# The State of Policy on Cyber Hate in the EU

## Bringing the Online in Line with Human Rights

INACH

2019

Supported by the Rights, Equality and Citizenship (REC) Programme of the European Union

bpb: Bundeszentrale für politische Bildung

## Legal Disclaimer

This publication has been produced with the financial support of the Rights, Equality and Citizenship (REC) Programme of the European Union. The contents of this publication are the sole responsibility of the International Network Against Cyber Hate and can in no way be taken to reflect the views of the European Commission.

# Table of Contents

# 1. Introduction

More and more politicians, policy makers, stakeholders and EU citizens realise nowadays that the major observable shifts in the political sphere within the European Union is, at least partially, the product of social media. Social media companies made it possible for politicians, parties, but even extremist organisations to target voters on a level that has been unheard of before. These targeting possibilities have brought forth an almost completely novel type of political messaging on the continent, which does not care for facts, does not care for truth, maybe only in a postmodern sense of the word, and does not really care for the common good. These changes mean that propaganda and so-called fake news, are flourishing in the online public sphere. The aim of this propaganda is to rile up people's most base instincts and feelings and then reap the political rewards.

The sliding of the political spectrum towards the right is not a new phenomenon in the EU, but it has been hastened in the past couple of years by technical developments. It has also become increasingly extreme, where so-called populist parties normalise and mainstream views that had been almost taboos since the 1940s. It cannot be ignored anymore that the right-wing parties of Europe are slowly becoming ever more fascistic in their tone, their messaging and their policies. One does not even have to think hard anymore to come up with examples of mainstream right-wing parties that are in government in different EU countries that are spewing views based on racism, xenophobia, Islamophobia and often even homophobia. Fidesz in Hungary, PiS in Poland, the FPÖ in Austria, the Lega Nord in Italy and the examples could be continued, showing that this is not even just an Eastern European issue (anymore).

There is another major issue caused by fake news and rampant online propaganda, the breakdown of people's trust in the mainstream media and the breakdown of the concept of truth and facts, things that were taken for granted by everybody for centuries. The deluge of misinformation and the breakdown in what societies mean by fact and truth could arguably lead to the total breakdown of Western-style democracy as we know it today.

Cyber hate, in other words hate speech that is being spread online, is intimately linked to propaganda and fake news. Far-right parties and extremists love attacking minority communities by using doctored pictures, spun stories, half-truths and bold-faced lies to gain voters and power. These all count as cyber hate and most of it constitute illegal hate speech. That is why both the EU and most member states have laws against hate speech (at least offline hate speech) and that is why multiple steps have been taken by the EU and some member states in the past couple of years to tackle these two interlinked issues.

INACH has already written a paper in 2017 on the policies of the EU and some of its members that try to tackle online hate speech. Its contents will be summarised in the first part of this paper. Not much has changed since, legally. However, some of the laws and legal suggestions that we examined in our previous paper were fairly new back then, thus there had not been enough time for them to have a substantial impact on these phenomena. Hence, in this paper, we will be looking at the same pieces of legislation, legal documents and policy suggestions as in our previous paper and see what impact they have made in the past two years. And just like in our previous paper, we will provide policy recommendations based on our findings, whilst examining how many of our previous recommendations became actual policy since 2017.

## 2. Short summary of our previous policy paper and reiterating our previous policy recommendations

As mentioned above, INACH has given policy recommendations before, so it is important to quickly summarize what was mentioned already, before diving deep into what has changed since. When looking at those previous recommendations, there were three main areas of interest; the EU level, the national level and Social Media companies.

Regarding what the EU was advised to focus on, there were four main goals; the first one was that a definition of hate speech needed to be introduced, as such a universal definition was yet to be created. This lack of clarity as to what hate speech pertains makes the fight against it ever more tedious. Finding some harmony in that department would be a steppingstone. Secondly, we advised that the Code of Conduct on Countering Illegal Hate Speech Online (CoC) should be developed further. Thirdly, we recommended that the Communication 'Tackling Illegal Content Online' published by the EC should be respected, enforced and abided by when fighting against cyber hate on an EU level. Lastly, regarding the monitoring exercise, a less biased approach was hoped for, as details as to when and who would be conducting the exercise was shared prior to the start of the exercise, rending the results somewhat flawed.

In regard to changes on a national level, the new German law regarding social media, the Act to Improve Enforcement of the Law in Social Networks (NEA) should be treated as an example. More so, increased effort was expected in terms of educational and awareness raising tools, with a focus on the youth and authority figures such as the police force. Tools such as campaigns, trainings, modules and so on are key here. More than only nationally, support for such initiatives should also be done at the level of the EC, and of the EU as a whole.

Last of all, changes were needed within social media companies. Those companies should have resolved one of their biggest issues, being the discrepancies between what is being removed and what is not, proof of inconsistencies in their guidelines. Moreover, another issue was that removal rates were influenced by the amount of complaints received, and by who the complainer was, which should obviously not be the case. We had therefore advised to work on harmonizing, detailing and clarifying the companies' content guidelines.

# 3. Legal environment

The three main legal documents mentioned above will now be summarized and their impacts in each department will be analysed. The first one that was mentioned above is the CoC. This code, which was created as an outcome of the EU Internet Forum, was meant to remedy the fear and hatred that were materializing by terrorist attacks as well as the increasing use of social media to radicalize the public. It allows those who combat hate speech to inform the IT companies about the illegal content whilst expecting a greater reaction and action from those companies. In May 2016, the Commission made an agreement with Facebook, Microsoft, Twitter and YouTube that they should abide by this Code. Its main tools for evaluating if those companies do abide by the Code are monitoring exercises. More in depth discussion about those exercises will be done in the following chapter.

Regarding outcomes of this Code, in 2018, Instagram, Google+, Snapchat and Dailymotion joined the CoC. More recently, Jeuxvideo.com also joined in January 2019, which makes the coverage of the EU market share of online platforms possibly affected by hateful content to be of 96%. Regarding actual results, the EC states on their website[1] that the companies are now assessing 89% of flagged content within 24 hours and 72% of the content deemed illegal hate speech is removed. They also state that companies have increased the number of employees monitoring potential illegal content, Facebook reporting itself that it has a network of about 15,000 people working on content review. The IT companies also state that they are involved with trainings and coaching for those content reviewers. Moreover, the number of 'trusted flaggers' in Europe is increasing. The EC states that "in the first year after the signature of the Code of conduct, Facebook reported to have taken 66 EU NGOs on board as trusted flaggers; and Twitter 40 NGOs in 21 EU countries. Out of a total of 38 training sessions provided in 2018 by YouTube to NGOs on their content policy and trusted flagger program, 18 were focused on hate speech and abusive content"[2]. Companies also established national points of contact which enhanced communication with relevant authorities at a national level. Transparency reports have also been published by Facebook, Twitter and YouTube leading to better and clearer information sharing. And lastly, regarding the communication between the complainant and the companies, the EC states that around two thirds of the notification get a systematic response regarding the next steps that will follow. However, as we will dive deeper into the advances social media companies have made, we will see that there is far from solely good news in that department.

The second legal document that should be analysed here is the EC's Communication. The pros and cons of this Communication were examined in depth in our previous Strategy Paper[3]. This Communication embodies the fact that the Commission felt it was necessary to give policy suggestions on an EU level regarding tackling cyber hate. Even though it is not a law per se and only recommendations, it was still a big step in the right direction and showed the issue is a top priority. This Communication became a part of the proposal for a Regulation preventing the

[1] https://ec.europa.eu/info/sites/info/files/hatespeech_infographic3_web.pdf

[2] https://ec.europa.eu/info/sites/info/files/hatespeech_infographic3_web.pdf

[3] http://www.inach.net/wp-content/uploads/kick-off-final-FINAL-version3-FINAL.pdf

dissemination of terrorist content online[4] from 2018, and should also play a key role in the upcoming “Digital Services Act”, a review of the E-Commerce Directive, which will represent legislations that will force social media and tech companies to remove illegal content under the threat of sanctions[5].

Lastly, the new German law will be examined. What was so exciting about this law was that it finally, like never before, forced companies to adhere to the law, with, amongst other things, the existence of fines, whilst at the same time offering access to a clear set of rules concerning the handling of reports. Regarding examples of the law being enforced, Facebook was fined 2.000.000€ because the company created two reporting channels, one for general reporting and one for NEA reporting; In their transparency report, Facebook stated that they received 1704 reports through the NEA reporting channel (compared to around 215.000 reports to YouTube and 260.000 reports to Twitter).[6] The Federal Office of Justice stated that many reports about content related to the NEA had been done through the general reporting channel, being more well known, leading the number of reports to probably being much higher. According to the Office, Facebook therefore, did not provide all necessary information in their transparency report. Facebook has appealed the fine[7]. Regarding the amount of complaints not removed by social media companies, according to the newspaper Handelsblatt[8], the number of said complaints have decreased. In 2019 (until the beginning of August) the Federal Office of Justice received 383 complaints, vs the 714 complaints in 2018.

However, the law is still controversial, and worries of over-removal, leading to a threat to freedom of speech persists. Examples of those threats materializing was the case of far-right politician Beatrix von Storch whose Tweet was removed and Twitter account temporarily suspended for criticizing the Cologne police force for tweeting in Arabic “to appease the barbaric, Muslim, rapist hordes of men”[9]. And this was not the only example. A review of the law will be made after a three-year operation period, which will give better insight regarding the law’s efficiency.

[4] https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-preventing-terrorist-content-online-regulation-640_en.pdf

[5] http://copyrightblog.kluweriplaw.com/2019/06/17/the-new-copyright-directive-a-tour-dhorizon-part-ii-of-press-publishers-upload-filters-and-the-real-value-gap/

[6] https://www.zdf.de/nachrichten/heute/facebook-soll-zwei-millionen-euro-strafe-zahlen-wegen-verstoss-gegen-netzdg-100.html

[7] https://www.heise.de/newsticker/meldung/Facebook-wehrt-sich-gegen-NetzDG-Bussgeld-4475699.html

[8] https://www.handelsblatt.com/politik/deutschland/netzdg-kaum-beschwerden-ueber-mangelhafte-loeschung-durch-soziale-netzwerke/24909042.html?ticket=ST-32826699-OWGygiIlELjU1dUM7QV0-ap1

[9] Joseph Nasr, Beatrix von Storch: German Police Accuse AFD Politician of Hate Incitement over Anti-Muslim Tweet, The Independent (02.01.2018), https://www.independent.co.uk/news/world/europe/beatrix-von-storchgermany-afd-anti-muslim-twitter-north-rhine-westphalia-new-years-eve-a8138086.html last accessed 14.12.2018

# 4 Monitoring Exercises

Not much has changed in the past two years regarding the monitoring exercises (ME) run by the European Commission (EC). Therefore, we cannot say much more than what we stated in our previous policy paper:

"*The monitoring exercise represents a major steppingstone for INACH and its fight against cyber hate, and the way in which it took place, including its results, success and limitations, were described in detail in our Annual Report. The monitoring exercise was an essential tool and should, without a doubt, be kept as a useful device in the future. Nevertheless, however successful the exercises have been, a few improvements could be made here and there for the future ones.*"[10]

Now, some improvements have been made. The EC decided to make the start of the ME confidential. In other words, social media companies were not aware of when the ME in 2018 started exactly. This was definitely a good step forward, and one that the participating organisations were requesting for quite some time in order to lower the possibility of skewed results, since, by knowing when the ME was starting exactly and how long it was running, social media platforms were able to focus more on reports on hate speech, especially from their trusted flaggers who were participating in the ME. NGOs feared that this approach was producing outcomes that were far to rosy for their experiences with hate speech reporting and removal on these platforms outside of the MEs' confines.

On the other hand, companies were still fairly aware of when the ME would be happening in 2018. They did not know the exact date, but they provided training to the NGOs on the changes in their reporting systems before the ME. Furthermore, we can assume that companies have the means to realise a sudden rise in reports within a six-week-long period from their trusted flaggers in a concerted manner in more than a dozen EU countries. Thus, even if they did not know the exact starting date, they most likely figured out that the ME was ongoing within a couple of days or maybe a week.

Hence, it would be highly beneficial to shroud the start and the run time of the MEs even further if possible. Maybe by doing it in a rolling manner, i.e. not all NGOs doing it at the same time but spread it out to a three-month-long period with different NGOs stepping in at different times. Obviously, this would mean a lot more work and organisation on the EC's side, but it would not mean more work for the NGOs.

Otherwise, INACH and its members are fairly content with the rest of the ME's methodology and we think that these exercises are an amazing tool in holding the companies to their word and check their adherence to the Code of Conduct.

One major development in the history of the MEs since 2017 is the independent unannounced (silent) monitoring exercise that was run by INACH and the sCAN project (a project solely run by

[10] http://test.inachbase.net/wp-content/uploads/Policy_Recommendations_to_Combat_Cyber_Hate.pdf

INACH members) jointly during the summer of 2019. This ME was kept a complete secret from social media companies, and it was run independently from the Commission. INACH and its participating members were extremely interested in seeing whether there would be a significant difference between the outcomes of the previous official MEs and this independent one. Naturally, there were some differences. This is what we wrote about them and the outcome of our independent ME:

"*Although the overall removal rate of 70,6 % turned out to be only slightly lower compared to the last* [official, EC-run] *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. [...]*

*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.*"[11]

As we observed, the biggest notable difference between the last official ME at the end of 2018 and our own ME was the decline in feedback from social media, a major issue both for users and NGOs that try to clean up the internet from hate and discrimination. (For the importance of substantial feedback and a more detailed description of the outcomes of our ME, please read our full monitoring report HERE).

Thus, we can argue that there is definitely room for improvement for most companies when it comes to providing timely and substantial feedback and the removal of cyber hate from their platforms. We can also argue that holding independent, unannounced MEs is useful and necessary to show that the official exercises - although they are very useful and paramount - might be a bit skewed by involving the companies too heavily in the organisation and preparation.

[11] http://www.inach.net/wp-content/uploads/INACH_sCAN_ME_press_release_12092019_fin-002.pdf

# 5 Social Media (and Transparency)

In 2017 INACH published a paper on how to improve reporting instances of cyber hate to social media.[12] In that paper we firstly argued that transparency should be a priority of social media when dealing with reports on cyber hate, and that they should simplify their reporting systems whilst also bridging the gap between the experiences of normal user reporters and trusted flaggers. Not much has changed since then. Social media companies have become a little bit more transparent towards NGOs on why they remove certain things and not others. The first big step in the right direction in this field happened during a workshop in Dublin in 2017. The event was organised by the Commission and the companies to nurture trust and better understanding between NGOs participating in the MEs and social media. During this training, all major companies gave presentations (with real life examples) on their policies of removal and non-removal. It was a good first step, but far off from being enough. NGOs and other stakeholders combating cyber hate still do not know whom normal users report to, how many moderators the companies have in different countries, who trains them, how they do their work, and the removal records of the companies are still fairly contradictory and often confusing. It has to be stated here that Facebook has made huge strides in removing hateful content (although they still struggle with keeping to the 24-hour timeframe), but YouTube and Twitter are still abysmal in their removals. The two latter companies also started to prefer geo-blocking rather than removing content, which is only a half-measure, especially if the content is in English or other widely spoken languages.

By now, we have some ideas about what it is like to be a Facebook moderator, at least in the United States, and the picture is not pretty.[13] [14] Based on the articles published by Verge this year, Facebook moderators are undertrained, overworked and they do not receive sufficient psychological support to deal with the violence, gore and hate they have to deal with, thus they often develop mental issues, such as PTSD and other forms of stress related traumas. A work environment like this most definitely does not help these people to do their jobs as best as they could otherwise, leading to rushed and bad decisions. It can also desensitise the moderators to hateful content, which could lead to a rise in their thresholds, leading to non-removal in cases that might seem less serious compared to more violent or disturbing content they also have to moderate. It has to be mentioned here that nobody really has any idea about the work environment of social media operators in EU countries. Nobody really knows how many moderators these companies have in different language areas either. Which is also an issue of transparency.

The other stubborn problem that has to be examined is the discrepancy between content removal when the content was reported as a normal user and when it was reported as a trusted flagger.

[12] http://test.inachbase.net/wp-content/uploads/HOW_TO_IMPROVE_REPORTING_INSTANCES_RELATED_TO_CYBER-HATE_.pdf

[13] https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona

[14] https://www.theverge.com/2019/6/19/18681845/facebook-moderator-interviews-video-trauma-ptsd-cognizant-tampa

(Social media companies have trusted flaggers. These are usually NGOs, who regularly report to them. They provide these NGOs with email addresses via which they can contact higher level moderators directly. Also, since they are considered as being experts by social media, their reports are taken more seriously). Since the beginning of the MEs, NGOs have been reporting cases as normal users and as trusted flaggers. This method highlighted the issue that quite often social media companies do not remove certain content when it is reported as a normal user, but they remove it when the same content is reported via a trusted flagger channel. Our latest ME found that Facebook removed 62.50% of cases that were escalated through trusted flagger channels, YouTube removed 26.09%, Twitter 43.9% and Instagram removed 50%. Now, one has to keep in mind that these were all cases that were NOT removed when reported via the platforms' regular reporting systems available for average users, yet they were removed when reported as trusted flagger. This discrepancy should not exist, or at least not to this extent. Normal user reports should be taken just as seriously or at least almost as seriously, and their removal rates should be much closer to trusted flagger removal rates.

Furthermore, the communication and feedback coming from the companies is highly lacking. Based on our latest ME, the companies do not provide any kind of feedback to reports in 41% of cases and very often the feedback they provide is not substantial. This approach does not promote engagement with reporters and disparages users from reporting to the companies on a regular basis. Also, companies arguably never explain their decision to remove or not remove certain content to the users that reported them. This makes their decisions often seem arbitrary and illogical, which also disparages reporting. Moreover, there are major differences in removal rates between countries and in what content is being removed and what is not. This shows that moderation is not centralised enough and that the trainings these moderators receive are insufficient. It also shows that companies let their employees be affected by local politics and the sensitivities of their countries' cultures. This should not happen on global platforms, especially not within the EU, where the legal environment and background is quite unified and uniform.

# 6 INACH’s Policy Recommendations

Before we jump into our recommendations, we have to examine the suggestions we made in 2017 and whether they were implemented, at least to a certain extent, or not.
Two years ago, we made the following policy recommendations to the EC and social media:

1. **Social media companies should find a solution to the problem of the discrepancies between what is being removed and what is not, by working on harmonizing, detailing and clarifying their content guidelines.**

This, sadly, has only been implemented partially. Facebook has made huge strides in removing hateful content and they managed to lower the discrepancy levels of their removals. However, other social media companies are very much behind and lacking in removing cyber hate. Furthermore, there are still often fundamental differences in all platforms what is being removed in different countries within the EU. Almost identical contents can be removed in one country and left online in another or only geo-blocked or otherwise restricted. Thus, we have to argue that this recommendation has only been implemented partially, and hence, we have to maintain it.

2. **On an EU level, work should be done to attain a more harmonized definition of hate speech, changes should be made to make the monitoring exercise less biased, and the code of conduct could be developed further.**

INACH and its members made tremendous strides in developing a uniform definition of hate speech. Most partners accept this definition, yet, its application is not that straightforward. Countries differ in their hate speech laws and our members have to take these differences into account. Still, we can say that our members subscribe to a unified definition of hate speech and they do their best to adhere to this definition as much as possible during their daily work and the MEs.

On the other side of this spectrum, EU member states are far from having a unified definition of hate speech. Even though their laws are based on EU directives, there are a lot of differences in what counts as illegal speech and what not. As far as the MEs go, some changes have been made to make them less skewed (we already mentioned these above), but there is definitely room for improvement. Furthermore, the code of conduct has not been changed since it was signed three years ago.

Therefore, just like the previous recommendation, this second one was also implemented (very) partially and has to be maintained.

3. **Social media’s adherence to the Code of Conduct should be kept in check through continuous monitoring exercises. The methodology of these exercises should be fine-tuned to mitigate bias.**

This policy recommendation was implemented by the EC thankfully. It seems that the Commission is committed to organising MEs on a yearly basis. Hopefully this will not change in the future, since there is a lot of room for improvement in hate speech removal, especially when it comes to Twitter and YouTube. The methodology of the MEs has only been changed slightly to make social media less aware of when they take place, yet there should be more done to shroud the ongoing MEs from the companies, and thus, further mitigate skewing the outcome.

Hence, we can say that this recommendation was taken to heart by the stakeholders and it was almost fully implemented. Yet, we have to maintain it as a recommendation, since it is absolutely paramount for the MEs to continue and their methodology to be further developed for even more positive outcomes in the coming years.

4. **The Communication published by the EC should be the minimum standard in the fight against cyber hate on an EU level.**

There is not much to say about this recommendation, but that INACH maintains this opinion. The Communication should be the very minimum standard in the fight against cyber hate.

5. **The EU should consider tougher approaches to policing illegal online content if the CoC and the Communication do not reach the intended goals in the coming years.**

INACH also maintains this recommendation. It seems that certain social media companies were more open and receptive to criticisms and changed their approach to cyber hate significantly in the past three years. Yet, other companies are still providing safe havens to extremists and trolls. It also seems that the CoC has less to do with certain companies paying more attention to cleaning up their houses than scandals. Facebook weathered multiple major scandals about privacy breaches and having too much influence on democratic elections by allowing fake news factories to target voters in a surgical way. Thus, they needed to do something to mitigate the fallout of these scandals and steer public opinion. Other companies were less under the microscope and had fewer or smaller scandals. Therefore, they have not felt the need to change their policies that much. This shows that the CoC and the MEs are useful, but arguably insufficient tools in getting companies to fight cyber hate vigorously. Therefore, INACH maintains that the EU should consider tougher approaches, i.e. fines for companies that do not adhere to the CoC.

6. **On a National level, the German law should be taken as an example in general terms, including the necessary development regarding its missing regulations on the deletion of legal content.**

INACH also maintains this recommendation, especially in the light of the issues described beforehand. The NEA should be taken as an example of how social media could be regulated in the future within the EU. Although its shortcomings should be mended first of course.

7. **More should be done in educating the public (hence the potential complainants), with a focus on younger people and authorities in charge of helping those complainants, such as the police.**

This recommendation still stays true and probably will stay true for the coming years. INACH cannot know how much work is being done in all EU countries to educate young people and authorities. However, the INACH Secretariat launched its counter-speech trainings in 2019. These trainings educate law enforcement agents and young people in how to recognise hate speech and discrimination and how to counter it online. We will continue this work in the coming years to spread awareness and to make the online more in line with human rights.

Having discussed our previous policy recommendations, it is time to add a few new ones:

8. **Since the start of the monitoring exercises, and even beforehand to become trusted flaggers, NGOs have received a lot of training and instructions from social media companies on how to use their reporting systems and lately on what they remove and what not. These exchanges of knowledge were useful, however, INACH thinks that there is a very strong argument for NGOs to train the moderators of social media. People who work for our members are experts in their fields and they have tremendous knowledge on hate speech and the laws that regulate it on national and EU levels. Yet, social media companies never allow them to train their moderators or to have discussions with them. Therefore, as we mentioned in a previous chapter, we can have no idea of who trains the moderators, based on what, what is the quality of that training, is it focused also on local hate speech laws or just the ToS of the companies, etc. Thus, we strongly recommend at least a pilot run of trainings for social media moderators given by NGO experts.**

9. **As a network of NGOs, INACH also has to be introspective and provide recommendations for civil society organisations. First of all, NGOs that combat cyber hate should move away from the cyber nanny approach. We should tell people less what not to do and more what to do. Switching to a more educational approach would lessen the likelihood of alienating our audience and would probably provide for a more fruitful discussion.**

10. **And as a final recommendation, also mainly for NGOs, building bridges helps. Thus, we recommend (and will try to implement ourselves) a more conscious approach to building rapport with the public, become more well-known and communicate in a way that is more consumer friendly and less argumentative. Combining our final two recommendations will hopefully lead to a more fruitful relationship with the public and in return to an online public sphere that is much more civil and acts as an engine for public debate, not as a platform for online shouting matches that often spew racism and other forms of hate.**

So, to summarise, these are INACH’s policy recommendations:

1. **Social media companies should find a solution to the problem of the discrepancies between what is being removed and what is not, by working on harmonizing, detailing and clarifying their content guidelines.**

2. **On an EU level, work should be done to attain a more harmonized definition of hate speech, changes should be made to make the monitoring exercise less biased, and the code of conduct could be developed further.**

3. **Social media’s adherence to the Code of Conduct should be kept in check through continuous monitoring exercises. The methodology of these exercises should be fine-tuned to mitigate bias.**

4. **The Communication published by the EC should be the minimum standard in the fight against cyber hate on an EU level.**

5. **The EU should consider tougher approaches to policing illegal online content if the CoC and the Communication do not reach the intended goals in the coming years.**

6. **On a National level, the German law should be taken as an example in general terms, including the necessary development regarding its missing regulations on the deletion of legal content.**

7. **More should be done in educating the public (hence the potential complainants), with a focus on younger people, the elderly and authorities in charge of helping those complainants, such as the police.**

8. **Social media companies should ask NGOs to train their moderators on hate speech and on the laws that regulate illegal speech in different EU countries.**

9. **NGOs should move away from the cyber nanny approach and gear their work more towards education, counter-speech and prevention.**

10. **NGOs should put a larger emphasis on building a relationship with the public, become better known and build an image that is easier to digest.**