bisher in ca. zwei Dritteln (65 %) der Fälle zur Entfernung des Inhalts durch die Plattform.¹⁴ Dies ist unerwartet, denn die Entfernung ist freiwillig und die Wirksamkeit der Streitbeilegung daher bisher unklar.

In der Praxis zeigen sich allerdings **deutliche Unterschiede in Qualität und Effektivität der Verfahren**: Während einzelne Stellen, allen voran die User Rights GmbH, belegen, dass Artikel 21 DSA

wirksam umgesetzt werden kann, verhindern **lange Bearbeitungszeiten, eingeschränkte Prüfungsmaßstäbe** und **fehlende Kommunikation** bei anderen Stellen die Rechtsdurchsetzung.

Der rechtliche Rahmen sieht nämlich vor, dass **Entscheidungen grundsätzlich binnen 90 Tagen**, in Ausnahmefällen binnen 180 Tagen zu ergehen habe. In 50 % der Fälle (25 von 50) wurde seit mindestens 90 Tagen nach Verfahrensbeginn keine Entscheidung zugestellt, in 7 Fällen laufen Verfahren sogar seit über 180 Tagen ohne Mitteilung einer Entscheidung. Allein bei ADROIT betrifft dies alle eingereichten Verfahren (11) und auch bei der österreichischen Streitbeilegungsstelle RTR blieben die Verfahren weitgehend unentschieden (13 von 14 Verfahren seit mindestens 90 Tagen nicht entschieden). Dies lässt potenzielle Verstöße gegen Artikel 21 Abs. 4 DSA erkennen. Besonders gravierend ist, dass in vielen Fällen über Monate hinweg keinerlei Rückmeldung erfolgte. Auch bei einem Fünftel der entschiedenen Fälle (20 %) wurde eine Entscheidung erst nach mehr als 90 Tagen seit Einreichung des Verfahrens zugestellt. Die Entfernung der Inhalte durch die Plattformen wurde wiederholt nicht kommuniziert und muss manuell überprüft werden, in mindestens zwei Fällen benötigte die Entfernung durch die Plattformen über einen Monat nach Zustellung der Entscheidung. Der Ausgang des Verfahrens ist daher für Nutzende oft nicht nachvollziehbar oder sehr zeitaufwändig.

Die **User Rights GmbH** bestätigte in allen abgeschlossenen Fällen die **Illegalität der Inhalte**, begründete die **Entscheidungen ausführlich transparent** und erreichte deren Umsetzung durch die Plattformen in der Regel innerhalb von drei Wochen nach Entscheidung. Insgesamt wurden 73 % der Inhalte, bezüglich derer ein Verfahren bei User Rights durchlaufen wurde, entfernt. Gut ein Viertel (27 %) der Inhalte blieb online, obwohl die Streitbeilegungsstelle die Plattform zur Entfernung aufforderte. Bei **ADROIT und RTR zeigen sich erhebliche**

Defizite: Teilweise informierten die Streitbeilegungsstellen über Verfahrensupdates, wie die Zustellung des Antrags an die Plattform. Doch Entscheidungen blieben in aller Regel trotz Nachfragen aus, bei ADROIT in mehreren Fällen seit über zehn Monaten. RTR traf nur in einem von vierzehn eingereichten Fällen eine Entscheidung, ADROIT fällte über keines der elf eingereichten Verfahren eine Entscheidung. Auch beim **Appeals Centre Europe** ist der Nutzen eingeschränkt, da diese Stelle auch bei der Meldung illegaler Inhalte nur die Vereinbarkeit mit den Plattform-AGB überprüft. Zwar konnte in einem Fall eine Entfernung erreicht werden, die Umsetzung durch TikTok erfolgte jedoch erst nach mehr als zweieinhalb Monaten.

Besonders problematisch ist, dass bei den Plattformen X und YouTube bis heute unklar bleibt, bei welcher Streitbeilegungsstelle überhaupt effektiv Verfahren zu illegalen Inhalten geführt werden können. Diese Unklarheit spiegelt sich auch in den Transparenzberichten wider: X gibt an, dass im letzten Berichtszeitraum keine Verfahren bei Streitbeilegungsstellen eingegangen seien,¹⁵ obwohl im Rahmen dieser Untersuchung während dieses Zeitraums **zwei Verfahren gegen X eingelegt wurden**, davon **eines bei ADROIT**, das **andere bei RTR**. Während ADROIT zurückmeldete, das Verfahren noch im Berichtszeitraum an X weitergeleitet zu haben, erfolgte bei RTR eine solche Weiterleitung erst über 5 Monate später und damit weit über der grundsätzlich 90-tägigen Frist. Google gibt bezüglich **YouTube** an, dass zwar Verfahren innerhalb des Berichtszeitraums eingegangen seien, aber keine Entscheidung getroffen wurde.¹⁶ Auch hier ist also die Effektivität der Verfahren noch unklar. Im Gegensatz dazu tätigen **Instagram und TikTok** konkrete Angaben zu anhängigen und bereits abgeschlossenen Streitbeilegungsverfahren. In der diesem Bericht zugrundeliegenden Untersuchung wurden 2024 über die User Rights GmbH Verfahren geführt, bei denen die beiden Plattformen die Entscheidung der Streitbeilegungsstelle jeweils freiwillig umgesetzt haben.¹⁷

¹² Die genannten Stellen wurden ausgewählt, weil sie bereits im Oktober 2024 zertifiziert waren und entweder Deutsch oder Englisch als Verfahrenssprachen zugelassen sind. Allerdings legte HateAid beim Appeals Centre Europe nur zwei Fälle ein und verzichtete auf weitere Einlegungen, da der Prüfungsmaßstab dort auf Verletzungen von Plattform-AGB begrenzt ist und keine Illegalität von Inhalten überprüft wird. Daher konzentrieren sich die Erkenntnisse auf die Stellen User Rights GmbH, ADROIT und RTR.

¹³ Diese abweisende Entscheidung vom Appeals Centre Europe wurde nicht weiter begründet. Naheliegender ist jedoch, dass der Antrag abgewiesen wurde, weil das Appeals Centre Europe nur Verfahren zu Plattform-AGB annimmt und unser Antrag sich auf die Illegalität eines Inhalts bezog.

¹⁴ Es kann nicht in jedem Fall zweifelsfrei festgestellt werden, dass der Inhalt gerade aufgrund des Streitbeilegungsverfahrens entfernt wurde. In einigen Fällen fehlt eine Bestätigung der Plattform, dass sie den Inhalt wegen der Entscheidung in der Streitbeilegung entfernt hätten, dieser ist jedoch nicht mehr online auffindbar.

¹⁵ X DSA Transparency Report, Oktober 2024 – März 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 Auch Facebook gibt im letzten Transparenzbericht an, dass die Plattform im Berichtszeitraum keine Verfahren von außergerichtlichen Streitbeilegungsstellen erhalten hätte, da nach dem Berichtszeitraum die User Rights GmbH ihren Prüfungsumfang auf Facebook erweiterte, dürfte sich dies jedoch in Zukunft verändern.

3.4 Moderationssysteme

Ein funktionierendes Moderationssystem ist zentral für die Eindämmung der Verbreitung illegaler Inhalte. Die aktuelle Regulierung fußt auf Haftungsprivilegien: Plattformen haften nicht für illegale Inhalte, sondern erst, wenn sie nach einem Hinweis nicht tätig werden, Art. 6 Abs. 1 DSA. Der Gesetzgeber hat sich entschieden, sich für den Umgang mit illegalen Inhalten auf Meldungen durch Nutzende zu verlassen. Umso entscheidender ist, dass Melde- und Moderationsprozesse verlässlich funktionieren und zugänglich sind.

Der DSA verpflichtet in Art. 16 Abs. 6 Plattformen dazu, anzugeben, **wenn eine Entscheidung mit Hilfe automatisierter Mittel** getroffen wurde. Bei der Bearbeitung von Beschwerden ist dabei gemäß Art. 20 Abs. 6 DSA eine rein automatisierte Entscheidungsfindung sogar gänzlich unzulässig.

Die Analyse weckt Zweifel an der Umsetzung dieser Regeln: Bei TikTok zeigen sich wiederkehrende und auffällig symmetrische Bearbeitungszeiten, die stark auf den Einsatz automatisierter Prozesse hindeuten, ohne dass die Plattform die Nutzenden, wie nach Artikel 16 Abs. 6 DSA vorgeschrieben, darüber informiert. So wurden im Jahr 2024 62 % der bearbeiteten Meldungen und 76 % der eingereichten Beschwerdeverfahren in auf die Minute exakt 30 Minuten bearbeitet. 2025 trat dieses Muster zwar seltener auf, war aber immer noch in

36 % aller Meldungen nachweisbar. Selbst wenn die Bearbeitungsdauer variierte, blieb das Muster erkennbar: Viele Entscheidungen erfolgten exakt nach Zeitabständen von 30 Minuten, etwa nach 1 Stunde 30 Minuten oder 1 Tag 7 Stunden 30 Minuten. Dennoch erfolgte **in keinem Fall ein Hinweis auf den Einsatz automatisierter Systeme**, wie ihn Art. 16 Abs. 6 DSA vorschreibt. Im zuletzt veröffentlichten Transparenzbericht¹⁸ gibt TikTok an, dass Meldungen illegaler Inhalte zunächst automatisiert auf mögliche Verstöße gegen die Community-Richtlinien (AGB) geprüft werden. Erst wenn dabei kein Verstoß festgestellt wird, soll eine menschliche Moderation prüfen, ob eine Rechtsverletzung vorliegt. Dies kann jedoch nicht als Rechtfertigung gelten.

Auch **X** entschied über Meldungen vermehrt auffällig schnell, was ebenfalls die Vermutung nahelegt, dass **automatisierte Mittel eingesetzt werden**: Insgesamt bearbeitete X 30 % der eingereichten Meldungen in weniger als 10 Minuten, 9 % in unter 5 Minuten und in beiden Projektjahren jeweils eine Meldung innerhalb von einer Minute. Über eine potenzielle Nutzung automatisierter Mittel klärt X an keiner Stelle auf. Auch im Transparenzbericht gibt die Plattform an, dass sämtliche Meldungen und Beschwerden durch menschliche Draufsicht überprüft wurden.¹⁹ Inwiefern dieser Prozess innerhalb weniger Minuten möglich ist, bleibt unklar.

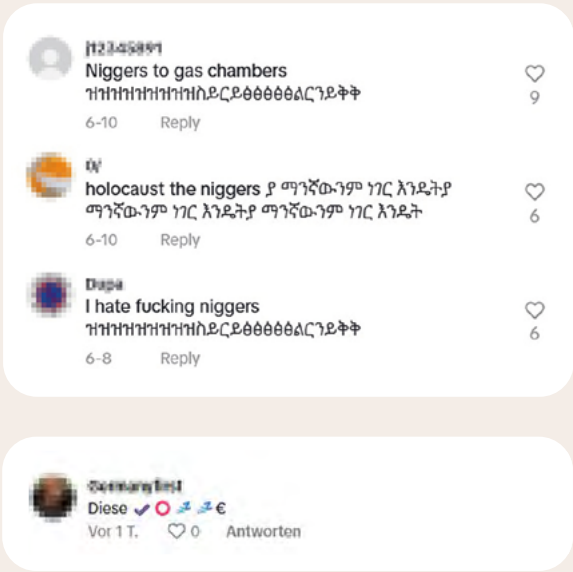
Ein besonders gravierendes Defizit zeigt sich auch bei der **Moderation von Inhalten auf TikTok in sogenanntem „Algospeak“**. Dabei handelt es sich um eine bewusste sprachliche Verschleierung, etwa durch die Nutzung nicht-römischer Schriftzeichen oder die Verwendung von Emojis, um Plattformfilter zu umgehen. Zwar sind die Inhalte für menschliche Lesende klar verständlich und häufig eindeutig illegal, doch die Untersuchung ergab, dass TikTok auf Meldungen solcher Inhalte oft überhaupt nicht reagierte. 2024 wurde in 71 % aller Fälle zu eindeutig illegalen Inhalten in Algospeak über mindestens drei Monate hinweg keine Entscheidung getroffen. Auch in den wenigen Fällen, in denen TikTok solche Meldungen entschied, kam es wiederholt zu nicht nachvollziehbaren Entscheidungen: Selbst offensichtliche Formalbeleidigungen oder volksverhetzende Inhalte wurden nicht als illegal eingestuft.

Die Folgen reichen weit über einzelne Fälle hinaus: Wenn illegale Inhalte in Algospeak von der Moderation systematisch übersehen werden, entsteht ein „Schutzraum“ für Tatpersonen. Sie können illegale Inhalte, wie Volksverhetzung, antisemitische Beleidigungen oder sexualisierte Gewaltandrohungen, ungestört verbreiten. Algospeak wird häufig gezielt in koordinierten Hasskampagnen eingesetzt, wenn gleiche oder ähnliche codierte Kommentare massenhaft in Kommentarspalten auftauchen.



Triggerwarnung: Darstellungen mit diskriminierender, rassistischer Sprache.

Abbildung 5:²⁰ Beispiele für illegale „Algospeak“-Kommentare auf TikTok.



¹⁸ TikTok DSA Transparency Report, Juli – Dezember 2024, S. 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, Oktober 2024 – März 2025, S. 7, <https://transparency.x.com/assets/dsa/transparency-report/dsa-transparency-report-april-2025.pdf>.

²⁰ Die Kommentare wurden auf der Plattform TikTok veröffentlicht und aus Datenschutzgründen anonymisiert. Sie sind mittlerweile nicht mehr online.

4 Sonstige Erkenntnisse zur DSA-Implementierung

Nach Artikel 53 können Nutzende und explizit auch Organisationen, Vereinigungen und Einrichtungen Beschwerden über Plattformen wegen Verstößen gegen den DSA bei den nationalen Digital Services Coordinators (DSC) einreichen. Dies umfasst das Recht, angehört und über den Sachstand der Beschwerde informiert zu werden. Hält der nationale DSC sich für unzuständig, leitet er die Beschwerde anschließend an den DSC des Sitzlandes der betreffenden Plattform weiter und versieht sie gegebenenfalls mit einer eigenen Stellungnahme zum Sachverhalt. Die Erfahrungen mit der Einreichung von Beschwerden bei der Bundesnetzagentur (BNetzA) als Digital Services Coordinator (DSC) zeigen allerdings: Die Zivilgesellschaft wird aktuell nicht ausreichend befähigt, ihren Rollen als Beschwerdeführende im Aufsichtsregime nachzukommen.

HateAid hat im Projektzeitraum insgesamt sieben Beschwerden größtenteils beim deutschen DSC, der Bundesnetzagentur, eingereicht. Die Ergebnisse: Beschwerden wiesen insgesamt sehr **lange Bearbeitungszeiten** von mehreren Monaten auf. Zudem war nicht nachvollziehbar, in welchem Stadium sich eine Beschwerde befand und ob sie möglicherweise kausal für die Einleitung eines Aufsichtsverfahrens war. In mehreren Fällen erhielt HateAid erst nach sechs bis sieben Monaten die Information, dass die Beschwerde an den zuständigen irischen DSC weitergeleitet wurde. Wiederholt teilte uns die deutsche DSC Monate nach Einreichung einer Beschwerde mit, man gehe davon aus, dass sich das Verfahren inzwischen erledigt habe, und bot nur bei ausdrücklicher Bestätigung eine Fortsetzung der Bearbeitung an. Dies ist möglicherweise auch Ausdruck der begrenzten personellen Ausstattung des deutschen DSC, der auch nach einer bereits erfolgten Aufstockung der Personalstellen bereichsübergreifend noch immer mit nur 24 Mitarbeitenden auskommen muss. Erstmals zeigte sich im Sommer 2025 ein positiveres Signal in Bezug auf Transparenz: Mitte August informierte der irische DSC HateAid, dass die im Juni beim deutschen DSC eingereichte und weitergeleitete Beschwerde gegen Meta in ein laufendes Verfahren der Europäischen Kommission nach Artikel 16 DSA eingeflossen sei.

Noch deutlicher traten **Herausforderungen im Bewerbungsverfahren als Trusted Flagger** zutage. Im Juni 2025 wurde HateAid gemäß Art. 22 DSA als Trusted Flagger zertifiziert. Der Prozess erwies sich mit über acht Monaten als langwierig und war von erheblichen Unsicherheiten bezüglich der inhaltlichen Anforderungen geprägt. Herausfordernd waren vor allem der Umfang sowie die mitunter fehlende Transparenz des Verfahrens: Es blieb über weite Strecken unklar, welche der abgefragten Informationen zwingende Voraussetzung für eine Zulassung waren und welche lediglich ergänzend abgefragt wurden. HateAid musste umfangreiche Unterlagen zur Struktur, Arbeitsweise und zu internen Dokumentationen einreichen, ohne zu wissen, welche Angaben tatsächlich entscheidungsrelevant waren. Zu den angeforderten Nachweisen gehörten unter anderem Zeugnisse von Mitarbeitenden. Obwohl es durchaus berechtigt und sogar wünschenswert ist, hohe Qualitätsanforderungen an Trusted Flagger zu stellen, gingen diese teilweise über den offiziellen Leitfaden hinaus.²¹ Hierdurch wurden zum Teil unnötig viele Ressourcen aufgewendet und das Verfahren verlängert. Zugleich ist anzuerkennen, dass das Zulassungsverfahren zum Zeitpunkt der Bewerbung noch weitgehend unerprobt war.

Ein weiterer Engpass ergibt sich aus der unerwartet **restriktiven Auslegung der Fachkenntnisanforderung**: Artikel 22 Abs. 2 fordert „besondere Sachkenntnis und Kompetenz in Bezug auf die Erkennung, Feststellung und Meldung rechtswidriger Inhalte“. Die BNetzA verlangte hingegen, dass jede Meldung entweder direkt von einer Person mit mindestens einem juristischen Staatsexamen durchgeführt oder durch eine solche freigegeben werde. Es ist unklar, ob dieser Maßstab europaweit derart hoch angelegt wird. Diese enge Auslegung schränkt nicht nur den Kreis potenzieller Trusted Flagger ein, sondern erschwert auch den Einsatz von auf bestimmte Themengebiete spezialisierten Organisationen, zum Beispiel Antisemitismus- oder Stalking-Beratungsstellen. Besonders problematisch ist die Diskrepanz im Vergleich zu Plattformen: Dort entscheiden Moderationsteams ohne jegliche juristische Ausbildung über die Entfernung von Inhalten mit zum Teil weltweiter Wirkung. An Trusted Flagger, die Inhalte nicht selbst entfernen, sondern nur melden können, werden somit höhere Anforderungen gesetzt als an die Plattformen, die Löschungen vornehmen. Hinzu kam im Bewerbungszeitraum eine spürbare öffentliche und mediale Sensibilität, die teilweise mit Sorge vor Angriffen verbunden war.

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

5 Politische Bewertung & Forderungen: Was jetzt zu tun ist

Damit der DSA seine Wirkung auch in der Praxis entfalten kann, müssen Aufsichtsbehörden und Politik entschlossen handeln. Die bisherigen Erfahrungen mit der Umsetzung des DSA zeigen deutlich: Ohne eine **entschlossene Aufsicht** und **strukturelle Nachbesserungen** droht das Gesetz in zentralen Bereichen hinter seinem eigenen

Anspruch zurückzufallen. Zwar hat der DSA wichtige Rechte für Nutzende geschaffen, die Praxis offenbart jedoch gravierende Lücken bei der Transparenz, der Durchsetzung und Zugang zum Recht. Für Nutzende und insbesondere Betroffene digitaler Gewalt ist es oft vom Zufall abhängig, ob sie Unterstützung erfahren.

5.1 Entschlossene Aufsicht und Durchsetzung

Bisher haben **trotz dutzender laufender Aufsichtsverfahren weder die nationalen DSCs noch die EU-Kommission eine Anordnung oder Sanktion gegen einen Anbieter verhängt**. Die konsequente Ahndung von Verstößen ist jedoch Voraussetzung für die Wirksamkeit des Gesetzes. Verzögerungen oder politische Erwägungen dürfen den Vollzug nicht beeinflussen. Zuletzt war die wohl bereits seit Januar 2025 vorbereitete Verhängung eines Bußgeldes gegen X zurückgestellt worden, um die

Verhandlungen im Zollstreit mit den USA abzuwarten.²² Es ist dringend davor zu warnen, die Durchsetzung des DSA von derartigen sachfremden Erwägungen abhängig zu machen. Die digitale Souveränität gebietet es, die demokratischen Werte der EU auch im digitalen Raum durchzusetzen. Zwar ist die Verhängung von Bußgeldern nicht das Ziel der Regulierung, dennoch kann dieser nur durch die Ahndung von Verstößen zu Geltung verholfen werden.

5.2 Klarstellungen und Nachbesserungen im Gesetzestext

Die anstehende Evaluation des DSA muss bestehende Unklarheiten bei der Auslegung des Gesetzes und Schlupflöcher adressieren: Plattformen nutzen gezielt Gestaltungsspielräume innerhalb des DSA sowie Unklarheiten bezüglich der Gesetzesauslegung, um die Wahrnehmung von Nutzendenrechten zu unterlaufen: So werden Nutzende in der Praxis von der Anzeige illegaler Inhalte durch das Design von Meldewegen und Meldeformularen abgehalten. Dies ist Auslegungsspielräumen im DSA selbst geschuldet, die im Gesetzestext klarzustellen sind.

Insbesondere müssen **anonyme Meldungen** ohne die Angabe des Klarnamens explizit ermöglicht werden, um Nutzende, gerade Betroffene schwerer

digitaler Gewalt, wirksam zu schützen. Hierfür bedarf es einer Klarstellung in Art. 16 Abs. 2 sowie den Erwägungsgründen 50 und 53, die derzeit unterschiedlichen Auslegungen zugänglich sind. Die Guidelines des European Data Protection Board²³ (EDPB) vom 11. September 2025 stellen bereits klar: Plattformen sollen die Möglichkeit bieten, Name und E-Mail-Adresse anzugeben, sie dürfen dies aber **nicht als zwingende Voraussetzung für eine Meldung nach Art. 16** verlangen. Diverse Plattformen fragen allerdings zwingend den vollständigen Namen ab und betrachten dies offenbar als vom Gesetz gedeckt. Nutzende schreckt dies von der Nutzung von Meldewegen ab. Ein Effekt, der den Plattformen entgegenkommt.

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>,
zuletzt abgerufen am 13.10.2025, 15:00 Uhr.

²³ Siehe: edpb_guidelines_202503_interplay-dsa-gdpr_v1_en.pdf.

5.3 Zugang zu Rechtsbehelfen sicherstellen

Interne Beschwerdeverfahren und die außergerichtliche Streitbeilegung müssen für alle Entscheidungen über gemeldete Inhalte zugänglich sein, unabhängig davon, ob sie auf AGB- oder Rechtsverstößen beruhen. Die Art. 20 Abs. 1 DSA und ErwG. 58 sehen dies bereits ausdrücklich vor. Damit Nutzende Zugang zu diesen Beschwerdeverfahren haben, müssen sie diese jedoch auch kennen. Hierfür bedarf es einer Rechtsbehelfsbelehrung als verpflichtenden Bestandteil der

Entscheidungsmitteilung auch bei Meldungen von Inhalten wegen AGB-Verstößen. Diese ist aktuell nicht unmissverständlich gesetzlich vorgeschrieben. Melden Nutzende also versehentlich wegen des irreführenden Designs der Meldewege Inhalte nach den AGB statt als illegalen Inhalt oder melden sie Desinformation, bleibt der Zugang zum Beschwerdeverfahren mangels Hinweises hierauf unbekannt, obwohl das Beschwerdeverfahren Abhilfe versprechen könnte.

5.4 Qualität und Einheitlichkeit der Streitbeilegung stärken

Auch bei der **außergerichtlichen Streitbeilegung** gibt es erheblichen Reformbedarf. Diese hat sich überraschend als nützliches Instrument erwiesen, um schnell und kostengünstig Streitigkeiten über Inhaltmoderation beilegen zu können. Dies ist unerwartet, da die Entscheidung der Streitbeilegungsstelle für die Plattformen nicht bindend ist. Die Unterschiede zwischen den bestehenden Stellen sind jedoch so groß, dass es für Betroffene oft vom Zufall abhängt, ob ihr Fall gemäß den gesetzlichen Vorgaben bearbeitet wird. Um hier Abhilfe zu schaffen, braucht es zwei sich ergänzende Instrumente: Zum einen sollte ein **Code of Conduct für Streitbeilegungsstellen nach Art. 45 DSA** entwickelt werden, der Standards zu Unabhängigkeit, Neutralität, Kommunikation mit Nutzenden und Transparenz von Entscheidungen festlegt. Ein solcher Kodex hätte den Vorteil, praxisnah unter Beteiligung von Streitbeilegungsstellen, Plattformen, Zivilgesellschaft und der Kommission erarbeitet zu werden. Er gilt jedoch nur für diejenigen Stellen, die ihn unterzeichnen, und schafft damit keine flächendeckende Verbindlich-

keit. Genau deshalb sollte ergänzend die Europäische Kommission verpflichtet werden, **Leitlinien zu Art. 21 DSA** herauszugeben. Leitlinien wirken als Auslegungshilfe **für alle Streitbeilegungsstellen** und sorgen EU-weit für einheitliche Standards, etwa zu Fristenkontrolle, Begründungspflichten und zur Verpflichtung, Nutzende über die Umsetzung von Entscheidungen durch die Plattformen zu informieren. Anders als Kodizes schaffen sie damit allgemeine Rechtsklarheit und gelten unabhängig von einer freiwilligen Unterzeichnung.

Die Aufsichtsbehörden müssen darüber hinaus in ihrer Tätigkeit den **Nutzendenrechten eine hohe Priorität einräumen** und Verstöße ahnden. Dies gilt für gesetzlich ausdrücklich geregelte Verfahrensweisen, wie die Kennzeichnung automatisierter Entscheidungen, aber auch für die Auslegung von nutzungsfreundlichen Vorgehensweisen. Darüber hinaus müssen Plattformen angehalten werden, **Referenznummern für Meldungen zu vergeben**, um so Entscheidungen nachvollziehbar zu machen.

5.5 Trusted-Flagger-Zertifizierungsverfahren reformieren

Auch im Bereich der **Trusted-Flagger-Zertifizierung** sind EU-weite Mindeststandards notwendig, insbesondere zu den Fachkenntnisanforderungen, die nicht unverhältnismäßig hoch angesetzt werden sollten. Einheitliche, praxistaugliche Kriterien würden hier Klarheit schaffen und den Zugang für

qualifizierte zivilgesellschaftliche Organisationen erleichtern. Statt starrer juristischer Qualifikationshürden müssen auch spezialisierte, geschulte, nicht-juristische Organisationen anerkannt werden, die über nachgewiesene Erfahrung in der Erkennung und Meldung illegaler Inhalte verfügen.

5.6 Transparenz und Nachvollziehbarkeit schaffen

Ein weiteres zentrales Defizit ist das **Fehlen eines europaweit einheitlichen Beschwerdetracking-systems** für Beschwerden bei den Aufsichtsbehörden nach Artikel 53 DSA. Derzeit existiert weder für Betroffene noch für zivilgesellschaftliche Organisationen eine verbindliche Möglichkeit, den Bearbeitungsstand von Meldungen und Beschwerden niedrigschwellig nachzuvollziehen. Diese Intransparenz erschwert nicht nur die Kontrolle des Verfahrensfortschritts, sondern auch die zivilgesellschaftliche Unterstützung der Behörden und die unabhängige Evaluation der Plattformpraxis. Zivilgesellschaftliche Organisationen, auf deren Expertise die Aufsichtsbehörden dringend ange-

wiesen sind, müssen gegenüber Förderern und der Öffentlichkeit belegen können, welche Wirkung ihre Arbeit erzielt. Ohne belastbare Rückmeldungen der Behörden oder Nachverfolgbarkeit von Beschwerdeverfahren fehlt dieser Nachweis und damit der Anreiz, begrenzte Ressourcen in aufwendige Melde- und Beschwerdeprozesse zu investieren.

Nur wenn Zivilgesellschaft, Aufsichtsbehörden und Plattformen in einem transparenten, überprüfbaren und ressourcengestützten Verfahren zusammenarbeiten, kann der DSA seinen Anspruch einlösen, Grundrechte im digitalen Raum wirksam zu schützen.

6 Anhang

Tabelle 1: Entfernung von Inhalten nach Plattformen und Verfahrensstadium (Stand: 16.10.2025)

	Anzahl der Meldungen	Inhalt entfernt (gesamt)	Inhalt nicht entfernt (gesamt, inkl. Fälle ohne Entscheidung)	Inhalt entfernt nach Meldung	> 1 Monat keine Entscheidung der Meldung	Anzahl Beschwerden	Inhalt entfernt nach Beschwerde	> 1 Monat keine Entscheidung der Beschwerde	Anzahl Streitbeilegung ²⁴	Inhalt entfernt nach Streitbeilegung	> 90 Tage keine Entscheidung der Streitbeilegung
Facebook gesamt	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 gesamt	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 gesamt	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 gesamt	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 gesamt	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 2024 reichten wir nur Streitbeilegungsfälle gegen Instagram und TikTok (und einen Versuchsfall gegen YouTube) ein, da User Rights, die einzige Plattform, von der wir Rückmeldungen erhielten, nur Verfahren gegen Instagram und TikTok bearbeitete. Auch bei den anderen Schlichtungsstellen blieben wir daher vorerst bei diesen Plattformen, um die Arbeit der Streitbeilegungsstellen miteinander vergleichen zu können. 2025 erweiterten wir die Untersuchung auf die übrigen Plattformen, um in Erfahrung zu bringen, ob eine effektive Streitbeilegung mit diesen überhaupt möglich ist.

²⁵ Die geringe Anzahl an Beschwerden gegen YouTube ist darin begründet, dass die Plattform kein Verfahren zur Beschwerde nach der Meldung eines Kommentars wegen Rechtsverletzung bereitstellt. Eine Beschwerde war möglich, weil in diesem Fall ein Video und kein Kommentar gemeldet wurde.

²⁶ Die geringe Anzahl an Streitbeilegungsfällen gegen YouTube erklärt sich ebenfalls damit, dass zur Einreichung eines Streitbeilegungsverfahrens erst ein Beschwerdeverfahren durchlaufen sein muss. Nur testweise riefen wir daher trotz fehlender Entscheidungen im Beschwerdeverfahren zweimal eine Streitbeilegungsstelle an – in einem Fall wurde der Antrag abgelehnt, im anderen Fall bestätigte die Streitbeilegungsstelle RTR die Rechtswidrigkeit und forderte YouTube zur Entfernung auf. Dieser Entscheidung kam die Plattform innerhalb von 28 Tagen nicht nach.

Tabelle 2: Bearbeitungszeiten von Meldungen und Beschwerden nach Plattformen (Stand: 16.10.2025)

	Anzahl der Meldungen/ Beschwerden	< 1 Std. (Davon in < 10 Min.)	1 – 24 Std.	1 – 2 T.	3 – 6 T.	7 – 14 T.	> 14 T.	Mind. 1 Monat keine Entscheidung
Facebook Meldungen gesamt	50	7 (1)	33	7	1	0	2	0
Meldungen 2024	23	7 (1)	13	2	1	0	0	0
Meldungen 2025	27	0	20	5	0	0	2	0
Facebook Beschwerden gesamt	24	1 (1)	12	3	3	3	2	0
Beschwerden 2024	8	0	1	0	3	3	1	0
Beschwerden 2025	16	1 (1)	11	3	0	0	1	0
Instagram Meldungen gesamt	56	3 (0)	32	19	2	0	0	0
Meldungen 2024	31	1 (0)	24	5	1	0	0	0
Meldungen 2025	25	2 (0)	8	14	1	0	0	0
Instagram Beschwerden gesamt	36	6 (2)	14	4	5	3	4	0
Beschwerden 2024	21	4 (2)	6	0	5	3	3	0
Beschwerden 2025	15	2 (0)	8	4	0	0	1	0
TikTok Meldungen gesamt	68	28 (0)	14	9	1	2		14
Meldungen 2024	42	20 (0)	8	1	1	2	0	10
Meldungen 2025	26	8 (0)	6	8	0	0	0	4
TikTok Beschwerden gesamt	31	22 (0)	5	3	1	0	0	0
Beschwerden 2024	21	16 (0)	1	3	1	0	0	0
Beschwerden 2025	10	6 (0)	4	0	0	0	0	0
X Meldungen gesamt	77	54 (23)	21	1	0	1	0	0
Meldungen 2024	34	21 (11)	11	1	0	1	0	0
Meldungen 2025	43	33 (12)	10	0	0	0	0	0
X Beschwerden gesamt	25	22 (12)	3	0	0	0	0	0
Beschwerden 2024	11	9 (7)	2	0	0	0	0	0
Beschwerden 2025	14	13 (5)	1	0	0	0	0	0
YouTube Meldungen gesamt	50	15 (3)	12	2	1	0	2	18
Meldungen 2024	21	4 (0)	7	1	0	0	0	9
Meldungen 2025	29	11 (3)	5	1	1	0	2	9
YouTube Beschwerden gesamt	1	0	0	0	0	0	0	1
Beschwerden 2024	1	0	0	0	0	0	0	1
Beschwerden 2025	0	0	0	0	0	0	0	0

Impressum

HateAid gGmbH
Greifswalder Straße 4
10405 Berlin

Telefon: +49 (0) 30 25208802
E-Mail: kontakt@hateaid.org
hateaid.org

Sitz der Gesellschaft: Berlin
Registergericht: Amtsgericht Charlottenburg
Handelsregisternummer: HRB 203883 B
USt-IdNr.: DE322705305

Geschäftsführerinnen: Anna-Lena von Hodenberg
und Josephine Ballon

