

September, 12<sup>th</sup> 2019

## Press Release

### **A silent monitoring by the International Network Against Cyber Hate finds drastic decline in YouTube's removal performance**

Although the overall removal rate of 70,6 % turned out to be only slightly lower compared to the last monitoring (-1,1 percentage points), this result is mostly owed to Facebook's consistently high removal rate of 84,5% (+0,9 percentage points) and Instagram's improvement to 77,2% (+6,6 percentage points). Twitter's performance remained low at 44,9% (+1,4 percentage points), and YouTube only removed 67,8% of illegal hate content, a major drop of 17,6 percent points compared to its last checked performance.

This monitoring exercise, which has been coordinated and conducted by the [International Network Against Cyber Hate \(INACH\)](#) and its partners of the project [sCAN \(Platforms, Experts, Tools: Specialised Cyber Activists Network\)](#) to check the compliance of social media platforms with the European Commission's Code of Conduct on Countering Illegal Hate Speech, was the first one in which the platforms have not been aware of the monitoring.

Between May, 6<sup>th</sup> and June, 21<sup>th</sup>, 12 organizations which are specialized in dealing with online hate have reported 432 cases<sup>1</sup> to the platforms through their public reporting channels, out of which 90 have been re-reported through reporting channels available to organisations recognized by the IT companies as "trusted flaggers"<sup>2</sup> after having been rejected by the platforms.

With the EU Code of Conduct, the companies have agreed to assess and remove illegal hate speech online that is against national law or their Terms of Services within 24 hours. Yet, only Facebook managed to reach a tolerable level in removing reported hate speech within that timeframe (64%). Instagram, Twitter and YouTube remained below 50%.

In addition, the companies' performance in providing feedback was poor: to 42% of reports the companies provided absolutely no feedback, reactions within the required 24 hours came to not even half of reports (46%). Again, only Facebook provided timely feedback to 70% of reports while YouTube remained silent to 97% of reports.

Providing no feedback, late feedback or meaningless feedback is a major issue that needs to be addressed by the companies as soon as possible. If people report online content that is hateful, discriminatory or inciting violence, it is not enough for platforms to send an automated reply stating that they received the report, or not even that. "Users need to know that their efforts in making the internet a friendlier place are taken seriously so they feel encouraged and valued" emphasises Ronald Eissens, General Director of INACH.

---

<sup>1</sup> The cases and numbers given are not representative of the prevalence and types of illegal hate speech online in absolute terms, and are based on the number of notifications and cases submitted by the participating organisations.

<sup>2</sup> A "trusted flagger status" is a special status granted by social media companies to organisations and institutions that have proven expertise in evaluation and classifying online hate. The status provides for direct communications and channels to the companies' respective departments.

Hence, INACH, together with the partners of the sCAN project and other member organizations that have participated in the monitoring, urges the social media companies, especially YouTube, Instagram and Twitter, to improve their removal practices further and to react and respond meaningfully to all users reporting hateful content.

**Press contact:**

**Tamás Berecz**, senior researcher at INACH

Phone: +36-20 326 85 77

E-Mail: [tamas@inach.net](mailto:tamas@inach.net)

**Camille Lhopitault**, project coordinator of project sCAN at LICRA

Phone: +33-1-45 08 08 08

E-Mail: [clhopitault@licra.org](mailto:clhopitault@licra.org)

**Background information:**

**Monitoring of the [EC Code of Conduct on countering illegal hate speech online](#)**

In 2016, the European Commission and the IT companies Facebook, Twitter, YouTube and Microsoft signed the Code of Conduct on Countering Illegal Hate Speech Online. Google+, Instagram, Snapchat and Dailymotion joined the Code of Conduct in 2018. Between 2016 and 2018 there have been four monitoring periods to evaluate the Code of Conduct.

The monitoring has been conducted by civil society organizations all over Europe that are experienced in identifying and combatting hate speech online. Resources, as provided for by the project sCAN, make it possible for NGOs to participate in these exceptional and time consuming activities. You can find more information on the Code of Conduct on countering illegal hate speech online and previous results here: <https://bit.ly/2HilhkM>

**[International Network Against Cyber Hate - INACH](#)**

Founded in 2002 by jugendschutz.net, Germany, and Magenta Foundation in the Netherlands, the International Network Against Cyber Hate (INACH) currently unites 29 organizations from Europe, Israel, Russia, South America and the United States. While starting as a network of online complaints offices, INACH today pursues a multi-dimensional approach of intervention and preventive strategies. The member organisations are united in a systematic fight against cyber hate, for example as complaints offices, monitoring offices or online help desks. In their respective countries, they provide important contacts for politicians, internet providers, educational institutions and users.

Funding for INACH is provided by:



European Union Rights, Equality and Citizenship Programme  
(2014-2020)

Funded by



Federal Ministry for  
Family Affairs, Senior Citizens,  
Women and Youth

as part of the federal programme

Demokratie **leben!**



Federal Agency for  
Civic Education

**Project sCAN:**

Coordinated by the LICRA (International League against Racism and Antisemitism), France, the [sCAN project](#) involves ten different European partners: ZARA – Zivilcourage und Anti-Rassismus-Arbeit, Austria, CEJI - A Jewish contribution to an inclusive Europe, Belgium, Human Rights House Zagreb, Croatia, ROMEA, Czech Republic, Respect Zone, France, jugendschutz.net, Germany, CESIE, Italy, Latvian Centre For Human Rights, Latvia and the University of Ljubljana, Faculty of Social Sciences, Slovenia. The project aims at gathering expertise, tools, methodology and knowledge on cyber hate and developing transnational comprehensive practices for identifying, analysing, reporting and counteracting online hate speech.

Funding for sCAN is provided by:



European Commission - Directorate General for Justice and Consumers, within the framework of the Rights, Equality and Citizenship Programme (2014-2020)

The content of this press release does not reflect the official opinion of the European Union. Responsibility for the information and views expressed lies entirely with the author(s).