

Ergebnisse der Tätigkeit als Trusted Flagger im Jahr 2025

REspect! im Netz

REspect! im Netz ist ein Angebot der Jugendstiftung Baden-Württemberg und bietet pädagogische Angebote und eine bundesweite Meldemöglichkeit. REspect! im Netz kooperiert mit der Bayerischen Staatsregierung und wird gefördert im Rahmen des Bundesprogramms „Demokratie leben!“ im Programmbereich „Entwicklung einer bundeszentralen Infrastruktur“ als Teil des Kooperationsverbundes „toneshift – Netzwerk gegen Hass im Netz und Desinformation“ durch das Bundesministerium für Bildung, Familie, Senioren, Frauen und Jugend sowie aus Mitteln des Bayerischen Staatsministeriums für Familie, Arbeit und Soziales.

(Für inhaltliche Aussagen und Meinungsäußerungen tragen die Publizierenden dieser Veröffentlichung die Verantwortung.)



in Kooperation mit

Bayerische Staatsregierung



Gefördert vom



Bundesministerium
für Bildung, Familie, Senioren,
Frauen und Jugend

im Rahmen des Bundesprogramms

Demokratie *leben!*

und

Bayerisches Staatsministerium für
Familie, Arbeit und Soziales



Einleitung

Trusted Flagger liefern Online-Plattformen Hinweise auf mutmaßlich rechtswidrige Inhalte. Bereits in der Vergangenheit haben Plattformen auf eigene Initiative Trusted Flagger ausgewählt, mit ihnen zusammengearbeitet und ihre Löschhinweise priorisiert. Neu ist, dass durch den Digital Service Act (DSA) das Arbeitsprinzip der Trusted Flagger in der gesamten Europäischen Union verpflichtend geregelt wird.

Die Jugendstiftung Baden-Württemberg ist mit dem Angebot „REspect! im Netz“ seit Oktober 2024 durch die Bundesnetzagentur als Trusted Flagger im Fachgebiet Hassrede, Terrorismuspropaganda und sonstige gewalttätige Inhalte nach dem DSA zertifiziert. Im Zuge des Zertifizierungsverfahrens hat REspect! nachgewiesen, dass sie über besondere Sachkenntnisse und Kompetenzen in Bezug auf die Erkennung, Feststellung und Meldung rechtswidriger Inhalte verfügt. REspect! ist unabhängig von jeglichen Anbietern von Online-Plattformen und übt seine Tätigkeit zur Übermittlung von Meldungen sorgfältig, genau und objektiv aus (vgl. Art. 22 Abs. 2 DSA).

Jede bei REspect! eingegangene Meldung wird von einem juristischen Team rechtlich eingeordnet. Meldungen, die als strafrechtlich relevant eingestuft werden, erhält das BKA zur weiteren Prüfung, damit die zuständige Polizeidienststelle und die zuständige Staatsanwaltschaft die Strafbarkeit prüfen können. Über mögliche strafrechtliche Konsequenzen entscheiden letztlich immer die unabhängigen Gerichte. Wird in der rechtlichen Prüfung durch das Bundeskriminalamt (BKA) die Strafbarkeit bestätigt und sollte der entsprechende Inhalt noch online aufrufbar sein, stellt REspect! eine Löschbitte an die Plattformen. Die Entscheidung, ob ein Inhalt gelöscht oder eingeschränkt wird, trifft die Plattform in eigener Verantwortung. Meldende Personen bekommen in jedem Fall eine Rückmeldung mit einer Einschätzung zu ihrer Meldung und bei Bedarf und auf Anfrage Hinweise für unterstützende Beratung.

Trusted Flagger sind laut Art. 22 Abs. 3 DSA verpflichtet, mindestens einmal jährlich leicht verständliche und ausführliche Berichte über die während des betreffenden Zeitraums eingereichten Meldungen zu veröffentlichen. Die Berichte sollen mindestens die Anzahl der Meldungen nach Identität der Plattform, Art der gemeldeten mutmaßlich rechtswidrigen Inhalte und der vom Anbieter ergriffenen Maßnahmen kategorisieren.

Nach der Zertifizierung im Oktober 2024 begann REspect! zunächst damit, bilateral mit den verschiedenen Anbietern von Onlineplattformen die in Art. 22 Abs. 1 DSA geforderten Meldewege zu etablieren sowie die geforderten Onboarding-Prozesse zu durchlaufen. Die eigentliche Tätigkeit, Meldungen als Trusted Flagger bei den Onlineplattformen einzureichen, konnte deshalb erst zum Januar 2025 beginnen.

Gesamtergebnisse

Im Zeitraum vom 01. Januar 2025 bis 31. Dezember 2025 wurden bei RESpect! insgesamt 21.813 Fälle gemeldet. Davon wurden 4.677 Beiträge als strafrechtlich relevant eingestuft und an das Bundeskriminalamt (BKA) weitergeleitet.

Bei Fällen, in denen der Anfangsverdacht einer Strafbarkeit innerhalb dieses Zeitraums bestätigt wurde und die Inhalte weiterhin online abrufbar waren, erfolgte eine Meldung durch RESpect! an die jeweiligen Plattformen gemäß den entsprechenden Paragraphen des Strafgesetzbuches (StGB). Insgesamt handelte es sich dabei um 1.576 mutmaßlich rechtswidrige Inhalte, die im Rahmen des Trusted Flagger Verfahrens an Hosting-Dienstanbieter übermittelt wurden.

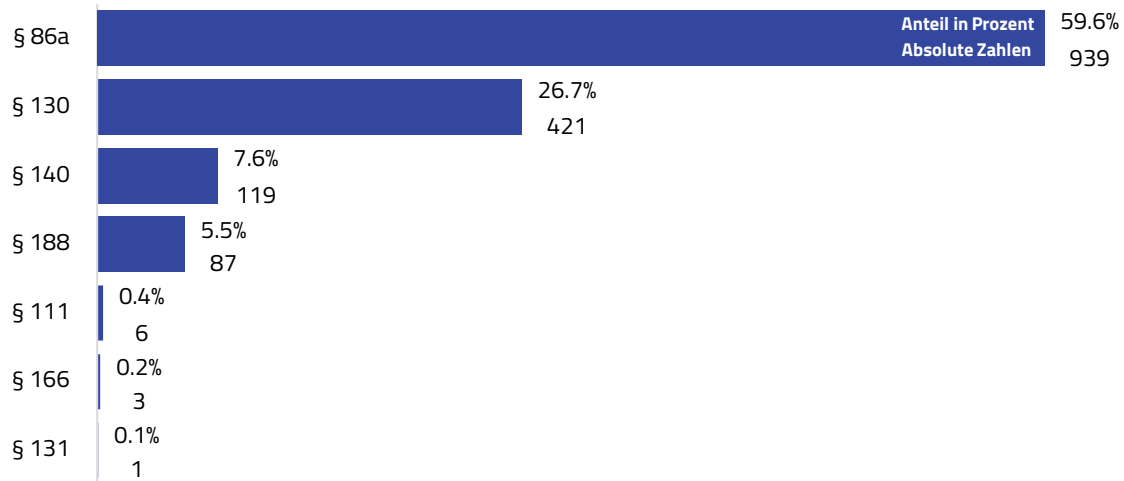
In 1.268 Fällen erfolgte daraufhin eine Löschung oder Sperrung des Inhalts durch den jeweiligen Hosting-Dienstanbieter. Das entspricht einer Quote von 80%. In 292 Fällen blieben die gemeldeten Inhalte weiterhin öffentlich zugänglich. Dies entspricht einer Quote von 19%. In weiteren 16 Fällen erhielt RESpect! auf die eingereichten Meldungen keine Rückmeldung.

Ergebnis der Löschbitten an die Plattformen

Plattform	Anzahl der Löschbitten (2025)	Entfernt oder eingeschränkt in DE	Abgelehnt	Keine Reaktion
X	671	493 (73%)	178 (27%)	–
Facebook	550	464 (84%)	85 (15%)	1
TikTok	163	154 (94%)	9 (6%)	–
Instagram	112	102 (91%)	10 (9%)	–
Telegram	15	–	–	15
YouTube	34	33 (97%)	1 (3%)	–
Reddit	7	–	7 (100%)	–
Soundcloud	10	10 (100%)	–	–
Threads	9	7 (78%)	2 (22%)	–
Pinterest	1	1 (100%)	–	–
Quora	3	3 (100%)	–	–
Steam	1	1 (100%)	–	–
Gesamt	1576	1268 (80%)	292 (19%)	16

Gesetzliche Grundlage der Löschbitten

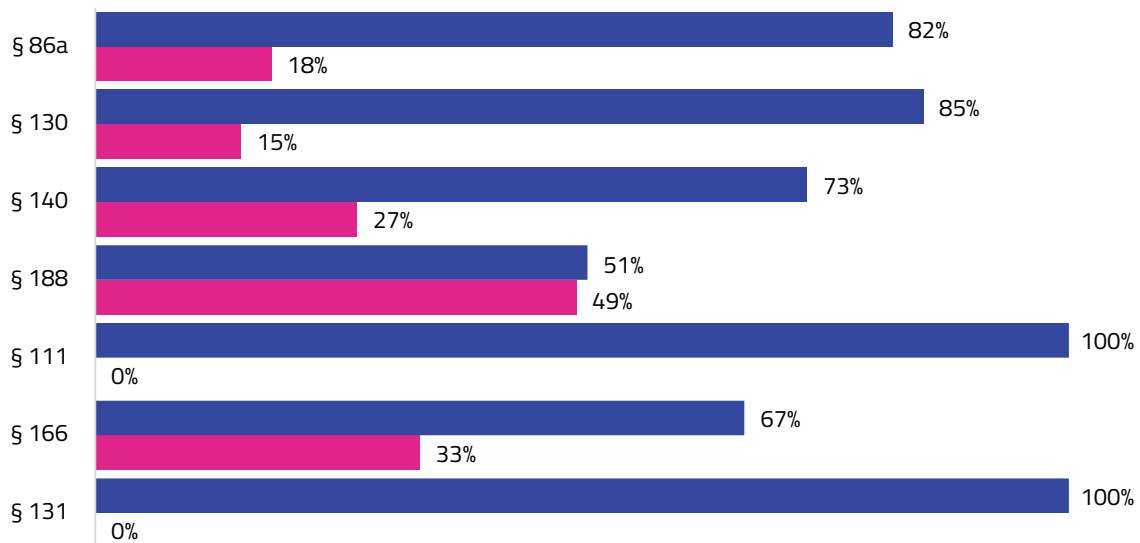
Abbildung 1 N = 1576



- § 86a StGB: Verwenden von Kennzeichen verfassungswidriger und terroristischer Organisationen
- § 111 StGB: Öffentliche Aufforderung zu Straftaten
- § 130 StGB: Volksverhetzung
- § 131 StGB: Gewaltdarstellung
- § 140 StGB: Belohnung und Billigung von Straftaten
- § 166 StGB: Beschimpfung von Bekenntnissen, Religionsgesellschaften und Weltanschauungsvereinigungen
- § 188 StGB: Gegen Personen des politischen Lebens gerichtete Beleidigung, üble Nachrede und Verleumdung

Löschquote der gemeldeten Inhalte nach Delikt

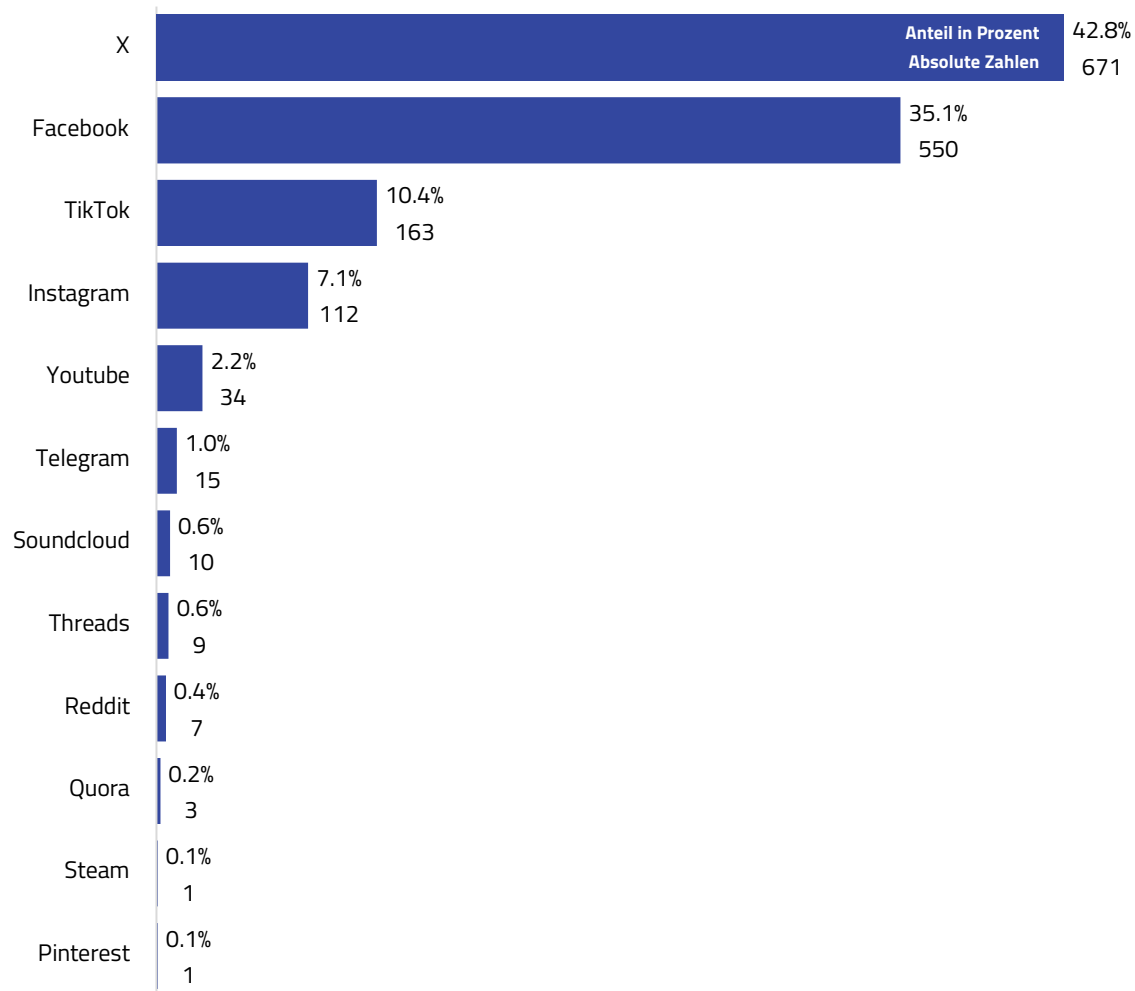
Abbildung 2 N=1567 ■ entfernt % ■ aufrufbar %



Im Folgenden werden die Ergebnisse der Tätigkeit als Trusted Flagger jeweils nach den verschiedenen Onlineplattformen aufgeschlüsselt dargelegt.

Meldungen gesondert nach Plattform

Abbildung 3 N=1567

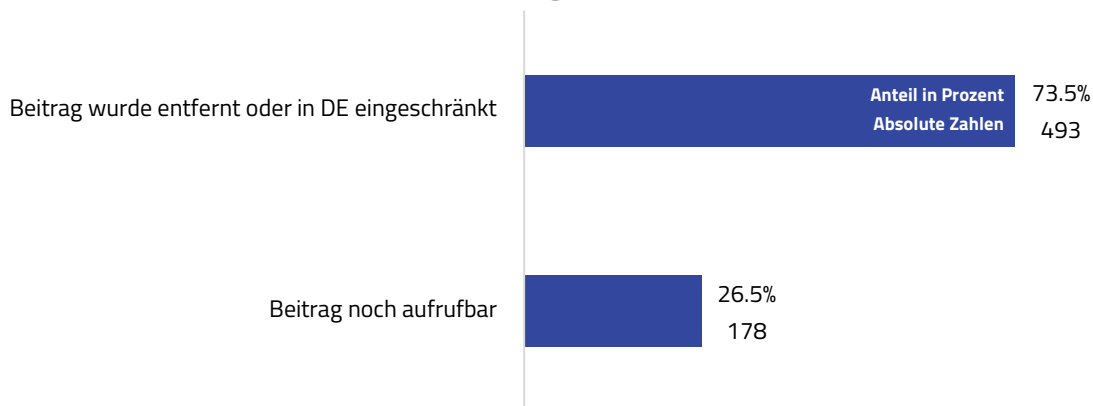


Löschbitten bei X

Die Rückmeldungen von X zu den eingereichten Löschbitten erfolgten insgesamt deutlich schneller im Vergleich zu anderen Plattformen. Sämtliche Antworten auf die eingereichten Löschanfragen wurden noch am selben Tag zugestellt – im Durchschnitt sogar innerhalb einer Stunde oder schneller.

Jedoch erwies sich die Möglichkeit, mehrere Inhalte beim Anbieter zu melden, als herausfordernd. Ein passwortgeschütztes Portal für die Tätigkeit als Trusted Flagger wurde nicht bereitgestellt. Nach der Meldung von drei Beiträgen war das Bestehen eines Captcha-Tests erforderlich. Nach sechs aufeinanderfolgenden Meldungen mussten sogar drei Captcha-Tests absolviert werden, was einen erheblichen Aufwand darstellt. Die Bestätigungsmail, die nach dem Versenden der Löschanfragen einging, enthielt keine Referenznummer, was die Nachvollziehbarkeit erschwerte.

Abbildung 4 N=671



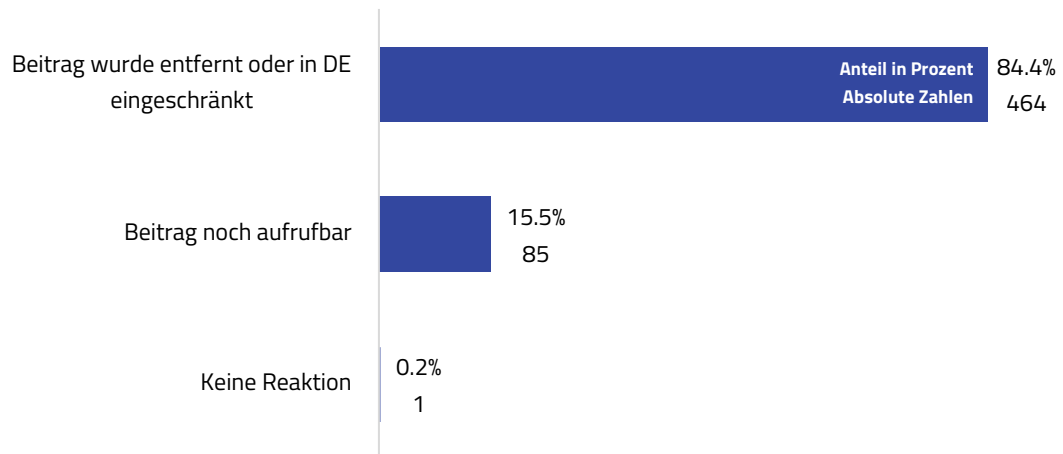
Löschbitten bei Meta

Die Meldung von Inhalten auf sämtlichen Plattformen, die zu Meta gehören, e.g. Facebook, Instagram und Threads, erfolgt über ein einheitliches Meldeformular. Dieses Formular ermöglicht es, Löschanfragen zu stellen, deren Bearbeitung und Ergebnisse zentral erfasst und zusammengestellt werden. Für sogenannte „Trusted Flagger“ – also vertrauenswürdige Hinweisgeber – bietet Meta eine strukturierte, geschützte und transparente Meldeplattform. Diese erlaubt es, Inhalte entweder einzeln oder in Bündeln von bis zu 20 URLs gleichzeitig zu melden.

Die Plattform bietet außerdem eine Übersicht über den Verlauf aller eingereichten Meldungen. Dadurch ist es möglich, jede Entscheidung im Zusammenhang mit einer Meldung nachzuvollziehen – inklusive der getroffenen Maßnahmen und der jeweiligen Beiträge. Die Bearbeitungszeit für Löschanfragen variiert stark. Sie kann von einem Tag bis zu zwei Wochen betragen.

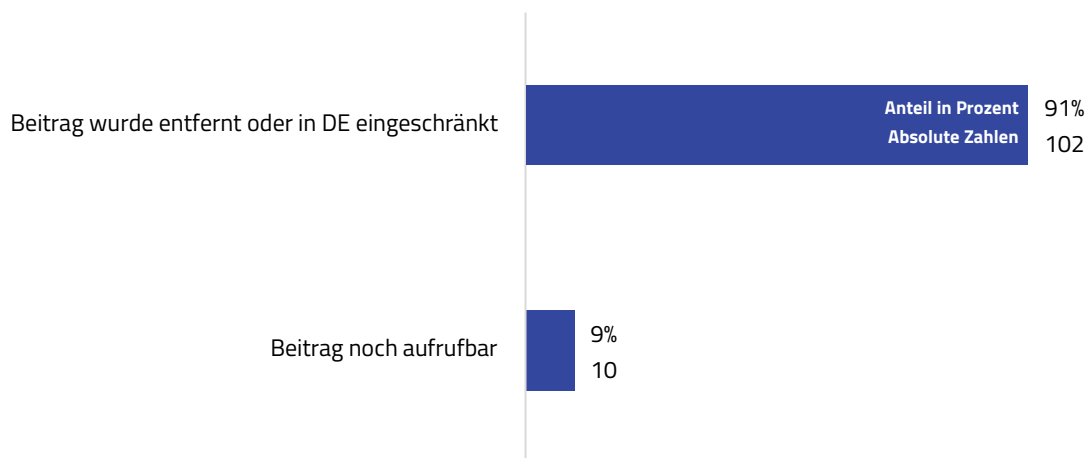
Ergebnis der Löschbitten bei Facebook

Abbildung 5 N=550



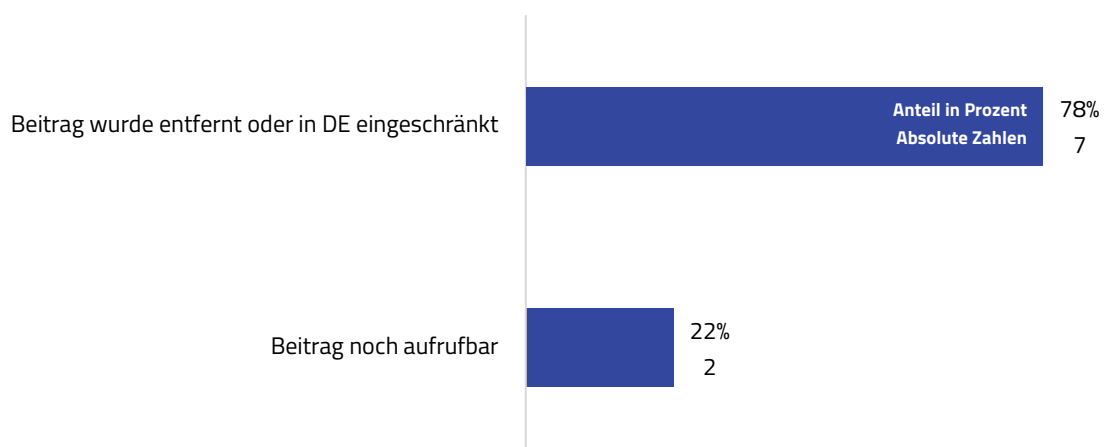
Ergebnis der Löschbitten bei Instagram

Abbildung 6 N=112



Ergebnis der Löschbitten bei Threads

Abbildung 7 N=9



Löschbitten bei TikTok

Die Meldung rechtswidriger Inhalte erfolgte bis April 2025 über ein spezielles Safety-Enforcement-Tool. Dieses ermöglichte die gleichzeitige Meldung mehrerer rechtswidriger Beiträge und verfügte über ein integriertes Verlaufsprotokoll, das eine übersichtliche Einsicht, Auswertung und Analyse der gemeldeten Inhalte hatte. Sämtliche Löschersuchen wurden bearbeitet. Die durchschnittliche Reaktionszeit betrug ein bis zwei Tage.

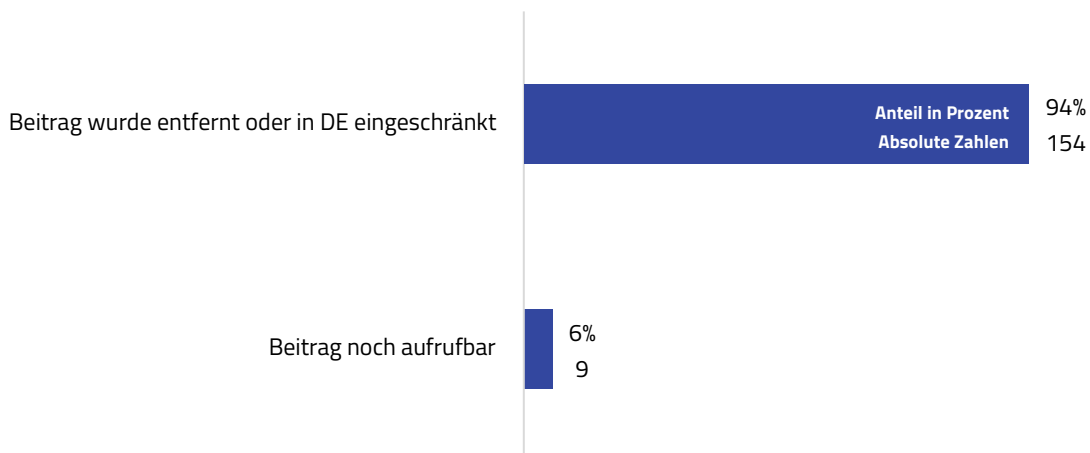
Ab April 2025 teilte für Trusted Flagger einen Meldeprozess mit. Hierfür müssen Trusted Flagger ein TikTok-Konto erstellen und dieses als Trusted-Flagger-Konto verifizieren lassen. Im Gegensatz zum bisherigen Safety-Enforcement-Tool bietet das neue Verfahren jedoch keine Möglichkeit mehr, mehrere Inhalte gesammelt zu melden. Stattdessen muss jeder rechtswidrige Beitrag bzw. Kommentar einzeln aufgerufen und gemeldet werden.

Darüber hinaus werden nach der Meldung keine Referenz- oder Vorgangsnummern vergeben, was die Nachvollziehbarkeit der Bearbeitung sowie die Dokumentation der Meldungen erheblich erschwert. Ebenso besteht keine Möglichkeit mehr, die Anzahl der eingereichten Meldungen oder deren Ergebnisse systematisch zu erfassen und auszuwerten. Die bislang vorhandenen Funktionen zur Extraktion und Analyse von Meldedaten, die das Safety-Enforcement-Tool bereitstellte, stehen somit nicht mehr zur Verfügung.

Eine weitere zentrale Herausforderung bei der Meldung rechtswidriger Kommentare auf TikTok besteht darin, den betreffenden Kommentar innerhalb von Beiträgen mit einer hohen Anzahl an Kommentaren aufzurufen. Da Kommentare häufig erst sukzessive geladen und angezeigt werden, ist es regelmäßig erforderlich, manuell durch eine Vielzahl von Einträgen zu scrollen, um den relevanten rechtswidrigen Kommentar aufzufinden. Dies führt insbesondere bei umfangreich kommentierten Beiträgen zu einem erheblichen zeitlichen Mehraufwand.

Ergebnis der Löschbitten bei TikTok

Abbildung 8 N=163

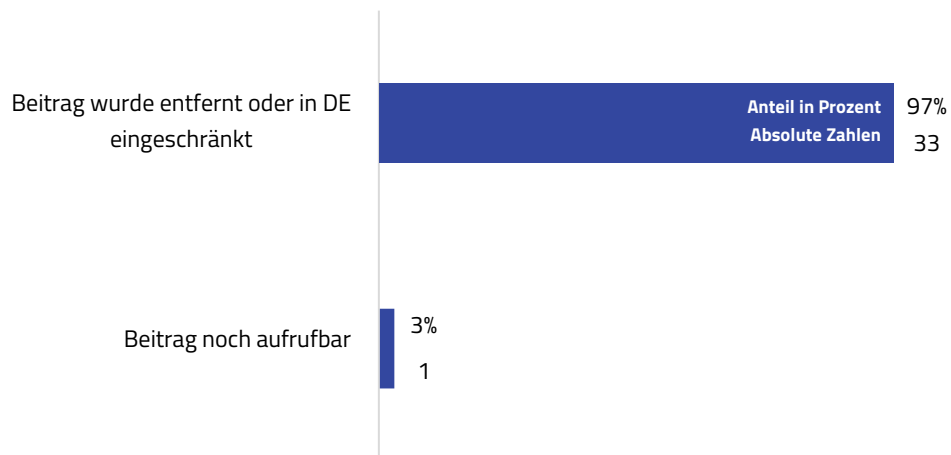


Löschbitten bei YouTube

Für die Meldung rechtswidriger Inhalte auf YouTube ist die Einrichtung eines gesonderten Kontos für die Tätigkeit als Trusted Flagger erforderlich. Nach erfolgreicher Verifizierung erfolgt die Meldung rechtswidriger Inhalte über die von YouTube bereitgestellten Meldewege. Dabei besteht die Möglichkeit, bis zu fünf Inhalte gleichzeitig zu melden.

Sämtliche Löschbitten wurden bearbeitet. Die Reaktionszeit auf eingereichte Meldungen betrug von wenigen Stunden bis zu zwei Tage.

Abbildung 9 N=34



Löschbitten bei Telegram

Die Meldung rechtswidriger Inhalte auf Telegram erfolgte im Jahr 2025 über die regulären Meldewege der Plattform. Ein gesonderter Meldekanal für Trusted Flagger stand in diesem Zeitraum noch nicht zur Verfügung und wurde erst im Jahr 2026 mitgeteilt.

Im Berichtszeitraum wurden insgesamt 15 rechtswidrige Inhalte an die Plattform gemeldet. Auf keine dieser Meldungen erfolgte eine Rückmeldung oder sonstige erkennbare Reaktion seitens Telegram. Ebenso konnten keine Maßnahmen hinsichtlich der gemeldeten Inhalte festgestellt werden. Die Löschrquote der eingereichten Meldungen lag damit im Jahr 2025 bei null Prozent.

Mangels Rückmeldungen der Plattform war eine Nachverfolgung des Bearbeitungsstands der Meldungen nicht möglich. Zudem konnten keine belastbaren Aussagen über die Bearbeitungsdauer oder die interne Behandlung der gemeldeten Inhalte getroffen werden.

Löschbitten bei Soundcloud

Die Meldung rechtswidriger Inhalte auf SoundCloud erfolgte im Jahr 2025 über die regulären Meldewege der Plattform. Ein gesonderter Meldekanal für Trusted Flagger steht nicht zur Verfügung. Im Berichtszeitraum wurden insgesamt zehn Inhalte an die Plattform gemeldet. Sämtliche Meldungen betrafen Inhalte mit strafrechtlicher Relevanz nach § 86a StGB. Die Plattform reagierte auf alle eingereichten Meldungen. Die durchschnittliche Bearbeitungszeit betrug einen Tag. Damit führte jede der eingereichten Meldungen zur Entfernung oder Sperrung der gemeldeten Inhalte.

Löschbitten bei Reddit

Die Meldung rechtswidriger Inhalte auf Reddit erfolgte im Jahr 2025 über die regulären Meldewege der Plattform. Ein gesonderter Meldekanal für Trusted Flagger steht nicht zur Verfügung.

Im Berichtszeitraum wurden insgesamt sieben Inhalte an die Plattform gemeldet. Bei fünf Meldungen erfolgte eine Bearbeitung innerhalb von 24 Stunden. Bei den beiden übrigen Meldungen betrug die Bearbeitungszeit jeweils mehr als zehn Tage.

Reddit reagierte auf alle eingereichten Meldungen. Sämtliche Löschbitten wurden jedoch abgelehnt, sodass keiner der gemeldeten Inhalte entfernt oder eingeschränkt wurde. Die Löschrquote der eingereichten Meldungen lag damit bei null Prozent.

Löschbitten bei Quora

Die Meldung rechtswidriger Inhalte auf Quora erfolgte im Jahr 2025 über die regulären Meldewege der Plattform. Im Rahmen des Meldeverfahrens besteht die Möglichkeit, den Status als Trusted Flagger anzugeben.

Im Berichtszeitraum wurden insgesamt drei Inhalte an die Plattform gemeldet. Die Plattform reagierte auf sämtliche Meldungen innerhalb eines Tages und entfernte beziehungsweise beschränkte die gemeldeten Inhalte. Damit führten alle eingereichten Meldungen zu Maßnahmen seitens der Plattform, sodass eine Löschrquote von 100 Prozent erreicht wurde.

Löschbitten bei Steam

Die Meldung rechtswidriger Inhalte auf Steam erfolgte im Jahr 2025 über die regulären Meldewege der Plattform. Ein gesonderter Meldekanal für Trusted Flagger steht nicht zur Verfügung.

Im Berichtszeitraum wurde ein Inhalt an die Plattform gemeldet. Nach Einreichung der Meldung erfolgte eine automatisierte Bestätigung des Eingangs. Eine weitere Rückmeldung oder Entscheidung der Plattform wurde jedoch nicht übermittelt.

Eine nachträgliche Überprüfung des gemeldeten Inhalts nach zwei Wochen ergab, dass dieser nicht mehr aufrufbar war. Es ist daher davon auszugehen, dass der beanstandete Inhalt entfernt oder anderweitig unzugänglich gemacht wurde. Die Bearbeitungszeit kann aufgrund der fehlenden Rückmeldung nicht bestimmt werden.

Verfahren zur Sicherung der Unabhängigkeit als Trusted Flagger

Die Jugendstiftung Baden-Württemberg ist eine unabhängige Stiftung. Der Vorstand der Stiftung besteht aus erfahrenen Fachleuten, die in keinem Beschäftigungs- oder Abhängigkeitsverhältnis zu Hostingdiensteanbietern stehen.

Die Finanzierung der Jugendstiftung erfolgt durch öffentliche Gelder, private Zuschüsse und Spenden. Die Jugendstiftung hat im Berichtszeitraum keinerlei Spenden von Hostingdiensteanbietern erhalten. REspect! im Netz finanziert sich durch öffentliche Gelder (90%) im Rahmen des Förderprogramms „Demokratie leben!“ durch das Bundesministerium für Bildung, Familie, Senioren, Frauen und Jugend, aus Mitteln des Bayerischen Staatsministeriums für Familie, Arbeit und Soziales und aus eigenen Mitteln der Jugendstiftung (10%). REspect! im Netz erhält keine Zuwendungen von Hostingdiensteanbietern und arbeitet unabhängig.