Minutes of the 4th Meeting of the Subgroup on Countering Hate Speech Online - Brussels, 25 October 2017

The Commission – DG Justice and Consumers – convened and chaired the **4th meeting of the Sub group on countering hate speech online** attended by delegates of **26 Member States** (all except Cyprus and Poland), representatives of **IT companies** (Facebook, Twitter and Google) as well as of **civil society organisations (CSOs)**, most of which involved in the monitoring of the Code of conduct**,** and the **European Council**.

Morning session

1. Guidance on tacking illegal content online

The Commission (DG CNECT) presented the Communication on Tackling Illegal Content Online, stressing the importance of the Code of conduct in this process. The **Commission** underlined that the work on the Code of conduct, in particular the results of the monitoring exercises, will feed into the upcoming impact assessment which may lead to further measures as referred to in the final part of the Communication. In addition, the Commission aims to further advance on transparency as regards hate speech online, in particular to increase the quality of feedback provided to notifiers.

The Communication was very positively welcomed by Member States, who appreciated the set out of harmonised standards for notice and action on illegal content as well as, generally, the commitment and the leadership of the Commission in this area. **Member States** also shared some challenges at national level, e.g. in establishing the illegality of the content; on identifying competent jurisdiction; on the delayed responses and on the links between hate speech and fake news; and pointed to the need, in some cases, of reinforced coordination among different Ministries and public bodies in order to establish the single contact point for all illegal content. There were also some questions on how to define trusted flaggers, some Member States flagged importance of certified trusted flaggers (e.g. equality bodies) whose notifications could lead to almost automatic take down.

Civil society organisations (CSOs) signalled the importance of ensuring equal treatment of the notifications from ordinary users and trusted flaggers (INACH, In Iustitia), stressing the need to reinforce the protection, including psychological support, of human rights defenders exposed to attacks (In Iustitia, DigiQ). There was also a general call by CSOs to enhance cooperation with IT Companies and increase the quality of feedback to notices (INACH).

IT Companies welcomed the Communication and expressed a firm intention to keep working with the Commission and stakeholders to foster and reinforce the progress made on the implementation of the Code of conduct. They all underlined the vital **importance of trusted flaggers**, not only for flagging content but also for counter-speech activities. IT Companies agreed on the need to enhance cooperation with CSOs also in relation to possible technological solutions to prevent psychological damages of notifiers repeatedly exposed to violent content.

The **Commission** stressed the importance of improving the response to notices and the quality of the feedback, to both trusted flaggers and ordinary users. It was announced that Google has agreed to host a workshop in Dublin to step up cooperation between CSOs involved in the monitoring of the Code of conduct and IT Companies.

2. State of play and next steps on the implementation of the Code of conduct

The **Commission** informed about the next steps in the implementation of the Code of conduct namely the imminent start of a third monitoring exercise which should bring results by January 2018 and the intention to work with IT Companies on increasing transparency, feedback to users, and proactive

measures to detect illegal content. It was also announced that discussions are ongoing with some additional online platforms with the aim to possibly enlarge the participation to the Code of conduct to new companies. Facebook offered to host an event in Dublin in November for a business-to-business discussion with potential new players.

The IT Companies informed participants of their ongoing activities on the implementation of the Code of conduct:

Facebook informed about reinforced staff, language capacities, and improved procedures and reporting tools to be used by trusted reporters as well as about their work with CSOs on counter-speech, including recurrent training events. Facebook shared global figures about hate speech posts per week removed. For example, in Germany in June 2017, Facebook has reviewed about 15000 notifications of hate speech per week and 3500 posts were removed.

Twitter presented new measures including proactive measures of shutting down abusive accounts, which in turn has proven effective in avoiding repeated abuses. Each day, 1000 accounts are temporally limited or suspended. Twitter internal data show that when an account is temporally suspended the number of the signalled abuses lowered substantially. A detailed justification on the reasons for suspending accounts seem to have helped in reducing reappearing of abusive content once accounts are reactivated. Measures are in place also to ensure appeal against suspension. Twitter also informed of improvements towards increased transparency by means of a publicly available calendar of measures taken and/or in the pipeline to reinforce the safety of the platform.

YouTube informed that since June 2017 actions focused on four key areas: (i) Better detection and faster removal by introducing proactive machine detection combined with human review. This has proven very effective: 83% of the removed video detected by machines proved to be against terms of services and led to removal. (ii): an enlarged and reinforced network of trusted flaggers; (iii) New policy on controversial content: if a content does not infringe policy, but is still considered controversial, YouTube can decide to limit the access, warn the users and limit the functionalities (e.g. block comments or sharing options); (iv) initiatives on counter-speech e.g. automatic search tool which redirect users searching for hateful content to videos promoting positive alternative messages (big investments on innovation on this area). On transparency progress has been made on feedback and reporting channels for trusted flaggers, while for normal users it is in the product roadmap.

There were some questions by **Member States** on the access to evidence after the removal of content for the investigative purpose. They expressed the need to reinforce the role of trusted flaggers outside the monitoring exercises, to focus on education of youth as well as on the protection not only for human right activists but also for law enforcement (Finland mentioned episodes of hateful attacks to police forces).

Afternoon sessions

Session I: National approaches, challenges, and experiences on countering illegal hate speech online (Member States only)

The Commission presented the results of the questionnaire submitted in May 2017 on the judicial cooperation between national authorities and IT Companies. These underline a general lack of data in the Member States mostly due to challenges in terms of competence/jurisdiction, limited disclosure of data; limited power to request removal and length of procedures. A number of Member States underlined that the establishment of national contact points has helped a better cooperation with IT Companies. The Commission (DG Justice) presented recent steps on improving cross-border access to electronic evidences. The measures cover both the practical measures that could help in the cooperation among the judicial authorities and to ensure the existence of a common platform for

sharing the evidences, as well as possible new legislative instruments including a European production order and direct access to evidence already accessible to law enforcement.

The discussion focused also on the role of certified trusted flaggers, some Member States pointed to challenges related to acting in absence of a judicial order, due to a risk of being perceived as an interference by the State into freedom of expression. Others argued that the State has a duty to ensure the swift removal of content to fulfil their positive obligation to protect vulnerable groups against illegal hate speech. Several Member States reiterated that content removal should not be seen as a replacement of judicial proceedings and the risks that when platforms are asked by a notice to remove illegal content, the content is not preserved as evidence for future potential criminal law investigations. In France, the Pharos platform has established a database of reported content. Several Member States indicated that while the establishment of single points of contact has been very positive, cross border request often are lengthy and that in some cases (in particular for illegal hate speech) such access is finally not granted due to double criminality requirements. In terms of needs any procedure that would guarantee the preservation of evidence and providing a faster and more efficient procedure would be welcome. The need to ensure sufficient safeguards for production orders was underlined too.

The **Commission** underlined that victims of hate speech online should be encouraged to report to law enforcement and to complement this by requesting removal to prevent the spread of illegal content. As concerned the scope and role of the national contact points the Commission offered to help as a facilitator should Member States wish to have a technical discussion and an exchange of good practices.

Session II: training and feedback on results of 2nd monitoring and preparation of 3rd round

IT Companies (Facebook, Twitter and YouTube) and CSOs involved in the monitoring gathered to discuss some challenges experienced during the 2nd monitoring exercise, with a view to clarify and avoid similar problems reoccurring during the upcoming 3rd round. The first part of the session focused on general issues (mostly about the lack of feedback by some of the IT Companies). Follow-up discussions with the enlarged network will take place in Dublin in the occasion of the workshop organised by Google. The second part of the meeting was operational with bilateral dialogues between CSOs and the IT Companies to clarify reasons behind lower performance in terms of take down rates, feedback or time of assessment during the 2nd monitoring exercise on the Code of conduct.