

Hate speech in digital space and online gender-based violence

Amirouche Nedjaa*
Erich Brost Institute - Dortmund
University, (Germany)
nedjaa@yahoo.fr

Submitted: 04/02/2023

Accepted: 01/03/2023

Abstract: This article explores the issue of hate speech and its opposition to freedom of expression. In addition to this opposition, it also does address two crucial issues: hate speech during elections and a very common form of hate speech expressed in gender-based violence in the digital space. It will bring an answer to the question of whether freedom of expression is absolute or not. It comes, as well, to show that the line between freedom of expression and hate speech is a very fine one. In order to deal with this growing phenomenon, there have been instruments and codes of conduct which were put in place. While on the one hand we still see online hate speech in general and violent rhetoric targeting women, more particularly those involved in politics, we do see, on the other hand, that many efforts have been made in this direction and that the results are palpable especially those made by the European Union.

Keywords: Hate speech, Digital space, Gender, Human rights, social media, Violence

1. Introduction

When it comes to the issue of hate speech in the media and digital space, a very important question must come in mind: When does freedom of expression, which obviously extends to the harshest criticism, become criminal hate speech and what better way to deal with it?

Usually, hate speech is classified as a derogatory or a discriminatory language towards an individual or a group based on their religious beliefs, ethnicity race or nationality. Hate speech, however, has no precise definition in international human rights terms. It is a term used to describe a speech that is general, extremely negative and constitutes a risk to social peace. According to the Committee of Ministers of the Council of Europe, hate speech covers any form of expression which spreads, justifies, incites or promotes racial hatred, xenophobia, anti-Semitism or any form of hatred based on intolerance. With the emergence of new forms of media, hate speech is also present online and requires further reflection and action to regulate and combat it.

Social media networks are the main vector of hate and disinformation as the types of content that circulate a lot is precisely that one which arouses emotional reactions, such as anger, fear and hate.

“Cyberspace offers freedom of communication and opinion expressions. However, the existing social media networks are regularly being misused to spread violent messages, comments, and hateful speech. This has been conceptualized as being an online hate speech,

* Corresponding author

defined as any communication that disparages a person or a group on the basis of characteristics such as race, color, ethnicity, gender, sexual orientation, nationality, religion, or political affiliation (Zhang & Luo, 2018). The urgency of this matter has been increasingly recognized (Gamback & Sikdar, 2017). In the European Union (EU), 80% of people have encountered hate speech online and 40% have felt attacked or threatened via Social Network Sites [SNS] (Gagliardone, Gal, Alves, & Martinez, 2015)".

The fact is that in many cases this hate speech targets women or is based on gender. Women remain privileged targets of this particular type of violence. It is a form of Gender Based Violence that is found on digital media.

The ultimate aim of this online violence is to make women silent especially those daring to participate in elections and politics. Hedy Fry, OSCE representative on Gender states: "We can no longer imagine public life without the internet and social media. Sadly, this digital realm has become a new space to promote gender-based violence, particularly violence against women journalists and politicians. Digital violence can happen any hour of the day and its perpetrators can act anonymously and without penalty. These perpetrators harass their targets through relentless spamming and cyberstalking, through threats of rape and physical harm to them or their families, or through the non-consensual distribution of sexual images, messages or personal information. These perpetrators know that damaging content posted online is almost impossible to erase".

It is important to underline that online hate speech and this online gender-based violence instill fear, violate privacy, and above all stifle the free speech. Also, the spreading of misinformation and fake news whose perpetrators particularly target minorities, marginalized people and women makes these groups subject to silence.

Hence, the spread of hate speech and violence on digital space can undermine the electoral process and democracy.

The UN strategy and plan of action on hate speech include sex and gender, and the international law dictates that some of this speech shouldn't be protected. In other words, some can be criminalized only when it is linked to incitement to violence hostility or discrimination and to racism.

In May 2016, aiming at preventing and countering the spread of illegal hate speech online, the Commission reached an agreement with Facebook, Microsoft, Twitter and YouTube on a "Code of conduct on countering illegal hate speech online".

In the course of 2018, Instagram, Snapchat and Dailymotion took part in the Code of Conduct, while Jeuxvideo.com joined in January 2019, and TikTok in September 2020. On 25 June 2021, LinkedIn also announced its participation in the Code of Conduct.

2. Is the right to freedom of expression absolute?

The scope of the right to freedom of expression is broad but not absolute as it requires States to ensure that all persons are free to seek, obtain or publish any kind of information or ideas regardless of frontiers and through any media of one's choice. A State may exceptionally restrict the right to freedom of expression in Article 19(3) of the International Covenant on Civil and Political Rights provided that such restriction is:

1. Provided by law with sufficient precision to enable individuals to act in accordance with the law,

2. Pursue legitimate objectives listed in an inventory including: (i) respect for the rights or reputations of others- (ii) protection of national security or public order.

3. Necessary for a democratic society, that is, the state has to establish specifically the nature of the threat it faces and the necessity and proportionality of the action it has taken, particularly by establishing a direct and immediate connection between the expression and the threat.

The triple test is considered an essential means in developing journalistic work and ensuring the compatibility of media contents with the principles related to the right to freedom of expression and information. Therefore, it is important for the journalist to always ask the following questions within the scope of discourse analysis:

- Does this expression affect a legitimate interest (the rights of others, the protection of national security, public order, the protection of health or public morals). Is there a legal text that prohibits this expression (international treaties, law against violence against women, Decrees...)?

- What is the behavior that would address hate speech in proportion to the format of the program, the type of guest and the content of the speech? It is this third question that determines the form of journalistic treatment whenever a journalist finds that a certain media content has left the circle of the right to freedom of expression.

Navigating the fine line between free speech and hate speech isn't such easy but in the interests of protecting human rights for all, there is a duty to work more intensively to address the problem without silencing free expression and the benefits it has for the development of democracy and human rights.

3. Freedom of expression and hate speech according to the International Law

There are three useful international instruments in this regard. These are the International Covenant on Civil and Political Rights (ICCPR) and the International Convention on the Elimination of All Forms of Racial Discrimination (CERD). Additionally, the United Nations' Rabat Plan of Action on the prohibition of incitement to hatred advises us at which point specific speech crosses the line into causing violence, hostility or discrimination.

3.1 The International Covenant on Civil and Political Rights (ICCPR) :

Article 19

1. Everyone shall have the right to hold opinions without interference.

2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.

3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:

(a) For respect of the rights or reputations of others;

(b) For the protection of national security or of public order (ordre public), or of public health or morals.

3.2 The International Convention on the Elimination of all forms of racial discrimination (CERD) :

Article 5

In compliance with the fundamental obligations laid down in article 2 of this Convention, States Parties undertake to prohibit and to eliminate racial discrimination in all its forms and to guarantee the right of everyone, without distinction as to race, color, or national or ethnic origin, to equality before the law, notably in the enjoyment of the following rights:

(a) The right to equal treatment before the tribunals and all other organs administering justice;

(b) The right to security of person and protection by the State against violence or bodily harm, whether inflicted by government officials or by any individual group or institution;

(c) Political rights, in particular the right to participate in elections-to vote and to stand for election-on the basis of universal and equal suffrage, to take part in the Government as well as in the conduct of public affairs at any level and to have equal access to public service;

(d) Economic, social and cultural rights, in particular:

(e) The right of access to any place or service intended for use by the general public, such as transport hotels, restaurants, cafes, theatres and parks.

(f) Other civil rights, in particular as the right to freedom of movement, right to freedom of peaceful assembly and association, right to freedom of opinion and expression;

3.3 The Rabat Plan of Action :

The Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence brings together the conclusions and recommendations from several OHCHR expert workshops (held in Geneva, Vienna, Nairobi, Bangkok and Santiago de Chile). By grounding the debate in international human rights law, the objective has been threefold:

- To gain a better understanding of legislative patterns, judicial practices and policies regarding the concept of incitement to national, racial, or religious hatred, while ensuring full respect for freedom of expression as outlined in articles 19 and 20 of the International Covenant on Civil and Political Rights (ICCPR);

- To arrive at a comprehensive assessment of the state of implementation of the prohibition of incitement in conformity with international human rights law and;

- To identify possible actions at all levels.

The Rabat Plan of Action was adopted by experts at the wrap-up meeting in Rabat on October 4-5, 2012.

In addition to these three important instruments, to prevent and counter the spread of illegal hate speech online, the European Commission agreed with Facebook, Microsoft, Twitter, YouTube Instagram, Snapchat, Dailymotion, Jeuxvideo.com, and TikTok on a “Code of conduct” on countering illegal hate speech online.

4. Effectiveness of these instruments

These three instruments and the EU Code of Conduct have helped lawmakers define and courts assess whether speech qualifies as hate speech and what to do about it. Particularly, it helped determine what to do when hateful expression is not directly linked to hate crimes involving incitement and or racism. While internet companies, today, have to obey laws prohibiting hate speech, there is the question of how do they interpret and implement these laws in practice.

In 2021, the IT companies informed the EU Commission that they took serious measures to tackle hate speech. Indeed, it did take down thousands of hate speeches.

Facebook took action on 31.5 million pieces of hate speech content, compared to 25.2 million in Q1 2021. By doing so, 97.6% of this content was proactively detected before it was reported;

- Action on 6.2 million pieces of organized hate content, compared to 9.8 million in Q1 2021 (marking a return to pre-Q1 levels as we update our proactive detection technology). By doing so, 97.8% of this content was proactively detected before it was reported;

- 1.4 million pieces of content which were subject to hate speech action were appealed. 88.1 million of these were restored.

Instagram took action on 9.8 million pieces of hate speech content on its platform, up from 6.3 million in Q1 2021. By doing so, 95.1% of this content was proactively detected before it was reported;

- Action on 367,000 pieces of organized hate content, up from 325,000 in Q1 2021. By doing so, 77.7% of this content was proactively detected before it was reported;

Twitter . Over 50% of Twitter enforcement is now done proactively through machine learning without having people carry the burden of reporting it.

- Twitter continues to invest in new product capabilities that allow to broaden the spectrum of enforcement decisions to be able to make beyond the binary ‘take down leave up’, including prompts, labelling Tweets that are misleading, labelling accounts with context and de-amplifying and limiting engagement on certain Tweets.

Abuse and harassment: There was a 34% decrease in the number of accounts actioned for violations of Twitter abuse policy.

From January - June 2020:

- Accounts actioned: 398.057 (number of unique accounts that were suspended or had some content removed for violating the Twitter Rules)
- Accounts suspended: 72,139 (number of unique accounts that were suspended for violating the Twitter Rules)
- Content removed: 609.253 (number of unique pieces of content, includes Tweets, profile image, banner, or bio, which Twitter required account owners to remove for violating the platform's Rules)

YouTube: Between January and March 2021, YouTube has removed over 85k videos and over 43.6 million comments for violating hate speech policies on the platform. In the same quarter, approximately 76% of the uploaded videos which were removed for violating the platform's Hate Speech policy were taken down before they had 10 views. During the same reporting period, there were 9.5M videos which YouTube removed for violating their Community Guidelines, and more than 15k other videos which were uploaded from an IP address in Europe were in violation of the platform's hate policies.

In October 2021, the European Commission has released the results of its sixth evaluation of the Code of Conduct on countering illegal hate speech online . The evaluation shows that on average:

- IT companies assessed 81% of the notifications in less than 24 hours, which is lower than the 2020's average of 90,4%.
- IT companies removed 62,5% of the content notified to them, which is lower than the average of 71% recorded in 2019 and 2020.
- Removal rates varied depending on the severity of hateful content. 69% of content calling for murder or violence against specific groups was removed, while 55% of the content using defamatory words or pictures aiming at certain groups was removed. Conversely, in 2020, the respective results were 83,5% and 57.8%.
- IT companies gave feedback to 60,3% of the notifications received, which is lower than during the previous monitoring exercise (67.1%).
- In this monitoring exercise, sexual orientation is the most commonly reported ground of hate speech (18,2%) followed by xenophobia (18%) and anti-gypsyism (12.5%).
- For the first time, the IT companies reported detailed information about measures taken to counter hate speech outside the monitoring exercise, including their actions to automatically detect and remove content.

In April this year, the European Union reached a major deal to tackle hate speech, disinformation and other harmful content online. The agreement would require from major technology companies to tighten their monitoring, make it easier for users to report problems and give regulators the power to punish breaches with billions in fines.

In the electoral context

Hate speech can undermine the whole electoral process

Hatred and inciting speech during the electoral process can potentially undermine the peaceful conduct of elections process. During such period and critical atmosphere of an election process especially during the campaign activities, inflammatory language can have the effect of inciting people to violence, thereby affecting freedom of expression and democratic human rights.

Unfortunately, many countries lack the legal basis to prevent a candidate or party with hate speech from registering for elections. But even when a legal framework is there and hate speech is strongly reprimanded, the dilemma consists in the fact that is precisely during an election that different political opinions must be expressed, and imposing restrictions on the expression of these opinions may limit both the rights of free expression and democratic rights.

As hate speech and ethnic contempt on social media and other platforms is known to catalyze environments of conflict and violence among groups and communities during elections, European center for electoral support (ECES) has started years ago working strongly on strengthening capacities of its partners to identify and monitor hate speech and ethnic contempt incidences and recommend appropriate remedial measures. This aims at avoiding electoral related violence by tracking media-transmitted incitement to violence, hate speech, fake news, and other messages that promote discrimination or social and political tensions.

Female political actors are more expose to hate speech than men if hate speech in the social media plays a significant role in contributing to electoral violence, it exacerbates gender-based violence in particular towards those women who have chosen to run for office, and increasingly becoming a threat to inclusive participation of women in electoral and political life. This type of violence against women aims at intimidating and silencing them in order to exclude them from both the public and political space.

According to The UN Women: Political violence can occur in many contexts, but it has a uniquely negative impact on women during an election period. Political tensions and fierce competition during elections can create vulnerabilities for the political participation of women, who in many cases are already at a disadvantage compared to men; this can lead to the exclusion of women from the process. Tactics such as physical assault, intimidation and psychological violence affect both men and women, but their impact varies by gender and some forms of gender-based violence are specific to women. Responses to prevent and mitigate political violence during elections must reflect these important gender dimensions.

To combat this kind of violence, the UN is continuously strengthening the legal framework for women's political and electoral participation. This is evident in the declarations, covenants and conventions on human and political rights, in the reports, resolutions and action plans of the UN in these areas, and in the electoral assistance policies developed and implemented by the UN. Citing as an illustration, the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) , adopted in 1979, reaffirms the right of women "to hold public office and to perform all public functions at all levels of government".

More Specifically, Article 7 states that:

States Parties shall take all appropriate measures to eliminate discrimination against women in the political and public life of the State.

In particular, they shall ensure to women, on equal terms with men :

(a) To vote in all elections and public referenda and to be eligible for election to all publicly elected bodies

(b) To participate in the formulation of government policy and the implementation thereof, to hold public office and to perform all public functions at all levels of government

At the end, it is a fact that hate speech constitutes a threat to societies and their cohesion and is a particular danger during election periods because of its complicity. While hate speech threatens both men and women, it has a greater impact on women as a vulnerable group and especially on those who have chosen to take part in political action. This phenomenon is exacerbated by social networks which provide a fluid and easy channel for the often anonymous dissemination of violence against women politicians.

The introduction of the article is written in a sequential and concise scientific way in which the researcher determines the problematic of his research and its purpose. In the Times News Roman, size 12, the line between lines 1 is written. ??(This format applies to the full content of the article).

5. References:

1. <https://www.coe.int/fr/web/freedom-expression/hate-speech#:~:text=Le%20discours%20de%20haine%20n,risque%20pour%20la%20paix%20sociale.>
2. Internet, social media and online hate speech. Systematic review - ScienceDirect
3. https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combating-discrimination/racism-and-xenophobia/eu-code-conduct-counteracting-illegal-hate-speech-online_en
4. <https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx>
5. <https://www.ohchr.org/en/professionalinterest/pages/cerd.aspx>
6. <https://www.ohchr.org/en/issues/freedomofopinion/articles19-20/pages/index.aspx#:~:text=The%20Rabat%20Plan%20of%20Action%20on%20the%20prohibition%20of%20advocacy,Bangkok%20and%20Santiago%20de%20Chile>
7. https://ec.europa.eu/commission/presscorner/detail/nl/ip_21_5082
8. <https://www.unwomen.org/sites/default/files/Headquarters/Attachments/Sections/Library/Publications/2017/Preventing-VAW-in-elections-fr.pdf>
9. <https://www.unwomen.org/sites/default/files/Headquarters/Attachments/Sections/Library/Publications/2017/Preventing-VAW-in-elections-fr .pdf> page 12.