

TAKING ACTION AGAINST HATE ON THE INTERNET IN A COLLABORATIVE SOCIETY

JULY 2017

WRITERS

Henri ISAAC, President of Renaissance Numérique Associate Professor at Université Paris-Dauphine

Camille VAZIAGA, Member and former Managing Director of Renaissance Numérique

Mike FEDIDA, Task Officer at Renaissance Numérique

EDITORIAL

THE DIFFICULT AWAKENING OF THE ELITES IN THE ERA OF CONVERSATIONAL DEMOCRACY – HENRI ISAAC, CEO OF RENAISSANCE NUMÉRIQUE

The emergence of the Internet, then the development of spaces for the production of text content (blogs), and more recently social networks, have considerably changed the way in which ideas are produced and circulate. By broadening the opportunity offered to each and every one of us to express ourselves and take part in a great many debates, society's digital networking paves the way for truly discursive democracy. However, while this constitutes unprecedented democratic progress, it is not without its difficulties.

The freeing up of speech on a scale never previously known in the history of humanity is also characterized by an infinite range of speech, from the most sickening to the most sophisticated, distancing us from the ideal of discursive democracy which came to light in 20th-century political philosophy, stemming from works by Habermas and others who came after him¹. "According to the ethics of discussion, a norm can only claim validity if all the people who might be concerned are in agreement (or could be) as participants in a practical discussion on the validity of this norm."²

As Alban Bouvier emphasizes, "While the principle is not in any way a political principle, we might nonetheless sense right away that this principle of discussion should be able to apply to legal norms, laws, various regulations, etc., in other words be the basis of a policy of discussion, or more precisely even of a democratic policy of discussion or of well-argued discourse or indeed of a 'discursive democracy,' as it is a question of participation in the discussion by all those concerned."³

It remains no less the case that to enable discursive or deliberative democracy to become a reality, citizens need to be equipped with methods and tools to reach a real ethical position on debate, without which debates are everything but well argued and are limited to abuse, insults and hate speech, a situation that is only facilitated and reinforced by the format and immediacy of electronic communications.

1 Habermas Jürgen, 1996, "Notes Programmatiques pour fonder en raison une Ethique de la discussion" in Habermas J., 1996, *Morale et Communication. Conscience et activité communicationnelle*, Paris, Cerf, pp. 63-130; Dryzek J.S., 1990, "Discursive Democracy: Politics, Policy, and Political Science," Cambridge University Press, 264 p.

2 Idem, Habermas J., 1996, p. 87.

3 Alban Bouvier, 2007, "Démocratie délibérative, démocratie débattante, démocratie participative", edited by Alban Bouvier and Samuel Bordreuil. *Revue européenne des sciences sociales*, Tome XLV, 2007, no. 136, Librairie Droz Genève Paris.

Internet n'a que 25 ans d'existence. Pour les élites politiques qui s'y déploient, jusqu'ici The Internet has only existed for 25 years. For the political elites who are active there, until now the sole upholders of representative democracy, this deliberative space that is the Internet seems mainly to be perceived from the sole perspective of hate speech. Worse still, they use it as an argument to limit, restrict, control and censor this deliberative space, frightened as they are by a register of speech that they are not accustomed to hearing in the mechanisms of discussion in which they participate.

At the think tank, an actor in the sphere of digital which is driven by public interest, we felt that, in order to facilitate and continue the debate and to make this discursive space a more pacified space, we should not merely wait for rules or solutions to come from institutions. Consequently, we propose a method and tool—Seriously—to develop a real ethical position on discussion and thus make it possible to have a discursive democracy that offers everyone the right to well-argued speech.

We hope this tool may be appropriated by all citizens, by all actors involved in defending rights and by those responsible for training in citizenship.

HENRI ISAAC PRÉSIDENT
DE RENAISSANCE NUMÉRIQUE



INTERNET AND SOCIAL NETWORKS: FORMIDABLE ACCELERATORS OF FREEDOM OF EXPRESSION, AND THE GLUE OF DEMOCRATIC DIALOGUE

There is no doubt that freedom of expression has entered into a new era, today forming an integral part of our daily life. It should be understood to mean each person's freedom to make themselves understood, take part in the debate, contribute their perspective and share ideas, so that these ideas can germinate and grow.

We are undeniably witnessing the emergence of an alternative to debates formerly hijacked by those who "know," by those who "have the right." Inherited power and credibility are being called into question.

However, month after month, we have to recognize that constructive contributions are being heard less and less on social networks. Extremists, hateful, violent people, conspiracy theorists, and racists of all stripes are hijacking the "debate." But what debate, exactly? The one they propose, that they impose, consisting of invective, violence and "alternative facts." (Verbal) violence has reached the digital realm, where it has then become a template which is exported beyond the confines of the Internet. This is reflected in the weakness and violence of current democratic public debate in the traditional media.

Of course, drawing a parallel between the violence of racist comments, of the drift of terrorists or conspiracy theorists, or of the tilting of political discourse, may seem shocking.

Words don't kill, terrorism does, and that's a fact! However, the terrain and the breeding grounds get mixed up, drift closer together, and feed off each other.

Freedom of expression has been a long battle. Today, it is considered to have been

won through democracy. But to preserve it, this freedom must be accompanied by a duty of expression. If I stay quiet about a situation, then I accept it. And yet, seeking out the guilty party or labeling the other as responsible doesn't change anything—on the contrary, it feeds divisions and removes responsibility.

In the digital era, it's up to each and every one of us to "mobilize" to the extent that we are able. We have acquired the means by which to act. So we must act, or at least contribute, at least try.

With this in mind, in the days following the January 2015 terrorist attacks one question dominated our thoughts at Renaissance Numérique. What can we, as a think tank on digital matters and above all as citizens, do to take concrete action to contain the hateful dynamic that prospers in our societies and finds a special catalyst on the Internet? How do we move forward from the essential—but insufficient—"think" to "do." If we don't try, we can be sure we won't get there...

And that's how Seriously.org was born.

It's a tool, a platform and a method, too. A different way of looking at the other, a different way of looking at ourselves.

Relearning how to engage in dialogue? And why not!

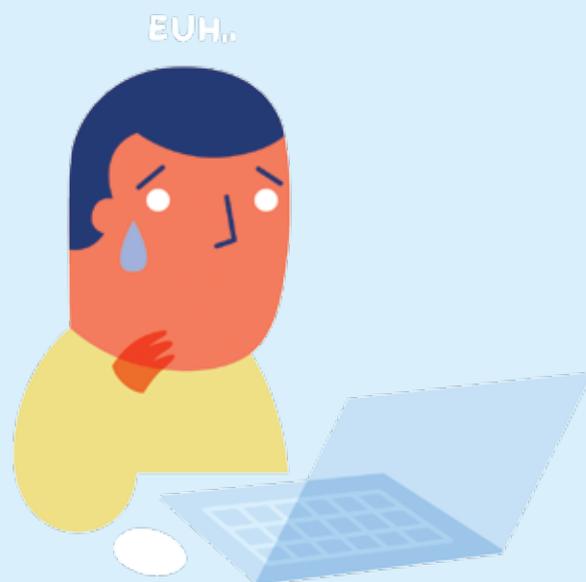


GUILLAUME BUFFET,
CEO OF U AND
CREATOR OF THE
SERIOUSLY PROJECT

CONTENTS

| | |
|---|-----------|
| INTRODUCTION: THE PARADOX OF HATE ON THE INTERNET | 6 |
| The Internet: The Perfect Scapegoat. | 7 |
| The Internet: A Complex Factor for the Law. | 8 |
| | |
| 1. HATE ON THE INTERNET: WHY HAS IT BECOME A SOCIAL ISSUE? | 10 |
| 1.1. The end of gatekeepers and the emergence of new public opinions | 11 |
| 1.2. Immediacy, virality and new relationships to facts | 12 |
| | |
| 2. HATE PHENOMENA ON THE INTERNET: THE LAW AND INSTITUTIONS GRIPPED BY THE PROBLEM | 16 |
| 2.1. “Hate speech”: What are the main applicable legal provisions? | 17 |
| 2.2. “Hate speech”: Who is responsible in the eyes of the judge? | 20 |
| 2.3. The tripartite arrangement of the response to the phenomenon of hate | 22 |
| 2.4. The law: A necessary but insufficient tool | 24 |
| | |
| 3. FIGHTING AGAINST HATE ON THE INTERNET: IN FAVOR OF THE COLLABORATIVE ORGANIZATION OF MEASURES | 27 |
| 3.1. The masses: Organizing the spontaneous citizens’ response | 29 |
| 3.2. The educational system | 32 |
| 3.3. Associations | 33 |
| 3.4. Major digital actors | 34 |
| 3.5. The justice system | 35 |
| 3.6. The media | 36 |
| 3.7. Research actors | 37 |
| | |
| CONCLUSION | 38 |
| APPENDIX | 40 |

INTRODUCTION: THE PARADOX OF HATE ON THE INTERNET



In this period of constantly shifting crisis—in the economy, with increasing unemployment; in politics, with the disintegration of trust between citizens and their elected representatives; and in society, where the opposition between social, ethnic and national groups is intensifying—we can observe a resurgence of hate and the expression of hate, of physical and verbal violence, as well as an increase in populism around the world. Hate is being normalized. Its manifestations are becoming increasingly frequent, and no “community” is spared, as recent surveys have shown:

- 85% of people have already witnessed hate speech in France in 2016⁴;
- 808 acts of anti-Semitism were recorded by the police in France in 2015, making an average of two acts per day⁵;
- Between 2014 and 2015, anti-Muslim incidents increased by 233% in France⁶;
- 100% of public transportation users have been victims, at least once in their life, of sexist harassment or sexual assault, whether they are aware or not that it comes under this name⁷;
- In 2015, the association SOS homophobie gathered 1318 witness statements of LGBT-phobic acts in France⁸;
- In August 2016, a petition was launched to condemn the normalization of anti-Asian attacks in France, which so far has gathered more than 16,000 signatures⁹.

This normalization of acts of hate is quite naturally reflected in the way citizens express themselves and therefore on the Internet and its social networks, where a large proportion of our daily self-expression takes place, as Facebook has 20 million daily users in France, and Twitter almost 6 million monthly visitors. Bearing in mind the intensity of this usage, it is not surprising to observe near-parity between how often we encounter hate speech in our daily lives offline and online: respectively 78% and 70%¹⁰ of French people say they have encountered hate speech on these two environments.

THE INTERNET: THE PERFECT SCAPEGOAT

However, the Internet in particular seems to crystallize the issues around the propagation of hate speech in our society. It is therefore perceived as the main hotbed for hate speech (58%), ahead of television (14%), family and friends (9%), the workplace (7%), the press (5%) and the radio¹¹.

4 “Les Français et les discours de haine” (The French and hate speech), a poll carried out in May 2016 by l’Institut CSA for Google France.

5 “Rapport sur la lutte contre le racisme, l’antisémitisme et la xénophobie” (Report on the fight against racism, anti-Semitism and xenophobia), 2015 annual report, CNCDH. Referenced on March 27, 2017. Available at: http://www.cncdh.fr/sites/default/files/les_essentiels_-_rapport_racisme_2015_page_a_page.pdf

6 Idem, 2015 annual report, CNCDH.

7 “Avis sur le harcèlement sexiste et les violences sexuelles dans les transports en commun” (Opinion on sexist harassment and sexual violence on public transportation), Avis n°2015-04-16-VIO-16, HCEfh, April 16, 2015. Referenced on March 27, 2017. Available at: <https://fr.scribd.com/doc/261942856/HCEfh-Avis-harcelement-2015-04-16-VIO-16-1>

8 “Rapport sur l’homophobie” (Report on homophobia), 2016 annual report, SOS Homophobie. Referenced on March 27, 2017. Available at: https://www.sos-homophobie.org/sites/default/files/rapport_annuel_2016.pdf

9 “Agression anti-asiatique : Lettre au Président de la République et au Gouvernement” (Anti-Asian attacks: Letter to the President of the Republic and Government). Petition, Change.org. Referenced on March 27, 2017. Available at: <https://www.change.org/p/agression-anti-asiatique-lettre-au-pr%C3%A9sident-de-la-r%C3%A9publique>

10 Idem, poll by Institut CSA / Google France.

11 Ibid.

It is paradoxical to note that the Internet is perceived as the main hotbed of hatred, as people say they are more exposed to insults and abuse in their offline environment. The Internet has the effect of a distorting mirror on the expression of hatred, discrimination and cruelty. The legislator has inscribed in law that posting messages on the Internet that are defamatory or that praise terrorism is an aggravating factor. By its viral, immediate and international nature, the Internet is therefore a mirror that magnifies insults and instances of defamation, which previously would have struggled to make it into the public realm.

At a time when people are more geared toward populist choices, as illustrated in 2016 by the election of Donald Trump and by the United Kingdom's referendum on leaving the European Union (Brexit), and when numerous Western countries are being hit by terrorist attacks, increasingly the finger is being pointed at online hatred as the phenomenon responsible for the disintegration of the social bond between citizens. Alain Juppé has spoken of the “*world's trashcan*”¹² and Najat Vallaud Belkacem advises wariness when it comes to social networks¹³. The problem is being looked at on a Europe-wide level, where pressure is mounting to make the main social networks evolve their mechanisms of operation.

THE INTERNET: A COMPLEX FACTOR FOR THE LAW

However, political actions carried out in the aim of restraining hatred on the web are having a limited effect, at this stage, on this diffuse phenomenon. The fiasco of the Dieudonné controversy and his YouTube videos during the 2010s demonstrates the limits of the law and the justice system to intervene effectively, with the removal of the videos being almost ineffective for all practical purposes, the media denunciation actually increasing the speed at which the content circulated, and the justice system finally invalidating the injunctions by the media and politicians to remove the video, in the end bolstering the position taken by YouTube, which did not want to take editorial responsibility for removing these videos.

Grasping the limits of freedom of expression is a very thorny matter, and already an extremely complex one even without the “Internet” factor, as we have seen during the high-profile lawsuits against Charlie Hebdo. There is no doubt that the viral and global dimension that the Internet gives to hate speech introduces further complexity and requires new approaches to be taken.

The aim of this introductory note is therefore to provide an outline of this complex phenomenon, which has become a major social problem for political decision-makers and the media, and yet remains just as poorly grasped as it has been for years.

12 “Juppé : « Mon plan pour réformer la France »” (Juppé: “My plan to reform France”), Article, Le Journal Du Dimanche, October 2, 2016. Referenced on March 27, 2017. Available at: <http://www.lejdd.fr/Politique/EXCLUSIF-Juppe-Mon-plan-pour-reformer-la-France-814099>

13 “Il faut se méfier des réseaux sociaux selon Najat Vallaud-Belkacem” (We should be wary of social networks according to Najat Vallaud-Belkacem) Article, TNTV News, October 24, 2016. Referenced on March 27, 2017. Available at: http://www.tntv.pf/Il-faut-se-mefier-des-reseaux-sociaux-selon-Najat-Vallaud-Belkacem_a14683.html

By setting the stage thus, Renaissance Numérique presents and provides the context for the Seriously project (www.seriously.org), launched in 2015. Drawing on its expertise in these issues, the think tank incubated its first “do-tank” project by developing a brand new tool: a platform that gives you arguments for neutralizing hate speech on the Internet and building dialogue, rather than sliding into the escalation of violence. Designed, supported and implemented by Renaissance Numérique, Seriously is the fruit of a collaborative approach in which around one hundred associations and experts worked together.

1. HATE ON THE INTERNET: WHY HAS IT BECOME A SOCIAL ISSUE?



1.1. THE END OF GATEKEEPERS AND THE EMERGENCE OF NEW PUBLIC OPINIONS

Firstly, it is the very nature of the Internet, which has a magnifying effect on the hate speech published on it: the open, networked architecture of Web 2.0 facilitates direct personal expression, sharing, commenting and repetition. The racist language used in a great many public spaces takes on a particular importance on the Internet. Since the arrival of Web 2.0 in the mid-2000s onward, social networks have become very widely popularized and everyone now knows that they can gain influence in a way that was inaccessible to them before the Internet. The 2007 French presidential campaign, with the rise of community campaign websites, demonstrated the arrival of an interactive Web on which activists could have their say. So, political speech and any other expression of opinion were freed up on forums and blogs first, and then more widely on social networks. The best and the worst of anything can be found on social networks, just like in life offline. Speech is freed up there, taking on a power that no media outlet or politician expected, as was shown by the immense surprise caused by the victory of the “no” vote in the 2005 referendum on the European Constitution. Although Internet experts had warned that “no” was dominating on chat sites, forums and social networks, for the first time the media discovered that the impact and power of conviction of their own communications had competition¹⁴, and were even ineffective against the many people discussing matters directly on the Internet.

The Internet thus marks the end of the gatekeepers who once decided who had the right to access media channels. The sociologist Dominique Wolton sums it up as follows: “if our democracies want to stay faithful to their project of political emancipation, they need to regulate the new media and thereby prevent freedom of communication from becoming synonymous with the law of the jungle.¹⁵” Not that the traditional channels of expression are the sole preserve of an irreproachable elite: France has plenty of singers, essayists and journalists whose discourse undoubtedly invites discrimination against groups of the population. However, the crowd of new voices on the Internet logically opens up access to public expression for groups of the population and associations that assert hatred and discrimination. Thus, principles which until then had enjoyed a consensus (at least in appearance) in the public realm, like secularism, republican values and peace between religions, are now criticized, mocked and disputed, if not publicly challenged.

These discussions are open and visible by everyone, without any limitation in terms of time or format. This is why groups that did not previously have access to public speech, like the Front National at one time, have developed a worrying strike force on the Internet: the website “Français de Souche” (Pure-blood French), and more generally speaking the online far right movement, have a very large community and power in this space because they had to turn to the Internet very early on, due to the inaccessibility of the traditional channels of expression. The Front National was in fact the first political party to have a website, in 1996.

14 “Ce que les campagnes Web ont changé dans la vie politique” (How Web campaigns have changed politics), Article, Contrepoints, April 28, 2015. Referenced on March 27, 2017. Available at: <https://www.contrepoints.org/2015/04/18/204959-ce-que-les-campagnes-web-ont-change-dans-la-vie-politique>

15 See Wolton, Dominique, 1999, Internet et après ? Une théorie critique des nouveaux médias, Flammarion.

Marine Le Pen recently explained this necessity: “our media ostracism and our poverty contributed to making us hyper-imaginative people.”¹⁶ This new access to public speech, sidestepping the traditional gatekeepers, explains why some political groups, associations and activists are particularly active on the Internet. At this point it seems worth emphasizing that neo-Nazis are more active than Islamic State (IS) terrorists: “neo-Nazi Twitter accounts today attract 22 times more followers than jihadist accounts, and tweet twice as much as sympathizers of IS.”¹⁷

IN SUMMARY: While we do not find greater numbers of people advocating hate and division on the Internet, their speech there is particularly easy to access. By bringing down the barriers to access to public speech, the Internet offers an echo chamber to activist groups which, without necessarily being illegal, had been refused access to spaces of public expression and quite rightly advocate on the Internet—with all the more vehemence—access to free speech. Consequently, they are able to spread their hateful, fascist, anti-Semitic, discriminating and insulting way of thinking.

1.2. IMMEDIACY, VIRALITY AND NEW RELATIONSHIPS TO FACTS

By responding to our clicks, likes and shares, the algorithms of Facebook, YouTube and Twitter can make some information more visible than others. This order is based on our past behavior and hence is a continuity of our tastes and opinions, which are betrayed by the digital tracks we leave. Journalists denounce this “*information bubble*,” i.e. the impossibility for people who get their information on the Internet and on social networks to access information that is different to their line of thinking.

The process that takes place on social networks undoubtedly involves a lack of understanding between the causes and the means. As Dominique Cardon puts it, to expand your access to encompass more diverse content on the Internet, via social networks, first expand your circle of friends¹⁸. Thus, we can understand that the problem is less about the algorithm that organizes what information which is shared by your contacts, and more about the capacity to incorporate divergent opinions into your circles. Similarly, in real life, social mixing is often considered a chimera, and left-wing voters will tend to prefer to subscribe to newspapers that correspond with their line of thinking. There is an information bias which depends on various sociological criteria, in other words what specialists in the cognitive sciences call “*confirmation bias*.”¹⁹

16 “Marine Le Pen a-t-elle gagné la bataille du web ?” (Has Marine Le Pen won the battle for the Web?), Article, Paris Match, October 9, 2016. Referenced on March 27, 2017. Available at: <http://www.parismatch.com/Actu/Politique/Marine-Le-Pen-a-t-elle-gagne-la-bataille-du-web-1090135>

17 “Les néonazis plus actifs que l’Etat islamique sur Twitter, selon une étude américaine” (Neo-Nazis are more active on Twitter than Islamic State, according to an American study), Article, Le Monde, September 5, 2016. Referenced on March 27, 2017. Available at: http://www.lemonde.fr/pixels/article/2016/09/05/les-neonazis-plus-actifs-que-l-etat-islamique-sur-twitter-selon-une-etude-americaine_4992917_4408996.html#fj34llHvq7qW0RB.99

18 “Sur Facebook : « Si vous êtes de gauche, ajoutez des gens de droite »” (On Facebook: “If you are from the left, add people from the right”), Article, Le Nouvel Observateur, November 13, 2016. Referenced on March 27, 2017. Available at: <http://tempsreel.nouvelobs.com/rue89/rue89-nos-vies-connectees/20161113.RUE4218/sur-facebook-si-vous-etes-de-gauche-ajoutez-des-gens-de-droite.html>

19 “Gérald Bronner : « Les croyances et le septicisme fragilisent la démocratie. »” (Gérald Bronner: “Beliefs and skepticism are weakening democracy”), Article, Les Echos, May 28, 2014. Available at: <https://www.lesechos.fr/28/05/2014/lesechos.fr/0202787138820-gerald-bronner----les-croyances-et-le-septicisme-fragilisent-la-democratie---.htm>

Journalists are talking about a “*post-truth*” era²⁰ to describe the fact that information shared on the Web is no longer validated by legitimate sources. For Katharine Viner, Editor-in-Chief of *The Guardian*, “[Brexit] was the first major vote in the era of post-truth politics: The listless Remain campaign attempted to fight fantasy with facts, but quickly found that the currency of fact had been badly debased.”²¹.

What we need to understand is that in this new era, the immediacy of information and its success on the social Web compete with what was once the value of a piece of information: its source, its author, its proof. Spheres whose expression was once in the minority within the public space are therefore organizing themselves in order to make their information go viral. Thus, by using techniques based on the functioning of the social Web, like astroturfing, communities such as the far right are able to make a piece of content highly visible.

THE TECHNIQUE OF ASTROTURFING, OR HOW TO BUILD AN ARTIFICIAL PUBLIC CONTROVERSY

Political propaganda¹, public controversy², poll hacking³, destabilization campaigns against a presidential candidate⁴... the different demonstrations of astroturfing having been growing in number of late. Let's decrypt this emerging practice which is defined as “all the techniques—manual or algorithmic—allowing to simulate the activity of a crowd on a social network.”⁵

In other words, it is a technique for circulating content, on social networks, that claims to be spontaneous but whose “buzz” is in reality orchestrated by coordinated persons (using accounts on Twitter, Facebook, etc.). This popularization of the event is also facilitated by the media, which prioritizes quantitative analysis of social networks when studying developments in opinion on the Web, neglecting the qualitative approach.

Astroturfing is therefore seen as a low-cost instrument for achieving relevant influence, for those who have an interest in influencing perceptions of opinion, particularly when it involves weakening social bonds. For example, in the “Reims bikini affair,” the far-right network organized itself online very quickly to politically seize the event and convey their propaganda.

20 The term was even voted “2016 International Word of the Year” by Oxford Dictionaries. “Post-Truth’ Defeats ‘Alt-Right’ as Oxford’s Word of the Year,” Article, The New York Times, Referenced on March 28, 2017. Available at: http://www.nytimes.com/2016/11/16/arts/post-truth-defeats-alt-right-as-oxfords-word-of-the-year.html?_r=1

21 “How technology disrupted the truth”, Article, The Guardian, July 12, 2016. Referenced on March 27, 2017. Available at: <https://www.theguardian.com/media/2016/jul/12/how-technology-disrupted-the-truth>. Article translated into French by Courrier International, September 9, 2016. Referenced on March 27, 2017. Available at: <http://www.courrierinternational.com/article/medias-comment-le-numerique-ebranle-notre-rapport-la-verite>

Nicolas Vanderbiest, a specialist in the issue, decrypts the process that took place¹, step by step:

1. **The news breaks.** A local newspaper announces that there has been an incident that “reeks of the religious police”;
2. **“Confused” phase and emergence of attempts at influence.** Uncertainty about the event feeds people’s fantasies. Some people then fill in the gaps (by intimating and distorting the details) in order to impose an interpretative framework on the news (anti-Islam);
3. **Creation of noise around the news.** Very quickly, the local newspaper is picked up by the “Français de souche” accounts, then the influential far right network (Gilbert Collard on this occasion) is alerted even before the information has been picked up by the national media, and that network reacts on a mass scale in order to popularize the news item;
4. **Response by opponents of the theory produced.** The association SOS Racisme then feeds the “noise,” even though it was seeking to denounce it;
5. **Media alert.** This “noise” on social networks, and not the actual news item itself, alerts the biggest national media outlet (Agence France Presse), then the others follow.
6. **The official version is disputed.** After inquiries, we finally learn that in reality it was an altercation with no religious overtones. However, this version is then disputed by nationalist activists, who base themselves on their “distorted intimations” to provide a version that corresponds with the group’s way of thinking.

More generally speaking, we can add that these digital activist actions are realized particularly through means such as:

- Creating fake accounts whose sole purpose is to (re)tweet, like and share to the maximum extent the message to be passed on;
- Creating and offering a number of “ready-to-use” tweets to make distributing propaganda easier;
- Buying likes to falsely inflate the intensity of the debate;
- Using shock images to make the message go viral more quickly.

1 “Philippe Platteau, le premier Astroturfing / Bot du FN de la présidentielle” (Philipp Platteau, the first Front National Astroturfing/Bot of the presidential election), Article, Reputatio Lab, January 16, 2017. Referenced on March 28, 2017. Available at: <http://www.reputatiolab.com/2017/01/philippe-plateau-premier-astroturfing-bot-fn-de-presidentielle/>

2 “#14h42 : du Hack d’un « sondage » de France 3 à l’astroturfing” (#14h42: from the hack of a ‘poll’ by France 3 to astroturfing), Article, Next Inpact, December 3, 2013. Referenced on March 28, 2017. Available at: <https://www.nextinpact.com/news/84687-14h42-hack-dun-sondage-france-3-a-astroturfing.htm>

3 “Il est trop tard pour s’alarmer d’une cyberguerre électorale” (It is too late to worry about an electoral cyberwar), Article, Slate, February 14, 2017. Referenced on March 28, 2017. Available at: <http://www.slate.fr/story/137273/presidentielle-cyber-guerre>

4 “Fake, manipulations et réseaux sociaux : pourquoi il faut vite comprendre ce qu’est l’astroturfing” (Fake, manipulations and social networks: why we need to understand what astroturfing is, and quickly), Fabrice Epelboin, Article, Les Inrocks, February 6, 2017. Referenced on March 28, 2017. Available at: <http://www.lesinrocks.com/2017/02/06/actualite/fake-manipulations-reseaux-sociaux-faut-vite-comprendre-quest-lastroturfing-11910209/>

5 “L’affaire du bikini de Reims ? Un astroturfing du FN et des médias à la rue” (The Reims bikini affair? Astroturfing by the Front National and the gutter media), Article, Reputatio Lab, July 30, 2015. Referenced on March 28, 2017. Available at: <http://www.reputatiolab.com/2015/07/laffaire-du-bikini-de-reims-un-astroturfing-du-fn-et-des-medias-a-la-rue/>

6 “#TelAvivSurSeine : d’un tweet à BFMTV, mode d’emploi” (#TelAvivSurSeine: from a tweet to BFMTV, quick guide), Article, Le Nouvel Observateur, August 11, 2015. Referenced on March 28, 2017? typo? Available at: <http://tempsreel.nouvelobs.com/rue89/rue89-explicateur/20150811.RUE0187/telavivsurseine-d-un-tweet-a-bfmtv-mode-d-emploi.html>

IN SUMMARY: Formerly regulated by gatekeepers, considered legitimate for the expression of opinions in the public space, the information market has undergone a revolution with the Internet. In this new environment, now unrestrained and open to all, the credibility of a piece of information, which was once its main value, now has to compete with immediacy and virality on the social Web. Information bubbles, astroturfing, the arrival of a new, so-called post-truth era and Donald Trump's alternative facts are emblematic examples of this. They indicate on the one hand a poor understanding of the problem by our elites, and on the other hand, in contrast, an excellent grip on the echo chamber that is the Internet by activist groups that did not previously have access to public spaces for dialogue.

2. HATE PHENOMENA ON THE INTERNET: THE LAW AND INSTITUTIONS GRIPPED BY THE PROBLEM



Despite legal definitions, the boundary between the reprehensible discourse of the right and freedom of expression has always been fragile. Recent cases such as Charlie Hebdo and Dieudonné show that judges are uneasy about penalizing people for their remarks, fearing that such a legal precedent would jeopardize the principle of freedom of expression, of which France and the Enlightenment claim to be the founders.

This search for balance proves complex and sometimes leads to strange legal decisions, such as one by the Paris Industrial Tribunal²² which ruled “that by considering the context of the hairdressing environment, the Tribunal considers the term ‘PD’ [roughly ‘faggot’ in English] used by the manager cannot be deemed homophobic speech, because it is recognized that hair salons regularly employ homosexuals [...] without this posing any problem.”²³

In other words, this rude, insulting and commonplace speech is at present legally positioned in an area that we can describe as “gray”: sometimes reprimanded, sometimes not, depending on whatever classification, contextualization and interpretation happens to be accepted by the judges. And yet, in the vast majority, this is the hate category on which general construction of prejudices rests.

If grasping the limits of the legality of hate speech is already a sensitive subject, on the Internet the difficulty is even greater. Putting comments, tweets or other forms of expression into context²⁴ is often very difficult: Is it humor? Provocation? Excessive anger? The assessment of the law is therefore made even more complex when the matter relates to the Internet. Furthermore, as the institutions emphasize, the non-professional nature of the remarks adds a further difficulty: “as most of the offenses are committed by anonymous non-professionals, their responsibility cannot be established.”²⁵

2.1. “HATE SPEECH”: WHAT ARE THE MAIN APPLICABLE LEGAL PROVISIONS?

When the nature of hate speech is public, we can identify two main categories of provisions that could apply and lead to the pronouncement of criminal penalties:

- Provisions under the French Law of 1881 on press freedom: Insults, libel, provocation of racial hate or discrimination, apologizing for war crimes, and questioning the existence of crimes against humanity are forms of speech which, when public²⁶, are perceived and punished as press offenses by means of the Law of 1881 on press freedom²⁷. The judge can

22 We should specify that this decision is based on an assessment of the insult by the Industrial Tribunal (which rules on labor law), and not in light of the Law of 1881, the Penal Code, or the Law on Confidence in the Digital Economy.

23 See Conseil de prud’hommes de Paris (Paris Industrial Tribunal), 4e chambre du commerce, décision RG N° F 14/14901, December 16, 2015. Referenced on March 28, 2017. Available at: <http://www.defenseurdesdroits.fr/decisions/ddd/jp/image19895.pdf>

24 For example, a Facebook Page may have an editorial policy, thereby providing the judge with some context.

25 See “La loi du 29 juillet 1881 à l’épreuve d’Internet : La mission d’information de la commission des lois dresse un bilan de la prévention des abus de la liberté d’expression sur Internet et propose un meilleur équilibre de la loi” (The law of July 29, 1881 withstands the test of the Internet: The information mission of the Law Commission reports on the prevention of the misuse of freedom of expression on the Internet and proposes better balancing of the law), Press release, French Senate, July 7, 2016. Referenced on March 28, 2017. Available at: <https://www.senat.fr/presse/cp20160707b.html>

26 When such comments are of a private nature, the provisions of the Penal Code may apply and also lead to the pronouncement of criminal penalties.

27 See the Law of July 29, 1881 on press freedom. Consolidated version as of March 28, 2017. Available at: <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000006070722>

then also apply a punishment extending as far as prison or a fine, and combine it with alternative measures (a citizenship program for some of the aforementioned offenses²⁸, for example). However, in the context of disintegrating social bonds, major lawsuits have created quite a din as they tested the limits of the interpretation of the law, with a certain degree of indulgence being granted to the media. This observation is illustrated by the 48 lawsuits brought against the satirical magazine Charlie Hebdo, which in the upshot has only been sentenced on nine occasions, mostly on the legal basis of insult²⁹.

- Provisions not related to the Law of 1881 on press freedom (apologizing for and provocation of terrorism³⁰): To say that apologizing for and provocation of terrorism fall under hate speech is to put it lightly. But it is important to place it back within the context of these issues, as it has amplified so much the gravity and urgency of channeling the liberation of hate, particularly on the Internet where jihadist propaganda enables IS to enlist new recruits. Indeed, terrorist propaganda is gaining in visibility thanks to a highly professional communication strategy on the Internet to occupy media space, communicate about their acts, and recruit new sympathizers. This is why, having been previously perceived as a press offense according to the general legal framework of the Law of 1881, remarks that apologize for terrorism have, as of November 2014 with the law strengthening provisions relating to the fight against terrorism, crossed into the category of common law criminal offense (Article 421-2-5 of the Penal Code³¹).

In this way, the penalties incurred were amended to five years of prison and a €75,000 fine³², and the procedures and methods of investigation pertaining to the fight against terrorism are now applied (for example, surveillance of electronic communications). We note that the writers of the law deemed it necessary to clarify in the explanatory statement for the draft law that “in this instance it is not a matter of restraining misuses of freedom of expression, but of penalizing events that are directly behind terrorist acts.”³³

28 This does not, however, apply to questioning the existence of crimes against humanity.

29 “« Charlie Hebdo », 22 ans de procès en tous genres” (“Charlie Hebdo”, 22 years of lawsuits of all kinds), Article, Le Monde, January 8, 2015. Referenced on March 28, 2017. Available at: http://www.lemonde.fr/societe/article/2015/01/08/charlie-hebdo-22-ans-de-proces-en-tous-genres_4551824_3224.html

30 Other provisions not related to the French law on press freedom can be applied to hate speech. For example, Article 227-24 of the Penal Code states: “The fact of either manufacturing, transporting, or distributing by any means and on any medium a message that is of a violent nature, inciting terrorism, pornographic, or seriously violating human dignity, or inciting minors to engage in games that put them in physical danger, or of trading in such a message, is punishable by three years of imprisonment and a 75,000 euro fine when this message might be seen or perceived by a minor.” At the European level, we also note the existence of the Additional Protocol to the Convention on Cybercrime, dated January 28, 2003, concerning the criminalization of acts of a racist and xenophobic nature committed through computer systems, which specifies that “any written material, any image or any other representation of ideas or theories, which advocates, promotes or incites hatred, discrimination or violence, against any individual or group of individuals, based on race, color, descent or national or ethnic origin, as well as religion if used as a pretext for any of these factors.” The main aim of this Convention being, for the member states of the Council of Europe and the other state signatories, the harmonization of the fight against racist and xenophobic propaganda. Referenced on May 17, 2017. Available at: <https://rm.coe.int/1680081610>.

31 See Article 421-2-5 of the Penal Code, established by Law no. 2014-1353 of November 13, 2014 – art. 5. Referenced on March 28, 2017. Available at: <https://www.legifrance.gouv.fr/affichCodeArticle.do?idArticle=LEGIARTI000029755573&cidTexte=LE-GITEXT000006070719>

32 When it is committed on an online public communication service, this offense is liable to a punishment of 7 years of imprisonment and a fine of 100,000 euros (which can be multiplied five-fold for legal persons). This constitutes an aggravating circumstance.

33 See “Projet de loi renforçant la prévention et la répression du terrorisme. Retour au dossier législatif” (Draft law strengthening the prevention and repression of terrorism. Feedback on legislative package), explanatory statement. Referenced on March 28, 2017. Available at: http://www.legifrance.gouv.fr/affichLoiPreparation.do;jsessionid=5C8442498B241696A27A805C3123A50B.tpdjo17v_1?idDocument=JORFDOLE000025673076&type=expose&typeLoi=proj&legislature=13

These legal concepts have therefore specified the nature of “hate speech” while assigning possible legal penalties to it. Without resolving the problem of subjectivity in the perception of such speech, this clear legal framework enables all actors from the public or private spheres, or from civil society, to be equipped to respond to or point out such speech.

CONSPIRACY THEORIES CAN EASILY ESCAPE THE LEGAL RADAR

“Conspiracy theories claim that the course of history, and its most notable moments, are not the result of chance, countless interactions or power relations, but in reality are consistently brought about by the secret activity of a small group of people who wish to see the realization of a project to control or dominate populations. These hidden characters, who pull the strings, may be groups internal to the country, agencies, minorities, internationalized groups which are organized on a worldwide level, or even extraterrestrials or monsters. Every event therefore becomes connected to a globalized mega-plot. Today, these conspiracy theories are highly developed both in the West and in the East, and are most often aligned with a political ideology that explains the world.”¹

Conspiracy theories proliferate on the Web, where they benefit from powerful methods of distribution like astroturfing. Such speech currently finds itself in a legal gray area, which is very complex to grasp by means of the law. That is to say, an instance of conspiracy theory discourse could fall within the scope of Article 24 of the Law of 1881 on the press, only if it provokes discrimination or hate toward a person, notably on grounds of their origin or their adherence to a particular religion.

Beyond such cases, conspiracy theories are not in themselves illegal: expressing theories on the veracity of facts is legal. The suppression of any hypothesis deviating from official sources (and who would define those?) would question the very founding principles of freedom of expression. However, specialists identify conspiracy theories as “a stepping stone to jihadist radicalization,”² and by extension to any other form of extremism or closed-mindedness (anti-Semitism, racism, etc.).

Faced with this kind of powerlessness in the law to prevent these threats, education—from the youngest age possible—in digital culture (economic model of online media, algorithms, astroturfing, etc.) and in critical thinking (Zetetic science, counter-speech, etc.) seems to be a viable solution. Some associations have realized this, and visit schools to show students how information and facts can be manipulated, and particularly so on the Internet.

1 Definition of conspiracy theories proposed by Emmanuel Taïeb during an interview on May 25, 2016 with Renaissance Numérique.

2 “La théorie du complot est un marchepied à la radicalisation djihadiste” (Conspiracy theories are a stepping stone to jihadist radicalization), Article, Le Journal du Dimanche, June 14, 2017. Referenced on March 28, 2017. Available at: <http://www.lejdd.fr/Societe/La-theorie-du-complot-est-un-marchepied-a-la-radicalisation-djihadiste-789945>

Certaines associations³ l'ont bien compris et interviennent dans les écoles pour montrer aux élèves comment l'information et les faits sont manipulables, et d'autant plus sur Internet.

3 To give some examples, Le Bal and Spicée visit high schools in France to carry out educational experiences with students, notably in decrypting the codes of conspiracy theory videos on the Internet in order to forearm them against such content. See Spicée's initiative: <https://www.spicée.com/fr/program/a-lecole-du-complot-847>; See the initiative by the Le Bal association: <https://vimeo.com/166931978>

IN SUMMARY: When hate speech leads to discrimination, possible problems with public order, or calls to commit terrorist acts, the law naturally and rightfully qualifies them and punishes them. In practice though, the boundary between freedom of expression and illegal comments is often fragile. Thus, the vast majority of hate speech, particularly that online, is not de facto illegal and finds itself in a subjective legal gray area, being difficult for the law and for reporting tools to grasp. And yet, it is on this basin of rude, insulting or conspiracy-theory speech that prejudices and mistrust of our institutions are built.

2.2. "HATE SPEECH": WHO IS RESPONSIBLE IN THE EYES OF THE JUDGE?

Following on from the definition of the boundary that marks out hate speech which oversteps legality, there is the issue of who is responsible for the distribution of such speech. Bearing in mind the social networks (or hosts like blogs or forums) which distribute the content and thus give it visibility, its authors, the people who share it... the law has had to evolve since the arrival of the Internet. And so, in June 2004, the law on confidence in the digital economy³⁴ (the "LCEN"³⁵), which transposed Directive 2000/31 (the Directive on electronic commerce³⁶) into French law, made the Law of 1881 on the press applicable to "all means of communication to the public by electronic means."

Through this same law (LCEN), the legislator also clarified the framework of liability of the actors involved³⁷: the publishers of the content are the main responsible parties for their hate speech on the Internet. They are the ones who are liable for punishment if their speech is recognized as illegal. Whether anonymous or not, illegal content can be punished through the provisions cited in the previous section.

34 See Law no. 2004-575 of June 21, 2004 on confidence in the digital economy (the "LCEN law"). Consolidated version as of March 28, 2017. Available at: <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000801164>

35 See LCEN law, idem.

36 See Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ("Directive on electronic commerce"). Referenced on March 28, 2017. Available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32000L0031&from=EN>

37 By transposing into the French law Articles 14 and 15 of the European Directive on electronic commerce (Directive 2000/31/EC)

THE SYSTEM OF RESPONSIBILITY OF INTERNET ACTORS

On the Internet, the law distinguishes between two main families of responsibility: the general system (the responsibility of the Internet user) and the specific system (the responsibility of certain Internet actors). In the latter case, the LCEN of 2004 distinguishes the system of responsibility according to which the Internet actor is considered a publisher or a host:

The publisher is defined by the law as the person whose activity is to publish “an online public communication service.” In other words, according to case law, the person who decides which content should be made available to the public on the service that the person set up or is responsible for⁷. This is the case, for example, for an Internet user who has a blog. The publisher is then held responsible for all the content (even if they are not the author) that appears on their site.

The host is defined as the person who provides, even if free of charge, a service of storing the information supplied by the people at whom the service is aimed. This actor benefits from a system of limited responsibility, that is, they can only be held responsible for content stored on their server if, and only if:

- The actor has been alerted about illegal content on their site;
- They have not removed it promptly, despite having been ordered to do so.

The principle of hosting is dealt with by the LCEN (see box above), which states that these actors do not bear civil and criminal responsibility for the messages distributed. However, the preeminent role played by social networks in the visibility of hate speech, through the scale of use of their tools, gives them a certain degree of responsibility in the eyes of the law. While they are not bound by a general obligation to monitor the content they store (unlike publishers of content), they are legally obliged to remove content that is clearly illegal when such content is pointed out to them. If the hate content is illegal (videos, images, messages, etc.), the complainant can act in accordance with two different procedures³⁸:

- **The non-judicial procedure:** This is the reporting procedure belonging to each host. For example, Facebook sets its own removal conditions via its “Community Standards”³⁹. And yet, these terms do not necessarily reflect the law applicable in France on the matter, rather they generally reflect the legal regulations in the United States, where these platforms mostly come from. However, while these standards are colored by certain cultural origins, we note that the social network does grant primacy to French law when the content respects its standards but contravenes French law.
- **The judicial procedure:** In this case, the procedure is stipulated by French law and not by the rules set by each host. This procedure invites the claimant to ask the author of the hate content to remove it, and if the author fails to do so, the claimant can advise the host of this. Thus, once the host has been duly informed by the claimant, they must

³⁸ “Responsabilité des contenus publiés sur internet” (Responsibility for content published on the Internet), Service-Public.fr, July 6, 2016. Referenced on March 28, 2017. Available at: <https://www.service-public.fr/particuliers/vosdroits/F32075>

³⁹ See “Community Standards”, Facebook. Referenced on March 28, 2017. Available at: <https://www.facebook.com/communitystandards>

remove the “clearly illegal” content promptly and, in some cases, inform the relevant authorities (see below), otherwise they may be deemed legally responsible in the event of willful negligence. The effectiveness of this procedure is dependent on the host’s ability both to remove the content reported to them and to establish whether or not it is “clearly illegal”.

IN SUMMARY: The issue of responsibility is primordial when the law seeks to regulate an area of activity. With regard to the activity of disseminating hate speech, the scheme of responsibility has had to evolve with the emergence of the new spaces for dialogue popularized by the Internet. Thus in 2004, the LCEN law, transposing the European Directive on e-commerce, included provisions to:

- *Make the legal provisions in this area applicable to digital technologies;*
- *Organize the system of responsibility of Internet actors, taking into account the specific features having arisen as a result of this network;*
- *Establish mechanisms for reporting content so that illegal content can be removed from the online environment.*

2.3. THE TRIPARTITE ARRANGEMENT OF THE RESPONSE TO THE PHENOMENON OF HATE

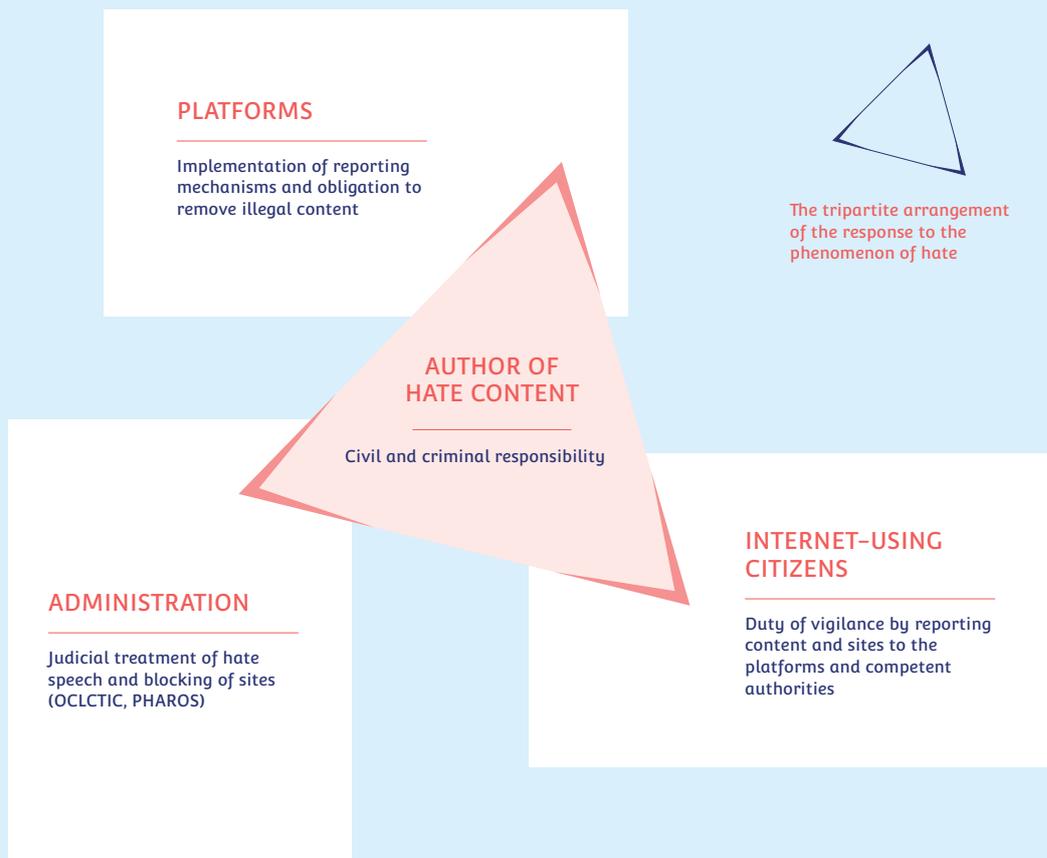
Beyond the tools proposed by hosts, the French state has strengthened the system for reporting hate speech on the Web by setting up its own platform dedicated to reporting: PHAROS⁴⁰. The status of competent authority in the fight against cybercrime was conferred to the Central Office for the Fight against Crime linked to Information and Communication Technologies (OCLCTIC), which is staffed by police officers and gendarmes. This authority was made responsible for examining any content that is reported, in order to verify whether it constitutes an offense and, if necessary, to transfer the case to the competent authorities in order to bring proceedings against its author, require the removal of the content, and possibly order a legal injunction or administrative freeze on the site concerned.

Beyond legal injunctions, the response to the phenomenon of distributing hate on the Internet is primarily organized around three main actors:

- Internet users, whether they are the authors of the speech in question and therefore responsible for it, or actors involved in regulation by reporting it either to the sites concerned or on the PHAROS platform;
- The OCLCTIC through the PHAROS platform, which receives requests and calls on the hosts concerned (notably social networks) in order to demand removal of the content;
- The hosts, who must provide the possibility to report content and to remove clearly illegal comments.

⁴⁰ “Questions et réponses sur le signalement” (Questions and answers on reporting), Internet-Signalement.gouv.fr. Referenced on March 28, 2017.

Available at: <https://www.internet-signalement.gouv.fr/PortailWeb/planets/Faq.action;jsessionid=4E90EEF40F983884A9504F-9D5EADE95D>



***IN SUMMARY:** The response to hate on the Internet is today based essentially on a tripartite arrangement of the actors in the fight against hate: Internet-using citizens as actors involved in regulation by means of reporting, hosts which process the reports and remove content, and the administration by means of the OCLCTIC and its PHAROS platform which also processes reports and, if necessary transfers the case to the competent authorities.*

2.4. THE LAW: A NECESSARY BUT INSUFFICIENT TOOL

The law's response is absolutely legitimate and necessary. And all the more so since legal pressure has played a role in boosting the speed of the implementation and effectiveness of mechanisms to fight against hate on the Internet.

From court summons to effective reporting mechanisms: A court summons seems to be the preferred means among human rights associations of obliging hosts to contribute and cooperate significantly in the fight against hate on the Internet.

- For example, the High Court (TGI) of Paris, in a referee's order dated April 13, 2016⁴¹, made a finding in favor of the associations UEJF, J'accuse, Licra, Mrap and SOS Racisme by ordering the website Egaliteetreconciliation.fr to implement a reporting mechanism which it was then lacking, even though the site regularly published illegal content.
- In addition, case law established in the “#UnBonJuif” (#AGoodJew) case by the UEJF against Twitter, on January 24, 2013 (TGI Paris), led to an improvement in the effectiveness of the social network's reporting mechanism. Although Twitter had such a mechanism, it was only available in English and therefore needed to evolve in order to be “easily accessible and visible⁴²” for French users.
- Finally, human rights associations (SOS Homophobie, SOS Racisme and UEJF) decided in May 2016 to launch “a mass testing phase” to evaluate the efficiency of these mechanisms. The results of a study carried out at the end of this revealed a lack of transparency and responsiveness in the platforms' management of reports⁴³. These associations then decided to adopt this same strategy of launching proceedings, taking the social networks to court again to obtain more transparency and effectiveness in their policies on reporting and moderation.

This pressure induced by an offensive mobilization of the judicial system has made it possible to improve the efficiency of solutions to restrain hate on the Internet. But these solutions do of course still need to gain further in effectiveness. With this in mind, “trusted reporter” mechanisms are implemented: they consist of giving third-party associations or people a form of priority in their reporting. For example, because Twitter knows that a report has come from SOS Homophobie, it will pay special attention to the request in order to process it more quickly, and will exercise more trust in making a judgment. Gilles Dehais, CEO of SOS Homophobie, emphasizes “*[unlike the current reporting system] an effort in processing when the hate speech is reported by the association.*”⁴⁴

41 “Egalité et réconciliation condamnée pour absence de dispositif de signalement” (Egalité et réconciliation punished for not having a reporting mechanism), Article, Legalis, April 15, 2016. Referenced on March 28, 2017. Available at: <https://www.legalis.net/actualite/egalite-et-reconciliation-condamnee-pour-absence-de-dispositif-de-signalement/>

42 Tribunal de Grande Instance de Paris (High Court of Paris), Referee's Order, N°RG 13/50262 13/50276, issued on January 24, 2013. Referenced on March 28, 2017. Available at: <https://cdn2.nextinpact.com/medias/ordonnance-tgi-paris-24-janvier-2013-uejf-vs-twitter.pdf>

43 “Des associations vont assigner en justice les trois géants de l'Internet américain” (Associations to take the three US Internet giants to court), Article, Le Monde, May 15, 2016. Referenced on March 28, 2017. Available at: http://www.lemonde.fr/pixels/article/2016/05/15/l-uejf-et-sos-racisme-vont-assigner-en-justice-twitter-youtube-et-facebook-pour-leur-manque-de-moderation_4919885_4408996.html

44 Press Kit, Les Assises de la lutte contre la haine sur Internet 2ème édition (The foundations of the fight against hate on the Internet, 2nd edition), May 11, 2016

While it remains the main instrument mobilized in these matters to influence behaviors, the legal system does face numerous obstacles in this area. Beyond the traditional difficulties inherent to the debates that can arise from the interpretation of the principle of freedom of expression, the very nature of the Internet has made the mechanism for regulating hate speech by legal order even more complicated:

- The virtual nature of the Internet versus legal interpretation: Whereas in real life it is easier to interpret the intention behind a message in the light of the identity and thoughts of the typical authors of the event⁴⁵, the digital aspect adds its own set of difficulties: anonymity, written and virtual nature, coded racist messages⁴⁶, etc.
- The immediacy and virality of the Internet versus legal time: The continuous flow of comments on the Internet, notably hate speech, naturally calls into question the capacity of a legal order to assess the entirety of illegal speech online.
- The ubiquity of the Internet versus legal territoriality: The international and networked nature of the Internet makes it less efficient to apply a territory's law in order to regulate the issue on its own and in an effective manner.

Furthermore, the reactive and coercive use of the justice system by public decision-makers, often due to the media emulation that surrounds these issues, has led them to promote policies whose potential consequences are cause for concern:

- The legislative inflation⁴⁷ that has underpinned the supervision of the Web for the past decade, and more particularly the supervision of the fight against terrorism on the Internet, is a good illustration of this. The Internet is identified by decision-makers, in the name of terrorism, as a cradle for recruitment and the planning of attacks, which leads them to promote a reactive, slippery and dangerous legislative policy (Law on military planning, Law on intelligence, etc.) aimed at censoring it and rolling back what we have won through democracy. One of the consequences of this policy is to move the problem to the dark net, which is the most difficult part of the Internet to access.
- The privatization of justice: As previously explained, platforms are not responsible for content published on the Internet, but they do have, on the other hand, an obligation to remove clearly illegal content that users notify them of via the reporting tools. We may note that in making them bear the load of assessing whether content is clearly illegal or not, we must at the same time consider the risks that might stem from a privatization of justice.

45 "Attentats : Zemmour accusé d'apologie du terrorisme par les familles de victimes" (Attacks: Zemmour accused of being an apologist for terrorism by the families of victims), Article, L'Express, October 6, 2016. Referenced on March 28, 2017. Available at: http://www.lexpress.fr/actualite/societe/zemmour-respecte-les-terroristes-prets-a-mourir-une-apologie-du-terrorisme_1838171.html

46 "Pour éviter la censure, les racistes américains remplacent leurs insultes par des codes" (To avoid censorship, American racists replace their insults with codes), Article, Slate, October 3, 2016. Referenced on March 28, 2017. Available at: <http://www.slate.fr/story/124475/racisme-google-yahoo-skype>

47 "Huit lois en dix ans pour encadrer le Web français" (Eight laws in ten years to supervise the French Web), Article, Les Décodeurs (Le Monde), April 15, 2015. Referenced on March 28, 2017 Available at: http://www.lemonde.fr/les-decodeurs/article/2015/04/15/sept-lois-en-dix-ans-pour-encadrer-le-web-francais_4615841_4355770.html

All the more so as all other things being equal, the difficulty in contextualizing and interpreting comments made online (recognition of black humor, the absence of tone, etc.) is just as valid for judges as for platforms' moderating teams. Although today there are systems that enable algorithmic recognition of content relating to pedophile pornography (images, videos, etc.) in order to facilitate their removal (which even then is not easy⁴⁸), the progress made in data mining is still insufficient to detect—automatically and effectively—all of the hate content in the online environment (for all of the previously cited reasons). Some experiments under way, however, suggest there may be positive developments in this area going forward, such as Google's "Perspective" project, which uses technology to help detect aggressive comments in order to moderate them⁴⁹.

***IN SUMMARY:** The law, through a combination of legislation and court proceedings brought by associations, has enabled the implementation and continuous improvement of reporting mechanisms to be made obligatory. However, the current approach, which consists of prioritizing regulation of the problem by focusing resources on the law, is running into major pitfalls. On the one hand, the coercive mobilization of the judicial system leads public actors to promote worrying policies (legislative inflation, risk of privatization of justice). On the other hand, the very nature of the Internet has complicated even further the regulatory mechanism of the law: thus, legal interpretation is confronted with the Internet's virtuality; legal time with its immediacy; and the territorial nature of law with its ubiquity. For these reasons, without calling into question respect for the law, whose common rules form the glue of society, it is necessary to encourage additional solutions to emerge and stand in for the role of the judge in this matter.*

48 "Après avoir censuré la Première ministre de Norvège, Facebook revoit sa politique" (After having censored the Prime Minister of Norway, Facebook reviews its policy), Article, Numerama, September 12, 2016. Referenced on March 28, 2017. Available at: <http://www.numerama.com/politique/193972-apres-avoir-censure-la-premiere-ministre-de-norvege-facebook-revoit-sa-politique.html>

49 "Modération des commentaires : Google propose un coup de pouce de l'intelligence artificielle" (Moderation of comments: Google proposes a little help from artificial intelligence), Article, Le Monde, February 23, 2017. Referenced on March 28, 2017. Available at: http://www.lemonde.fr/pixels/article/2017/02/23/moderation-des-commentaires-google-propose-un-coup-de-pouce-de-l-intelligence-artificielle_5084356_4408996.html

3. FIGHTING AGAINST HATE ON THE INTERNET: IN FAVOR OF THE COLLABORATIVE ORGANIZATION OF MEASURES



Assert a vision of social togetherness to get all actors of civil society on board

Provide civil society with the means (financial means in particular) to coordinate a citizens' response in the face of hate

PUBLIC ACTORS

Punish the authors of illegal comments

Offer citizenship modules

JUSTICE

CITIZENS

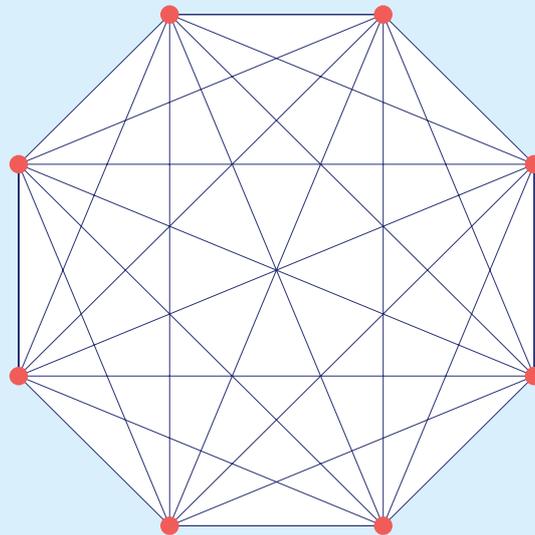
Beyond freedom of expression, a duty of expression when faced with hate

Seize channels of expression to convey positive messages

NATIONAL EDUCATION

Digital literacy and assimilation for the very young

Encourage digital civic-mindedness and critical thinking by mobilizing Internet



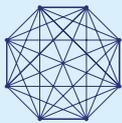
ASSOCIATIONS

Produce connection and positive content to facilitate the organization of civil society and good use of counter-speech

Encourage feedback and vigilance in the field

MEDIA

Assume a social role, notably of the community managers



In favor of the collaborative organization of measures

PLATFORMS

Inform people about their platform's codes of communication in order to facilitate the virality of positive content

Offer modules reminding people of democratic principles

Facilitate access to reporting tools

Establish an approach of cooperation with the justice system

RESEARCH ACTORS

Provide long-term reflections, shed light on the debate

Provide tools to enable civil society

Publish and monitor the development of hate speech (propagation, ideologies, risks, etc.)

Measure the effectiveness of measures implemented

Advise the public authorities and civil society on which strategies to adopt

“To put an end to the propagation of hate on the Internet, let's close the Internet down!” This is often the reaction from the authorities with regard to various subjects relating to the Internet. But focusing on the consequences of a problem rather than its causes is an inadequate strategy, which consists more in hiding the problem than resolving it. At best, such an approach moves the problem to the “deep Web” or “dark net” or toward physical organizations, which are much more difficult to identify and counter. Either way, sooner or later it will be necessary to treat the causes that lead citizens to insult each other on the Internet or develop an interest in jihadist dogmas.

Rather than focusing solely on a repressive approach, it would be preferable to favor a multi-actor method, by relying on digital codes and a “Community Organizing” type of approach, in order to coordinate and benefit from the complementarity of the actors and actions.

3.1. THE MASSES: ORGANIZING THE SPONTANEOUS CITIZENS' RESPONSE

At a time when an overly anxious vision of digital dominates, wouldn't it be better to rely on the Internet's assets, on the premise of a simple observation: tolerance is increasing in the majority of the population. Indeed, "it seems that, since the recent wave of terrorist attacks, and despite the discourse of some public figures, French society refuses conflation and values the acceptance of the other"⁵⁰ reports the French National Advisory Committee on Human Rights (CNCDDH).

The remarkable post-Charlie citizen momentum of January 7, 2015 is evidence of this. The sharing of the "je suis Charlie" drawing—the work of a then-anonymous illustrator—and the sharing of the hashtag #JeSuisCharlie was a historic milestone in the history of the social Web. Citizens have the ability to mobilize and act as a community when a simple channel to express themselves is offered to them (a hashtag, drawing, etc.).

There is one community that has understood particularly well this formidable cohesive force which is specific to social networks, and that is the online far-right movement. Thus, to serve out their ideas, they have the capacity to mobilize Facebook and Twitter accounts on a mass level, and to organize particularly formidable disinformation campaigns. We saw this, for example, with the false accusations and smear campaign against Alain Juppé's policy on Islam during the mainstream right's campaign for the presidential primary. So, the strong capacity to mobilize Web 2.0 benefits first and foremost from this community, to the detriment of the republican, secular, respectful forces, which are undoubtedly greater in number but less well organized.

The "Renard du Net" Twitter account illustrates this strong capacity to mobilize. Under the cover of fighting against "anti-white racism and anti-France" comments, the "Renard du Net" Twitter account has developed a highly effective digital strategy. Monitoring, curating content, astroturfing, creating dedicated hashtags, this anonymous team has accumulated numerous techniques to engage its community to list and then "chew up" (from reporting it⁵¹ to PHAROS to online lynchings) any content that it considers does not correspond with patriotic values. In the virtual world, it is a real "civilian militia," with all the questions of legitimacy and ineffective public action that this situation poses. Their denunciations and pressurizing have no valid legal basis and are largely guided and supported by extremist messages and communities. So far, the account has gained more than 19,000 followers in little over a year of existence, and has spawned a real troop of net vigilantes: #TeamRDN (recognizable by this symbol: V), which includes in its number, for example, "Cigogne du Net," who "busts" "those who show contempt for life, the family and desecrancy," in other words, comments deemed "pro-abortion."

50 Ibid, Rapport Annuel 2015 (2015 Annual Report), p.4, CNCDDH.

51 The excellent understanding of the mechanisms of digital enables this type of net activist to skillfully hijack the purpose of some of the systems in place, notably reporting. For instance, in order to process reporting more effectively, accounts and inappropriate content are suspended when the number of reports surpasses a certain figure. In regard to this, Gérald Bronner reminds us that "this approach to reporting enables a downward spiral because the most motivated people are given leave to act. For example, researchers proposing rational studies on vaccines have had their account deleted following swarm attacks by anti-vaccine campaigners." Comments by Gérald Bronner while participating in the event "Contre-discours : comment lutter contre les discours de haine et l'extrémisme violent ?" (Counter-speech: how to fight against hate speech and violent extremism) organized by Facebook in Paris on March 3, 2017.



Screenshot of Florian Philippot's Twitter account on May 10, 2017
https://twitter.com/f_philippot/status/784804365005619201



Screenshot of Renard du Net's Twitter account on May 10, 2017
<https://twitter.com/RenardDuNet/status/803863058548461568>

While a part of the French population is capable of such cohesion, even without the support of the media, public actors or the social networks themselves, imagine what could be done by an entire, much larger community, which would be recognized behind a distinctive sign symbolizing republican values. It is therefore high time to awaken and reawaken the silent voices that advocate cohesion. The engagement of millions of citizens needs to be understood as a determining factor, as it is the source of viral phenomena. Like the well-structured organization of the authors of hate speech, a citizens' response needs to be put in place. We note that the French are aware of their role, as 59% of them consider themselves to be the most legitimate actors to fight against hate speech⁵²

EXPERIMENTING WITH A CITIZENS' RESPONSE UNIT ON THE INTERNET

When the victims of a terrorist attack are commemorated, it is unfortunately a time conducive for the expression of online hate. For this reason, on January 7, 2017, on the occasion of the second anniversary of the January 2015 attacks, Renaissance Numérique experimented with organizing a citizens' response unit to engage in mobilization actions on social networks.

This citizen's unit, made up of several associations engaged in counter-speech, had the objective of:

- Monitoring the relevant Twitter accounts and Facebook pages;
- Promoting positive content circulating on the net;
- Responding in a calm and synchronized manner (likes, retweets, comments) to people promoting hate and terrorism, with support from the Seriously tool.

Having been convinced by the relevance of this approach and the day's results, our ambition is now to increase and structure this online community in the aim of having these republican mobilizations target social networks on a more permanent basis¹.

1 The experience of this citizens' response unit was repeated, structured under the hashtag "#FraterniTeam" for a second experiment this time in a different context: a quiet time, on Friday, May 5, 2017.

The difficulty is solely in developing incitements, both short and long term, enabling the emergence of a sense of individual responsibility in citizens within their close circles (both digital and physical). The popularization of spaces for speech, made possible by digital, invites us as citizens to make freedom of expression evolve into a duty: the duty to grasp digital channels of expression in order to convey positive counter-messages. However, this empowerment of society will depend on their tools and training, and so the role of the education system would appear naturally as a decisive factor.

52 Ibid, Google France poll by Institut CSA.

3.2. THE EDUCATIONAL SYSTEM

Education is the corner stone of the fight against the prevalence of stereotypes and prejudices. All the more so since studies in neuroscience have established that the neuronal structure, which is connected to fear and emotion, and therefore prejudices, is not innate in humans, but rather its activity develops during adolescence⁵³.

The national education initiative “*Onze mesures pour une grande mobilisation de l'École pour les valeurs de la République*” (*Eleven measures for a great mobilization of schools in favor of the values of the Republic*) marks the start of a growing awareness, particularly the measure that aims to create an “*educational path to citizenship*” by means of “*education in the media and information, fully taking into account the challenges of digital and its uses*.”⁵⁴

Digital is an extraordinary democratic advance in terms of access to knowledge, as it has never been so easy for us to discuss and distribute our ideas. The rise of an unrestrained freedom of expression thanks to social networks and Civic Tech tools thus makes it possible to open the way to Jürgen Habermas's discursive democracy⁵⁵. That is, citizens' informed and active participation in the public space and in political decision-making, in other words the possibility to negotiate in order to “reach a consensus on ethical or political norms through well-argued debate⁵⁶”. To grasp the opportunities of these new conversational spaces, a more ambitious digital education is needed, including:

- The technical architecture of the network: its horizontal and decentralized nature, its resilience;
- The organization of information on the Internet: algorithms that govern the visibility of content;
- The economic model of online media and how they organize information;
- And beyond digital, shining a light on the cognitive processes associated with processing a piece of information.

The whole challenge lies in supporting and equipping children's training in order to awaken their critical thinking. To this end, digital offers new educational facilities (scientific You-Tubers, MOOCs, practical educational tools for critical thinking, serious games, etc.), which national education should take advantage of in order to try new ways to teach and practice civic education.

53 “Les préjugés nous collent à la peau, mais tout n'est pas écrit...” (Prejudice sticks to us like glue, but isn't pre-destined...), Article, The Conversation, October 16, 2016. Referenced on March 28, 2017. Available at: <https://theconversation.com/les-prejuges-nous-collent-a-la-peau-mais-tout-nest-pas-ecrit-65669>

54 “Onze mesures pour une grande mobilisation de l'École pour les valeurs de la République” (Eleven measures for a great mobilization of schools in favor of the values of the Republic), Press kit, Najat Vallaud-Belkacem – January 22, 2015. Referenced on March 28, 2017. Available at: <http://www.education.gouv.fr/cid85644/onze-mesures-pour-un-grande-mobilisation-de-l-ecole-pour-les-valeurs-de-la-republique.html>

55 Ibid. Habermas J., 1996

56 “Théorie de l'agir communicationnel” (The Theory of Communicative Action), Jürgen Habermas, 1981, Fr. trans. 1987, republished. Fayard, 2001 for the trans. I, Fayard, 1997 for the trans. II. C. H. Article, Sciences Humaines, September 1, 2003. Referenced on March 28, 2017. Available at: http://www.scienceshumaines.com/theorie-de-l-agir-communicationnel_fr_13095.html

3.3. ASSOCIATIONS

The work carried out in the field on a daily basis by associations makes these organizations key actors in living together, “doing together” and the affirmation of republican values. Furthermore, associations benefit from a high level of trust from public actors and citizens. According to a recent survey⁵⁷, 52% of individuals think associations are best placed to provide useful information to fight against hate.

Associations have three levers that can easily be activated in order to increase the effectiveness and strike force of their missions:

1. **Observe the field in order to feed research.** From a collaborative perspective, the space they occupy on a daily basis in the field makes them a powerful observatory for the analysis of trends in hate, which can be used in the work of research organizations.
2. **Equip citizens in order to remobilize them on the Internet.** For these actors, this transition also implies fully taking into account the uses of digital by producing positive content which would then be relayed on a mass scale by Internet-using citizens.
3. **Form networks in order to scale up and reach a critical mass.** Working in silos is one of the great pitfalls of the non-profit sector. Initiatives abound, but all too often remain little known by their peers, as they are confined to detrimental cliques. A more collaborative approach would enable them to pool resources for some similar and/or complementary initiatives, and to establish networks in order to magnify the power and level of distribution of their actions. To do so, it therefore seems important for digital actors to continue their work of training associations by making their skills available to them in order to gradually support them in their digital transition.

57 Ibid, Sondage, Google France by Institut CSA.

3.4. MAJOR DIGITAL ACTORS

The LCEN law, which transposes the European Directive on Electronic Commerce, implemented the principle of reduced civil and criminal responsibility for hosts of content (see *Hosts of content and the obligation to remove illegal content published on the Internet*). These actors have been compelled to put in place reporting buttons, legally obliging them to remove illegal content that has been reported to them, thereby making them responsible in the event that they do not take action promptly or if they leave such problematic content online.

On May 31, 2016, four major Web actors (Facebook, Microsoft, Twitter and Google) agreed before the European Commission to propose a series of measures to fight the propagation of hate speech in Europe⁵⁸. Less than one year following this announcement, the European Commission again criticizing these actors on the basis that they were not taking action quickly enough to delete reported messages⁵⁹.

PLATFORMS FAVOR A COLLABORATIVE APPROACH WITH ASSOCIATIONS

Training associations in digital codes: For example, via its “Google Grants” program, the company makes its technical know-how available free of charge to associations in order to maximize the impact of their counter-speech campaigns:

- The Google Ad Grants program provides training in AdWords for associations to promote their website through keyword targeting.
- The YouTube mentoring program provides training in the conventions for communicating with the tool, and makes the YouTube Space filming studio available so that associations can create positive content for their campaigns.
- As well as the Google Ad Grants program, another program in collaboration with the start-up Moonshot CVE¹ aims to redirect aspiring jihadists toward counter-propaganda content when they enter one of 1,700 selected keywords. This initiative will soon be extended to include other types of extremism.

1 “Après les djihadistes, Google souhaite dissuader les extrémistes de droite tentés par la violence” (After jihadists, Google wants to dissuade far-right extremists tempted by violence), Article, Le Monde, September 9, 2016. Referenced on March 28, 2017. Available at: http://www.lemonde.fr/pixels/article/2016/09/09/apres-les-djihadistes-google-souhaite-dissuader-les-extremistes-de-droite-tentes-par-la-violence_4995071_4408996.html

58 See “Code of conduct on countering illegal hate speech online”, European Commission. Referenced on March 28, 2017. Available at: http://ec.europa.eu/justice/fundamental-rights/files/hate_speech_code_of_conduct_en.pdf

59 “Discours haineux : la Commission européenne rappelle à l’ordre les réseaux sociaux” (Hate speech: the European Commission calls social networks to order), Article, Le Monde, December 5, 2016. Referenced on March 28, 2017. Available at: http://www.lemonde.fr/pixels/article/2016/12/05/discours-haineux-la-commission-europeenne-rappelle-a-l-ordre-les-reseaux-sociaux_5043390_4408996.html

Bring associations together by organizing events: Whether it's Facebook with its Counter-speech Lab, Google and the Net Positive project, or Twitter with its TweetUP event, digital actors are organizing more and more events devoted to these issues in order to promote synergies between the actors and their initiatives.

Improve access to, speed and effectiveness of moderation tools. As part of their commitments to the European Commission, agreed with the signing of the Code of Conduct, the four digital actors committed to delete illegal hate content within 24 hours. They also promised to help associations become "trusted third parties" so they can "contribute high-quality reporting" as part of a network at the European level. The implementation of this process—based on trust—therefore requires collaboration with the associations (notably in order to train them in the reporting tools). We note in particular the approach taken by Facebook in this matter, as this trusted status which is granted to associations enables the social network to target specific problems within the French framework, and then roll out their analyses more widely (when similar reports are made by other users). These mechanisms are currently still finding their feet, but are destined to be improved further following testing, thanks to new relationships of cooperation, notably with the justice system (see below). Indeed, in response to the Commission's warning, these platforms announced their alliance revolving around a shared database of illegal content associated with terrorism in order to improve moderation².

2 "Contenus terroristes : Facebook, Microsoft, Twitter et YouTube font base de données commune" (Terrorist content: Facebook, Microsoft, Twitter and YouTube set up a shared database), Article, ZDNet, December 6, 2016. Referenced on March 28, 2017. Available at: <http://www.zdnet.fr/actualites/contenus-terroristes-facebook-microsoft-twitter-et-youtube-font-base-de-donnees-commune-39845584.htm>

A margin of progress is certainly possible for the Web actors, which will always prefer to take action toward the empowerment of actors capable of contributing counter-speech, rather than toward censorship. Thus, for two years now, Google, Facebook and Twitter have met frequently with associations in order to make their tools available to the non-profit community: training in social networks and in reporting mechanisms, access to free advertising, etc.

3.5. THE JUSTICE SYSTEM

Although this paper places an emphasis on all the alternatives to regulation by law for this issue, the justice system's role in penalizing authors remains essential. However, the usefulness of alternative punishments such as citizenship programs should not become obscured by these necessary but repressive measures.

The justice system has been invited to promote collaborative approaches. Last December, at a conference organized by the Association of Community Internet Services (ASIC)⁶⁰, repre-

60 See report from the morning of talks "Hébergeurs : arbitres ou gardiens de la liberté d'expression en ligne ?" (Hosts: arbitrators or guardians of freedom of expression online?) dated December 1, 2016. Referenced on March 28, 2017. Available at: <http://www.lasic.fr/?p=911>

representatives of the platforms emphasized the difficulty for hosts in assessing the legality of some hate content, notably because the human aspect of the reporting process should not be underestimated. Taking into account these observations, Benoit Tabaka, Senior Policy Manager at Google Europe, called for moves to boost the cooperation and joint establishment of solutions with the justice system in order to ensure their effectiveness.

3.6. THE MEDIA

The rise of the Web brought about deep changes in our relationship with information, as explained previously. Journalism as a profession is evolving too, keeping pace with the Internet. By journalists' own admission:

- 90% of them say they prioritize speed of analysis in their work on social networks.
- On average, 62% of French journalists feel that social networks have contributed to the degradation of traditional journalistic values, like objectivity⁶¹.

These figures partly explain why journalists may lament some of the effects of the social Web. But disengaging the media from its responsibility by focusing criticism solely on the emergence of social networks would be too simplistic.

Indeed, due to their mistrust of the Internet, which was fed into by their legitimate concerns about their economic model, the media did not take up digital tools early enough or quickly enough. They did not succeed in taking the influence they had in the real world and recreating it in the digital world. Today, the media needs to make up for this deficit of credibility and conviction toward the population who express political choices that differ from theirs and who do so justly in order to express their mistrust of media institutions and politics.

Some attempts are made to inform the public differently, taking these changes into account, for example *factchecking*. This trend began a few years ago, but recent votes in favor of Donald Trump and Brexit unfortunately constitute evidence of its low level of effectiveness.

The media needs to take on the fight for facts again and find new ways to measure public opinion in order to decipher trends that are visible on the Internet (to counter a mounting controversy, avoid falling into the trap of astroturfing, etc.). More specifically, these actors should observe public opinion via the Web, by mobilizing tools for observing social networks and putting in place tools for *monitoring* the popularity and propagation of content.

61 "Plus de la moitié des journalistes ne peut plus se passer des réseaux sociaux" (More than half of journalists can no longer do without social networks), Article, French Web, September 20, 2016. Referenced on March 28, 2017. Available at: <http://www.frenchweb.fr/plus-de-la-moitie-des-journalistes-ne-peut-plus-se-passer-des-reseaux-sociaux/256956>

3.7. RESEARCH ACTORS

Research organizations should put their capacity to analyze social phenomena in the service of social actors. For them, this implies:

- Measuring the propagation and dynamics of hate speech on the basis of information collected through daily observation by associations in the field;
- Evaluating the influence and effectiveness of counter-speech actions;
- Shedding light on the debate by putting it into the perspective of the long term and moderation of reflections;
- Assuming an advisory role for public authorities and civil society by identifying the right forms of behavior and policies to adopt in order to keep hate in check;
- Assessing the impact and effectiveness of actions implemented;
- Providing the frame of reference and specific tools allowing to organize the citizens' response;
- Setting up spaces for discussion and relationships of collaboration with media and digital actors in order to build solutions together that will allow to stem disinformation⁶².

62 "A Call for Cooperation Against Fake News", Jeff Jarvis, Medium, November 18, 2016. Referenced on March 28, 2017. Available at: <https://medium.com/whither-news/a-call-for-cooperation-against-fake-news-d7d94bb6e0d4#.8pcmsx11l>

CONCLUSION



When talking about the future, French people seem, both individually and collectively, to fall into a profound pessimism. For instance, only “13% of them anticipate that the different groups making up society will be able, over the next ten years, to live together with good relationships, whereas 73% expect tensions,” even though the reality of the figures (inequality, unemployment, etc.) “do not justify such a dark outlook⁶³.” These figures are worrying in terms of our ability to come together as a society.

This social fatigue finds justification in the growing mistrust toward political, economic and media institutions. Even well before the Internet, it was this fertile ground for disillusion that would lead some of our frustrated youth to get involved in violent political counter-movements. At this point, let us remember that terrorism is first and foremost a means of action, a military mode of operation, which serves a political ideology. The security arsenal can only be a short-term response. To find long-lasting stability, a comprehensive solution needs to be envisaged.

Antonio Gramsci defines crisis as “the old world dies, the new world takes time to appear, and in this half-light the monsters emerge⁶⁴.” The best way to fight the monsters is therefore on the field of ideas, of opportunities available. Thus, the urgent issue for political decision-makers consists in gathering together, giving back meaning to living together in the long term.

To achieve this, it is necessary to fully measure the strengths and opportunities of digital, and stop stigmatizing the Internet. At a time when everyone is now free to express their opinions and take part in many debates, society’s digital networking opens the way forward to a truly deliberative democracy.

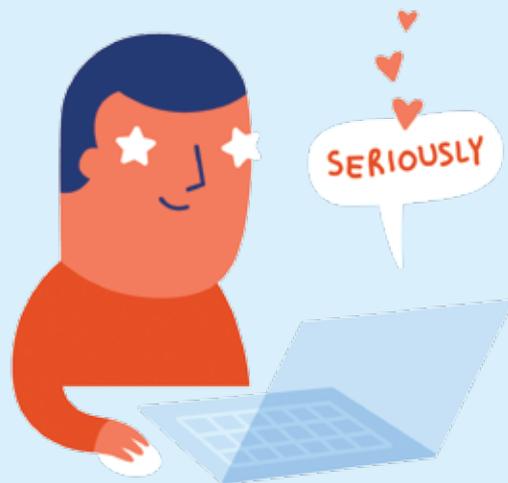
This new dimension of the debate will be a major democratic step forward as long as it is accompanied by training, particularly for the youngest people, in well-argued speech, so that all the actors in society are able to participate. This necessarily involves making available tools and methods that make it possible to remove the current difficulties in order to make this space for discussion a calmer space.

In this context, the state needs to be able to trust in civil society to become aware of its responsibilities. As such, its role should be that of a strategic state which gives civil society the means, notably the financial means, to organize and coordinate a citizens’ response to make living and doing things together meaningful and effective again.

63 See the report “Lignes de faille. Une société à réunifier” (Fault lines. A society to reunify), France Stratégie, October 2016. Referenced on March 28, 2017. Available at: <http://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/rapport-lignes-de-faille-ok.pdf>

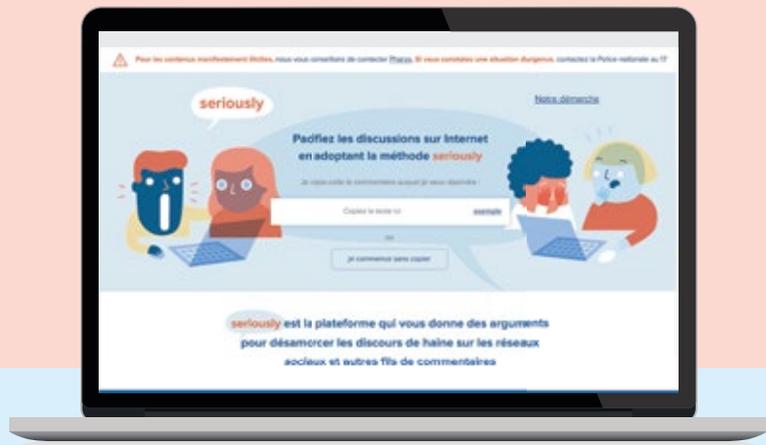
64 Antonio Gramsci, “Les Cahiers de Prison”, Cahiers 3, Ed. Gallimard Paris, 1983

APPENDIX



THE SERIOUSLY PROJECT

WWW.SERIOUSLY.ORG



THE ORIGINS OF THE PROJECT

Launched in 2015 after the January terrorist attacks, the Seriously project is the result of a twofold objective for the Renaissance Numérique think tank:

- Take action to restrain the worrying hate dynamic that prospers in our societies, and is particularly visible on the Internet;
- Make the role of the think tank evolve toward the “do” tank by making available tools that encourage citizen participation and thus promote the transition from a collective to a collaborative society.

THE SERIOUSLY APPROACH

Seriously is not only a digital tool (www.seriously.org), but also a method of support that makes it possible, through argumentation, to pacify discussions online. Seriously proposes to convert hate speech into a discursive process leading to de-escalation of invective and to the awakening of critical thinking.

In complement to the reporting mechanisms, this method is based on a tool that offers concrete arguments and diverse behavioral approaches (reasoned, neutral, humorous) through an intuitive user experience built around three sets of functionality:

- Factual elements (key figures, data, graphs, etc.) to objectify the debate and arouse critical thinking;
- Expert advice to support users emotionally and psychologically in order to temper the discussion;
- Resources suitable for a digital format (videos, studies, images, etc.) to illustrate the argumentation.

THE PARTNER ECOSYSTEM

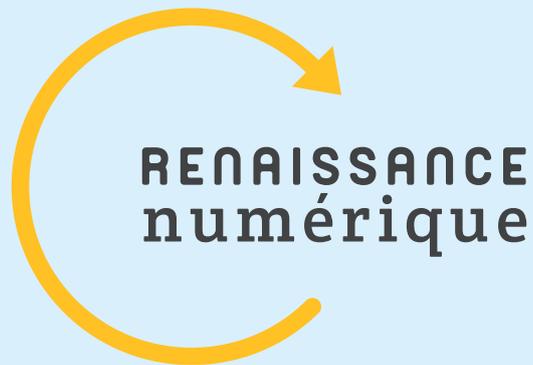
GOVERNING BODIES

- A steering committee: made up of partner associations, it approves the project's operational and strategic directions;
- A scientific council: made up of partner experts, it ensures the scientific basis of the content and method.

MEMBERS

| SCIENTIFIC COUNCIL | STEERING COMMITTEE |
|---|--|
| – Romain Badouard, Université de Cergy-Pontoise – Lecturer and a researcher at the AGORA laboratory, specialist in shifts of opinion on the Internet | – Kahina Ba, Parle-moi d'Islam – President |
| – Catherine Blaya, Université de Nice Sophia-Antipolis – Professor of Educational Sciences and President of the International Observatory of Violence in School | – Guillaume Brossard, Hoaxbuster – Co-founder |
| – Aurélie Dumond, Université de Nice Sophia-Antipolis – Doctoral student and expert in cyber-sexism subjects | – Stéphane Lacombe, AfVT.org – Assistant Director responsible for prevention |
| – Marc Hecker, IFRI – Researcher at the IFRI Centre for Security Studies, Director of Publications at IFRI and Editor-in-Chief of Politique étrangère. | – Maryna Shcherbyna, SOS Racisme – Head of communications |
| – Emmanuel Taïeb, Sciences Po Lyon – Teacher, specialist in conspiracy theories. | – Edouard Portefaix, Council of Europe |
| | – French Representative of the No Hate Speech Movement |
| | – Philippe Potentini, Council of Europe – Head of communications |

In order to guarantee independent decision-making, Seriously is financed through a balanced multi-actor arrangement, by both private and public actors.



ABOUT RENAISSANCE NUMÉRIQUE

Renaissance Numérique is the think tank on digital society. It brings together large Internet companies, start-ups, universities and representatives of civil society to participate in defining a new economic, social and political model arising from the digital revolution.

It currently has more than 50 members bringing to life thoughts on digital matters all over the country. Henri Isaac, Associate Professor Université Paris-Dauphine, is the President of the think tank.

www.renaissancenumerique.org - @RNumerique