



**Strategic Paper –
Policy
Recommendations to
Combat Cyber Hate**

By Tamás Berecz and Charlotte Devinat

**Project Research - Report -
Remove: Countering Cyber Hate
Phenomena**

INACH



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

Executive Foreword

This publication was written within the framework of the ***Research – Report – Remove: Countering Cyber Hate Phenomena*** project of the International Network Against Cyber Hate (INACH); funded by the European Commission Directorate-General for Justice and Consumers. The duration of the project is 2016-2017, and its aim is to study, document and report on online hate speech in a comparative and comprehensive way; and to establish structures for a transnational complaints system for instances of cyber hate.

Hate speech is intentional or unintentional public discriminatory and/or defamatory statements; intentional incitement to hatred and/or violence and/or segregation based on a person's or a group's real or perceived race, ethnicity, language, nationality, skin colour, religious beliefs or lack thereof, gender, gender identity, sex, sexual orientation, political beliefs, social status, property, birth, age, mental health, disability, disease.

This report was completed with the participation of the different members of the Network and partners in the project, namely the Zivilcourage und Anti-Rassismus-Arbeit (ZARA) from **Austria**, the Movimiento contra la Intolerancia (MCI) from **Spain**, jugendschutz.net from **Germany**, the Ligue Internationale Contre le Racisme et l'Antisémitisme (LICRA) from **France**, the Inter-Federal Centre For Equal Opportunities and Opposition to Racism from **Belgium** (now called Unia), and the Magenta Foundation from the **Netherlands** (MDI); who provided most of the data this report is based upon.

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 content

I.	Introduction	-	-	-	-	-	-	-	-	-	p. 3
II.	International level recommendations	-	-	-	-	-	-	-	-	-	p. 4
	1. Focus on Social Media companies and the Monitoring Exercise	-	-	-	-	-	-	-	-	-	p. 4
	2. The future of the Code of Conduct	-	-	-	-	-	-	-	-	-	p. 6
	3. Analysis and critique of the new Communication from the European Commission ‘Tackling Illegal Content Online’	-	-	-	-	-	-	-	-	-	p. 7
III.	National level recommendations	-	-	-	-	-	-	-	-	-	p. 11
	1. Suggestion for further harmonisation with a focus on the German law	-	-	-	-	-	-	-	-	-	p. 11
	2. Counter strategies on an educational level	-	-	-	-	-	-	-	-	-	p. 13
VI.	Conclusion	-	-	-	-	-	-	-	-	-	p. 14
VII.	Bibliography	-	-	-	-	-	-	-	-	-	p. 16

I. Introduction

The internet, being one of the most innovative discoveries of the 20th century, is multi-faceted. It is a stimulating and thrilling medium that people use to connect with each other all over the world, enabling the sharing of information to be close to unlimited. Nevertheless, by consequence, it bears a great complication. There is close to no limitation as to what information people choose to share online, for better and, unfortunately, for worse. Namely, cyber hate stems from this unlimited access and appears as one of the most destructive aspect of the whole issue. As people feel safe behind their screens that they use as shields, their boundaries and restraint disappear, leading online hate speech to become a daily occurrence for internet users. Furthermore, this shield also gives a sense of protection for those whose unique purpose is to spread their hateful propaganda.

Consequently, the most dangerous angle, which some people tend to be in denial about, is that online also hurts offline. The screen gives a false idea that since all the discourse is taking place in the virtual world, it stays within that realm, but that is, of course, never the case. Hate in the virtual world entices hate in the real world, arguably leading to discrimination and, thereupon, hate crimes. It is therefore undeniably paramount to find a solution to this growing problem that is putting us all in harm's way.

INACH has been working for two years on its Project "Research - Report - Remove: Countering Cyber Hate Phenomena", meticulously collecting data on cyber hate since the middle of 2016, as well as monitoring cyber hate and trying to get instances of illegal and/or harmful hate speech removed from online spaces, especially social media. This enables us to have a clear picture of what the situation looks like, and where the focus should be in order to ameliorate it.

Arguably the biggest issue is that, as mentioned in our ["Annual Report 2016-2017"](#) and our article ["Legislation related to cyber hate"](#), there is no universal definition of hate speech. This lack of harmonization illustrates the current approach towards cyber hate. This disparity can be seen, amongst other things, through the example of the way in which Holocaust denial is treated differently within European Countries. For example, there are clear provisions made in France, Spain, Austria, Belgium and Germany that make Holocaust denial unequivocally illegal, but in the Netherlands there is no specific and explicit law on that (all though the Supreme Court ruling of 1995 stated that Article 137 of the Criminal Code, which makes illegal the defamation of groups based on their religion or ethnicity, also bans Holocaust denial). In addition, not every country in Europe even has a clear definition of hate speech in their criminal code, namely the Netherlands and Croatia.

This specific issue came to light as throughout our two years of monitoring and data analysis a massive discrepancy arose between what material was considered legal and illegal. A very large amount of material considered hate speech by users and by ourselves was not taken down and not

considered illegal by social media companies. This lack of consensus is an undeniable obstacle which gives social media companies too much leeway to interpret their own rules and guidelines subjectively and arbitrarily. This is a real constraint for NGOs like us and it dampens our ability to bring the online in line with human rights.

Nonetheless, transnational institutions and governments have produced a number of treaties and legislation to remedy the issue to some extent, a list of those were published on our website in an article titled [“Legislation related to cyber hate”](#). In this article, a compilation of UN Declarations, Conventions and Treaties, OSCE Decision (No. 9/09), and Council of Europe's legislation related to the subject is listed as well as multiple European and National laws.

All in all, this does show that, even though there is a lack of international consensus regarding the legal approaches to cyber hate, there is nonetheless a clear consensus that cyber hate is a serious matter that needs remedy. Civil society organisations and public institutions are willing to work towards the same goal in order to reach the balance between protecting the public from online discrimination and defending the freedom of speech. A possible harmonization would definitely lead the path to even more positive change.

In this paper we will go in depth in the understanding of specific issues that will be looked at one by one in detail along with clear recommendations of what should be done as a remedy. We will discuss possible policy changes and advancements regarding more specific counter strategies on a political, legal and educational level. The issue will be addressed on an international, i.e. EU level as well as focusing on social media companies and on a national level, to provide an all-encompassing view on how INACH imagines tackling the issue of cyber hate in the coming years.

II. International level recommendations

1. Focus on Social Media companies and the monitoring exercise

We will first take a look at the international recommendations regarding social media companies and EU policies. In September 2017, the European Commission produced an extremely useful Communication titled “Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions - Tackling Illegal Content Online - Towards an enhanced responsibility of online platforms”. The Commission listed numerous ways in which social media could enhance their effectiveness in the fight against cyber hate, and many good points were made. Those points will be discussed in part 3 of this chapter. In any case, to start with, we want to add a few observations to that already long list that we have made through our data analysis and monitoring work during the past two years.

As it was briefly mentioned above, the lack of harmonization in a definition of hate speech leads to a worrying issue for social media companies, namely, the discrepancy between what is deemed illegal and not and what is being deleted and what is not. There was a lot of inconsistencies between what was being removed and what was not, which denotes an issue coming from the content guidelines of the social media platforms which seem inconsistent themselves.

Similarly, another issue that was previously mentioned in our Annual Report as well, is that removal rates were greatly influenced by the amount of complaints received, and by who the complainer was. If it was an authority or a very well established local NGO, or other civil society organization that was a trusted reporter or flagger, it was much more likely that the hateful content was removed, as it was the case for the monitoring exercise discussed above; just like when a lot of people complained about a certain content.

Consequently, some modification should be made in regards to this, as illegal content and content that violates the guidelines should be removed globally and universally, irrespectively of the number of complainers or who the flagger is. Otherwise, just like the issue of a lack of harmonization, this will additionally hinder our goal of a safer and cleaner internet, which should promote respect for human rights. This is especially true for the three major social media companies, namely Facebook, YouTube and Twitter. These companies must recognise that their unmatched reach that affects enormous amount of lives bestows upon them a social responsibility that they need to live up to.

The above represents a direct recommendation for the social media companies themselves, but the following recommendations are directed more at the EU, as we will now focus on the so-called monitoring exercise. The monitoring exercise represents a major stepping stone 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.

The main issue was that a bias was put in place mainly due to the fact that social media companies were informed about many details of the exercise in advance, such as when it was going to take place and who was involved. A more restrictive approach should therefore be taken during the future monitoring exercises. It would, for instance, be advised not to inform social media beforehand about the timeframe of the exercise. Moreover, the ongoing exercise should be masked in some way, making the companies unaware that it is ongoing. This could be done by using low level monitoring for a longer period of time or by carrying out the exercise in a rolling or snowball manner, where NGOs do not start it at the same time and end it at the same time, but spread it out for a longer period and start one after another. The idea to make the ratio of flagging content as

normal users and not as trusted flaggers bigger is also one that should be thoroughly considered. This could be done by using anonymous or fake accounts. Using these methods the elevated activity could be a bit more hidden. Moreover, these measures could ensure less bias during future exercises and thus their results would be even more representative.

2. The future of the Code of Conduct

Another great advancement for the international community was the development of the Code of Conduct on Countering Illegal Hate Speech Online (2016) (CoC). This code emerged as an outcome of the EU Internet Forum, due to the climate of fear and hatred in the context of the terrorist attacks in Brussels and the use of social media to radicalize the public, and notably young people. This code of conduct 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 who have a clearer responsibility. INACH has already discussed the Code of Conduct in our paper on [“Legislation related to cyber hate”](#). Here are some examples of the commitments made by social media in the CoC:

-The IT Companies to have in place clear and effective processes to review notifications regarding illegal hate speech on their services so they can remove or disable access to such content. The IT companies to have in place Rules or Community Guidelines clarifying that they prohibit the promotion of incitement to violence and hateful conduct.

-Upon receipt of a valid removal notification, the IT Companies to review such requests against their rules and community guidelines and where necessary national laws transposing the Framework Decision 2008/913/JHA, with dedicated teams reviewing requests.

-The IT Companies to review the majority of valid notifications for removal of illegal hate speech in less than 24 hours and remove or disable access to such content, if necessary.

-The IT Companies to encourage the provision of notices and flagging of content that promotes incitement to violence and hateful conduct at scale by experts, particularly via partnerships with CSOS, by providing clear information on individual company Rules and Community Guidelines and rules on the 13 reporting and notification processes. The IT Companies to endeavor to strengthen partnerships with CSOs by widening the geographical spread of such partnerships and, where appropriate, to provide support and training to enable CSO partners to fulfil the role of a ‘trusted reporter’ or equivalent, with due respect to the need of maintaining their independence and credibility."

However positive this might be, it does have some limitations and we could argue as to what the future of the CoC could look like. Since we have already discussed the future of the monitoring exercises, the only other thing that has to be elucidated is whether further and more stringent steps should be taken by the EC. Those steps would be taken if the voluntary pledges taken by the companies in the CoC paired with regular monitoring exercises prove to be insufficient to force the sector to raise their removal rates and also to remove illegal content more promptly. In that case, INACH recommends that the EC enacts a directive based on the Code of Conduct that can be actually enforced through different already existing EU bodies and also forces member states to put the issue on their agendas and start combating online hate speech more effectively. Also, the use of fines should not be off the table if and when this directive will be drafted. The system of fines could be based on the system that the German parliament enacted in the Act to Improve Enforcement of the Law in Social Networks (the NEA), described in chapter III. However, we are not suggesting that the system of fines should necessarily be identical to the NEA, although we do recommend high fines, since smaller ones will not be sufficient to deter social media companies. A directive similar to the above described is a logical next step if the Code of Conduct fails to live up to the expectations.

3. Analysis and critique of the new Communication from the European Commission ‘Tackling Illegal Content Online’

The focus of this chapter will be the Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on “tackling illegal content online” and moving “towards an enhanced responsibility of online platforms” mentioned earlier on.

Even though this Communication is not a law and does not enact any measures of enforcement, it definitely sets the tone of policy that the Commission wants to pursue in the coming years. The Commission, observing the developments in the fight against cyber hate in certain members states, and realising that the issue needs addressing on a supranational scale, felt necessary to draw up policy suggestions on EU level that take several serious steps towards tackling online hate speech.

With this policy document, the Commission tries to take necessary steps towards ensuring human dignity online, whilst sidestepping the criticisms and backlash that the NEA received internationally. However, this Communication is basically a wish list of principles and rules that the EC would like social media companies to adhere to when it comes to all illegal hate speech such as for instance the glorification of terrorism, cases of copyright infringement and child pornography. In certain ways it has even less weight than the Code of Conduct, because that was signed by four major (social media) companies: Facebook, Twitter, YouTube and Microsoft, thus taking responsibility to adhere to the provisions set forth in that document. The Communication

by the Commission on the other hand is just an advisory document, however, it absolutely signals the direction the EU will be taking on illegal content online.

Let us examine the main points of this Communication. The Communication states that the timely and speedy removal of illegal content is pivotal to hinder the wider dissemination of hateful and/or illegal materials and content. It also stipulates that social media platforms should have policies in place that allow NGOs and state bodies to become so-called trusted flaggers, because trusted flaggers speed up the reporting and removal procedures by providing expert complaints to the companies that do not have to be checked as thoroughly as complaints from average users.

It also argues that “In order to ensure a high quality of notices and faster removal of illegal content, criteria based notably on respect for fundamental rights and of democratic values could be agreed by the industry at EU level” (page 8). Then it further stipulates that trusted flaggers should be regularly tested against these agreed criteria in order to prevent the misuse and exploitation of the system that could lead to “over-removal”. The Communication also states that the trusted flagger status should be revoked in cases where NGOs misused that status.

The Communication goes further to say that all social media platforms should have user-friendly and contextual reporting methods and systems in place that users can use anonymously. Furthermore, that companies should stay in touch with the complainers and give follow-up on the status of their complaint(s).

Furthermore, the EC made huge strides in combating online hate speech by stating in this Communication that social media companies should publish transparency reports on a regular basis. In these reports the companies should give an overview of how many complaints they received in a certain period of time, what content they removed, what content they did not remove and provide the reasons behind their decision making processes. This measure is very similar to the one in the NEA.

The EC also argues that companies should have proactive measures in place to filter and remove illegal content, maybe even by using software that recognises such content and removes it automatically. However, besides the need for speedy removal and proactive measures, the Communication also states that platforms should do everything in their power not to accidentally remove legal content and should have systems in place through which their decisions to remove content can be appealed and contested.

According to the Communication social media should decide if content is illegal based on EU law, national laws and of course their own content policies and terms of service. As far as illegal content online goes in this Communication, the EC conflates hate speech, inciting to terrorism or other

violent acts and even copyright infringements. Hence, these policy suggestions pertain to all these types of content.

Naturally, the summary given of the Communication is fairly skeletal. However, it is enough to paint a picture of the direction the EC is aiming towards and the issues they are most concerned with.

These are the following:

- pertaining to illegal online content (incitement to terrorism, hate speech, copyright infringements),
- speedy removals by social media platforms through - preferably uniform - user-friendly and transparent reporting systems,
- publishing transparency reports,
- utilising trusted flaggers as much as possible;
- using proactive measures, such as filtering softwares or algorithms that block or remove illegal content automatically and
- building in safeguards to protect free speech, which include certain penalties for trusted flaggers if they abuse or misuse the system.

This Communication underpins the Commission's tireless work that they have done in the past years to tackle online hate and discrimination. Like previously mentioned, it is a major step towards the right direction and we hope that the EC will do everything in its power to get the social media companies to internalise the policies enshrined in it. Furthermore, the Communication is a wonderful signal to the Member States and the whole European NGO sphere that the EC is not backing down and cyber hate will be kept as a top priority in the coming years by EU bodies.

Nevertheless, INACH would like to raise a couple of concerns about the Communication. Our first issue is the conflation of different type of illegal online content. The fact that the EC is handling the glorification of terrorism and incitement to terrorism together with online hate speech is somewhat understandable. A lot of illegal online content related to terrorism is considered hate speech, and thus it has very similar aspects as other types of cyber hate. However, the methods to fight against terrorist propaganda online just partially intersect with the methods to fight other kinds of online hate speech. Furthermore, the inclusion of copyright infringement issues in a policy Communication that otherwise deals with serious crimes related to human rights issues and terrorism is a bit misguided. We understand that this Communication is about illegal online content and not online hate speech exclusively, but that is exactly the issue. Online hate speech should definitely be handled through separate Communications and directives. That way, the focus of the policies would be less fragmented and social media companies would lose a tool that enables them right now to shift the focus from the plethora of vile and hateful content they host to content that infringes the rights of other major companies.

Our second concern is the Communication's overly anxious approach to the dangers of so-called over-removal of content. It is paramount to protect the freedom of speech and have safeguards against the removal of content that should not have been removed. There should also be appellate systems in place through which users and companies can contest decisions of removal and have their content reinstated if it is proven legal. However, most trusted flaggers are NGOs or government bodies full of experts on cyber hate. They work tirelessly to make the internet a safer place for vulnerable communities and force online platforms to clean up the virtual public space that they had created. They are obviously pushing hard and trying to pressure social media to adhere to national and international laws or at least provide transparent and consistent interpretations of their own terms of service, but they are still far from being a looming threat against free speech online.

INACH appreciates that laws have to be very comprehensive, all-encompassing and detailed in order to cover every possible aspect of an issue, eschew ambiguity and leave no loopholes. However, devoting such a big part of a Communication from the EC on illegal online content to the dangers of over-removal and the malicious use of hate speech laws by trusted flaggers signals an approach that is a bit over-cautious. Over-removal is a phenomenon that has not been proven to occur due to NGOs exploiting their trusted flagger status. There are definitely disagreements between social media and trusted flaggers about what should be removed or not, the exploitation of the trusted flagger system by NGOs, however, is a pseudo-problem that unnecessarily shifts the limelight from the responsibility of social media companies and provides a fertile ground for 'whataboutism'. We hope that social media will not try to shift the focus and reset the agenda completely to over-removal from speedy removal based on the policies set forth by this Communication.

If and when in the future the Commission will feel the need to step up the fight against online hate even further and enact a law somewhat similar to the German NEA, we hope that that law will be focused a bit less on the probability of misuse and misconduct by trusted flaggers and more on forcing the social media sector to adhere to the law.

Nonetheless, all other policies and initiatives set forth in the Communication are magnificent steps in the right direction and INACH hopes that they will enhance the already fruitful cooperation between the Commission, NGOs and social media, so we can continue the fight even more effectively against the toxicity that is cyber hate, developing remedies and maybe even coming up with a panacea in the long run.

III. National level recommendations

1. Suggestion for further harmonisation with a focus on the German law

Regarding national level recommendations, an aforementioned example of a tremendous advancement in the field was the new German law regarding social media, the Act to Improve Enforcement of the Law in Social Networks (NEA). The ground-breaking aspect of this new law is that it forces companies to adhere to the law for the first time whilst giving them a clear set of rules concerning the handling of reports. Moreover, a system of fines is set into place, which gives an extra incentive and pressure for the companies to take this seriously.

Firstly let's take a closer look as to what is written in detail. In Section 4 the Provisions on regulatory fines is listed. The following offenses are subject to fines;

- “1. In contravention of section 2(1) sentence 1¹, **fails to produce a report**, to produce it correctly, to produce it completely or to produce it in due time, or fails to publish it, to publish it correctly, to publish it completely, to publish it in the prescribed form or to publish it in due time, 2.
2. In contravention of section 3(1) sentence 1², **fails to provide**, to provide correctly or to provide completely, a procedure mentioned therein for dealing with complaints submitted by complaints bodies or by users whose place of residence or seat is located in the Federal Republic of Germany, 3.
3. In contravention of section 3(1) sentence 2³, **fails to supply a procedure** mentioned therein or to supply it correctly, 4.
4. In contravention of section 3(4) sentence 1⁴, **fails to monitor the handling of complaints or to monitor it correctly**, 5.
5. In contravention of section 3(4) sentence 2⁵, **fails to rectify an organisational deficiency or to rectify it in due time**, 6.
6. In contravention of section 3(4) sentence 3⁶, **fails to offer training or support** or to offer them in due time, or 7.
7. In contravention of section 5, **fails to name a person authorised to receive service in the Federal Republic of Germany or fails to name a person in the Federal Republic**

¹ Providers of social networks which receive more than 100 complaints per calendar year about unlawful content shall be obliged to produce half-yearly German-language reports on the handling of complaints about unlawful content on their platforms, covering the points enumerated in subsection (2), and shall be obliged to publish these reports in the Federal Gazette and on their own website no later than one month after the half-year concerned has ended

² The provider of a social network shall maintain an effective and transparent procedure for handling complaints about unlawful content in accordance with subsections (2) and (3)

³ The provider shall supply users with an easily recognisable, directly accessible and permanently available procedure for submitting complaints about unlawful content.

⁴ The handling of complaints shall be monitored via monthly checks by the social network's management.

⁵ Any organisational deficiencies in dealing with incoming complaints shall be immediately rectified.

⁶ The social network's management shall offer the persons tasked with the processing of complaints training courses and support programmes delivered in the German language on a regular basis, this being no less than once every six months.

of Germany authorised to receive information requests from German law enforcement authorities, or

8. In contravention of section 5 subsection (2), second sentence⁷, fails to respond to requests for information while acting as the person authorised to receive service.”

Subsection 4 (1) states that "2. in contravention of section 3(1) sentence 1, fails to provide, to provide correctly or to provide completely, a procedure mentioned therein for dealing with complaints submitted by complaints bodies or by users whose place of residence or seat is located in the Federal Republic of Germany, [...]" and Section 3(1) sentence 1 has a direct reference to Section 3(2). Section 3 (2) states: "The procedure shall ensure that the provider of the social network: [...] 2. removes or blocks access to content that is manifestly unlawful within 24 hours of receiving the complaint; [...] 3. removes or blocks access to all unlawful content immediately, this generally being within 7 days of receiving the complaint; [...]" Therefore, regulatory fines can be given if the provider doesn't provide a complaints handling procedure that ensures they'll block or remove unlawful content within the given time periods.

This new law raises the following question: Should this be the direction the harmonised policy takes, or is there room for improvement? Some critics of this law found that it might encourage censorship (even of legal content) as companies will perform 'blanket-removals' in order to avoid fines, which could lead to the censorship of legal content that is just controversial.

Attorney Joachim Steinhöfel makes some valid points regarding the limitations of the law, namely due to his distrust of the companies in doing the right thing. In an article from the BBC⁸, he uses examples from his blog that show Facebook deleting posts which, according to him, are not illegal, as well as leaving posts that he deemed illegal. He uses the case where the post "Jews to the gas chambers" was not removed, even after request for removal. He states that "it goes to show how unprofessional they are in dealing with this content", while explaining that he does not trust companies such as Facebook to decide in an efficient and fair way what posts are legal and those which are not, perhaps leading to a "mass deletion of legal posts that Facebook doesn't want to take any risks with, out of an abundance of caution". This point was supported by many others, such as, Miko Hohmann, from the Global Public Policy Institute, who stated "there are real concerns that the law will incentivise social media companies to excessively delete content," and "faced with fines of up to €50 million, social media platforms will likely err on the side of caution and delete lawful content when in doubt."⁹ Similarly, Daniel Calingaert, executive vice-president

⁷ The person so authorised shall be obliged to respond to such requests for information pursuant to the first sentence within 48 hours of receipt.

⁸ Patrick Evans, *Will Germany's new law kill free speech online?*, BBC, found on <http://www.bbc.com/news/blogs-trending-41042266>

⁹ [Cara McGoogan](http://www.telegraph.co.uk/technology/2017/06/30/germany-fine-facebook-youtube-50m-fail-delete-hate-speech/), Germany to fine Facebook and YouTube €50m if they fail to delete hate speech, The Telegraph, found on <http://www.telegraph.co.uk/technology/2017/06/30/germany-fine-facebook-youtube-50m-fail-delete-hate-speech/>

at Freedom House also stated that “the legislation would give social networking companies strong motivation to censor first and ask questions later”¹⁰.

We do agree with the fact that social media companies have a very lackluster approach to the identification and removal of illegal content, whilst they also have a checkered past when it comes to over-policing nudity or controversial content. They also like ambiguity in their terms of service and opacity when it comes to what they remove and why. Hence, they might indeed resort to blanket-removal of controversial contents just to avoid fines and also using more funds to assess content for legality. Thus, we arrive to our sole criticism of the NEA: it lacks proper safeguards against the removal of legal content and it does not have a clear appellate system enshrined in it through which people could contest a decision of removal. Such safeguards are paramount to mitigate the possible infringement of free speech and the NEA should be amended accordingly. Nonetheless, this should not take away from the fact that this is a tremendous step in the right direction, and should be used as a model for future policies on the matter.

2. Counter strategies on an educational level

Lastly, educating people from different backgrounds and ages is a major asset in the fight against cyber hate. It is essential to keep the focus on all those affected or who have an effect on the fight. For instance, educating the police as to what is legal and what is not makes the process much more efficient for everybody. Educating young people also takes care of the fact that those people are the most susceptible to cyber hate or to suffer from its effects. Giving them the power back and the knowledge of what is acceptable and what is not and what to do about it is an immense step forward. Those are a few examples of where the focus should be kept on in order to help make the internet a better place.

When it comes to people recognising hate speech, internalising critical thinking, recognising fake news items and propaganda, the ultimate solution is more educational courses and trainings delivered via more effective educational methods. If we can educate people so they know what hate speech is, they can recognise it in the online public space; if we manage to teach them what can be done to eliminate cyber hate and how, it would represent a real breakthrough.

Therefore, it is absolutely pivotal for counter-speech and counter-narrative projects and campaigns to be available to the general public on a permanent basis, in order to spread their message and education as wide as possible. These projects - such as the No Hate Speech Movement, Facing Facts, Get The Trolls Out, the Dangerous Speech Project - are doing great work to educate people, to raise awareness, to fight the deluge of negative coverage of vulnerable communities in the media and to make both offline and online public spheres safer for everybody.

¹⁰ Natalia Smolentceva, German Law on Social Media May Threaten Freedom of Expression, Legal Dialogue, found on <http://legal-dialogue.org/german-law-social-media-may-threaten-freedom-expression>

Thus, we recommend that the EU, the EC and national governments continue funding such projects and on top of that, help new projects to get off the ground. The online public sphere has become just as dominant as the offline one if it has not surpassed it in the past 10 years. Yet, we have done hardly anything to arm people and especially adolescents with sufficient tools not to be manipulated by it. Therefore we also recommend that EU member states incorporate courses in their national curricula that teach children from a young age how to utilise the internet in a safe way and how not to be manipulated by racist propaganda, fake memes, fake news and other types of cyber hate.

Knowing all of the above, we will ourselves focus on attaining this goal in the coming years by developing workshops, counter-speech modules and curricula, and trainings. We will also work hard in order to persuade all our members and national governments to develop such measures themselves for as many European countries as possible.

IV. Conclusion

To conclude and summarize the recommendations in short, they are the following;

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

Cyber hate grows more destructive every day and due to the cheap and easy measure of circulating cyber hate on social media, it is becoming omnipresent and normalised. This normalisation has

been shown in multiple studies published by INACH and its members, such as the '[Manifestations of Online Hate – Reports on Antisemitic, Antiziganistic, homophobic and anti-Muslim Hate Speech](#)' and '[“Kick Them Back Into The Sea” – Online Hate Speech Against Refugees](#)'. This normalisation process leads to people becoming desensitized to human rights abuses and growingly accept them as truths. That is why it is pivotal for stakeholders such as authorities, EU institutions and European governments to recognize the dangers presented by cyber hate and develop policies such as the ones we mention above to counter it.

Moreover, we are not just making propositions for everyone to change their ways but, like previously mentioned, intend to follow our own recommendations throughout the next years. We are aware that we will never truly have a clean internet, it has grown into a powerful monster difficultly tameable, but by all working hand in hand and looking at all the different aspects of the problem, where and why does cyber hate start, and how it's effect can be diminished, there is a chance for us to, if not fully stop it, limit it.

The future of internet should not be a bleak one, and this astonishing tool should be developed and hopefully used in a positive light, by bringing us all closer rather than building walls between us.

VII. Bibliography

Evans, Patrick, *Will Germany's new law kill free speech online?*, BBC, found on <http://www.bbc.com/news/blogs-trending-41042266>

INACH, [*Manifestations of Online Hate – Reports on Antisemitic, Antiziganistic, homophobic and anti-Muslim Hate Speech*](#)

INACH, [*“Kick Them Back Into The Sea” – Online Hate Speech Against Refugees’*](#)

INACH, [*Annual Report 2016-2017*](#)

INACH, [*Legislation related to cyber hate*](#)

McGoogan, Cara, *Germany to fine Facebook and YouTube €50m if they fail to delete hate speech*, The Telegraph, found on <http://www.telegraph.co.uk/technology/2017/06/30/germany-fine-facebook-youtube-50m-fail-delete-hate-speech/>

Smolentceva, Natalia, *German Law on Social Media May Threaten Freedom of Expression*, Legal Dialogue, found on <http://legal-dialogue.org/german-law-social-media-may-threaten-freedom-expression>