

Dear United Nations Office,

Please accept this letter as a recommendation for what we believe should be added to the standards already existing for preventing, eradicating, and prosecuting hate speech, including through digital technologies.

In today’s political climate, with major wars being fought in the Occupied Palestinian territories and the Ukraine, the discrimination which is the seedling of hate speech has evolved into racial discrimination. This is not a new phenomenon and we have seen how it also unfortunately prevails in modern countries such as the United States. The major instruments available to guide the tackling of such discrimination are the UN Convention on the Elimination of All Forms of Racial Discrimination (CERD)[[1]](#footnote-1) which was adopted in 1965 and entered into force in 1969, and Universal Declaration of Human Rights (UDHR)[[2]](#footnote-2).

Hate speech indicates that there is a characteristic in a person or group of people, be it inherent or acquired, which is somehow so divergent from the “norm” of somebody else that it can neither be accepted nor tolerated. Consequently, in an effort to show how unacceptable that trait is, negativity in the form of hate speech becomes directed towards the person or group carrying “the difference”. With time, and if not counteracted, hate speech turns into discrimination which is somehow overlooked by the other members of society and inevitably undermines the person’s or the group’s right to equal treatment in it.

As is pointed out in General Recommendation No. 35 of the Committee on the Elimination of Racial Discrimination Article II. 7, “Racist hate speech can take many forms and is not confined to explicitly racial remarks”[[3]](#footnote-3). It can very much be intersectional because no human or group is one-dimensional. As defined by Delgado, intersectionality means “the predicament of women of color and others who sit at the intersection of two or more categories”[[4]](#footnote-4). For example, a person could be a receiver of hate speech for being both a woman and a member of a minority group. The document also promotes “teaching, education, culture and information” in Article 7 as some of the measures which could be practically taken to combat hate speech.

Notwithstanding the above main documents, there are additional provisions put in place for safeguarding against hate speech. For example, in order to broaden the scope of what hate speech might mean and constitute, recommendation no. 97(20) of the Council of Europe, Committee of Ministers[[5]](#footnote-5) has proposed a definition of “hate speech” which covers all forms of expression of hatred on the basis of race, religion, xenophobia and intolerance. This adds onto article 7 of the Universal Declaration of Human Rights[[6]](#footnote-6) which deals with the discrimination as an infringement of equality and protection.

Of course, perhaps CERD provides the most comprehensive definition of hate speech in article 4[[7]](#footnote-7) and obliges states to render any conduct contrary to its provisions punishable. Recommendation No.35[[8]](#footnote-8) on combating racist hate speech specifically also exists to safeguard individuals and groups from becoming victims and gives effective remedies in case such form of discrimination does happen. The Intergovernmental Working Group on the Effective Implementation of the Durban Declaration and Programme of Action (DDPA)[[9]](#footnote-9) also provides comprehensive and direct practical proposals as to how any type of discrimination can be tackled.

Yet, despite existing morally strong anti-discrimination and hate speech documents which despite not all being legally binding have still helped shape national laws against hate speech, one must recognise that occurrences of discrimination, especially in today’s climate, still exist. Perhaps this is most evident from the mere fact that the above mentioned documents exist because the need to have them and so many speaks to the difficulty in trying to eradicate discrimination.

When it comes to the role of the media and the digital space in general where arguably most instances of discrimination occur through slurs, hateful comments and debates directed against certain persons or groups of people, Recommendation No. R(97) 21 adopted by the Committee of Ministers[[10]](#footnote-10), outlines the positive impact the media can have on preventing hate speech even in political debates. This has had a positive impact on further movements which aim to spread the no hate campaign online. For example, the No Hate Speech Movement[[11]](#footnote-11) is a series of online campaigns which rests on the documents and recommendations so far discussed and seeks to implement the objectives laid down in their provisions.

Additionally, Article 6 of Directive 2010/13/EU also directs the general audiovisual field of the media of Member States to: “ensure by appropriate means that audiovisual media services provided by media service providers under their jurisdiction do not contain any incitement to hatred based on race, sex, religion or nationality”[[12]](#footnote-12). Logically, that would also include the digital providers of this audiovisual media which is streamed online. In a report by the Law Reform Commission from 2016 called “Harmful Communications and Digital Safety”[[13]](#footnote-13), the type of dangers which social media and online communications pose if not adequately and timely addressed are outlined in detail.

It is important to note that hate speech and discriminatory language could arguably be seen as more harmful when used on digital platforms than in any other mode of delivery, including in person. This is based on 2 things specific to the online space, namely: the speed with which messages get delivered in real time from one user to another, and the ability of social media to spread information in a matter of seconds across the vast online space. Therefore, reaching a large group of people in no time while using hate speech could have more detrimental impacts on the target of that discriminatory attack and because of its speed and reach, it could be harder to vet out and be stopped.

Recently, digital platforms have become places of hatred, places of persecution. For both the United Nations and the Member States, the objectives are therefore to prevent xenophobia and racial discrimination in digital spaces, but also to use these platforms in another way: to raise awareness and promote integration.

We have tried to apply the principles of Critical Race Theory to find solutions to these problems.

Firstly, one could advocate for the drafting of a general definition of hate speech that would be the same for all Member States in order to harmonize cooperation between them. Having such a common definition would enable states to react more systematically to punish this type of behaviour. Such a common definition would also make it possible to act on the same basis as regards to criminal sanctions. Certain types of behaviour could therefore be 'de-banalised' and criminalized by all Member States. Penalizing discriminatory comments on social networks also implies the need for cooperation between states, as it is sometimes difficult to locate the authors of such comments behind their screens. It may even be necessary to develop a definition of "online hate speech".

It is therefore necessary to force digitals platforms to provide protection equivalent to that demanded by international human rights law against discrimination based on race, gender, sex or religion, even if it is not required in national laws. Social media platforms will have to respect international law protecting discrimination, even if this means hindering the right to freedom of expression. If a balance is to be struck, the criteria of necessity and proportionality must be defined by the United Nations and respected uniformly by all Member states. Strict criteria need to be set so that social media companies can't justify their lack of action on the grounds of freedom of expression.

A definition must be chosen and validated by the member states and the United Nations. This definition must be broad enough to include all forms of discrimination across cultures. Social networks will then have to incorporate this definition into their user policies made available in every language and be held responsible in the event of any excesses.

A first solution could be to oblige the social and digital platforms used in the various Member States to set up automatic deletions of discriminatory comments. This could be set up by an AI system, detracting in several languages and widely deleted such comments before they are even seen. As said by many experts, now “it is feasible to generate automatic detectors of online racist/xenophobic hate speech using machine learning with solid performances”[[14]](#footnote-14). But this automated moderation must be accompanied by a human element, which must be carried out in conjunction with the populations concerned.

This idea is linked with the critical race theory concept of  ordinariness, which means for Delgado “that racism is difficult to address or cure because it is not acknowledged".[[15]](#footnote-15) Acting as if racism didn't exist is counterproductive. Punishing discriminatory behaviour alone, while necessary, is not enough. Today, it is necessary to highlight different racial conceptions and different histories in order to enable greater integration and greater inter-racial understanding. Indeed, as Nellie Wong’s poem “When I was growing up”[[16]](#footnote-16) points out, mainstream representation can lead to hatred of others who don't look like that “mainstream person”, there is a need to use social networks to adequately represent human diversity. It is a question of educating and acting against stereotypes. To achieve such a thing, it is necessary to shake up the representations and stories known as "mainstream" in order to leave room and give their fair value to the "counter-narrative". As noted by the “Noise Project”[[17]](#footnote-17) It's important to see all the sides behind a story. It's essential to challenge the main stories, in order to offer a more open-minded approach. Setting up awareness campaigns in the various Member States and on the different digital platforms is essential in this fight.

Social media companies will have to submit reports to the United Nations on discriminatory comments, so that awareness-raising campaigns can be adapted accordingly. In the same spirit, these reports should be accessible to the public, so that they can be used for educational purposes. Also, Member states need to set up support systems for victims of hate speech.

Also as pointed out in Kimberlé Crenshaw's article[[18]](#footnote-18), it is of course essential to vary education and teaching methods. “There is a clear mandate for State education to address the issue of hate speech in order to prepare students to understand, analyse and resist the call of hate groups but also to equip students to exercise informed judgment about the nature of conspiracy theories circulated by social media platforms like Facebook”[[19]](#footnote-19). Then, from a very early age, children need to be exposed to different narratives and learn about the impact that hatred can have. Each Member State should ensure that its curriculum contains provisions against hate and the misuse of social networks. Today's social networks make it possible to act with little consequence and give a kind of feeling of impunity. One can therefore believe that raising young people's awareness of this type of behaviour is essential in the fight against xenophobia and discriminatory comments.

Earlier education on the different realities will help to understand the concept of “different racialization”. As defined by Delgado as “the idea that each race has its own origins and ever-evolving history" this concept is linked with the idea of “intersectionality and antiessentialism”[[20]](#footnote-20). No one has only one identity, everyone has different stories, beliefs, and ideas. Educating people about diversity could help overcome stereotypes and racial hatred.

The produced and attached is part of our assignment for the module Critical Race Theory from the University of Galway.

As such we hope that our voice will be heard and taken into consideration.

Kind Regards,

Alexandra COOPER and Camille RAVARY
LLM Students in the Irish Center for Human Rights, National University of Galway.

1. United Nations, Convention on the Elimination of All forms of Racial discrimination (CERD), adopted in 1965, entered into force 1969 (CERD) [↑](#footnote-ref-1)
2. UN General Assembly, Universal Declaration of Human Rights (UDHR), 10 December 1948, 217 A (III) [↑](#footnote-ref-2)
3. CERD General recommendation No. 35 on Combating racist hate speech (2013), article II.7 [↑](#footnote-ref-3)
4. Critical Race Theory, Fourth Edition : An Introduction, Richard Delgado, Jean Stefancic, and Angela Harris, New York University Press 2023, p7 [↑](#footnote-ref-4)
5. Council of Europe, Committee of Ministers, [Recommendation No. R (97) 20 to Member States on “hate speech” adopted on 30 October 1997](https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=0900001680505d5b) [↑](#footnote-ref-5)
6. UDHR, article 7 [↑](#footnote-ref-6)
7. CERD, article 4 [↑](#footnote-ref-7)
8. Recommendation n°35 [↑](#footnote-ref-8)
9. The Intergovernmental Working Group on the Effective Implementation of the Durban Declaration and Programme of Action (DDPA) [↑](#footnote-ref-9)
10. Council of Europe, Committee of Ministers, Recommendation No. R (97) 21 to member states on the media and the promotion of a culture of tolerance adopted on 30 October 1997 [↑](#footnote-ref-10)
11. The Council of Europe, No Hate Speech Youth Campaign, available at No Hate Speech Youth Campaign Website - No Hate Speech Youth Campaign (coe.int) [↑](#footnote-ref-11)
12. Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services, article 6 [↑](#footnote-ref-12)
13. Law Reform Commission, 27th September 2016, Report “Harmful Communications And Digital Safety” [↑](#footnote-ref-13)
14. Arcila-Calderón, C.; Amores, J.J.; Sánchez-Holgado, P.; Vrysis, L.; Vryzas, N.; Oller Alonso, M. How to Detect Online Hate towards Migrants and Refugees? Developing and Evaluating a Classifier of Racist and Xenophobic Hate Speech Using Shallow and Deep Learning. Sustainability 2022, p12 [↑](#footnote-ref-14)
15. Delgado, p8 [↑](#footnote-ref-15)
16. When I was growing up, Nellie Wong, from “This Bridge Called My Back", Kitchen Table: Women of Color Press, 1981, accessed 27 February 2013. [↑](#footnote-ref-16)
17. The Noise Project, available at [The Noise Project (thenoiseprojectlv.com)](https://www.thenoiseprojectlv.com/) [↑](#footnote-ref-17)
18. Kimberlé Crenshaw, Twenty Years of Critical Race Theories of Critical Race Theory: Looking Back to Move Forward, Conn.L.Rev. 43 (2010) 1253-1352 [↑](#footnote-ref-18)
19. Michael A. Peters, Educational Philosophy And Theory 2022, VOL. 54, NO. 14, Limiting the capacity for hate: Hate speech, hate groups and the philosophy of hate, Educational Philosophy and Theory, 2325–2330, p2326 [↑](#footnote-ref-19)
20. Delgado, p14 [↑](#footnote-ref-20)