Journal of Social and Human Science Studies (مجلة دراسات إنسانية واجتماعية)
University of Oran^2 / Vol: 11 N:03. 16/06/ 2022.
ISNN: 2253-0592 EISSN: 2588-199X / Prefix:10.46315

The Role of User-Generated Content in Increasing Hate Speech
-State of the issue of Racism, Xenophobia and Minorities on Social Media-

دور المحتوى الذي ينشئه المستخدم في تصعيد خطاب الكراهية
-حالة قضية العنصرية وكره الأجانب والأقليات على منصّات التواصل الاجتماعي-

GHERABI AbdEsselem* غرابي عبدالسّلام abdesselem.gherabi@univ-msila.dz	Communication	Sociological laboratory of the quality of public service, University Mohamed Boudiaf of M'sila, Algeria
Dr .Naima BERARDI د. براردي نعيمة naima.berardi@univ-msila.dz	Communication	Sociological laboratory of the quality of public service, University Mohamed Boudiaf of M'sila, Algeria

DOI : 10.46315/1714-011-003-060

Received: 05/ 01/ 2021 Accepted: 09/ 05/ 2021 Published : 16/06/ 2022

Abstract: Social media are rife with hate speech, although most major social media companies have their own policies regarding whether and what kinds of hate speech are permitted on their sites, the policies are often inconsistently applied and can be difficult for users to understand.

This paper will address the topic of hate speech as a key concept in social media today, in an effort to identify solutions for curtailing hate speech in social media especially with regard to racism, xenophobia and minorities.

Keywords: User-Generated Content; Hate Speech; Racism; Xenophobia and Minorities; Social Media.

ملخّص : تعجّ وسائل التّواصل الاجتماعي بخطاب الكراهية، وعلى الرغم من أنّ معظم شركات وسائل التواصل الاجتماعي الكبرى لديها سياساتها الخاصّة بأنواع الخطاب التي يتمّ قبولها على مواقعها، إلاّ أنّ هذه السّياسات غالباً ما يتمّ تطبيقها بشكل غير مُتّسق ويصعب على المستخدمين فهمها.

ستتناول هذه الورقة موضوع خطاب الكراهية اليوم كمفهوم أساسي في وسائل التّواصل الاجتماعي، في محاولة لتحديد حلول للحدّ منها خاصّة فيما تعلّق بالعنصريّة وكره الأجانب والأقليّات.

كلمات مفتاحية: المحتوى الذي ينشئه المستخدم؛ خطاب الكراهيّة؛ العنصريّة؛ كراهية الأجانب والأقليات؛ منصّات التّواصل الاجتماعي.

* - Corresponding author: abdesselem.gherabi@univ-msila.dz.

Journal of Social and Human Science Studies (مجلة دراسات إنسانية واجتماعية)
University of Oran^2 / Vol: 11 N:03. 16/06/ 2022.
ISNN: 2253-0592 EISSN: 2588-199X / Prefix:10.46315

Introduction

With the rise of digital and mobile technologies, interaction on a large scale became easier for individuals than ever before; and as such, a new media age was born where interactivity was placed at the center of new media functions. One individual could now speak to many, and instant feedback was a possibility. Where citizens and consumers used to have limited and somewhat muted voices, now they could share their opinions with many. The low cost and accessibility of new technology also allowed more options for media consumption than ever before -and so instead of only a few news outlets, individuals now have the ability to seek information from several sources and to dialogue with others via message forums about the information posted.

Around the world, we are seeing a disturbing groundswell of xenophobia, racism and intolerance — including rising anti-Semitism, anti-Muslim hatred. Social media and other forms of communication are being exploited as platforms for bigotry. Neo-Nazi and white supremacy movements are on the march. Public discourse is being weaponized for political gain with incendiary rhetoric that stigmatizes and dehumanizes minorities, migrants, refugees, women and any so-called "other".

This is not an isolated phenomenon or the loud voices of a few people on the fringe of society. Hate is moving into the mainstream — in liberal democracies and authoritarian systems alike. And with each broken norm, the pillars of our common humanity are weakened.

This paper will look at **why Hate Speech has pervaded Social media platforms, what has contributed to enemy images of foreigners, and how, if at all, such images can be removed from the international consciousnessand we can better the current situation؟**

The main goal of this paper is the design of and monitoring and analyzing hate in Social media.

1- Definition of Key Terms

1-1- User-Generated Content

User-generated content (UGC), sometimes also referred to as user-created content (UGC), is a generic term that encompasses a wide range of media and creative content types that were created or at least substantially cocreated by "users" that is, by contributors working outside of Conventional professional environments. Although UGC in digital formats is as old as computing technology itself, and UGC in nondigital formats has an even longer history, the term emerged to widespread recognition especially with the participative turn in Web design and practices that took place in the

early years of the new millennium and is often referred to as the emergence of "Web 2.0". (Holt , 2015, 1799)

1-2- Hate Speech

There is no international legal definition of hate speech, and the characterization of what is "hateful" is controversial and disputed. In The Context of this document, the term hate speech is understood as any kind of communication in speech, writing or Behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor. This is often rooted in, and generates intolerance and hatred and, in certain contexts, can be demeaning and divisive. (Wermiel, 2018, 3)

We summarize leading definitions of hate speech from varying sources, as well as some aspects of the definitions that make the detection of hate speech difficult:

- Encyclopedia of the American Constitution: "Hate speech is speech that attacks a person or group on the basis of attributes such as race, religion, ethnic origin, national origin, sex, disability, sexual orientation, or gender identity". (Buyse, 2014, 783)

- Facebook: "We define hate speech as a direct attack on people based on what we call protected characteristics-race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disease or disability. We also provide some protections for immigration status. We define attack as violent or dehumanizing speech, statements of inferiority, or calls for exclusion or segregation". (Bleich, 2011, 925)

- Twitter: "Hateful conduct: You may not promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease." (Bleich, 2011, 918)

Another definition is based on an analysis of the following characteristics from other definitions : (Cole, J, 2009, 29)

- Hate speech is to incite violence or hate. - Hate speech is to attack or diminish.
- Hate speech has specific targets. - Whether humor can be considered hate speech.

1-3- Racism

Racism is a global hierarchy of superiority and inferiority along the line of the human that have been politically, culturally and economically produced and reproduced for centuries by the

institutions of the "capitalist/patriarchal western-centric/Christian-centric modern/colonial world-system". (Grosfoguel, 2011, 07)

This definition of racism allows us to conceive of diverse forms of racism, evading the reductionisms of many existing definitions. Depending on the different colonial histories in diverse regions of the world, the hierarchy of superiority/inferiority along the lines of the human can be constructed through diverse racial markers. Racism can be marked by color, ethnicity, language, culture and/or religion.

Although since colonial times color racism has been the dominant marker of racism in most parts of the world, it is not the only or exclusive form of racist marker. On many occasions we confuse the particular/concrete social marker of racism in one region of the world with what is taken to be as the exclusive form or universal definition of racism. This has created an enormous amount of conceptual and theoretical problems.

1-4- Xenophobia and Minorities

Xenophobia can be defined as "attitudes, prejudices and behavior that reject, exclude and often vilify persons, based on the perception that they are outsiders or foreigners to the community, society or national identity". (Campbell, E, 2003, 15)

And form minority group is a subgroup of the population with unique social, religious, ethnic, racial, and/or other characteristics that differ from those of a majority group. The term usually refers to any group that is subjected to oppression and discrimination by those in more powerful social positions, whether or not the group is a numerical minority. Examples of groups that have been labeled minorities include African Americans, women, and immigrants among others.(Gleason, 1991, 395)

1-5- Social Media

Social media is the term often used to refer to new forms of media that involve interactive participation. Often the development of media is divided into two different ages, the broadcast age and the interactive age. In the broadcast age, media were almost exclusively centralized where one entity-such as a radio or television station, newspaper company, or a movie production studio-distributed messages to many people. Feedback to media outlets was often indirect, delayed, and impersonal. Mediated communication between individuals typically happened on a much smaller level, usually via personal letters, telephone calls, or sometimes on a slightly larger scale through means such as photocopied family newsletters.(Wright, B, 2011, 23)

2- Hate Speech in Media Discourse

2-1-Hate speech in the context of freedom of speech

It is a richly documented work on hate speech in the right-wing media. The book exposes the analytical productivity of the perspective which is imposed by hate speech. The analysis, however, has a fundamental flaw, as it identifies hate speech as typical solely for right-wing

Views. The authors of this book use this concept – mostly describing the activities of the right-wing media – at the same time not paying attention to the use of language, which corresponds with the definition of hate speech, by other political parties, mostly left-wing ones. This attitude reflects the general tendency, in which the "hate speech" concept is often used – especially by some groups of politicians – unreliably, instrumentally, and in a manipulative way, which is manifested in that the same word spoken by political opponents is stigmatised, whereas when spoken by supporters it is approved of as a symptom of eloquence and being lettered.(Yulia, 2003, 45)

It is a richly documented work on hate speech in the right-wing media. The book exposes the analytical productivity of the perspective which is imposed by hate speech. The analysis, however, has a fundamental flaw, as it identifies hate speech as typical solely for right-wing views. The authors of this book use this concept – mostly describing the activities of the right-wing media-at the same time not paying attention to the use of language, which corresponds with the definition of hate speech, by other political parties, mostly left-wing ones. This attitude reflects the general tendency,

in which the "hate speech" concept is often used – especially by some groups of politicians unreliably, instrumentally, and in a manipulative way, which is manifested in that the same word spoken by political opponents is stigmatised, whereas when spoken by supporters it is approved of as a symptom of eloquence and being lettered.

In the discourse analysis of hate speech the following questions arise :(Gerstenfeld, 2003, 31)

- Is hate speech an expression of individual opinions or a political device fuelling hatred?
- Is conveying every, even the most extreme opinion, good for public debate?

The debate about hate speech covers concepts related to a conflict of two values: freedom of speech and respect for human dignity.

In the discussion about media freedom, there sometimes appear voices of ignorance and miscomprehension. An utterance of a publicist of one of the most prominent daily newspapers can serve as an example, I think, however, that if someone is a supporter of freedom of speech, they also have to be supporting freedom of filthy, foul, stupid and harmful speech. Someone who states that

they are a supporter of freedom of speech, however, on the condition that this will be beautiful, wise and noble speech, is, as a matter of fact, a supporter of censorship.

No freedom, including freedom of speech, is absolute: it encounters a boundary in the shape of the duty to respect others' dignity and their legitimate freedom. One should not write, create and broadcast programmes if they damage the truth: and I mean not only the factual truth that is conveyed, but also the "truth about a human being", a person's dignity in all dimensions.

This presumption is not an absolute, indefeasible norm. There are obvious instances – for example, libel and slander, messages that seek to foster hatred and conflict among individuals and groups, obscenity and pornography, the morbid depiction of violence – where No right to communicate exists. Plainly, too, free expression should always observe principles like truth, fairness, and respect for privacy.

2-2- The media creation of an artificial reality of hatred

The assessment of the hate speech which is present in the media must involve knowledge of the media functionality.

Along with the original function of reflecting the world, the media work more and more actively in creating reality. As a result of the process, one finds a change in the competences of the creators and the recipients of media communications. The creator and sender of the communication does not convey an objective meaning in it, but is reduced to an inventor of contexts for the receiving creation of the world, and becomes one of the interpretative contexts. (Waldron, J, 2012, 13)

The media visualisation of reality consists of a permanent confrontation of The "quotidian reality" with the "media reality". People constantly experience artificial Worlds by means of the media. This experience begins to question the Exclusivity of the real world, and later on it blurs the sense of reality, to the extent That the "clear distinction between the quotidian reality and the media reality is no longer possible". A long-term effect of media hate speech is the creation of artificial Realities of hatred, which is accomplished through different media processes and through several stages:(Pariser, E, 2011, 37)

- **Firstly**, the true reality is shaped according to the rules of the media. Media hate speech adopts performative features, which create an artificial world of media hatred. This world permeates the real world of interpersonal relations (the media is responsible for the emotions connected with hatred, for example in the evaluation of political opponents).

- **Secondly**, the media influence the very shape of reality. Today many real political Events are staged from the very beginning because of the possibility of being presented in the media. The media shape real emotional stages of hatred within the reality that exists beyond the media. This reality is increasingly sated with elements of the media hate speech.
- **Thirdly**, the media change the time and space conditioning of real human Life and human communication possibilities. Thanks to the media, a structure of Omnipresence without a distinguished presence is created. Time and space, traditionally Fundamental coordinates of our world, become vestigial. The change Within these conditions proves the thesis that the reality of media hatred is largely A media structure. (Nakamura, 2014, 259)
- **Fourthly**, the media blur the line between reality and a staged event or a simulation,The experience of simulation increasingly becomes a model for real events, and the reality is more and more often assessed along with the mental image of the Media representation. Hence, one may believe that the simulated hatred, which is presented in the media, is a part of real life.

2-3- Some varieties of linguistic destruction

Some negative tendencies of cultural mentality are reflected in the mass media Culture, and these promote banality, vulgarity, crudeness, absolute liberty, etc.

They undoubtedly contribute to the blurring of the lines between good and evil, Deepening the state of ethical confusion. This tendency has a negative influence On the moral sensitivity of a human being, in such that many people do not use the Concept of evil in everyday speech, but instead they describe it with substitutes Like: incongruity, lack of propriety, inverity, insubordination, order violation, etc, Which partly justify the evil or trivialise it. The media use the so-called visibility Method, often excessively exposing the negative phenomena, and making them The main feature of their own appeal. The advancing ignorance and trivialization of evil can be a consequence of such activity. The process can lead to a hypertrophy of the insensibility to evil, which means there appears a systematic substitution of the good for the worse and the worse for the bad, a substitution of the Excess for ordinariness and ordinariness for abnormality, etc. It is a replacement Which, in the end, blurs the lines between good and evil. Such a substitution is so Easy because of the assistance of language which trivialises evil, for example the Word “to kill” is replaced by the expression “to cause death out of compassion”, “A lie” is replaced by “an unexplainable matter”, “subjugation” becomes “creating A new imagination”, “pornography” becomes “the language of the body”.

This linguistic practice does not testify that the awareness of the evil declines, But rather that the evil is trivialised, and that it is the capability of a correct ethical Evaluation that recedes. It is sometimes observed that tragedies and scandals Awaken the conscience from this ethical confusion in evaluation. (Shepherd, 2015, 57)

Media hate speech is also a part of the media show. Media communications (Conveying information) are created within the entertainment convention: radio, the dynamics of fact simplification, television, the show, press, tabloids and Social Media.

There is a danger that every content of a message can be submitted to the absolute entertainment process, which is based on commercialism, a good show and on cheap amusement. It does not put love, truth and upbringing in the first place, but instead – as Hans Arp calls it – it praises the "collective ecstasy and fast money making". It is worth noticing that the processes of degradation and elimination of the rational discourse from the media described by Postman are not determined by the very "nature" of media communications, and thus are not determined by the linguistic forms, but rather by the messages of the communications and the processes of media commercialisation. Postman shows that in the omnipotent era of the media and "show business" the life presented or impersonated on the screen appears to be more important than the reality, and that media emotions effectively reduce the world of human experiences. One of the most frequent tools of the media used in order to play with the emotions is astonishing the recipient with extreme forms and messages full of negative values, which consist primarily of hostility and hatred.(Whine,1999, 234)

2-4- A conscience responsible for speech quality

The search for tools and measures to restrict media hate speech indicates The need for media ethics, which no longer serves only as a postulate of moral Reflection upon the media, but also as a necessity conditioned by many different Factors, and which serves the truth and the well being of the human and the Community. The human being and the community experience many cultural and Media challenges, and all the more need a clear orientation not to get lost within The aspects which are "hostile towards humans" (violence, aggression, fanaticism, Human dignity degradation, hatred), as well as to be reassured about the validity of the owned or created world of values.

The moment one seeks grounds and conditionings for freedom of speech, which forms the boundaries in which also hate speech can appear, one has to clearly state that it is achieved primarily within the inner human sphere, in the sphere of individual decision-making and of the choices conditioned by ethical rules upholding values. It is achieved in the inner human sphere, within one's

conscience. Freedom of speech is thus a matter of human conscience. It is indicated by different terms such as: "subjective moral consciousness", "individual responsibility", "acting according to one's conscience", "acting according to one's inner belief". (Mogekwu, 2005, 07)

3-Internet Regulation by Social Media Companies

With the emergence of social media, hate groups have added platforms such as Facebook and Twitter to their communicative networks. However, unlike the regulation of hate speech on websites by Internet service providers, social networking platforms enjoy greater freedom to decide whether and how to address expressions of hate:

3-1- Hate Speech on YouTube

Of the popular social media companies, some experts have found that Google's YouTube struggles the most to effectively craft and execute a consistent policy toward removing content and comments filled with hate speech from their site.

In addition to the low marks YouTube receives for its ability to deter bigots from using its site to spread hateful propaganda, the sheer amount of user-generated content featuring hate speech on YouTube makes it an excellent case study. Every minute, users upload 13 hours of content onto the site. Although the site prohibits hate speech, which it defines as speech that "attacks or demeans a group based on race or ethnic origin, religion, disability, gender, age, veteran status and sexual orientation/gender identity," a sizable portion of its channels, content and comments contain hateful rhetoric.(Vis, 2013, 27)

In fact, there are channels aimed at degrading Hispanics, blacks, women and others. It seems on YouTube no ethnic or minority group is exempt from exposure to hateful rhetoric. In addition to channels and content featuring hate speech, the comments section on YouTube regularly contains hate speech.

Although YouTube has a community-based system that allows users to flag potentially inappropriate content, it does not prevent the proliferation of hate speech on the site. Once flagged, a video is not removed, it is simply preceded by a message stating that "the following content has been identified by the YouTube community as being potentially Offensive or inappropriate. Viewer discretion is advised".(Gerstenfeld,2003, 41)

3-2- Hate Speech on Facebook

Facebook both sets and enforces the criteria for the removal of hateful content. In its terms of service agreement, Facebook users agree to "not post content that: is hate speech, threatening, or

pornographic; incites violence; or contains nudity or graphic or gratuitous violence" Facebook's definition of hate speech is further specified in the platform's community standards, as "content that directly attacks people based on their: Race, Ethnicity, National origin, Religious affiliation, Sexual orientation, Sex, gender, or gender identity, or Serious disabilities or diseases".(Burnap, 2016, 68)

Although the platform's community standards specifically state that "organizations and people dedicated to promoting hatred against these protected groups are not allowed a presence on Facebook" , it also distinguishes between humorous and serious speech, and advocates for the freedom to challenge ideas, institutions, and practices.

Critics of the platform argueagainst the lack of transparency of its content removal policy .Facebook encourages its users to report content they consider harmful under various criteria, and the platform determines, based on a set of internal rules, whether or not the reported content violated its community standards . However, the decision to remove or keep reported content is not explained to the users. Moreover, Facebook uses a country-specific blocking system that acts in accordance with each country's legislation in terms of removing undesirable pages. For instance, Nazi content is forbidden in Germany but allowed in the United States. Social networking platforms thus play a significant role as cultural intermediaries because their capacity to decide what content should be allowed is a "real and substantive" intervention into our understanding of public discourse and freedom of expression.(Erjavec, 2012, 901)

Regulators' views are divided between those who predict that strict enforcement of the platforms' terms of use would be more effective than the law to fight hate practices online and between those who argue the opposite" argues that "community standards will never protect speech as scrupulously as unelected judges enforcing strict rules about when speech can be viewed as a form of dangerous conduct".As a result, the platform hosts controversial content it does not consider harmful.(Williams, 2016, 213)

We argue that although Facebook's terms of use function as a gatekeeper, the platform's corporate logic in deciding what content should be allowed allows for the circulation of covert discrimination through its technological affordances and the communicative acts that it hosts.

3-3- Hate Speech on Twitter

Among the many existing social networks, Twitter currently ranks as one of the leading platformsand is one of the most important data sources for researchers. Twitter is a defensible and logical source of data for such analysis given that users of social media are more likely to express

emotional content due to deindividuation (anonymity, lack of self-awareness in groups, disinhibition). (Burnap, 2015, 225)

Twitter is a well-known real-time public microblogging network where, frequently, news appear before than on official news media.Characterized by its short message limit and unfiltered feed, its usage has quicklyescalated, especially amid events, with an average of 500 million tweets posted per day.(Kwok, 2013, 1625)

Twitter has updated its policies to now include links to hateful content within its parameters to unacceptable activity. As outlined by Twitter: (Shlapentokh, 2007, 139)

"At times, Twitter will take action to limit or prevent the spread of URL links to content outside Twitter. This is done by displaying a warning notice when the link is clicked, or by blocking the link so that it can't be Tweeted at all."

Among the URLs that Twitter may block, it now includes: (Sutton, 2013, 881)

"Content that promotes violence against, threatens or harasses other people on the basis of race, ethnicity, national origin, caste, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease."

This is in addition to malicious/spammy links, terrorism, illegal activity and private information.

Twitter detects such violations via user reports, automated detection systems and third-party reviewers, so it's entirely possible that some of this type of content will still make it through. But Twitter now has more specific rules against posted links to hateful conduct, in addition to policing hate speech directly posted on its own platform.

4- Potential Solutions

Several media law scholars have addressed the question, "What, if anything, should be done to curtail hate speech on the Internet?". While social media Web sites are the specific focus of this inquiry, it is still valuable to explore the various solutions offered to date to address hate speech across the Internet. Therefore, this chapter will examine the major solutions offered by various Scholars about how to best solve the problem of hate speech online. Each section of this chapter will provide an overview of one of the approaches, which include: legislative action, international regulation, industry self-regulation, filtering by end-users and not regulating hate speech on social media sites.

4-1- Hate speech and international legislation

Hate speech is a type of discriminatory speech that arises when people from different social, ethnic, or religious groups interact with one another, or when one such group asserts its power over others.

The above description is a thorough explanation of how hate speech can be understood. Establishing a definition, however, is one element in understanding hate speech. How it is negotiated within a particular context, in a given society and at a particular point in time is equally important. An understanding of the political and socio-economic context in which the hate speech act occurs should also be supplemented with an analysis of the speaker and audience to fully gauge the likely impact of the discourse. For example, dangerous speech framework also allows for analysis of the speaker and the degree of influence they have over the audience; the grievances and fears that the audience may have that the speaker is able to cultivate in the message; and the mode of dissemination, which may be influential in itself Negotiating hate speech is a delicate matter because, "from a human rights perspective, the right to life and the prohibition of discrimination are to be balanced against the freedom of expression" and the sometimes Consequential need for tolerance of these multiple expressions.

In this way, a controversial case could be made for the protection of speech acts that are often divisive. Protecting hate speech, however, presents the risk of prejudices becoming entrenched in pluralistic societies, which then compromise concepts of human dignity, defamation and human rights. Still, protecting hate speech does not only protect the speaker's rights but also allows the target of these speech acts to "speak back" .Freedom of speech principles then need to be balanced by considering whether or not these speech acts are offensive or incite violence, and so the question of Legalisation comes into play.(Nobata,2016, 148)

When we think about legislation, established laws and judicial systems areheavily reliant on Western paradigms, frameworks and institutions. American courts have been contending with issues on free speech for a few hundred years, whereas the European courts have been dealing with them within the last seven decades. When considering hate speech, there is a need to remember thathuman rights law does not dictate that freedom of expression is an unconditional right. Freedom of expression can be limited by protocols determined by documents like the International Covenant on Civil and Political Rights (ICPPR), the American Convention on Human Rights (ACHR), and the European Convention for Human Rights (ECHR). With regards to discrimination, Article 20(2) of the

ICPPR states: "Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law". Further clauses on racial discrimination are also found in the International Convention for the Elimination of All Forms of Racial Discrimination. It is also evident that much has been written in the American and European contexts, resulting in a need for more discussion in the non-Western contexts.(Hopkins, 2006, 247)

4-2- Self -Regulation by Social Media Companies

Today, social media companies such as YouTube, Facebook and Twitter are Responsible for creating and enforcing their own hate speech policies. As previously Discussed, the result is often an inconsistent application of rules that are in many Instances far less speech-protective than U.S. hate speech jurisprudence would mandate.

In addition, social media companies, in particular Google, which owns YouTube, has said That it would prefer not to act as an arbiter of free speech One of the primary benefits to this approach is that it does not implicate the First Amendment because the government is not involved. However, the situation that would Be created by this lack of oversight is one in which social media companies have Unprecedented power to censor online content. According to Alexander Tsesis, it is a mistake to exclusively place the power to decide whether and to what extent hate speech should be blocked in the hands of commercial interests. Keep in mind, says Tsesis, forprofit companies are not beholden to humanistic principles.

However, the most compelling argument against maintaining the current approach To self-regulation on the part of social media companies is that it is simply not working. Despite their best efforts to flag and even remove some hateful content, social media Web sites remain saturated with hate speech. (Djuric, 2015, 33)

4-3- Filtering Software

Filtering software installed by end-users to block unwanted content is called censor ware. Much like its name suggests, this software filters out or censors undesirable Web sites. Commercial software programs such as Cyber sitter, N2H2, Net nanny, Surf watch and Wise choice are designed to restrict an individual's ability to send or receive certain types of information, such as sexual or other obscene content. Users install them on their computers and then select users on that machine, such as children, are restricted from accessing certain Web sites. However, this software does not eliminate hate speech that appears on social media Web sites. To combat the more specific issue of hate speech online, the Anti-Defamation League (ADL) has developed free Hate Filter software,

which blocks access to sites that advocate hatred, bigotry or violence toward groups based on their race, religion, ethnicity, sexual orientation or other immutable characteristics.(Burch, 2001, 179)

There are several benefits to filtering software. First and foremost, this approach puts the power to regulate content in the hands, or mouse, of the end-user. As Judge Harlan notoriously pointed out in the Cohen case, "one man's vulgarity is another's lyric." Filtering software allows each individual, not the government or a corporation, to decide what kind of Internet sites are and are not appropriate for their family. Here, the individual maintains control over message transmissions and receptions. When used properly, it is possible for this software to protect young children from being exposed to offensive or even obscene content online. Finally, and perhaps most importantly, this approach to the problem of hate speech online does not interfere with the First Amendment.(Boromisza,H, 2013, 49)

However, it is essential to keep in mind that these filters often do not block hate Speech that appears on social media sites and therefore may not be a viable solution. In

Fact, there are several drawbacks to using filtering software to curtail hate speech on social media sites. First, end-users putting filters in place to restrict the content available to them will not minimize the amount of hate speech content that exists online. As a result, women and minorities will continue to feel marginalized, excluded and at worst, unwilling to participate in political or social discourse. In addition, those without access to the filtering devices will continue to be exposed to bigotry online. Filtering software also often makes mistakes by casting too wide a net and accidentally or inadvertently

Blocking out nondiscriminatory Web sites. For example, the filtering software Cyber Patrol classifies the National Academy of Clinical Biochemistry as "full nude" and Prevents access to its content. Conversely, they may also inadvertently let undesirable content through. For example, this software does not have the capability to filter out and deliver to the user only that YouTube or Facebook content that does not include certain Words or phrases, such as "fag" or "nigger". (Buyse, 2014, 783)

Finally, organizations such as the American Civil Liberties Union (ACLU) and the Electronic Privacy Information Center (EPIC) argue that the use of filtering software conflicts with individual rights to freedom of expression and freedom of association, as mandated by the United States Constitution. For some, even the warnings or blocking statements, which on YouTube warn users they are about to view flagged content, represent a kind of crowd sourcing or groupthink that may somehow hinder or impede individual thought. Lastly, opponents of commercial filtering software warn that because

these tools could someday be misused, potentially transformed into public filters that would allow for massive government censorship of the Internet, they should be avoided at all costs.(American Library Association, 2017)

Conclusion

Hate speech is a menace to democratic values, social stability and peace. As a matter of principle, the United Nations must confront hate speech at every turn. Silence can signal indifference to bigotry and intolerance, even as a situation escalates and the vulnerable become victims.

Addressing hate speech does not mean limiting or prohibiting freedom of speech. It means keeping hate speech from escalating into something more dangerous, particularly incitement to discrimination, hostility and violence, which is prohibited under international law.

An advocacy of hate speech legislation is most incomprehensible in the light of available alternatives. A number of traditional as well as more innovative concepts are far more promising to effectively prevent harm emanating from speech without jeopardising FoE and media freedom. Responsibility is returned to the wider public while the state's possibilities to manipulate public discourse are reduced. The enforcement of those alternative approaches would prospectively be the first step towards the prevention of harm as a result of speech.

Although there are viable arguments both for and against efforts to try to curtail hate speech, the fact remains that hate speech online or offlineisfully protected, The second prong of the recommended approach calls for the creation of a new generic top-level domain,"social" that would require adherence to uniform hate speech policies and procedures. This approach would:

- Minimize hate speech on social media sites
- Benefit social media organizations currently struggling to implement their own hate speech policies.
- Prevent either the government or social media companies from having total control over determining which offensive content will and will not be permitted on these platforms

Provide user swith more clear information about what kind of videos, images, text or comments they can expect to see on these sites

- Limit the number of women, minorities and members of other protected classes who are exposed to hateful rhetoric on these sites, which may lead to an increase in the number and types of voices heard in the public sphere.

ISNN: 2253-0592 EISSN: 2588-199X / Prefix:10.46315

Still, the reare potential drawbacks to the recommended approach, including the fact that social media organizations may not be willing to participate in the online civility forum and subsequent move to ".social." In addition, it is likely that critics of any hate speech regulation would claim that the actions proposed hereto reduce hate speech on social media sites will do very little to impact people. In fact, efforts to curtail hate speech on social media sites may even send this expression offline, where it would be even more difficult to monitor.

By enhancing global resilience against this insidious phenomenon, we can strengthen the bonds of society and build a better world for all.

******

****- List of References**

- American Library Association. (2017).Hate speech and hate crime .Retrieved from http:// www.ala.org/advocacy/intfreedom/hate.

- Bleich, E. (2011). 2011.The Rise of Hate Speech and Hate Crime Laws in Liberal Democracies. Journal of Ethnic and Migration Studies, 37(03), 917-934.

- Boromisza, H. (2013). Speaking hatefully: Culture, communication, and political action in Hungary. University Park, PA: Penn State University Press.

- Burch, E. (2011). Censoring Hate Speech in Cyberspace: A New Debate in New America. 3 N. C. L.J. & TECH, 175(04) ,188- 211.

- Burnap, P. (2015). Cyber hate speech on twitter: an application of machine classi-fication and statistical modeling for policy and decision making. Policy and Internet, 07(01), 223-242.

- Burnap, P.(2016). Us and them: Identifying cyber hate on Twitter across multiple protected characteristics.EPJ Data Science, 05(01), 67- 89.

- Buyse, A. (2014). Words of Violence"Fear Speech," or How Violent Conflict Escalation Relates to the Freedom of Expression. Human Rights Quarterly, 36(04), 779-797.

- Buyse, A. (2014). Words of violence: "Fear speech," or how violent conflict escalation relates to the freedom of expression. Human Rights Quarterly, 36(04), 779-797.

- Campbell, E. (2003). Attitudes of Botswana Citizens toward Immigrants: Signs of Xenophobia, Oxford: Blackwell's.

- Cole, J. (2009). Martyrdom: Radicalisation and Terrorist Violence Among British Muslims. London: Pennant Books.

- Djuric, N. (2015). Hate speech detection with comment embeddings". inProc. WWW Companion, 01(03), 29-49.

- Erjavec, K.(2012). You don't understand, this is a new war!": Analysis of hate speech in news web sites' comments. Mass Communication & Society, 15(04), 899-920.

- Gerstenfeld, P. (2003). Hate online: A content analysis of extremist Internet sites. New Media & Society, 03(01), 29-44.

- Gillespie, T. (2010).The politics of "platforms.New Media & Society, 12(03), 347-364.

- Gleason, P.(1991). Minorities (almost) all: The minority concept in American social thought. American Quarterly, 43(02), 392-424.

- Grosfoguel, R.(2011).Decolonizing post-colonial studies and paradigms of political-economy: Transmodernity, decolonial thinking and global coloniality. Transmodernity. Journal of Peripheral Cultural Production of the Luso-Hispanic World, 01(01), 1-38.
- Holt, K . (2015). Random acts of journalism?"How citizen journalists tell the news in Sweden".New Media & Society, 17(02), 1795-1810.
- Hopkins, N.(2006).Minority group members 'theories of intergroup contact: A case study of British Muslims' conceptualisations of 'Islamophobia' and social change. The British Journal of Social Psychology, 45(03), 245-264.
- Kwok, Y.(2013) Locate the hate: Detecting tweets against blacks. inProc,AAAI, 13(03), 1621-1622.
- Mogekwu, M. (2005). African Union: Xenophobia as poor intercultural information. Ecquid Novi, 26(01), 05-20.
- Nakamura, L.(2014).I WILL DO EVERYthing that am asked: Scambaiting, digital show-space, and the racial violence of social media. Journal of Visual Culture, 13(03), 257-274.
- Nobata, J.(2016). Abusivelanguage detection in online user content. inProc. WWW, , 03(02), 145-153.
- Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. New York, NY: Penguin
- Shepherd, T.(2015).Histories of hating.Social Media and Society, 01(02), 53-71.
- Shlapentokh, V. (2007).The Hatred of Others. The Kremlin's Powerful but Risky Weapon. World Affairs, 169(03), 134-142.
- Sutton, G.(2013).High times for hate crimes: Explaining thetemporal clustering of hate-motivated offending.Criminology,51(04), 871-894.
- Vis, F. (2013).Twitter as a reporting tool for breaking news: Journalists tweeting the 2011 UK riots".Digital Journalism, 01(01), 27-47.
- Waldron, J. (2012).The harm in hate speech. Cambridge: Harvard University Press.
- Wermiel, S. (2018).The Ongoing Challenge to Define Free Speech. Human Rights Magazine, 43(04), 21-44.
- Whine, M. (1999). Cyberspace A new medium for communication command and control by extremists. Studies in Conflict & Terrorism, 22(03), 231-245.
- Williams, M.(2016).Cyber Hate on social media in the aftermath of Woolwich: A case study in computational criminology and big data. British Journal of Criminology, 56(02), 211-238.
- Wright, B. (2011).Computer-Mediated Communication in Personal Relationships. New York: Peter Lang.
- Yulia, A. (2003). Hate speech online: Restricted or Protected?. Journal of Transnational Law & Policy, 12(02), 43-59.