



### 3. ZARA – Monitoring der Löschraktiken illegaler Online-Hetze von IT-Unternehmen - Ergebnisse Österreich im Detail

#### *Einleitung*

Ende Mai 2016 vereinbarte die [Europäische Kommission](#) mit den vier IT-Unternehmen Facebook, Microsoft, Twitter und YouTube einen Verhaltenskodex zur Begegnung illegaler online Hetze<sup>1</sup>, den sogenannten „[Code of Conduct on Countering Illegal Hate Speech Online](#)“. In diesem Kodex verpflichteten sich die Unternehmen u.a. dazu:

- Klare und wirksame Mechanismen einzurichten, die eine Meldung und das Entfernen von illegaler online Hetze ermöglichen.
- In ihren Nutzungsbedingungen hervorzuheben, dass Aufrufe zur Gewalt und online Hetze nicht geduldet werden.
- Gemeldete Verstöße sowohl bezüglich der Einhaltung der Nutzungsbedingungen als auch der nationalen Gesetzeslage zu überprüfen.
- Meldungen innerhalb von 24 Stunden zu überprüfen und Verstöße zu entfernen bzw. für die Nutzer\*innen nicht mehr zugänglich zu machen.

Im Verhaltenskodex wurde auch vereinbart, dass diese öffentliche Verpflichtung regelmäßig überprüft und evaluiert werden solle. Zu diesem Zweck wurde die „Untergruppe zur Begegnung von online Hass und Hetze“ eingerichtet, bestehend aus Vertreter\*innen der EU-Mitgliedstaaten, der IT-Unternehmen und verschiedener, in diesem Bereich tätigen zivilgesellschaftlichen Organisationen.<sup>2</sup>

Im Herbst 2016 entwickelte diese Gruppe eine Methode zur Überprüfung des Umgangs von IT-Unternehmen mit Meldungen von Beiträgen, die zu Hass und Gewalt aufrufen und organisierte unter Einbeziehung zivilgesellschaftlicher Organisationen ein Monitoring.

Am ersten Monitoring vom 10.10. – 20.11.2016 haben 12 Organisationen aus neun EU-Staaten teilgenommen, darunter ZARA – Zivilcourage und Anti-Rassismus-Arbeit in Österreich. Dabei wurde eine durchschnittliche Löschrquote von 28 % festgestellt. In Österreich lag das Ergebnis weit darunter: lediglich 11,4 % der gemeldeten illegalen Inhalte wurden von den IT-Unternehmen gelöscht.

Beim zweiten Monitoring überprüften 34 Organisationen aus 24 Ländern die Löschraktiken der IT-Unternehmen im Zeitraum 20.03. - 05.05.2017. Das Ergebnis für Österreich war im Vergleich zum ersten Durchgang erheblich besser - die Löschrquote der gemeldeten Hassinhalte betrug 76,1 %. Das Gesamtergebnis aller beteiligten Länder belief sich auf eine Löschrquote von 59.1 %.

---

<sup>1</sup> Der Verhaltenskodex kann hier heruntergeladen werden: [http://ec.europa.eu/justice/fundamental-rights/files/hate\\_speech\\_code\\_of\\_conduct\\_en.pdf](http://ec.europa.eu/justice/fundamental-rights/files/hate_speech_code_of_conduct_en.pdf) (letzter Zugriff 19.12.2017)

<sup>2</sup> Untergruppe zur Begegnung von online Hass und Hetze: <http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetailDoc&id=26468&no=1> (letzter Zugriff 19.12.2017)



Im Herbst 2017 wurde ein drittes Monitoring durchgeführt. Insgesamt beteiligten sich daran 33 zivilgesellschaftliche Organisationen und zwei staatliche Stellen in 27 EU-Mitgliedsstaaten.

## ***Ergebnisse Österreich/ 3. Monitoring 6. November – 15. Dezember 2017:***

### **Übersicht**

ZARA – Zivilcourage und Anti-Rassismus-Arbeit hat im Monitoring Zeitraum, 06.11. - 15.12.2017 insgesamt 105 illegale Hassinhalte an die drei Social-Media Plattformen Facebook, YouTube und Twitter gemeldet, davon 69 an Facebook, 18 an YouTube und 18 an Twitter. Dabei waren alle gemeldeten Inhalte im Vorfeld von den Rechtsexpert\*innen von ZARA – Zivilcourage und Anti-Rassismus-Arbeit als klar strafrechtswidrig eingestuft worden.

### **Inhalt der gemeldeten Beiträge**

Gemeldet wurden Beiträge, die zu Gewalt aufrufen (z.B., dass Flüchtlinge, Muslim\*innen, Schwarze, Ausländer\*innen, Jüd\*innen vergast/erschossen/verbrannt/vom Panzer überrollt/in die Luft gesprengt, usw. werden sollen), die rassistische beleidigende Ausdrücke verwenden oder in den Bereich der nationalsozialistischen Widerbetätigung fallen (z.B. ein Hakenkreuz als Profilbild auf Twitter, eine Rede Hitlers mit zustimmender Videoüberschrift). Als illegal wurden Beiträge angesehen, die einen Straftatbestand nach §283 öStGB zu Verhetzung<sup>3</sup> und/oder nach dem Verbotsgesetz von 1947<sup>4</sup> darstellen.

Durch die thematische Ausrichtung von ZARA auf Anti-Rassismus-Arbeit lag der Schwerpunkt der gemeldeten Beiträge auf den verschiedenen Formen rassistischer Hassrede. 43 der gemeldeten Beiträge enthielten Aussagen, die gegen Muslim\*innen hetzen. In 36 Beiträgen haben die Poster\*innen zu Hass oder Gewalt gegen Geflüchtete und Migrant\*innen aufgerufen, 11 enthielten generell rassistische Aussagen, in 9 Fällen antisemitische Botschaften und 6 weitere Postings richteten sich gegen Personen aufgrund ihrer nationalen oder ethnischen Herkunft. In 11 Fällen wurde nationalsozialistisches Gedankengut verbreitet oder der Nationalsozialismus verherrlicht und fielen damit unter das Verbotsgesetz. In 94 Fällen war der Tatbestand der Verhetzung nach § 283 öStGB verwirklicht. Anzumerken ist, dass die Zuordnung zu den einzelnen Gruppen aufgrund der undifferenzierten Hassbotschaften gegen Geflüchtete, Muslim\*innen oder Migrant\*innen mitunter schwierig war. Letztlich wurden die Beiträge der jeweils speziellsten Gruppe zugeordnet (z.B. der Gruppe der Muslim\*innen, wobei diese Gruppe häufig mit Geflüchteten bzw. mit Migrant\*innen gleichgesetzt wird). Grundsätzlich bilden diese Zahlen nicht die Verteilung von Hass- und Hetzbotschaften im Internet ab, sondern basieren auf der Herangehensweise an das Monitoring von ZARA – Zivilcourage und Anti-Rassismus-Arbeit und wie entsprechende Beiträge gesucht und gefunden wurden.

---

<sup>3</sup><https://www.ris.bka.gv.at/Dokumente/Bundesnormen/NOR40177271/NOR40177271.pdf> (letzter Zugriff 20.12.2017).

<sup>4</sup><https://www.ris.bka.gv.at/GeltendeFassung/Bundesnormen/10000207/Verbotsgesetz%201947%2c%20Fassung%20vom%2020.12.2017.pdf> (letzter Zugriff 20.12.2017)

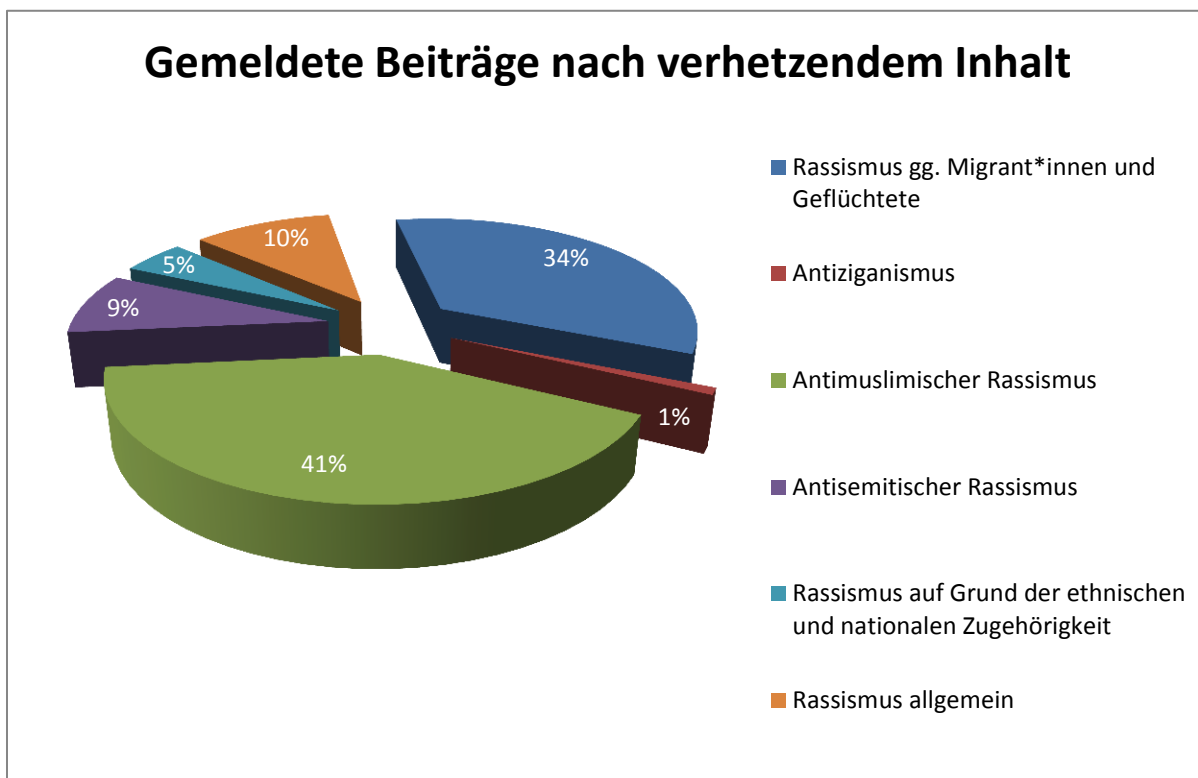


Abbildung 1 Gemeldete Beiträge nach Inhalt. Die Zahlen geben die Anzahl der Beiträge an, die einen entsprechenden verhetzenden Inhalt enthalten.

#### Löschraten

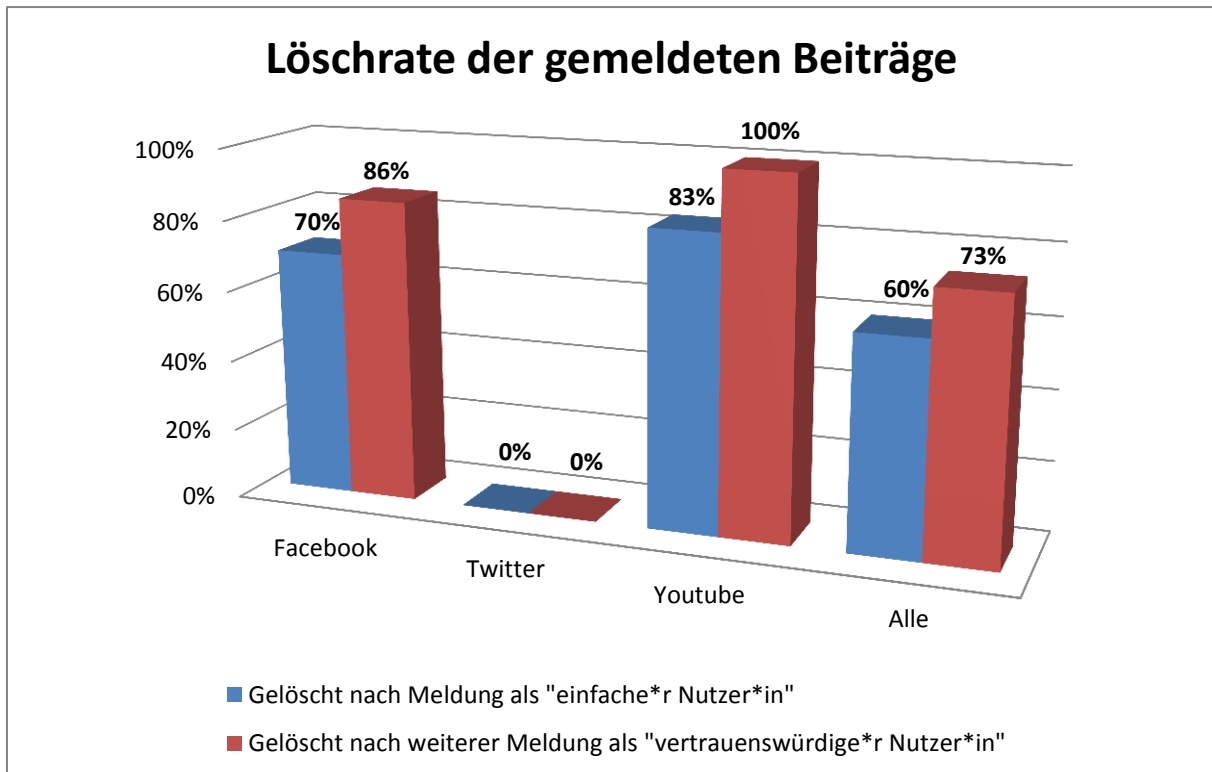
Insgesamt wurden 105 Hassinhalte von ZARA an die drei Plattformen Facebook, Twitter und YouTube geschickt. Davon wurden 63 zunächst anhand der Meldemöglichkeiten, die allen Nutzer\*innen zur Verfügung stehen, gemeldet. Bei Facebook und YouTube wurden 26 der im vereinbarten Zeitraum nicht gelöschten Beiträge ein zweites Mail als sog. „vertrauenswürdiger Nutzer“ gemeldet. Insgesamt hat dieses zweistufige Verfahren zu einer Löschrates von 73,3 % geführt.

Plattform	Hassinhalte	Gelöscht	Löschrates
Facebook	69	59	86%
Twitter	18	0	0%
Youtube	18	18	100%
<b>Gesamt</b>	105	77	<b>73,3 %</b>

Tabelle 1 Übersicht zu den Meldungen.

Die nochmalige Meldung als „vertrauenswürdiger Nutzer“ hat dabei zu einer erheblichen Verbesserung der Löschrates geführt, insbesondere bei YouTube, die damit eine Löschrates von 100%

erreichten. Facebook reagierte zu 70 % auf Meldungen als „normale Nutzer\*innen“; nach einer nochmaligen Meldung der nicht gelöschten Beiträge reagierte das Unternehmen insgesamt zu 86 % mit einer Löschung.



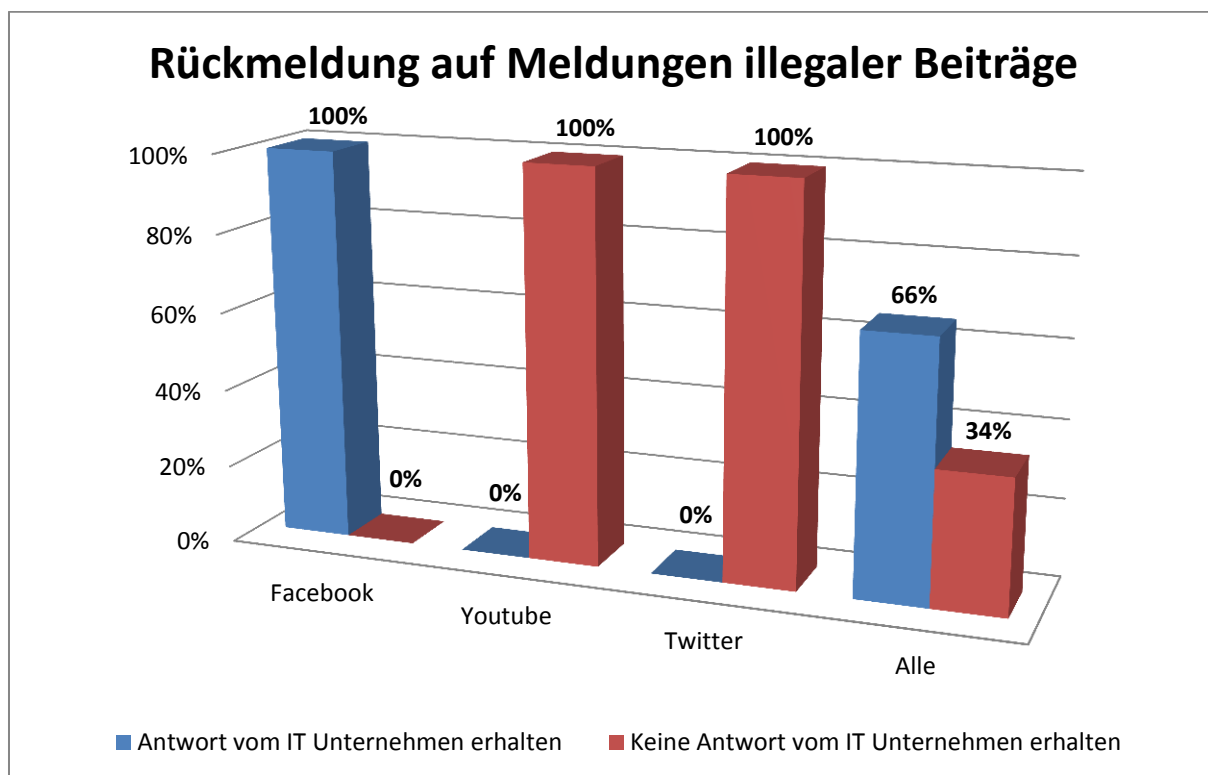
**Abbildung 2 Löschräte aller von ZARA gemeldeten Fälle in Prozent. Als gelöscht wurden alle Fälle gezählt, die im Beobachtungszeitraum von Österreich aus nicht mehr abrufbar waren. \*Bei Twitter wurden keine Fälle als „vertrauenswürdige\*r Nutzer\*in“ gemeldet.**

### Rückmeldung durch IT-Unternehmen

Rückmeldungen, ob Meldungen bearbeitet und ob Beiträge gelöscht wurden oder nicht, erhielt ZARA lediglich von Facebook zu allen der 69 als „normale Nutzer\*innen“ gemeldeten Beiträge. Facebook informiert Nutzer\*innen, die Beiträge als Hasspostings melden, ob die gemeldeten Beiträge gelöscht wurden oder nicht, direkt über den eigenen Nutzer\*innen Account. Bei Meldungen als „vertrauenswürdiger Nutzer“ erhielt ZARA von Facebook zu allen Meldungen eine Antwort.

Twitter reagierte auf gemeldete Inhalte lediglich mit einer Bestätigung für den Erhalt. Im Anschluss darauf gab es keinen Hinweis darauf, dass die Meldung weiterbearbeitet wurde. Die Inhalte waren zum Zeitpunkt der Berichtslegung nach wie vor online abrufbar.

In ihrer Begründung, warum gemeldete Beiträge nicht gelöscht wurden, verwies Facebook auf seine Nutzungsbedingungen und dass die geprüften Beiträge nicht gegen diese verstoßen würden. YouTube informierte bei Meldungen als „normaler Nutzer“ weder über den Eingang einer Meldung, ob diese geprüft wurden, noch über den Ausgang einer eventuellen Prüfung. Auch bei Meldungen als „vertrauenswürdiger Nutzer“ gab es keine Rückmeldung von YouTube.



**Abbildung 3** Anzahl der Rückmeldungen der Social-Media Unternehmen auf gemeldete illegale Beiträge in Prozent. Abgebildet werden nur die Meldungen, die als „normale Nutzer\*innen“ geschickt wurden. Als Rückmeldung wurde gezählt, wenn nach der Meldung aktiv die Nutzer\*innen informiert wurden, dass die Meldung bearbeitet wurde unabhängig vom Ergebnis.

#### ***Erläuterung zur Herangehensweise/Methode***

ZARA profitierte beim Monitoring stark von der jahrelangen Erfahrung und Expertise der Mitarbeiter\*innen der [ZARA-Beratungsstelle für Opfer und Zeug\\*innen von Rassismus](#) und der [Beratungsstelle #GegenHassimNetz](#) im Umgang mit online Hass und Hetze. Aufgrund ihrer Arbeit sind den Mitarbeiter\*innen einschlägige Gruppen und Nutzer\*innen bekannt, in deren Umfeld illegale online Hetze besonders häufig veröffentlicht wird und die deshalb für das Monitoring aufgesucht werden konnten. Darüber hinaus konnte ZARA auf eine Liste rassistischer Schlüsselwörter zurückgreifen, die das Finden von entsprechenden Hasspostings erleichterte.

Gemeldet und dokumentiert wurden ausschließlich Beiträge, die als illegal eingeschätzt wurden. Die Entscheidung, welche Beiträge als mutmaßlich illegal anzusehen waren, oblag der juristischen Einschätzung der Mitarbeiter\*innen von ZARA.

Um eine Vergleichbarkeit der Monitoring Daten gewährleisten zu können, entwickelte die „Untergruppe zur Begegnung von online Hass und Hetze“ einen Leitfaden für die Dokumentation der Löschrouten illegaler Beiträge auf Social-Media Plattformen. Darin wurde, neben der grundlegenden Vorgabe, dass nur illegale Beiträge gemeldet werden sollen, festgelegt, auf welche Art und Weise Beiträge gemeldet und in welchen Zeitabständen kontrolliert werden soll, ob gemeldete Beiträge noch abrufbar sind oder nicht.



Mehrere Social-Media Plattformen bieten zivilgesellschaftlichen Organisationen und weiteren ausgewählten Nutzer\*innen an, Beiträge über separate Wege und eigene Kanäle zu melden. Um den Unterschied einer Meldung über die angebotenen Meldemechanismen, die allen Nutzer\*innen zur Verfügung stehen, zur Meldung als „vertrauenswürdige Nutzer\*innen“ erheben zu können, wurden beide Möglichkeiten aufeinander aufbauend genutzt. Zuerst wurden alle Beiträge als „normale User\*in“ gemeldet, nicht gelöschte Beiträge nach einer vereinbarten Zeit dann erneut als „vertrauenswürdige Nutzer\*innen“ gemeldet.

Im Frühjahr 2017 wurde ZARA – Zivilcourage und Anti-Rassismus-Arbeit von Facebook und YouTube in deren Programme für Policy Casework aufgenommen, über die „vertrauenswürdige Nutzer\*innen“ illegale Inhalte über gesonderte Kanäle melden können.

Für Fälle, bei denen die Social Media Plattformen nicht darüber informieren, ob gemeldete Beiträge gelöscht wurden oder nicht, wurde in vorgegebenen Abständen kontrolliert, ob die Beiträge noch online sind oder nicht. Auch wurden am Ende des Monitoring noch einmal abschließend alle Beiträge kontrolliert, ob sie noch abrufbar sind. Außer der Zeit zwischen Meldung und Löschung wurde dokumentiert, ob es eine Rückmeldung vonseiten des IT-Unternehmens gab, Gründe, warum ein Beitrag nicht gelöscht wurde, gegen wen sich das Hassposting richtet, das Datum der Meldung und ob und wann der Beitrag entfernt wurde oder nicht.

Abschließend möchte ZARA – Zivilcourage und Anti-Rassismus-Arbeit den ehrenamtlichen Mitarbeiter\*innen danken, die eine Teilnahme am 3. Monitoring im erforderlichen Umfang erst möglich gemacht haben.