**European Regional Forum on Hate Speech, Social Media and Minorities**

Keynote speech by

Mr Guilherme Canela

Chief of Freedom of Expression and Safety of Journalists,

UNESCO

21 September 2020

**[Introduction]**

**[**To read depending on protocol of other speakers for the webinar]

Ms. Anna-Mária Bíró*,* Director of the Tom Lantos Institute

H.E. Mr. Alexander Schallenberg, Minister of Foreign Affairs Austria

Ms. Snežana Samardžić-Marković,Director General of Democracy, Council of Europe

Ms. Birgit Van Hout, OHCHR Regional Representative for Europe

Mr. Michael O’Flaherty, Director of the EU Agency for Fundamental Rights

Ms. Maria Daniella Marouda, Chair of European Commission against Racism and Intolerance

Mr. Christophe Kamp, Director of the Office of the High Commissioner on National Minorities

Ms. Katarzyna Gardapkhadze, First Deputy Director, OSCE Office for Democratic Institutions and Human Rights

Mr. Fernand de Varennes, UN Special Rapporteur on minority issues

Experts, participants,

Good morning!

Although I’m very pleased to join you on behalf of UNESCO, I must say that sad are the times when we still need to talk about the issue of hate. I would prefer that, just as in that Emily Dickson poem, we all could say: ***I had no time to hate***.

Yet, here we are. These are complex and serious issues, and simple solutions to address them tend to be too good to be true.

I’m therefore very thankful to the organizers for the opportunity to discuss how we can tackle hate speech without limiting the human rights that also hold the key to fighting it.

**[Historical perspective]**

Throughout the history of human rights violations and atrocities, we see a common pattern: perpetrators first dehumanize those groups whose rights they wish to violate. They didn’t torture *women* in medieval ages, they tortured *witches*.

They didn’t commit atrocities against *men*, *women* and *children* in the plantations, they did that against *slaves*.

To dehumanize an individual or a group, it is necessary, first, to build a narrative of hate against them. Hate speech is the first step to further human rights violations. There is hate against the unknown, there is hate against the different, there is hate against the ‘other’, there is hate to amplify or keep power.

That’s why we see hateful discourses peak around times of crisis, change and hardship - at times when accurate information is scarce, and people are searching for answers on how to solve problems. We see this happening over and over again in times of elections and now, on an unprecedented scale, as we’re tackling a pandemic.

With this in mind, allow me to share a couple of key elements trying to offer some inputs for the working groups, which will discuss these issues in detail today and tomorrow.

**[Key stakeholders]**

First, we need to acknowledge that there are key stakeholders with an important role to play to counter hate speech. These include:

* Journalists, who play a crucial role in conveying reliable information;
* Judicial operators, whose work in apply international human rights standards when dealing with hate speech cases is paramount;
* Internet platforms, since no scalable solution is possible without them; and
* Regular citizens as we go about our every-day life, since no long term solution is possible without empowering them

With education, information, transparency, accountability and the tools both to detect hateful narratives and content, as well as to understand how to navigate the various pitfalls of unnecessary restrictions, these stakeholders can together turn the tide of our public discourse. As international institutions, policy makers, civil society and research institutions, our role is to figure out how to enable this collective effort.

**[hate speech targeting minorities in social media and the need for greater transparency]**

As we know, it is difficult to gauge the full extent of hate speech online or the effectiveness of responses to hate speech due to numerous factors. Improving our understanding is nevertheless indispensable and will increase in urgency as the advancement of Artificial Intelligence brings additional challenges and opportunities.

Therefore, for evidence-based policy making, transparency is essential. Particularly, transparency regarding how hate speech spreads within social media platforms and how they are dealing with this phenomenon. Unfortunately, we are not there yet and much greater progress is needed.

**[International legal and institutional framework]**

Another challenge in tackling hate speech is the lack of a universally accepted definition. For guidance here, we can turn to international human rights law and standards, particularly in regard to illegal hate speech.

Articles 19 and 20 of the International Covenant on Civil and Political Rights lays out the conditions for which speech is to be legitimately restricted. - Hate speech is considered illegal if it constitutes incitement to discrimination, hostility or violence. Understanding how to apply these principles in specific cases can draw on the six-part threshold of the “The Rabat Plan of Action”.

Regulatory, self-regulatory and co-regulatory laws, policies and codes of practice to counter hate speech have been on the rise, and it is fundamental that the adoption and implementation of these laws align with international standards.

I would like to emphasize a point from the UN Strategy and Plan of Action: **the most effective method for countering hate speech is through *more speech*, not less.**

**[Regulation of online hate speech: The role and responsibility of intergovernmental organizations, States, internet companies and social media platforms]**

In light of the massive scale of the problem and recent cases of hate speech that incite violence on social media, there is a growing consensus of the need for large internet companies to do more to combat hate speech.

As we are seeing during the pandemic, the speed and scale of content upload increasingly makes companies turn to automated processes and artificial intelligence for content moderation. Such technologies allow for detecting and responding to content that would otherwise be impossible given the hundreds of millions of posts per day.

Yet greater transparency into the use of such automated processes is needed. This is particularly important as distinguishing between content that should be legitimately removed requires greater nuance and context than what algorithms can guarantee.

A commitment is necessary by these companies and all other actors to a human rights-based, open, accessible and multi-stakeholder governed internet. This is in line with UNESCO’s framework of internet universality.

**[Towards a safer space for minorities: positive initiatives to address online hate speech: The role of NHRIs, human rights organizations, civil society and other stakeholders**

So, what can we do? Curtailing harmful content through blanket regulations or automated processes is tempting, but rarely the best approach. Such general regulations are blunt tools that often directly backfire against the very groups that are targets of hate speech -and that we are trying to protect. Infringing on the right to freedom of expression is a very dangerous path to enter.

The UNESCO report *Countering Online Hate Speech* finds that laws and regulations on this topic risk being overly broad, and that social and non-regulatory responses may often be preferable to legislation.

Recommendations include: developing early warning systems and methods to distinguish among different typologies of speech; coordination by members of civil society to address emergent threats that connect online hatred and violence offline; and media literacy campaigns and initiatives aimed at preparing users to interpret and react to hateful messages.

These approaches inform the work of UNESCO, which is the UN agency with a mandate to promote both freedom of expression and quality education for all. Our work in countering hate speech also falls within the framework of the UN Strategy and Plan of Action on Hate Speech.

Our project on building trust in media in South East Europe, with the support of the European Commission, has a two-fold strategy to address ‘potential harmful content’ online, for example.

The first is to support media and their commitment to professional standards and media self-regulation. In that work, UNESCO supports press and media councils of the region, most of which are very active in combating online hate speech.

Secondly, the project reinforces the critical thinking of citizens, through piloting media and information literacy in schools and youth organizations.

Media and information literacy is also a key part of another UNESCO project, in cooperation with the European Commission, Twitter, and the World Jewish Congress, on the related problem of conspiracy theories. The project just launched an awareness-raising campaign of the existence and consequences of conspiracy theories linked to the COVID-19 crisis.

An important part of building young people’s resilience to discourses that seek to stigmatize, divide and threaten the stability of our societies lies in long term education policies. In this respect, UNESCO promotes global citizenship education to reinforce young people’s commitment to peace, human rights and the values of democracy. This provides a framework to address hate speech towards minorities, such as through a partnership with OSCE to tackle anti-semitism.

A further area of combatting hate speech while respecting human rights is through UNESCO’s judges’ initiative, which has empowered more than 17,000 judicial operators across Latin America and Africa. This has given them the knowledge to make rulings in line with international standards, for example in applying the six-part threshold test of the Rabat Plan of Action.

**[Conclusion]**

Ladies and gentlemen,

I hope that these remarks will add some reflections for the thematic working groups and the recommendations that they will produce over the coming two days.

In your discussions on this topic, I invite you to keep in mind the words of Martin Luther King Jr: “Darkness cannot drive out darkness: only light can do that. Hate cannot drive out hate: only love can do that.”

Thank you.