



European
Commission



Code of Conduct on countering illegal hate speech online: One year after

Fact sheet | June 2017

Věra Jourová

Commissioner for Justice,
Consumers and Gender Equality



Directorate-General for
Justice and Consumers



One year after its adoption, the *Code of Conduct on countering illegal hate speech online*⁽¹⁾ has delivered significant progress.

After one year, the results of the second monitoring exercise, which involved a larger sample of organisations located in 24 EU countries, show that **significant progress has been made by the social media platforms participating in the Code of Conduct. They improved both their efficiency and speed in assessing notifications.**

The IT companies have strengthened their **reporting systems** to make it easier for users to report illegal hate speech. They have also **trained their staff** and **increased cooperation with civil society**. By implementing the Code of Conduct, the IT companies have strengthened and widened their networks of trusted flaggers throughout Europe. This is important since the monitoring process has shown that cooperation between IT companies and civil society organisations leads to a **higher quality of notifications, more effective handling times** and **better reactions to notifications**.

Results also show **some improvement in the coherence of treatment** between the notifications coming from trusted reporters/flaggers or from general users.

In terms of **transparency and feed-back provided** to users in response to notifications, the monitoring exercise reveals that **there is still scope for progress**. In this context, the **mid-term review of the Digital Single Market Strategy** confirmed the need to continue working towards **minimum procedural requirements** for the 'notice and action' procedures of online intermediaries. These would include quality criteria for notices, counter-notice procedures, reporting obligations, third-party consultation mechanisms and dispute resolution systems.

Ensuring **efficient cooperation between the IT companies and national authorities** is another important objective of the Code of Conduct. The establishment of **national contact points** in the Member States is progressing.

In this context, the Commission will continue to support the monitoring of the implementation of the Code of Conduct by civil society organisations.

⁽¹⁾ http://ec.europa.eu/justice/fundamental-rights/files/hate_speech_code_of_conduct_en.pdf

Results of the 2nd monitoring exercise of the implementation of the Code of Conduct

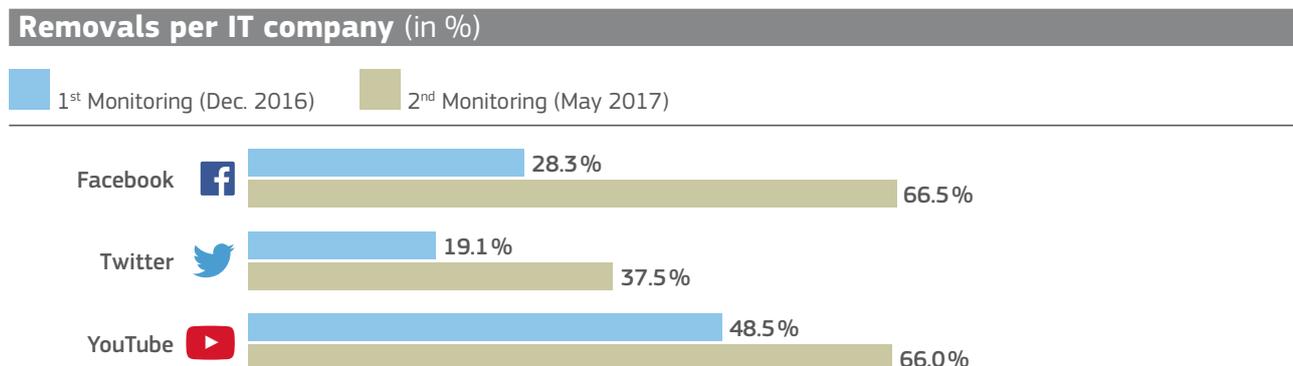
1. Notifications of illegal hate speech

- > In the second monitoring exercise, **2575 notifications were submitted to the IT companies taking part in the Code of Conduct**. This represents a fourfold increase compared to the first monitoring exercise in December 2016.
- > **The geographical coverage of the exercise substantially increased: 31 civil society organisations and 3 national authorities, located in 24 EU countries**, sent notifications relating to hate speech deemed illegal to the IT companies during a period of 7 weeks, (20 March to 5 May 2017). In order to establish trends, this exercise used the same methodology as the first monitoring exercise (see Annex).
- > Out of the total number of notifications, 1830 cases were submitted through the reporting channels available to general users, while 745 cases were submitted through specific channels available only to trusted flaggers/reporters.
- > **Facebook received the largest amount of notifications (1273 cases), followed by YouTube (658 cases) and Twitter (644 cases)**. Microsoft did not receive any notification.
- > In addition to flagging the content to the IT companies, the organisations taking part in the monitoring exercise **submitted 212 of the cases to the police, public prosecutor's bodies or other national authorities**.

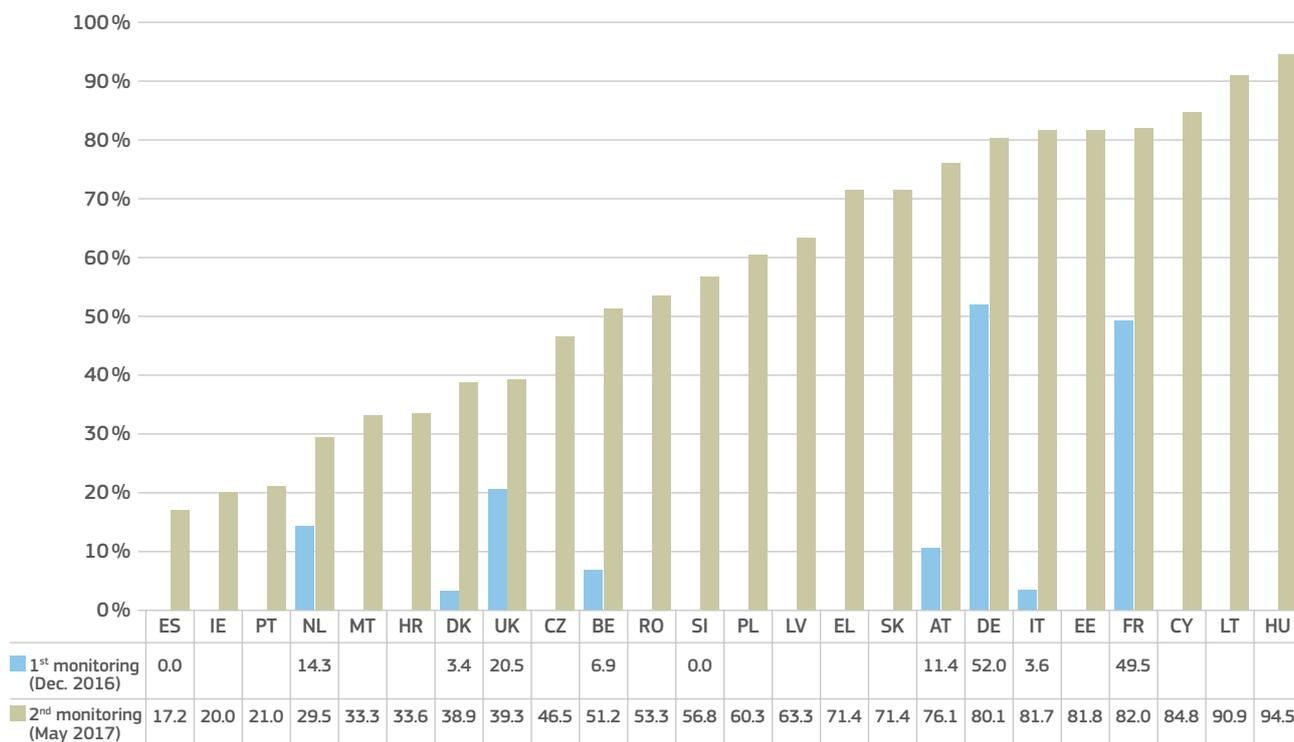
2. Removal rates

Overall, **1 522 of the notifications (59.1%) led to the removal of the notified content, while in 1 053 cases (40.9%) the content remained online**.

Facebook removed the content in 66.5 % of cases, Twitter in 37.4 % and YouTube in 66 % of the cases. This represents a **substantial improvement for all three companies compared to the results presented in December 2016**, where the overall rate was 28.2 %.



Rate of removals per EU country (in %) ^(?)



3. Time of assessment of notifications

- > In **51.4% of cases** IT companies assessed notifications **in less than 24 hours**, in 20.7% in less than 48 hours, in 14.7% in less than a week and in 13.2% it took more than a week.
- > Facebook assessed the notifications in less than 24 hours in 57.9% of the cases and in less than 48 hours 24.9% of cases. The corresponding figures for YouTube are 42.6% and 14.3% and for Twitter 39% and 13.7%, respectively.

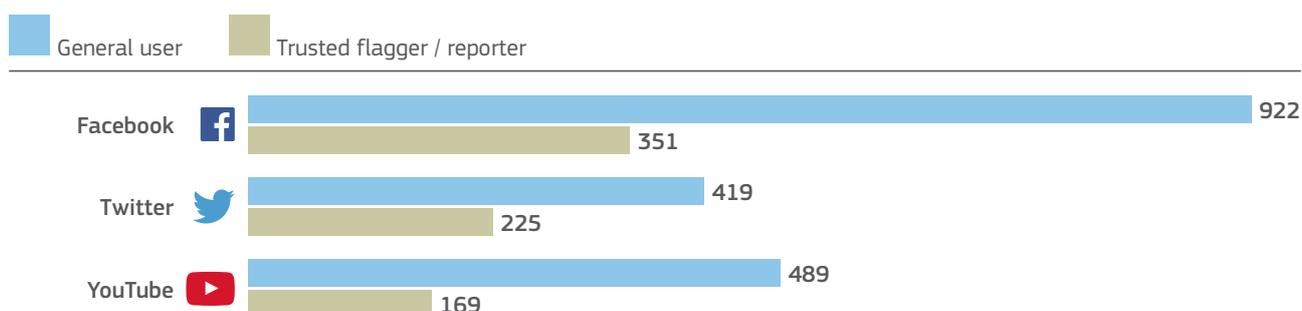
There is a **positive overall trend in the time of assessment compared to the results of the first monitoring exercise in December 2016**. 40% of all responses were received within 24 hours while another 43% arrived after 48 hours.

^(?) The table does not reflect the global issue on illegal hate speech online in a specific country and it is based on the number of notifications sent by each individual organisation.

4. Coherence of treatment of notifications irrespective of the reporting channels

> Out of the total number of notifications, 71.1 % of cases were submitted through the channels available to general users, while 28.9% of the cases were notified through the channels only available to trusted flaggers/reporters.

Amount of notifications received through different reporting channels (number of cases)

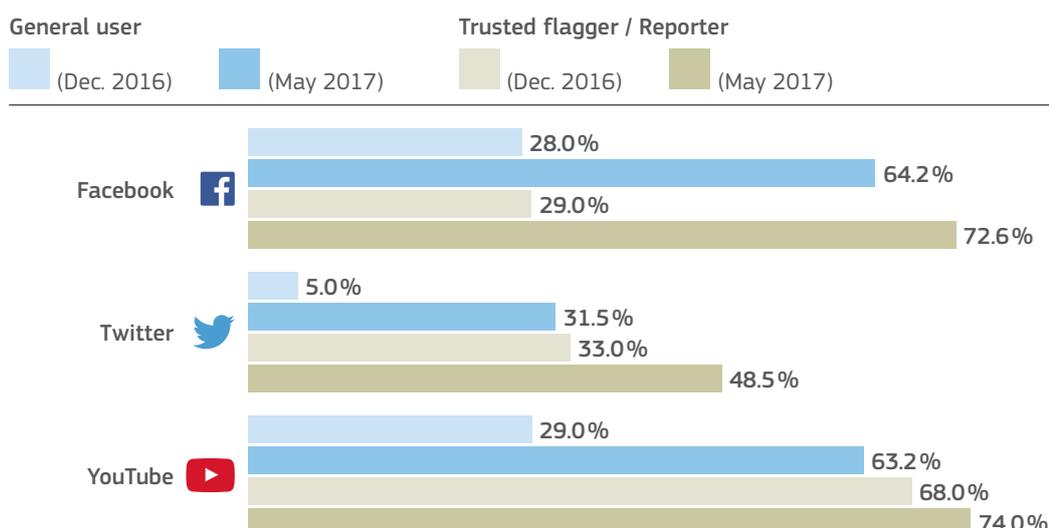


> Breaking down the removals by reporting channel, **56.5% notifications made using channels available to general users** led to the removal of the notified content, while a **higher removal rate of 65.6% was recorded for notifications made using the trusted flaggers/reporters channel.**

> Compared to the first monitoring exercise, **the removal rates between the two reporting channels are converging**, narrowing the gap in **difference of treatment depending on the source of the notification (trusted flaggers or general users).**

> Some differences in treatment still persist.

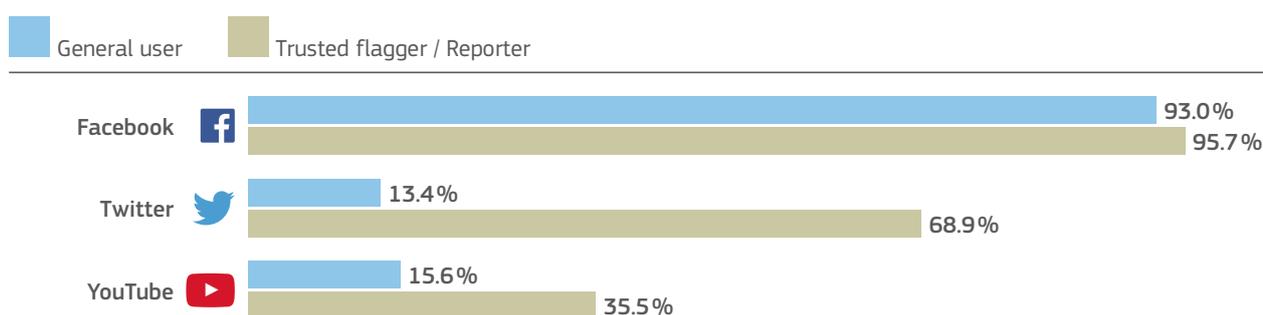
Removal rates according to reporting channel (in %)



5. Feedback to users and transparency

- > **Data shows a large disparity between IT companies when giving feedback to notifications made.** While Facebook sent feedback in 93.7% of the cases, Twitter did so in only 32.8% of cases and YouTube in 20.7% of the cases.
- > **Twitter and YouTube provide more feedback when reporting comes from trusted flaggers:** Twitter provided feedback to 68.9% of notifications made using the trusted flaggers' channel, but only gave feedback to 13.4% of those notifications made by general users. For YouTube the corresponding figures were 35.5% and 15.6% respectively.

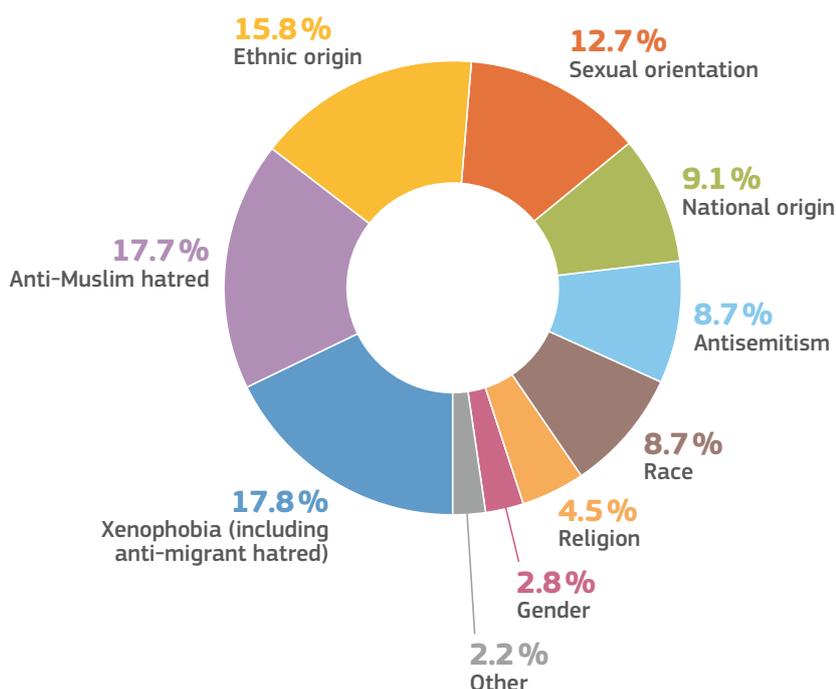
Feedback provided to different types of user (in %)



6. Grounds for reporting hatred

- > Xenophobia (17.8%), which includes anti-migrant hatred, has been reported, together with anti-Muslim hatred (17.7%), as the most recurrent ground of hate speech, followed by ethnic origin (15.8%).
- > The results confirm the predominance of hatred against migrants and refugees.

Notifications per ground of hate speech (in %)



ANNEX

Methodology of the exercise

- The second exercise was carried out for a period of 7 weeks, from 20 March to 5 May 2017, using the same methodology as the first monitoring exercise.
- **31 organisations and 3 public bodies (France, Romania and Spain)** reported a total sample of **2 575 notifications from all the Member States except for Finland, Sweden, Bulgaria and Luxembourg**. An additional 25 cases were reported to other social platforms.
- The organisations only notified the IT companies about **content deemed to be “illegal hate speech” under national laws transposing the EU Council Framework Decision 2008/913/JHA** ⁽³⁾ on combating certain forms and expressions of racism and xenophobia by means of criminal law.
- Notifications were submitted either through reporting channels available to all users, or via dedicated channels only accessible to trusted flaggers/reporters.
- The organisations having the status of trusted reporter/flagger often used the dedicated channels to report content which they previously notified anonymously (using the channels for all users) to check if the outcomes could diverge. Typically, this happened in cases when the IT companies did not send feedback to a first notification and content was kept online.
- The organisations participating in the second monitoring exercise are the following:

BELGIUM		CYPRUS	
CEJI - A Jewish contribution to an inclusive Europe	11 cases	Aequitas	92 cases
Centre interfédéral pour l'égalité des chances (UNIA)	41 cases	LATVIA	
CZECH REPUBLIC		Mozaika	124 cases
In Iustitia	99 cases	Latvian Centre for Human Rights	161 cases
DENMARK		LITHUANIA	
Anmeldhad.dk / Reporthate.dk	72 cases	National LGBT Rights Organisation (LGL)	109 cases
GERMANY		HUNGARY	
jugendschutz.net – 112 cases		Hättér Society	103 cases
Freiwillige Selbstkontrolle Multimedia-Diensteanbieter e.V. (FSM)	53 cases	MALTA	
ESTONIA		Malta LGBTIQ Right Movement (MGRM)	15 cases
Estonian Human Rights Centre	99 cases	NETHERLANDS	
IRELAND		Meldpunt Internet Discriminatie (MiND)	6 cases
ENAR Ireland	35 cases	Magenta Foundation	55 cases
GREECE		AUSTRIA	
SafeLine / Forth	7 cases	Zivilcourage und Anti-Rassismus-Arbeit (ZARA)	142 cases
SPAIN		POLAND	
Movimiento contra la intolerancia (MCI)	59 cases	HejtStop / Projekt: Polska	121 cases
Fundación Secretariado Gitano	51 cases	PORTUGAL	
Federación Estatal de Lesbianas, Gais, Transexuales y Bisexuales (FELGTB)	39 cases	Associação ILGA Portugal	100 cases
Spanish Observatory on Racism and Xenophobia (OBERAXE)	110 cases	ROMANIA	
FRANCE		Active Watch - 93 cases	
Ligue Internationale Contre le Racisme et l'Antisémitisme (LICRA) – 115 cases		Romanian Police	46 cases
Plateforme PHAROS – 31 cases		SLOVENIA	
CROATIA		Spletno oko	81 cases
Centre for Peace Studies	128 cases	SLOVAKIA	
ITALY		eSlovensko	14 cases
Ufficio Nazionale Antidiscriminazioni Razziali (UNAR)	197 cases	UNITED KINGDOM	
		Community Security Trust (CST)	69 cases
		Tell Mama/Faith Matters	10 cases

⁽³⁾ <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:328:0055:0058:en:PDF>