Online Hate Speech

Online hate speech is a type of speech that takes place online with the purpose of attacking a person or a group based on their race, religion, ethnic origin, sexual orientation, disability, or gender. Online hate speech is the expression of conflicts between different groups within and across societies. Online hate speech is a vivid example of how the Internet brings both opportunities and challenges regarding the freedom of expression and speech while also defending human dignity. Multilateral treaties such as the International Covenant on Civil and Political Rights (ICCPR) have sought to define its contours. Multi-stakeholders processes (e.g. the Rabat Plan of Action) have tried to bring greater clarity and suggested mechanisms to identify hateful messages. Yet, hate speech is still a generic term in everyday discourse, mixing concrete threats to individuals and/or groups with cases in which people may be simply venting their anger against authority. Internet intermediaries—organizations that mediate online communication such as Facebook, Twitter, and Google—have advanced their own definitions of hate speech that bind users to a set of rules and allow companies to limit certain forms of expression. National and regional bodies have sought to promote understandings of the term that are more rooted in local traditions. The Internet's speed and reach makes it difficult for governments to enforce national legislation in the virtual world. Social media is a private space for public expression, which makes it difficult for regulators. Some of the companies owning these spaces have become more responsive towards tackling the problem of hate speech online. Politicians, activists, and academics discuss the character of online hate speech and its relation to offline speech and action, but the debates tend to be removed from systematic empirical evidence. The character of perceived hate speech and its possible consequences has led to placing much emphasis on the solutions to the problem and on how they should be grounded in international human rights norms. Yet this very focus has also limited deeper attempts to understand the causes underlying the phenomenon and the dynamics through which certain types of content emerge, diffuse and lead— or not—to actual discrimination, hostility or violence. Online hate speech has been on the rise since the start of 2020, with COVID-19 tensions, Anti-Asian rhetoric, ongoing racial injustice, mass civil unrest, violence, and the 2020 Presidential Election. Yet, many instances of hate speech have been refuted with the First Amendment, which allows online hate speech to continue.

Keywords: online hate ; online communication ; national legislation

1. Definitions

1.1. Hate Speech

Hate Speech is the controversial clash of freedom of expression and individual, group and minority rights, as well as concepts of dignity, liberty and equality.[1]

In national and international legislation, hate speech refers to expressions that advocate incitement to harm, particularly discrimination, hostility or violence, based upon the targets' social and/or demographic identity. Hate Speech may include, but is not limited to, speech that advocates, threatens, or encourages violent acts. The concept may extend also to expressions that foster a climate of prejudice and intolerance on the assumption that this may fuel targeted discrimination, hostility and violent attacks. At critical times, such as during elections, the concept of hate speech may be prone to manipulation; accusations of fomenting hate speech may be traded among political opponents or used by those in power to curb dissent and criticism. Hate speech (be it conveyed through text, images or sound) can be identified by approximation through the degrading or dehumanizing functions that it serves. Hate Speech contains two types of messages. The first message is addressed to the targeted group and functions to dehumanize and diminish members of this group. The following is an example of this type of message:

"Don't be fooled into thinking you are welcome here. [...] You are not wanted, and you and your families will be shunned, excluded, beaten, and driven out, whenever we can get away with it. We may have to keep a low profile right now. But don't get too comfortable. [...] Be afraid."[2]

The other message in hate speech is to let others with similar opinions know they are not alone, and to reinforce a sense of an in-group that is (purportedly) under threat. The following is an example of this type of message, sent to like-minded individuals:

"We know some of you agree that these people are not wanted here. We know that some of you feel that they are dirty (or dangerous or criminal or terrorist). Know now that you are not alone. [...] There are enough of us around to make sure these people are not welcome. There are enough of us around to draw attention to what these people are really like."[2]

1.2. Characteristics

The proliferation of hate speech online, observed by the UN Human Rights Council Special Rapporteur on Minority Issues poses a new set of challenges.[3] Both social networking platforms and organizations created to combat hate speech have recognized that hateful messages disseminated online are increasingly common and have elicited unprecedented attention to develop adequate responses.[4] According to HateBase, a web-based application that collects instances of hate speech online worldwide, the majority of cases of hate speech target individuals based on ethnicity and nationality, but incitements to hatred focusing on religion and class have also been on the rise.[5]

While hate speech online is not intrinsically different from similar expressions found offline, there are peculiar challenges unique to online content and its regulation. Those challenges related to its permanence, itinerancy, anonymity and complex cross-jurisdictional character.

Hate speech can stay online for a long time in different formats across multiple platforms, which can be linked repeatedly. As Andre Oboler, the CEO of the Online Hate Prevention Institute, has noted, "The longer the content stays available, the more damage it can inflict on the victims and empower the perpetrators. If you remove the content at an early stage you can limit the exposure. This is just like cleaning litter, it doesn't stop people from littering but if you do not take care of the problem it just piles up and further exacerbates."[6] Twitter's conversations organized around trending topics may facilitate the quick and wide spreading of hateful messages,[7] but they also offer the opportunity for influential speakers to shun messages and possibly end popular threads inciting violence. Facebook, on the contrary, may allow multiple threads to continue in parallel and go unnoticed; creating longer lasting spaces that offend, discriminate, and ridicule certain individuals and groups.[1]

Hate speech online can be itinerant. Even when content is removed, it may find expression elsewhere, possibly on the same platform under a different name or on different online spaces. If a website is shut down, it can quickly reopen using a web-hosting service with less stringent regulations or via the reallocation to a country with laws imposing higher threshold for hate speech. The itinerant nature of hate speech also means that poorly formulated thoughts, or under-the-influence behavior, that would have not found public expression and support in the past may now land on spaces where they can be visible to large audiences.[1]

Anonymity can also present a challenge to dealing with online hate speech. "(T)he internet facilitates anonymous and pseudonymous discourse, which can just as easily accelerate destructive behavior as it can fuel public discourse."[8] As Drew Boyd, Director of Operations at The Sentinel Project, has stated, "the Internet grants individuals the ability to say horrific things because they think they will not be discovered. This is what makes online hate speech so unique, because people feel much more comfortable speaking hate as opposed to real life when they have to deal with the consequences of what they say."[9] *China* and *South Korea* enforce real-name policies for social media. Facebook, LinkedIn, and Quora have sought to activate a real-name system to have more control on online hate speech. Such measures have been deeply contested as they hit at the right to privacy and its intersection with free expression.

Many instances of online hate speech are posted by Internet "trolls," which are typically pseudonymous users who post shocking, vulgar, and overall untrue content that is meant to trigger a negative reaction out of people, as well as trick, influence, and sometimes recruit them, if they share the same opinion.[10] Social media has provided a platform for radical or extremist political groups to form, network, and collaborate to spread their messages of anti-establishment and anti-political correctness, and promote ideologies that are racist, anti-feminist, homophobic, etc.[11] Fully-anonymous online communication is rare, as it requires the user to employ highly technical measures to ensure that they cannot be easily identified.[1]

A further complication is the transnational reach of the Internet, raising issues of cross jurisdictional co-operation in regard to legal mechanisms for combating hate speech. While there are Mutual Legal Assistance treaties in place across Europe, Asia, and North America, these are characteristically slow to work. The transnational reach of many private-sector Internet intermediaries may provide a more effective channel for resolving issues in some cases, although these bodies are also often impacted upon by cross-jurisdictional appeals for data (such as revealing the identity of the author(s) of a particular

content).[1] Each country has a different understanding of what defines as hate speech, making it difficult to prosecute perpetrators of online hate speech, especially in The US where there is a deep-rooted constitutional commitment to the Freedom of Speech.[12]

Unlike the dissemination of hate speech through conventional channels, online hate speech dissemination involves multiple actors, whether knowingly or not. When perpetrators disseminate their hateful message on social media, they do not only hurt their victims, but may also violate terms of service in that platform and at times even state law, depending on their location. The victims, on their part, may feel helpless in the face of online harassment, not knowing to whom they should turn to for help. Nongovernmental organizations and lobby groups have been raising awareness and encourage different stakeholders to take action.[1]

Some tech companies, such as Facebook, use Artificial Intelligence (AI) systems to monitor hate speech.[13] However, Artificial Intelligence may not always be an effective way of monitoring hate speech since the systems lack emotion and judgment skills that humans have.[14] For example, a user might post or comment something that classifies as hate speech and/or violates community guidelines, but if the target word is misspelled, or some letters are replaced with symbols, the AI systems will not recognize it. Therefore, humans still have to monitor the AI systems that monitor hate speech, which is a concept referred to as "Automation's Last Mile."[14]

2. Frameworks

2.1. *Stormfront* Precedent

In the aftermath of 2014's dramatic incidents, calls for more restrictive or intrusive measures to contain the Internet's potential to spread hate and violence are common, as if the links between online and offline violence were well known. On the contrary, as the following example indicates, appearances may often be deceiving. Stormfront is considered the first "hate website."[15] Launched in March 1995 by a former Ku Klux Klan leader, it quickly became a popular space for discussing ideas related to Neo-Nazism, White nationalism and White separatism, first in the *United States* and then globally.[16] The forum hosts calls for a racial holy war and incitement to use violence to resist immigration.[16] and is considered a space for recruiting activists and possibly coordinating violent acts.[17] The few studies that have explored the identities of Stormfront actually are depict a more complex picture. Rather than seeing it as a space for coordinating actions. Well-known extreme right activists have accused the forum to be just a gathering for "keyboard warriors." One of them for example, as reported by De Koster and Houtman, stated, "I have read quite a few pieces around the forum, and it strikes me that a great fuss is made, whereas little happens. The section activism/politics itself is plainly ridiculous. [...] Not to mention the assemblies where just four people turn up."[18] Even more revealing are some of the responses to these accusations provided by regular members of the website. As one of them argued, "Surely, I am entitled to have an opinion without actively carrying it out. [...] I do not attend demonstrations and I neither join a political party. If this makes me a keyboard warrior, that is all right. I feel good this way. [...] I am not ashamed of it."[18] De Koster and Houtman surveyed only one national chapter of Stormfront and a non-representative sample of users, but answers like those above should at least invite to caution towards hypotheses connecting expressions and actions, even in spaces whose main function is to host extremist views.[1] The Southern Poverty Law Center published a study in 2014 that found users of the site "were allegedly responsible for the murders of nearly 100 people in the preceding five years."[19]

2.2. International Principles

Hate speech is not explicitly mentioned in many international human rights documents and treaties, but it is indirectly called upon by some of the principles related to human dignity and freedom of expression. For example, the 1948 Universal Declaration of Human Rights (UDHR), which was drafted as a response to the atrocities of the World War II, contains the right to equal protection under the law in Article 7, which proclaims that: "All are entitled to equal protection against any discrimination in violation of this Declaration and against any incitement to such discrimination."[20] The UDHR also states that everyone has the right to freedom of expression, which includes "freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers."[20]

The UDHR was decisive in setting a framework and agenda for human rights protection, but the Declaration is non-binding. A series of binding documents have been subsequently created to offer a more robust protection for freedom of expression and protection against discrimination. The International Covenant on Civil and Political Rights (ICCPR) addresses hate speech and contains the right to freedom of expression in Article 19[20] and the prohibition of advocacy to hatred that constitutes incitement to discrimination, hostility or violence in Article 20.[20] Other more tailored international legal instruments contain provisions that have repercussions for the definition of hate speech and identification of

responses to it, such as: the Convention on the Prevention and Punishment of the Crime of Genocide (1951), the International Convention on the Elimination of All Forms of Racial Discrimination, ICERD (1969), and, to a lesser extent, the Convention on the Elimination of All Forms of Discrimination against Women, CEDAW (1981).[1]

2.3. Hate speech and the ICCPR

The ICCPR is the legal instrument most commonly referred to in debates on hate speech and its regulation, although it does not explicitly use the term "hate speech." Article 19, which is often referred to as part of the "core of the Covenant", [21] provides for the right to freedom of expression. This sets out the right, and it also includes general strictures to which any limitation of the right must conform in order to be legitimate. Article 19 is followed by Article 20 that expressly limits freedom of expression in cases of "advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence."[22] The decision to include this provision, which can be characterized as embodying a particular conceptualization of hate speech, has been deeply contested. The Human Rights Committee, the *United Nations* body created by the ICCPR to oversee its implementation, cognizant of the tension, has sought to stress that Article 20 is fully compatible with the right to freedom of expression.[23] In the ICCPR, the right to freedom of expression is not an absolute right. It can legitimately be limited by states under restricted circumstances:

"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."[24]

Between Article 19 (3) and Article 20, there is a distinction between optional and obligatory limitations to the right to freedom of expression. Article 19 (3) states that limitations on freedom of expression "may therefore be subject to certain restrictions", as long as they are provided by law and necessary to certain legitimate purposes. Article 20 states that any advocacy of (certain kinds of) hatred that constitutes incitement to discrimination, hostility or violence "shall be prohibited by law." Despite indications on the gravity of speech offenses that should be prohibited by law under Article 20, there remains complexity.[25] In particular there is a grey area in conceptualizing clear distinctions between (i) expressions of hatred, (ii) expression that advocate hatred, and (iii) hateful speech that specifically constitutes incitement to the practical harms of discrimination, hostility or violence. While states have an obligation to prohibit speech conceived as "advocacy to hatred that constitutes incitement to discrimination, hostility or violence", as consistent with Article 20 (2),[26] how to interpret such is not clearly defined.[27]

2.4. Other International Legal Instruments

ICERD

The International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), which came into force in 1969, has also implications for conceptualizing forms of hate speech. The ICERD differs from the ICCPR in three respects.[1] Its conceptualization of hate speech is specifically limited to speech that refers to race and ethnicity. It asserts in Article 4, paragraph (a), that state parties:

"Shall declare as an offense punishable by law all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another color or ethnic origin, and also the provision of any assistance to racist activities, including the financing thereof; This obligation imposed by the ICERD on state parties is also stricter than the case of Article 20 of the ICCPR covering the criminalization of racist ideas that are not necessarily inciting discrimination, hostility or violence."

An important difference is in the issue of intent. The concept of "advocacy of hatred" introduced in the ICCPR is more specific than discriminatory speech described in the ICERD, since it is taken to require consideration of the intent of author and not the expression in isolation—this is because "advocacy" is interpreted in the ICCPR as requiring the intent to sow hatred.[28] The Committee on the Elimination of Racial Discrimination has actively addressed hate speech in its General Recommendation 29, in which the Committee recommends state parties to:

"(r) Take measures against any dissemination of ideas of caste superiority and inferiority or which attempt to justify violence, hatred or discrimination against descent-based communities; (s) Take strict measures against any incitement to discrimination or violence against the communities, including through the Internet; (t) Take measures to raise awareness among media professionals of the nature and incidence of descent-based discrimination;"[29]

These points, which reflect the ICERD's reference to the dissemination of expression, have significance for the Internet. The expression of ideas in some online contexts may immediately amount to spreading them. This is especially relevant for private spaces that have begun to play a public role, as in the case of many social networking platforms.[1]

Genocide Convention

Similarly to the ICERD, the Genocide Convention aims to protect groups defined by race, nationality or ethnicity, although it also extends its provisions to religious groups. When it comes to hate speech the Genocide Convention is limited only to acts that publicly incite to genocide, recognized as "acts committed with intent to destroy, in whole or in part, a national, ethnical, racial or religious group", regardless of whether such acts are undertaken in peacetime or in wartime.[1] Specifically gender-based hate speech (as distinct from discriminatory actions) is not covered in depth in international law.

CEDAW

The Convention on the Elimination of All Forms of Discrimination against Women (CEDAW), which entered into force in 1981, imposes obligations on states to condemn discrimination against women[30] and "prevent, investigate, prosecute and punish" acts of gender-based violence.[31]

2.5. Regional Responses

Most regional instruments do not have specific articles prescribing prohibition of hate speech, but they more generally allow states to limit freedom of expression—which provisions can be applied to specific cases.[1]

American Convention on Human Rights

The American Convention on Human Rights describes limitations on freedom of expression in a manner similar to the ICCPR in Article 19 (3). The Organization of American States has also adopted another declaration on the principles of freedom of expression, which includes a specific clause stating that "prior conditioning of expressions, such as truthfulness, timeliness or impartiality is incompatible with the right to freedom of expression recognized in international instruments."[32] The Inter-American Court has advised that "(a)buse of freedom of information thus cannot be controlled by preventive measures but only through the subsequent imposition of sanctions on those who are guilty of the abuses." [33] The Court also imposes a test for States willing to enact restrictions on freedom of expression, as they need to observe the following requirements: "a) the existence of previously established grounds for liability; b) the express and precise definition of these grounds by law; c) the legitimacy of the ends sought to be achieved; d) a showing that these grounds of liability are 'necessary to ensure' the aforementioned ends."[34] The Inter-American System has a Special Rapporteur on Freedom of Expression who conducted a comprehensive study on hate speech. He concluded that the Inter-American Human Rights System differs from the United Nations and the European approach on a key point: The Inter-American system covers and restricts hate speech that actually leads to violence, and solely such speech can be restricted.[34]

African Charter on Human Rights and Peoples' Rights

The African Charter on Human Rights and Peoples' Rights takes a different approach in Article 9 (2), allowing for restrictions on rights as long as they are "within the law." This concept has been criticized and there is a vast amount of legal scholarship on the so-called "claw-back" clauses and their interpretation.[35] The criticism is mainly aimed at the fact that countries can manipulate their own legislation and weaken the essence of the right to freedom of expression. The Declaration of Principles on Freedom of Expression in Africa elaborates a higher standard for limitations on freedom of expression. It declares that the right "should not be restricted on public order or national security grounds unless there is a real risk of harm to a legitimate interest and there is a close causal link between the risk of harm and the expression."[36]

Cairo Declaration on Human Rights in Islam

In 1990, the Organization of the Islamic Conference (which was later renamed Organization of Islamic Cooperation, OIC) adopted the Cairo Declaration on Human Rights in Islam (CDHRI), which calls for criminalization of speech that extends beyond cases of imminent violence to encompass "acts or speech that denote manifest intolerance and hate."[37]

Arab Charter on Human Rights

The Arab Charter on Human Rights, which was adopted by the Council of the League of Arab States in 2004, includes in Article 32 provisions that are relevant also for online communication as it guarantees the right to "freedom of opinion and expression, and the right to seek, receive and impart information and ideas through any medium, regardless of geographical boundaries."[38] It allows a limitation on a broad basis in paragraph 2 "Such rights and freedoms shall be exercised in conformity with the fundamental values of society."[39]

ASEAN Human Rights Declaration

The ASEAN Human Rights Declaration includes the right to freedom of expression in Article 23. Article 7 of the Declaration provides for general limitations, affirming, "the realization of human rights must be considered in the regional and national context bearing in mind different political, economic, legal, social, cultural, historical and religious backgrounds."[40]

Charter of Fundamental Rights of the European Union

The Charter of Fundamental Rights of the European Union which declares the right to freedom of expression in Article 11, has a clause which prohibits abuse of rights. It asserts that the Charter must not be interpreted as implying any "limitation to a greater extent than is provided for therein."[41] An example of a limitation which implies a strict test of necessity and proportionality is the provision on freedom of expression in the European Convention on Human Rights, which underlines that the exercise of freedom of expression carries duties and responsibilities. It "may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary."[42]

The European Court of Human Rights is careful to distinguish between hate speech and the right of individuals to express their views freely, even if others take offense.[43] There are regional instances relevant specifically to online hate speech. The Council of Europe (CoE) in 2000 issued a General Policy Recommendation on Combating the Dissemination of Racist, Xenophobic and Anti-Semitic Material via the Internet.[44] The creation of the CoE Convention on Cybercrime in 2001, which regulates mutual assistance regarding investigative powers, provides signatory countries with a mechanism to deal with computer data, which would include transnational hate speech online.[45] In 2003 the CoE launched an additional protocol to the Convention on Cybercrime which addresses online expression of racism and xenophobia. The convention and its protocol were opened for signature and ratification of countries outside Europe, and other countries, such as *Canada* and *South Africa* , are already part of this convention. The Protocol imposes an obligation on Member States to criminalize racist and xenophobic insults online of "(i) persons for the reason that they belong to a group distinguished by race, color, descent or national or ethnic origin, as well as religion, if used as a pretext for any of these factors; or (ii) a group of persons which is distinguished by any of these characteristics."[46]

2.6. Private Spaces

Internet intermediaries such as social networking platforms, Internet Service Providers or Search Engines, stipulate in their terms of service how they may intervene in allowing, restricting, or channelling the creation and access to specific content. A vast amount of online interactions occur on social networking platforms that transcend national jurisdictions and which platforms have also developed their own definitions of hate speech and measures to respond to it. For a user who violates the terms of service, the content he or she has posted may be removed from the platform, or its access may be restricted to be viewed only by a certain category of users (e.g. users living outside a specific country).[1]

The principles that inspire terms of service agreements and the mechanisms that each company develops to ensure their implementation have significant repercussions on the ability that people have to express themselves online as well as to be protected from hate speech. Most intermediaries have to enter in negotiations with national governments to an extent that varies according to the type of intermediary, areas where the company is registered, and the legal regime that applies. As Tsesis explains, "(i)f transmissions on the Internet are sent and received in particular locations, then specific fora retain jurisdiction to prosecute illegal activities transacted on the Internet."[1] Internet Service Providers are the most directly affected by national legislation because they have to be located in a specific country to operate. Search Engines, while they can modify search results for self-regulatory or commercial reasons, have increasingly tended to adapt to the intermediary liability regime of both their registered home jurisdictions and other jurisdictions in which they provide their services, either removing links to content proactively or upon request by authorities.[47]

All Internet intermediaries operated by private companies are also expected to respect human rights. This is set out in the Guiding Principles on Business and Human Rights elaborated by the United Nations Office of the High Commissioner for Human Rights. The document emphasizes corporate responsibility in upholding human rights. In principle 11, it declares that: "Business enterprises should respect human rights. This means that they should avoid infringing on the human rights of others and should address adverse human rights impacts with which they are involved."[48] The United Nations Guiding Principles also indicate that in cases in which human rights are violated, companies should "provide for or cooperate in their remediation through legitimate processes."[48] In the case of Internet intermediaries and conceptions of hate speech, this means that they should ensure that measures are in place to provide a commensurate response.[1]

▌3. Social Responses

3.1. Case Studies

The American Trends Panel

Pew Research Center surveyed over 10,000 adults in July 2020 to study social media's effect on politics and social justice activism. 23% of respondents, who are adult social media users, reported that social media content has caused them to change their opinion, positively or negatively, on a political or social justice issue.[49] 35% of those respondents cited the Black Lives Matter movement, police reform, and/or race relations.[49] 18% of respondents reported a change of opinion on political parties, ideologies, politicians, and/or President Donald Trump.[49] 9% of respondents cited social justice issues, such as LGBTQIA+ rights, feminism, immigration, etc.[49] 8% of respondents changed their opinion on the COVID-19 pandemic, and 10% cited other opinions.[49] Based on these results, social media plays an important role in influencing public opinion.

Media Manipulation and Disinformation Online

A study conducted by researchers Alice Marwick and Rebecca Lewis observed media manipulation and explored how the alt-right marketed, networked, and collaborated to influence their controversial beliefs that could have potentially helped influence President Trump's victory in the 2016 election. Unlike mainstream media, the alt-right does not need to comply to any rules when it comes to influence and do not need to worry about network ratings, audience reviews, or sensationalism.[50] Alt-right groups can share and persuade others of their controversial beliefs as bluntly and brashly as they desire, on any platform, which may have played a role in the 2016 election. Although the study could not conclude what exactly the effect was on the election, but did provide extensive research on the characteristics of media manipulation and trolling.[50]

Hate Speech and Linguistic Profiling in Online Gaming

Professor and gamer Kishonna Gray studied intersectional oppressions in the online gaming community and called on Microsoft and game developers to "critically assess the experiences of non-traditional gamers in online communities...recognize diversity...[and that] the default gaming population are deploying hegemonic whiteness and masculinity to the detriment of non-white and/or non-male users within the space."[51] Gray examined sexism and racism in the online gaming community. Gamers attempt to identify the gender, sexuality, and ethnic background of their teammates and opponents through linguistic profiling, when the other players cannot be seen.[51] Due to the intense atmosphere of the virtual gaming sphere, and the inability to not be seen, located, or physically confronted, gamers tend to say things on the virtual game that they likely would not have said in a public setting. Many gamers of marginalized communities have branched off from the global gaming network and joined "clans," that consist of only gamers of the same gender, sexuality, and/or ethnic identity, to avoid discrimination while gaming. A study found that 78 percent of all online gamers play in "guilds" which are smaller groups of players, similar to "clans."[52] One of the most notable "clans," Puerto Reekan Killaz, have created an online gaming space where Black and Latina women of the LGBTQIA+ community can play without risk of racism, nativism, homophobia, sexism, and sexual harassment.[51]

In addition to hate speech, Professor and gamer Lisa Nakamura found that many gamers have experienced identity tourism- which is when a person or group appropriate and pretend to members of another group- as Nakamura observed white male gamers play as Japanese "geisha" women.[53] Identity Tourism often leads to stereotyping, discrimination, and cultural appropriation.[53] Nakamura called on the online gaming community to recognize Cybertyping- "the way the Internet propagates, disseminates, and commodifies images of race and racism."[54]

Anti-Chinese Rhetoric Employed by Perpetrators of Anti-Asian Hate

As of August 2020, over 2,500 Asian-Americans have reported experiencing racism fueled by COVID-19, with 30.5% of those cases containing anti-Chinese rhetoric, according to Stop AAPI (Asian-American/Pacific Islander) Hate. The language used in these incidents are divided into five categories: virulent animosity, scapegoating of China, anti-immigrant nativism, racist characterizations of Chinese, and racial slurs. 60.4% of these reported incidents fit into the virulent animosity category, which includes phrases such as "get your Chinese a** away from me!"[55]

Myanmar

The Internet has grown at unprecedented rates. Myanmar is transitioning towards greater openness and access, which leads to social media negatives, such as using hate speech and calls to violence.[56] In 2014, the UN Human Rights Council Special Rapporteur on Minority Issues expressed her concern over the spread of misinformation, hate speech and incitement to violence, discrimination and hostility in the media and Internet, particularly targeted against a minority

community.[3] The growing tension online has gone parallel with cases of actual violence leaving hundreds dead and thousands displaced.[57] One challenge in this process has concerned ethnic and religious minorities. In 2013, 43 people were killed due to clashes that erupted after a dispute in the Rakhine state in the Western Part of the country.[57] A year earlier, more than 200 people were killed and thousands displaced 37 because of ethnic violence, which erupted after an alleged rape case.[58] Against this backdrop, the rapid emergence of new online spaces, albeit for a fraction of the population, has reflected some of these deeply rooted tensions in a new form.[1]

Dealing with intolerance and hate speech online is an emerging issue. Facebook has rapidly become the platform of choice for those citizens making their first steps online. In this environment there have been individual and groups, which have championed a more aggressive use of the medium, especially when feeling protected by a sense of righteousness and by claims to be acting in defense of the national interest. Political figures have also used online media for particular causes. In social media, there has been the use of derogatory terms in reference to minorities.[59] In this complex situation, a variety of actors has begun to mobilize, seeking to offer responses that can avoid further violence. Facebook has sought to take a more active role in monitoring the uses of the social network platform in Myanmar, developing partnerships with local organizations and making guidelines on reporting problems accessible in Burmese.[1][60]

The local civil society has constituted a strong voice in openly condemning the spread of online hate speech, but at the same time calling for alternatives to censorship. Among the most innovative responses has been Panzagar, which in Burmese means "flower speech", a campaign launched by blogger and activist Nay Phone Latt to openly oppose hate speech. The goal of the initiative was offering a joyful example of how people can interact, both online and offline.[61] Local activists have been focussed upon local solutions, rather than trying to mobilize global civil society on these issues. This is in contrast to some other online campaigns that have been able to attract the world's attention towards relatively neglected problems. Initiatives such as those promoted by the Save Darfur Coalition for the civil war in Sudan, or the organization Invisible Children with the Kony2012 campaign that denounced the atrocities committed by the Lord Resistance Army, are popular examples. As commentaries on these campaigns have pointed out, such global responses may have negative repercussions on the ability for local solutions to be found.[62]

Ethiopia

The long-lived ethnic rivalry in Ethiopia between the Oromo people and the Amhara people found battleground on Facebook, leading to hate speech, threats, disinformation, and even deaths. Facebook does not have fact-checkers that speak either of the dominant languages spoken in Ethiopia nor do they provide translations of the Community Standards, therefore hate speech on Facebook is largely unmonitored in Ethiopia. Instead, Facebook relies on activists to flag potential hate speech and disinformation, but the many burnout activists feel mistreated.[63]

In October 2019, Ethiopian activist Jawar Mohammed falsely announced on Facebook that the police were going to detain him, citing religious and ethnic tension. This prompted the community to protest his alleged detainment and the racial and ethnic tensions, which lead to over 70 deaths.[64]

A disinformation campaign originated on Facebook centering on popular Ethiopian singer, Hachalu Hundessa, of the Oromo ethnic group. The posts accused Hundessa of supporting their controversial Prime Minister Abiy Ahmed, whom Oromo nationalists disproved of for his catering to other ethnic groups. Hundessa was assassinated in June 2020 following these hateful Facebook posts, prompting public outrage. Facebook users blamed the Amhara people for Hundessa's assassination without any evidence in a long thread of hateful content.[63] According to The Network Against Hate Speech, many Facebook posts called for "genocidal attacks against an ethnic group or a religion — or both at the same time; and ordering people to burn civilians' properties, kill them brutally, and displace them."[63] The violence in the streets and on Facebook escalated to the point that the Ethiopian Government had to shut down the Internet for three weeks. However, neighboring countries could still post and access the hateful content of the matter, and the volunteer activists could not access the Internet to flag hate speech. Therefore "there are hours of video that came from the diaspora community, extremist content, saying we need to exterminate this ethnic group," according to Professor Endalk Chala of Hamline University.[63]

Facebook officials ventured to Ethiopia to investigate, but did not release their findings. Facebook announced that they are hiring moderators who can speak Amharic and other languages in Ethiopia but did not provide extensive detail.[63]

3.2. Private Companies

Internet intermediaries have developed disparate definitions of hate speech and guidelines to regulate it. Some companies do not use the term hate speech, but have a descriptive list of terms related to it.[1]

Yahoo!

Yahoo!'s terms of service prohibit the posting of "content that is unlawful, harmful, threatening, abusive, harassing, tortuous, defamatory, vulgar, obscene, libellous, invasive of another's privacy, hateful, or racially, ethnically or otherwise objectionable."[65]

Twitter

In December 2017, Twitter began enforcing new policies towards hate speech, banning multiple accounts as well as setting new guidelines for what will be allowed on their platform.[66] There is an entire page in the Twitter Help Center devoted to describing their Hateful Conduct Policy, as well as their enforcement procedures. The top of this page states "Freedom of expression means little if voices are silenced because people are afraid to speak up. We do not tolerate behavior that harasses, intimidates, or uses fear to silence another person's voice. If you see something on Twitter that violates these rules, please report it to us." Twitter's definition of hate speech ranges from "violent threats" and "wishes for the physical harm, death, or disease of individuals or groups" to "repeated and/or non-consensual slurs, epithets, racist and sexist tropes, or other content that degrades someone."

Punishments for violations range from suspending a user's ability to tweet until they take down their offensive/ hateful post to the removal of an account entirely. In a statement following the implementation of their new policies, Twitter said "In our efforts to be more aggressive here, we may make some mistakes and are working on a robust appeals process" . . . "We'll evaluate and iterate on these changes in the coming days and weeks, and will keep you posted on progress along the way".[67] These changes come amidst a time where action is being taken to prevent hate speech around the globe, including new laws in Europe which pose fines for sites unable to address hate speech reports within 24 hours.[68]

YouTube

YouTube, a subsidiary of the tech company Google, allows for easy content distribution and access for any content creator, which creates opportunity for the audience to access content that shifts right or left of the 'moderate' ideology common in mainstream media.[69] YouTube provides incentives to popular content creators, prompting some creators to optimize the YouTuber experience and post shock-valued content that may promote extremist, hateful ideas.[69][70] Content diversity and monetization on YouTube directs a broad audience to the potentially harmful content from extremists.[69][70] YouTube allows creators to personally brand themselves, making it easy for young subscribers to form a parasocial relationship with them and act as "regular" customers.[70] In 2019, YouTube demonetized political accounts,[71] but radical content creators still have their channels and subscribers to keep them culturally relevant and financially afloat.[70]

YouTube has outlined a clear "Hate Speech Policy" amidst several other user policies on their website.[72] The policy is worded as such: "We encourage free speech and try to defend your right to express unpopular points of view, but we don't permit hate speech. Hate speech refers to content that promotes violence against or has the primary purpose of inciting hatred against individuals or groups based on certain attributes, such as: race or ethnic origin, religion, disability, gender, age, veteran status, sexual orientation/gender identity". YouTube has built in a user reporting system in order to counteract the growing trend of hate speech.[73] Among the most popular deterrents against hate speech, users are able to anonymously report another user for content they deem inappropriate. The content is then reviewed against YouTube policy and age restrictions, and either taken down or left alone.

Facebook

Facebook's terms forbid content that is harmful, threatening or which has potential to stir hatred and incite violence. In its community standards, Facebook elaborates that "Facebook removes hate speech, which includes 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."[74] It further states that "We allow humour, satire or social commentary related to these topics, and we believe that when people use their authentic identity, they are more responsible when they share this kind of commentary. For that reason, we ask that Page owners associate their name and Facebook Profile with any content that is insensitive, even if that content does not violate our policies. As always, we urge people to be conscious of their audience when sharing this type of content."[74]

Instagram, a photo and video-sharing platform owned by Facebook, has similar hate speech guidelines as Facebook, but is not divided in tiers. Instagram's Community Guidelines also forbid misinformation, nudity, self-injury glorification, and posting copyrighted content without authorization.[75]

Facebook's hate speech policies are enforced by 7,500 content reviewers, as well as many Artificial Intelligence monitors. Because this requires difficult decision making, controversy arises among content reviewers over enforcement of policies. Some users seem to feel as though the enforcement is inconsistent. One apt past example is two separate but similarly graphic postings that wished death to members of a specific religion. Both post were flagged by users and reviewed by

Facebook staff. However, only one was removed even though they carried almost identical sentiments.[76] In a quote regarding hate speech on the platform, Facebook Vice President of Global Operations, Justin Osofky stated, "We're sorry for the mistakes we have made — they do not reflect the community we want to help build…We must do better."[77]

There has been additional controversy due to the specificity of Facebook's hate speech policies. On many occasions there have been reports of status updates and comments that users feel are insensitive and convey hatred. However these posts do not technically breach any Facebook policies because their speech does not attack others based on the company's list of protected classes. For example, the statement "Female sports reporters need to be hit in the head with hockey pucks," would not be considered hate speech on Facebook's platform and therefore would not be removed.[78] While the company protects against gender based hatred, it does not protect against hatred based on occupation.

Facebook has been accused of holding bias when policing hate speech, citing political campaign ads that may promote hate or misinformation that have made an impact on the platform.[13] Facebook adjusted their policies after receiving backlash, accusations, and large corporations pulled their ads from the platform to protest the platform's loose handling of hate speech and misinformation.[13] As of 2020, political campaign ads have a "flag" feature that notes that the content is newsworthy but that it may violate some community guidelines.[13]

Facebook also tries to accommodate users who promote other hate speech content with the intent of criticizing it. In these cases, users are required make it clear that their intention is to educate others. If this intention is unclear then Facebook reserves the right to censor the content.[79] When Facebook initially flags content that may contain hate speech, they then designate it to a Tier 1, 2, and 3 scale, based on the content's severity. Tier 1 is the most severe and Tier 3 is the least. Tier 1 includes anything that conveys "violent speech or support for death/disease/harm."[80] Tier 2 is classified as content that slanders another user's image mentally, physically, or morally.[81] Tier 3 includes anything that can potentially exclude or discriminate against others, or that uses slurs about protected groups, but does not necessarily apply to arguments to restrict immigration or criticism of existing immigration policies.[81]

In March 2019, Facebook banned content supporting white nationalism and white separatism, extending a previous ban of white supremacy content.[82] In May 2019, it announced bans on several prominent people for violations of its prohibition on hate speech, including Alex Jones, Louis Farrakhan, Milo Yiannopoulos, Laura Loomer, and Paul Nehlen.[83]

In 2020, Facebook added guidelines to Tier 1 that forbid blackface, racial comparisons to animals, racial or religious stereotypes, denial of historical events, and objectification of women and the LGBTQIA+ community.[84]

Hate Speech on Facebook and Instagram quadrupled in 2020, leading to the removal of 22.5 million posts on Facebook and 3.3 million posts on Instagram in the second quarter of 2020 alone.[13]

Microsoft

Microsoft has specific rules concerning hate speech for a variety of its applications. Its policy for mobile phones prohibits applications that "contain any content that advocates discrimination, hatred, or violence based on considerations of race, ethnicity, national origin, language, gender, age, disability, religion, sexual orientation, status as a veteran, or membership in any other social group."[85] The company has also rules regarding online gaming, which prohibit any communication that is indicative of "hate speech, controversial religious topics and sensitive current or historical events."[86]

TikTok

Your account has limited access

For security reasons, your account has limited access to the site for a few days. If you have any question, please contact our Help Center.

Okay

"Facebook Hates Google+ ???" by Frederick Md Publicity is licensed with CC BY 2.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/2.0/

TikTok lacks clear guidelines and control on hate speech, which allows bullying, harassment, propaganda, and hate speech to become part of normal discourse on TikTok. Far-right hate groups, terrorist organizations, and pedophiles thrive on TikTok by spreading and encouraging hate to an audience as young as 13 years old.[87] Children are naive and easily influenced by other people and messages, therefore they are more likely to listen and repeat what they are being shown

or told.[88] The Internet does not have an excessively monitored space that guarantees safety for children, so as long as the internet is public, children and teenagers are bound to come across hate speech.[88] From there, young teenagers have a tendency to let their curiosity lead them into furthering their interest and research into radical ideas.[88]

However, children cannot take accountability for their actions in the way that adults can and should,[88] placing the blame on not only the person who posted the vulgar content, but the social media platform itself. Therefore, TikTok is criticized for their handling of hate speech on the platform. While TikTok prohibits bullying, harassment, and any vulgar or hateful speech in its Terms & Conditions, TikTok has not been active long enough to have developed an effective method to monitor this content.[87] Other social media platforms such as Instagram, Twitter, and Facebook have been active long enough to know how to battle online hate speech and vulgar content,[87] but the audience on those platforms are old enough to take accountability for the messages they spread.[88]

TikTok, on the other hand, has to take some responsibility for the content distributed to its young audience.[87] TikTok users are required to be of at least 13 years of age, however that requirement can be easily waived, as apps cannot physically access users' age. Researcher Robert Mark Simpson concluded that combatting hate speech on youth-targeted media "might bear more of a resemblance to regulations governing adult entertainment than to prohibitions on Holocaust denial."[88]

3.3. Media and Information Literacy

Media and Information Literacy aims to help people to engage in a digital society by being able to use, understand, inquire, create, communicate and think critically; while being able to effectively access, organize, analyze, evaluate, and create messages in a variety of forms.[89]

Citizenship education focuses on preparing individuals to be informed and responsible citizens through the study of rights, freedoms, and responsibilities and has been variously employed in societies emerging from violent conflict.[90] One of its main objectives is raising awareness on the political, social and cultural rights of individuals and groups, including freedom of speech and the responsibilities and social implications that emerge from it. The concern of citizenship education with hate speech is twofold: it encompasses the knowledge and skills to identify hate speech, and should enable individuals to counteract messages of hatred.[91] One of its current challenges is adapting its goals and strategies to the digital world, providing not only argumentative but also technological knowledge and skills that a citizen may need to counteract online hate speech.[1]

Information literacy cannot avoid issues such as rights to free expression and privacy, critical citizenship and fostering empowerment for political participation.[92] Multiple and complementary literacies become critical. The emergence of new technologies and social media has played an important role in this shift. Individuals have evolved from being only consumers of media messages to producers, creators and curator of information, resulting in new models of participation that interact with traditional ones, like voting or joining a political party. Teaching strategies are changing accordingly, from fostering critical reception of media messages to include empowering the creation of media content.[93]

The concept of media and information literacy itself continues to evolve, being augmented by the dynamics of the Internet. It is beginning to embrace issues of identity, ethics and rights in cyberspace.[94] Some of these skills can be particularly important when identifying and responding to hate speech online.

Series of initiatives aimed both at providing information and practical tools for Internet users to be active digital citizens:

- 'No place for hate' by Anti-Defamation League (ADL), *United States* ;
- 'In other words' project by Provincia di Mantova and the European Commission;
- 'Facing online hate' by MediaSmarts, *Canada* ;
- 'No hate speech movement' by Youth Department of the Council of Europe, Europe;[95]
- 'Online hate' by the Online Hate Prevention Institute, *Australia* .

Education is also seen as being a tool against hate speech. Laura Geraghty from the 'No Hate Speech Movement' affirmed: "Education is key to prevent hate speech online. It is necessary to raise awareness and empower people to get online in a responsible way; however, you still need the legal background and instruments to prosecute hate crimes, including hate speech online, otherwise the preventive aspect won't help."[1][96]

▌4. Sources

This article incorporates text from a free content work. Licensed under CC BY SA 3.0 IGO License statement: _Countering Online Hate Speech_, 73, Iginio Gagliardone, Danit Gal, Thiago Alves, Gabriela Martinez, UNESCO. UNESCO. To learn how to add open license text to HandWiki articles, please see this how-to page. For information on reusing text from HandWiki, please see the terms of use.

References

1. Gagliardone, Iginio; Gal, Danit; Alves, Thiago; Martinez, Gabriela (2015). Countering Online Hate Speech. UNESCO. pp. 75. http://unesdoc.unesco.org/images/0023/002332/233231e.pdf.

2. Waldron, Jeremy (2012). The Harm in Hate Speech. Harvard University Press. ISBN 978-0-674-06589-5. https://www.jstor.org/stable/j.ctt2jbrjd.

3. Izsak, Rita (2015). Report of the Special Rapporteur on minority issues, Rita Izsák. Human Rights Council.

4. See Council of Europe, "Mapping study on projects against hate speech online", 15 April 2012. See also interviews: Christine Chen, Senior Manager for Public Policy, Google, 2 March 2015; Monika Bickert, Head of Global Policy Management, Facebook, 14 January 2015

5. See HateBase, Hate speech statistics, http://www.hatebase.org/popular

6. Interview: Andre Oboler, CEO, Online Hate Prevention Institute, 31 October 2014.

7. Mathew, Binny; Dutt, Ritam; Goyal, Pawan; Mukherjee, Animesh. "Spread of hate speech in online social media". ACM WebSci 2019. Boston, MA, USA: ACM.

8. Citron, Danielle Keats; Norton, Helen L. (2011). "Intermediaries and Hate Speech: Fostering Digital Citizenship for Our Information Age". Boston University Law Review (Rochester, NY) 91. https://papers.ssrn.com/abstract=1764004.

9. Interview: Drew Boyd, Director of Operations, The Sentinel Project for Genocide Prevention, 24 October 2014.

10. Phillips, Whitney (2015). This Is Why We Can't Have Nice Things: Mapping the Relationship between Online Trolling and Mainstream Culture. MIT Press.

11. Marwick, Alice; Lewis, Rebecca (2017). Media Manipulation and Disinformation Online.. Data & Society Research Institute.

12. Banks, James (Nov 2010). "Regulating hate speech online". International Review of Law, Computers & Technology 24 (3): 4–5. doi:10.1080/13600869.2010.522323. http://shura.shu.ac.uk/6901/1/Banks_regulating_hate_speech.pdf.

13. "Hateful posts on Facebook and Instagram soar" (in en). https://fortune.com/2020/08/11/facebook-instagram-hate-speech-violations-q2/.

14. Gray, Mary; Suri, Siddharth (2019). GHOST WORK: How to Stop Silicon Valley from Building a New Global Underclass. New York: Houghton Mifflin Harcourt.

15. Meddaugh, Priscilla Marie; Kay, Jack (2009-10-30). "Hate Speech or "Reasonable Racism?" The Other in Stormfront". Journal of Mass Media Ethics 24 (4): 251–268. doi:10.1080/08900520903320936. ISSN 0890-0523. https://doi.org/10.1080/08900520903320936.

16. Bowman-Grieve, Lorraine (2009-10-30). "Exploring "Stormfront": A Virtual Community of the Radical Right". Studies in Conflict & Terrorism 32 (11): 989–1007. doi:10.1080/10576100903259951. ISSN 1057-610X. https://doi.org/10.1080/10576100903259951.

17. Nobata, Chikashi; Tetreault, J.; Thomas, A.; Mehdad, Yashar; Chang, Yi (2016). "Abusive Language Detection in Online User Content". WWW: 145–153. doi:10.1145/2872427.2883062. ISBN 9781450341431. https://www.semanticscholar.org/paper/Abusive-Language-Detection-in-Online-User-Content-Nobata-Tetreault/e39b586e561b36a3b71fa3d9ee7cb15c35d84203.

18. Koster, Willem De; Houtman, Dick (2008-12-01). "Stormfront Is Like a Second Home to Me". Information, Communication & Society 11 (8): 1155–1176. doi:10.1080/13691180802266665. ISSN 1369-118X. https://doi.org/10.1080/13691180802266665.

19. Cohen-Almagor, Raphael (2018). "Taking North American white supremacist groups seriously: The scope and challenge of hate speech on the Internet". The International Journal for Crime, Justice and Social Democracy 7 (2): 38–57. doi:10.5204/ijcjsd.v7i2.517. https://dx.doi.org/10.5204%2Fijcjsd.v7i2.517

20. "The Universal Declaration of Human Rights". 1948. https://www.un.org/en/universal-declaration-human-rights/#:~:text=Article%207.,any%20incitement%20to%20such%20discrimination..

21. Lillich, Richard B. (April 1995). "U.N. Covenant on Civil and Political Rights. CCPR Commentary. By Manfred Nowak. Kehl, Strasbourg, Arlington VA: N. P. Engel, Publisher, 1993. Pp. xxviii, 939. Index, $176; £112;DM/sfr. 262." (in en). American Journal of International Law 89 (2): 460–461. doi:10.2307/2204221. ISSN 0002-9300. https://www.cambridge.org/core/journals/american-journal-of-international-law/article/un-covenant-on-civil-and-political-rights-ccpr-commentary-by-manfred-nowak-kehl-strasbourg-arlington-va-n-p-engel-publisher-1993-pp-xxviii-939-index-176-112dmsfr-262/9599915F22324C9F23B72A988B04ABF7.

22. Leo, Leonard A.; Gaer, Felice D.; Cassidy, Elizabeth K. (2011). "Protecting Religions from Defamation: A Threat to Universal Human Rights Standards". Harvard Journal of Law & Public Policy 34: 769. https://heinonline.org/HOL/Page?handle=hein.journals/hjlpp34&id=779&div=&collection=.

23. Human Rights Committee. General Comment no. 11, Article 20: Prohibition of Propaganda for War and Inciting National, Racial or Religious Hatred, 29 July 1983, para. 2. In 2011, The Committee elucidated its views on the relationship of Article 19 and 20 when it reaffirmed that the provisions complement each other and that Article 20 "may be considered as lex specialis with regard to Article 19". Human Rights Committee. General Comment no. 34, Article 19: Freedoms of opinion and expression, CCPR/C/GC/34, 12 September 2011, paras. 48-52.

24. Article 19 (3) if the ICCPR.

25. Even the Human Rights Committee, which has decided on cases concerning Article 20, has avoided providing a definition of incitement to hatred. Human Rights Council. Incitement to Racial and Religious Hatred and the Promotion of Tolerance: Report of the High Commissioner for Human Rights, A/HRC/2/6, 20 September 2006, para. 36.

26. Faurisson v. France, C. Individual opinion by Elizabeth Evatt and David Kretzmer, co-signed by Eckart Klein (concurring), para. 4.

27. Human Rights Council. Report of the United Nations High Commissioner for Human Rights Addendum, Expert seminar on the links between articles 19 and 20 of the International Covenant on Civil and Political Rights, A/HRC/10/31/Add.3, 16 January 2009, para. 1.

28. Report of the High Commissioner for Human Rights, A/HRC/2/6, para. 39;

29. Committee on the Elimination of Racial Discrimination, General Recommendation 29, Discrimination Based on Descent (Sixty-first session, 2002), U.N. Doc. A/57/18 at 111 (2002), reprinted in Compilation of General Comments and General Recommendations Adopted by Human Rights Treaty Bodies, U.N.Doc. HRI\GEN\1\Rev.6 at 223 (2003), paras. r, s and t

30. Article 2 of the CEDAW.

31. General recommendation No. 28 on the core obligations of States parties under article 2 of the Convention on the Elimination of All Forms of Discrimination against Women Para. 19

32. Inter-American Commission on Human Rights. Inter-American Declaration of Principles on Freedom of Expression, 20 October 2000, para. 7.

33. Inter-American Commission on Human Rights, Advisory Opinion OC-5/85, 13 November 1985, para. 39

34. Inter-American Commission on Human Rights, Advisory Opinion OC-5/85, 13 November 1985, para 39.

35. Viljoen, Frans (2007). International Human Rights Law in Africa. Oxford: Oxford University Press.

36. African Commission on Human and Peoples' Rights. Declaration of Principles on Freedom of Expression in Africa, 32nd Session, Banjul, 17–23 October 2002.

37. Organization of Islamic Cooperation, Sixth OIC Observatory Report on Islamophobia, Presented to the 40th Council of Foreign Ministers, Conakry, Republic of Guinea, December 2013, p 31.

38. League of Arab States, Arab Charter on Human Rights, 22 May 2004, entered into force 15 March 2008, para. 32 (1)

39. League of Arab States, Arab Charter on Human Rights, 22 May 2004, entered into force 15 March 2008, para. 32 (2).

40. Article 7 of the ASEAN Human Rights Declaration.

41. Article 54 of the Charter of Fundamental Rights of the European Union.

42. Article 10 of the European Convention on Human Rights.

43. Handyside v. the United Kingdom, 7 December 1976, para. 49. More cases of hate Speech under the European Court can be found at: http://www.echr.coe.int/Documents/FS_Hate_speech_ENG.pdf 49

44. ECRI General Policy Recommendation No. 6, On Combating the Dissemination of Racist, Xenophobic and Antisemitic Material via the Internet, adopted on 15 December 2000.

45. Council of Europe, Convention on Cybercrime, 23 November 2001, paras 31–34.

46. Council of Europe, Additional Protocol to the Convention on cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems, 28 January 2003, art 5 para 1.

47. Mackinnon, David; Lemieux, Christopher; Beazley, Karen; Woodley, Stephen (November 2015). "Canada and Aichi Biodiversity Target 11: understanding 'other effective area-based conservation measures' in the context of the broader target". Biodiversity and Conservation 24 (14): 3559–3581. doi:10.1007/s10531-015-1018-1. https://www.researchgate.net/publication/284077686.

48. United Nations (2011). Guiding Principles on Business and Human Rights. New York: Office of the New Commissioner.

49. Perrin, Andrew (15 October 2020). "23% of users in U.S. say social media led them to change views on an issue; some cite Black Lives Matter" (in en-US). https://www.pewresearch.org/fact-tank/2020/10/15/23-of-users-in-us-say-social-media-led-them-to-change-views-on-issue-some-cite-black-lives-matter/.

50. Marwick, Alice; Lewis, Rebecca (2017). Media Manipulation and Disinformation Online. Data & Society Research Institute.

51. Gray, Kishonna (2012). "Intersecting Oppressions and Online Communities". Information, Communication & Society 15 (3): 411–428. doi:10.1080/1369118X.2011.642401. https://www.tandfonline.com/doi/abs/10.1080/1369118X.2011.642401.

52. Seay, A. Fleming; Jerome, William J.; Lee, Kevin Sang; Kraut, Robert E. (2004-04-24). "Project massive: a study of online gaming communities". CHI '04 Extended Abstracts on Human Factors in Computing Systems. CHI EA '04 (Vienna, Austria: Association for Computing Machinery): 1421–1424. doi:10.1145/985921.986080. ISBN 978-1-58113-703-3. https://doi.org/10.1145/985921.986080.

53. Nakamura, Lisa (2002). "After Images of Identity: Gender, Technology, and Identity Politics". Reload: Rethinking Women + Cyberculture: 321–331.

54. Nakamura, Lisa (2002). Cybertypes: Race, Ethnicity, and Identity on the Internet. New York: Routledge.

55. Jeung, Russell; Popovic, Tara; Lim, Richard; Lin, Nelson (2020). "ANTI-CHINESE RHETORIC EMPLOYED BY PERPETRATORS OF ANTI-ASIAN HATE". http://www.asianpacificpolicyandplanningcouncil.org/wp-content/uploads/REPORT_ANTI-CHINESE-RHETORIC_EMPLOYED_10_08_20.pdf.

56. Hereward Holland, "Facebook in Myanmar: Amplifying Hate Speech?," Al Jazeera, 14 June 2014, http://www.aljazeera.com/indepth/features/2014/06/facebook-myanmar-rohingya-amplifying-hatespeech-2014612112834290144.html

57. "Why Is There Communal Violence in Myanmar?", BBC, 3 July 2014. https://www.bbc.co.uk/news/worldasia-18395788

58. Andrew Marshall, "Special Report: Plight of Muslim Minority Threatens Myanmar Spring," Reuters, 15 June 2012, https://www.reuters.com/article/us-myanmar-rohingya-idUSBRE85E06A20120615

59. Erika Kinetz, "New Numerology of Hate Grows in Burma", Irrawaddy, 29 April 2013 http://www.irrawaddy.org/religion/new-numerology-of-hate-grows-in-burma.html. Hereward Holland, "Facebook in Myanmar: Amplifying Hate Speech?," Al Jazeera, 14 June 2014; Steven Kiersons, "The Colonial Origins of Hate Speech in Burma", The Sentinel Project, 28 October 2013, https://thesentinelproject.org/2013/10/28/the-colonial-origins-of-hate-speech-in-burma/ ., http://www.aljazeera.com/indepth/features/2014/06/facebook-myanmar-rohingya-amplifying-hate-speech-2014612112834290144.html

60. Tim McLaughlin, "Facebook takes steps to combat hate speech", The Myanmar Times, 25 July 2014. http://www.mmtimes.com/index.php/national-news/11114-facebook-standards-marked-for-translation.html

61. San Yamin Aung, "Burmese Online Activist Discusses Campaign Against Hate Speech", Irrawaddy, http://www.irrawaddy.org/interview/hate-speech-pours-poison-heart.html

62. Georg, Schomerus (13 January 2012). "Evolution of public attitudes about mental illness: a systematic review and meta-analysis". Acta Psychiatrica Scandinavica 125 (6): 423–504. doi:10.1111/j.1600-0447.2012.01826.x. PMID 22242976. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1600-0447.2012.01826.x.

63. Gilbert, David (14 September 2020). "Hate Speech on Facebook Is Pushing Ethiopia Dangerously Close to a Genocide" (in en). https://www.vice.com/en/article/xg897a/hate-speech-on-facebook-is-pushing-ethiopia-dangerously-close-to-a-genocide.

64. Lashitew, Addisu (8 November 2019). "Ethiopia Will Explode if It Doesn't Move Beyond Ethnic-Based Politics" (in en-US). https://foreignpolicy.com/2019/11/08/ethiopia-will-explode-if-abiy-ahmed-doesnt-move-beyond-ethnic-based-politics/.

65. "Help for Yahoo Account" (in en-US). https://help.yahoo.com/kb/account.

66. "Twitter starts enforcing new policies on violence, abuse, and hateful conduct". The Verge. https://www.theverge.com/2017/12/18/16789606/twitter-new-safety-policies-hate-groups.

67. "Hateful conduct policy" (in en). https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy.

68. "Germany to enforce hate speech law" (in en-GB). BBC News. 2018. https://www.bbc.com/news/technology-42510868.

69. Munn, Luke (July 2020). "Angry by design: toxic communication and technical architectures". Humanities and Social Sciences Communications 7: 1–11. doi:10.1057/s41599-020-00550-7. https://www.researchgate.net/publication/343321214.

70. Munger, Kevin; Phillips, Joseph (2019). A Supply and Demand Framework for YouTube Politics. University Park: Penn State Political Science. pp. 1–38. https://osf.io/73jys/.

71. "Our ongoing work to tackle hate" (in en-us). https://blog.youtube/news-and-events/our-ongoing-work-to-tackle-hate/.

72. "Hate speech policy - YouTube Help" (in en). https://support.google.com/youtube/answer/2801939?hl=en&ref_topic=2803176.

73. "Report inappropriate content - Android - YouTube Help" (in en). https://support.google.com/youtube/answer/2802027.

74. "Community Standards | Facebook". https://www.facebook.com/communitystandards.

75. "Community Guidelines | Instagram Help Center". https://www.facebook.com/help/instagram/477434105621119.

76. Tobin, Ariana. "Facebook's Uneven Enforcement of Hate Speech Rules Allows Vile Posts to Stay Up". Propublica. https://www.propublica.org/article/facebook-enforcement-hate-speech-rules-mistakes.

77. Tobin, Ariana. "Facebook's Uneven Enforcement of Hate Speech Rules Allows Vile Posts to Stay Up". https://www.propublica.org/article/facebook-enforcement-hate-speech-rules-mistakes.

78. Carlsen, Audrey. "What Does Facebook Consider Hate Speech? Take Our Quiz". The New York Times. https://www.nytimes.com/interactive/2017/10/13/technology/facebook-hate-speech-quiz.html.

79. "Community Standards: Objectionable Content". https://www.facebook.com/communitystandards/objectionable_content/.

80. Mills, Chris. "This is what Facebook won't let you post". http://bgr.com/2018/04/24/facebook-community-standards-content-policy-nudity-sexual-content-hate-speech/.

81. "Community Guidelines: Objectionable Content". https://www.facebook.com/communitystandards/objectionable_content/hate_speech/.

82. "Facebook Bans White Nationalism And Separatism Content From Its Platforms" (in en). https://www.npr.org/2019/03/27/707258353/facebook-bans-white-nationalism-and-separatism-content-from-its-platforms.

83. "Facebook Bans Alex Jones, Louis Farrakhan And Other 'Dangerous' Individuals" (in en). https://www.npr.org/2019/05/03/719897599/facebook-bans-alex-jones-louis-farrakhan-and-other-dangerous-individuals.

84. "Community Standards Recent Updates | Facebook". https://www.facebook.com/communitystandards/recentupdates/hate_speech/.

85. "Content Policies". https://msdn.microsoft.com/en-us/library/windows/apps/hh184842(v=vs.105).aspx.

86. "Xbox Community Standards | Xbox" (in en). https://www.xbox.com/en-US/legal/community-standards.

87. Weimann, Gabriel; Masri, Natalie (2020-06-19). "Research Note: Spreading Hate on TikTok". Studies in Conflict & Terrorism 0: 1–14. doi:10.1080/1057610X.2020.1780027. ISSN 1057-610X. https://doi.org/10.1080/1057610X.2020.1780027.

88. Simpson, Robert Mark (2019-02-01). "'Won't Somebody Please Think of the Children?' Hate Speech, Harm, and Childhood" (in en). Law and Philosophy 38 (1): 79–108. doi:10.1007/s10982-018-9339-3. ISSN 1573-0522. https://dx.doi.org/10.1007%2Fs10982-018-9339-3

89. "Media and Information Literacy" (in en). 2016-09-01. https://en.unesco.org/themes/media-and-information-literacy.

90. Osler, Audrey; Starkey, Hugh (2006). "Education for Democratic Citizenship: a review of research, policy and practice 1995-2005". Research Papers in Education 24 (4): 433–466. doi:10.1080/02671520600942438. https://www.researchgate.net/publication/228396402.

91. Mathew, Binny; Saha, Punyajoy; Tharad, Hardik; Rajgaria, Shubham; Singhania, Prajwal; Goyal, Pawan; Mukherjee, Animesh. "Thou shalt not hate: Countering Online Hate Speech". ICWSM 2019. Munich, Germany: AAAI.

92. Mossberger, Karen; Tolbert, Caroline; McNeal, Ramona (2007). Digital Citizenship: The Internet, Society, and Participation. MIT Press.

93. Hoechsmann, Michael; Poyntz, Stuart (2012). Media Literacies: A Critical Introduction. West Sussex: Blackwell Publishing.

94. See Paris Declaration: Paris Declaration on MIL in the Digital Era. http://www.unesco.org/new/en/communication-andinformation/resources/news-and-in-focus-articles/in-focus-articles/2014/paris-declaration-on-mediaand-information-literacy-adopted/

95. The 'No Hate Speech Movement' is a regional campaign that encompasses 50 countries far beyond the European Continent. Although the campaign has common goals and develops joint strategies, the particular projects and initiatives rune in every country are the responsibility of the National coordinators and subject to the capacity and resources in each country.

96. Interview: Laura Geraghty, No Hate Speech Movement, 25 November 2014.