

European Regional Forum on Hate Speech, Social Media and Minorities

ONLINE EVENT:
SEPTEMBER 21-22, 2020



Preamble

Across the world today, discriminatory practices and hateful messages serve to stigmatise and vilify those perceived to be the “other”, the “foreigner” or the “one who does not belong”. Most of those targeted by these messages are minorities – overwhelmingly so. Among the key findings of the most recent 2018 Hate Crime Data covering European and other countries of the Organisation for Security and Cooperation in Europe indicate that more than 76% of hate crimes involve Jewish, Muslim, and other ethnic and religious minorities, with 4405 out of 5735 reported incidents.

Such expressions of hate and discrimination increasingly dominate political agendas and discourses, and are mainstreamed through public life, creating a climate of fear amongst individuals and communities. They can at times also create a climate of rejection, exclusion and even intolerance, threatening societal values and undermining the respect of human dignity and the protection of human rights. This type of speech, often called ‘hate speech’, in most cases target persons belonging to minorities, who are portrayed as a threat to national unity, societal harmony, national security and public order, or who are subjected to discrimination because of their distinct ethno-cultural identities, religions or languages.

Digital technologies and social media platform owners may play a role in contributing to hate speech and undermining human rights. Indeed, in recent times, there have been numerous and flagrant examples of the “rallying power” of social media platforms being abused to spread hatred, unfounded and stigmatising rumours, fostering a climate of insecurity and animosity, and in the most extreme cases, leading to violent campaigns against members of minorities. Such unregulated online expressions of hate can result in or increase the chances of human rights violations taking place offline against some of the most vulnerable segments of society.

In response to this issue and as part of a series of regional forums on hate speech, social media and minorities, convened by the Special Rapporteur on minority issues, Dr Fernand de Varennnes, the European Regional Forum on Hate Speech, Social Media and Minorities was held virtually on 21 and 22 September 2020. The following set of recommendations were developed as part of the European Regional Forum. The recommendations were compiled through the submissions of experts and participants of the Regional Forum and were collated and edited by the Special Rapporteur on minority issues and the Regional Forum organising team. Within the recommendations, there are several key terms that merit explanation.

First, the term “minorities” refers to national, or ethnic, religious and linguistic minorities as specified in the United Nations Declaration on the Rights of Persons Belonging to National or Ethnic, Religious and Linguistic Minorities [1992] and further defined in the Special Rapporteur’s last two reports to the UN General Assembly in connection of the meaning of the term ‘minority’

[2019] and on the scope and significance of the four categories of minorities recognised at the United Nations (national or ethnic, religious and linguistic) [2020]. In addition, within the text, certain specific forms of discrimination or hatred are enumerated, such as antisemitism, Islamophobia, anti-Gypsyism, caste, as well as other distinctive forms of hatred. This list is intended to be non-exhaustive and open-ended to include all protected characteristics in accordance with international human rights law.

It is also important to emphasise that there is no clear and internationally accepted definition of the term “hate speech”, as states, social media platforms, or other stakeholders have not come to a consensus as to what “hate speech” means. However, under international law, various expressions may fall under three different levels of hate speech, which require different responses by states.

The most **severe forms of “hate speech”** should be prohibited by states and include the following: (1) direct and public incitement to genocide as defined by international criminal law under the Genocide Convention and the Rome Statute; (2) any advocacy of racial, national or religious hatred, which constitutes incitement to discrimination, hostility or violence under Art. 20(2) of the ICCPR; or (3) the dissemination of ideas based on racial superiority or hatred and incitement to discrimination under ICERD. The [Rabat Plan of Action](#) establishes the 6-prong threshold test for incitement (context; speaker; intent; content and form; extent of the speech act; likelihood, including imminence).

There are certain types of **intermediate “hate speech”** that may be prohibited by states under international law provisions such as Article 18 of the International Covenant on Civil and Political Rights even if they do not meet the Rabat Plan of Action threshold test for incitement. Certain types of biased expressions may be restricted if such restrictions: (1) are provided by law, (2) pursue a legitimate aim, (3) are necessary, and (4) are proportionate (for example, threats of violence or identity-based harassment especially prior to elections may be restricted if they meet these requirements).

Finally, there are expressions that may be considered as the **least severe forms of “hate speech”** that should not be subjected to legal restrictions under international law, but may require non-legal responses from various stakeholders. This may include expressions that are shocking or disturbing, blasphemy, condoning or denying historical events, and disinformation.

Recommendations from the European Regional Forum on “Hate Speech”, Social Media and Minorities¹

Session 1: Causes, scale and impact of hate speech targeting minorities in social media

1. States should effectively implement their obligations and responsibilities to protect the human rights of minorities. States should address concerns about the situations of minorities in other states on the basis of international human rights law, international cooperation and the conduct of friendly relations; such concerns should not be regarded as a threat to state sovereignty.
2. Bilateral and multilateral agreements should include the protection of minority rights as a means to enhance the full and effective realisation of human rights.
3. Intergovernmental organisations and states should provide full recognition of minorities and distinct grounds of discrimination, such as antisemitism, Islamophobia, anti-Gypsyism, caste and other grounds² necessary for the full and effective realisation of minority rights.
4. Intergovernmental organisations, such as the UN and the EU, should directly and specifically recognise that hate speech includes group-specific incitement to discrimination, hostility or violence, such as antisemitism, Islamophobia, anti-Gypsyism, caste, and other distinctive forms of hatred that mainly target minorities.
5. States should counter hate speech and other harmful manifestations of hatred on social media by drawing on comprehensive educational policies, including education on critical thinking, democratic citizenship, human rights, and digital literacy.
6. Social and economic inequality, particularly in relation to minorities, should be addressed as a root cause of hate speech, including scapegoating and incitement to discrimination, hostility, or violence.
7. States should develop long-term, comprehensive policies for combating negative stereotypes of, and discrimination against, minority individuals and groups, and promote intercultural understanding by, among other things, the teaching of the culture and history of minorities in the national curriculum. Majority students should be explicitly encouraged to learn the culture and language of minority communities to develop better understanding of, and empathy for, minority situations. The content of existing curricula, including textbooks, should be evaluated on the basis of clear standards on intercultural education, including standardized content analysis.
8. States, together with all relevant stakeholders, should ensure the right of minorities to mother-tongue education, and promote the benefits of multilingualism as a means of reducing tensions and preventing hate speech, conflicts and discrimination based on language.
9. States should provide relevant education to ensure that headteachers, teachers and students have the necessary competencies and skills to recognise and respond to bullying, intimidation, and discrimination against adolescents and

¹ The term ‘minorities’ is used to refer to national or ethnic, religious and linguistic minorities as specified in the UN Declaration on the Rights of Persons Belonging to National or Ethnic, Religious and Linguistic Minorities [1992], and clarified by the UN Special Rapporteur in his annual report to the United Nations General Assembly in terms of the meaning of the word ‘minority’ [2019], and the scope and significance of the four categories of minorities recognised at the UN [2020 – to be published].

² This list is meant to be non-exhaustive, and open-ended to include all protected characteristics and minority categories in accordance with international human rights law.

- children belonging to minorities in secondary educational institutions (including such actions on the Internet, e.g. cyber bullying).
10. Media pluralism should be encouraged, including through the facilitation of access to and ownership of media by minority, indigenous and other groups, including media in their own languages. Local empowerment through media pluralism facilitates the emergence of speech capable of countering hate speech.
 11. States should provide education to establish at an early age the digital and technological literacy and critical thinking of students, in order to counter online hate speech.

Session 2: International legal and institutional framework (norms, institutions, mechanisms)

12. International organisations should allocate more resources to and use existing norms, institutions and mechanisms more effectively for the protection of minority rights.
13. A multi-stakeholder approach – including States, regulatory bodies, private Internet and tech companies, and the representatives of minorities - should be established for the effective implementation of a comprehensive regulatory and policy framework that provides for a range of differentiated and complementary strategies needed to effectively combat all types of hate speech in all its manifestations; it should comprise civil and criminal law measures, as well as information, educational and cultural measures; it should also include suitable preventive, remedial, and, where necessary, punitive measures to counter and respond to hate speech, in accordance with international human rights law.
14. State authorities and the private sector should protect and respect the human rights of persons belonging to minorities in the face of incitement and ‘hate speech’ in accordance with international human rights law, including on the right to freedom of opinion and freedom of expression.
15. States should report on measures taken to combat hate speech, including hate speech targeting minorities, as part of their commitments to achieving the 2030 agenda and Sustainable Development Goals.
16. In accordance with Article 9 of the UN Declaration on the Rights of Persons Belonging to National or Ethnic, Religious and Linguistic Minorities (UNDM), UN agencies and bodies should explicitly and emphatically urge states to implement fully and effectively the human rights of minorities.
17. Governments and tech companies should implement and expressly reference the [Rabat Plan of Action](#) in addressing incitement to hostility, discrimination or violence as part of their strategies to deal with hate speech, including hate speech targeting minorities, at the global level. The [Rabat six-part threshold test](#) provides the criteria and conditions under which content that constitutes incitement should be criminalised in national law and the standards for when content ought to be removed from social media platforms. These criteria are: context, the speaker, intent, content and form, extent of the speech, and likelihood of the [incitement](#).
18. States should implement hate speech-related recommendations of the OSCE, including those contained in its recent “[OSCE Human Dimension Commitments and State Responses to the Covid-19 Pandemic](#)”, and its “[Tallinn Guidelines on National Minorities and the Media in the Digital Age](#)”
19. The European Union should create a systematic legal and policy framework for automated decision-making systems, online platforms, and data-based business

models that ensures the protection and promotion of freedom of expression and other fundamental rights. This should address algorithms that drive the content distribution across platforms, and which amplify hate speech and other forms of potentially harmful legal content.

20. [Open Internet Regulation EU 2015/2120](#) Article 3.2 should be amended to remove protections for content which would be considered as incitement to hatred under international human rights law (particularly Article 20(2) of the ICCPR and the Rabat Plan of Action) and to allow for the establishment of an institutional framework by which national regulatory authorities and the Body of European Regulators for Electronic Communications would create single points of contact for NGOs and human rights organisations to bring to their attention websites that would undermine this body of international human rights law.
21. The European Commission Digital Services Act should ensure that 'hate speech' as defined by the existing EU Code of Conduct on Countering Illegal Hate Speech is not permitted on any online platform, including smaller, alternative ones.
22. The European Commission Digital Services Act should provide guidance and encouragement for platforms to adopt and use the International Holocaust Remembrance Alliance (IHRA) definition of antisemitism to identify, address and counter antisemitic content.
23. The European Commission Digital Services Act should ensure the availability of data to better understand the mechanisms by which hate speech spreads online, while respecting privacy and the protection of personal data. Catalogues of symbols and tropes, and mapping of the flow of conspiracies online can be useful starting points.
24. States should support the International Holocaust Remembrance Alliance (IHRA) work on the creation of materials that may better inform national and international bodies on the development, enactment, and interpretation of Holocaust-denial laws.

Session 3: Regulation of online hate speech: The role and responsibility of intergovernmental organisations, states, internet companies and social media platforms (regulate what and how?)

General

25. All stakeholders should ensure the 'governance alignment' between laws and regulations of governmental authorities, on the one hand, and the content moderation policies and oversight mechanisms of internet platforms, on the other.
26. Initiatives to address hate speech should use a holistic approach that minimizes the distinction between online and offline forms of hate speech, while keeping in mind that such differences do exist. Such initiatives should be constructive, positive, and mindful of the local context.
27. Hate speech regulations should distinguish between spontaneous, individual manifestations of hate speech and purposeful hate speech campaigns.
28. All stakeholders should contribute to the improved coverage, comparability and quality of government statistics and encourage standardized reporting from large tech companies, including search platforms.
29. All stakeholders should develop methods for identifying expressions that amount to advocacy of hatred that constitutes incitement to discrimination, hostility or violence.

30. All stakeholders should identify emergent minority categories, including protected characteristics, that are targeted by hateful content.
31. In cases where authors and disseminators of hate speech are anonymous, states and internet service providers should ensure the availability of avenues for the disclosure of their identity only in cases where they constitute prohibited forms of hate speech under international law, i.e. (a) direct and public incitement to genocide as defined by international criminal law under the Genocide Convention and the Rome Statute; (b) any advocacy of racial, national or religious hatred, which constitutes incitement to discrimination, hostility or violence under Art. 20(2) of the ICCPR; (c) the dissemination of ideas based on racial superiority or hatred and incitement to discrimination under ICERD, or (d) permissible restrictions to freedom of expression which are provided by law, pursue a legitimate aim, are necessary and proportionate and only to ensure respect of the rights or reputations of others or for the protection of national security or of public order (ordre public), or of public health or morals.

International Organisations

32. International organisations should use existing soft law principles and standards, such as the [Guiding Principles on Business and Human Rights \(2011\)](#) and the [Report of the Special Rapporteur on the promotion and protection of freedom of opinion and expression on the regulation of online 'hate speech', A/HRC/38/35, 6 April 2018 \(see, from para. 41\)](#), that imposes responsibility and due diligence requirements on Internet companies, social media platforms owners, media organisations, as well as on educational institutions and civil society in order to prevent the spread of hatred, the stigmatisation of minority communities and in the most extreme cases, incitement to violence against members of minorities.
33. International organisations should involve, empower, and support the work of minority youth and NGOs working on the promotion and protection of minority rights, and on addressing and countering hate speech.
34. International organisations should provide relevant technical assistance to states to counter hate speech.
35. The Body of European Regulators for Electronic Communications (BEREC) or other relevant bodies, such as National Regulatory Authorities, should continually work on lists of internet addresses that should be blocked by search engines in Europe in accordance with online hate speech laws, and share such lists with search engine companies.
36. Intergovernmental organisations, notably the UN, EU, and the Council of Europe, should engage in dialogue with the International Holocaust Remembrance Alliance on issues related to denial and distortion of Holocaust-era crimes.

States

37. States should adopt a broader set of policy measures to combat hate speech in a manner that is consistent with international human rights law and standards, including by policies aimed at dismantling prejudices, promoting diversity and protecting, and advancing minority rights.
38. States should develop national legislation and institutional mechanisms, including properly resourced NHRIs or non-discrimination bodies, to counter hate speech

and address incitement to hatred with the meaningful participation of minority representatives.

39. States should launch formal and regular training programs related to Holocaust denial laws for the law enforcement, the judiciary, and civil society stakeholders, such as journalists and educators. Protections should be provided for scholars, educators, journalists, writers, and general researchers who engage in, present, and publish bona fide, data-driven research on Holocaust-era crimes.
40. National, regional and local government representatives should refrain from making statements that may promote discrimination or undermine equality, including rhetoric that normalises and encourages racist, xenophobic, or other hateful reactions. [They should take advantage of their positions to promote intercultural understanding, including by contesting, where appropriate, discriminatory statements or behavior.](#)
41. States should determine the responsibilities of internet service providers and social media platforms in relation to hate speech in accordance with international human rights law, particularly on the right to freedom of expression.
42. States should ensure that the authors of content that amounts to incitement to hatred face appropriate criminal and/or civil sanctions in line with international human rights law (particularly Article 20(2) ICCPR and the Rabat Plan of Action).
43. States should provide in their legal systems standing and support for those targeted by hate speech to bring proceedings that seek to remove, within a reasonable timeframe, content that would be considered as incitement to discrimination, hostility or violence under international human rights law.
44. States should ensure that individuals who have suffered cognizable harm as a result of hate speech that is prohibited under international human law (particularly Article 20(2) ICCPR and the Rabat Plan of Action) have access to justice and appropriate remedies.
45. States should ensure special attention is paid to addressing and countering hate speech targeting minorities during election cycles guided by existing recommendations such as those of [Equinet](#).

Internet Companies and Social Media Platforms accordance with international human rights law, particularly on the right to freedom of expression.

General Policies

46. Social media companies should put human rights at the centre of their content moderation policies and practices, as well as oversight mechanisms. Freedom of expression should have a central role, alongside the principles of equality and non-discrimination, with a specific focus on protected characteristics, such as antisemitism, Islamophobia, anti-Gypsyism, caste and other grounds.
47. As laid out in the UN Guiding Principles on Business and Human Rights, social media companies should evaluate how their products and services affect human rights, including the rights of persons belonging to minorities, and make that information available through public and transparent periodic human rights impact assessments (HRIAs). Such HRIAs must address hate speech and its impact on minority communities and companies must show how they are addressing recommendations made by the HRIAs.
48. Social media companies should ensure that their hate speech policies include an evaluation of context, including the harm to persons belonging to minorities, by

- ensuring that human beings are involved in the application of any use of automation or artificial intelligence tools;
49. Social media companies should establish Councils/Committees of minority representatives from those groups most at risk of hate speech and violence to advise on their content policies and moderation practices.
 50. Fact-checkers and content moderators should be educated about antisemitism, Islamophobia, anti-Gypsyism, caste-based hate and other forms of hate.
 51. Platforms should design “consent” and privacy policies in a way that facilitates informed choice for users and is compliant with data protection laws. Users have to be able to exercise control over information they receive and impart. In practice, they should have tools to exclude certain content from algorithmic personalisation as well as certain sources of content recommended to them.
 52. In order to return the agency back to online users, online platforms should properly disclose that a user is or will be subjected to algorithmic decision making, including personalised content curation.
 53. Linguistic diversity of over 3 billion social media users, including minorities, should be accommodated. All social media platforms’ content policies (however they are called, whether values, standard or rules) should be translated in at least all supported languages of the particular platform. Automated detections should include more languages, especially the most commonly used. Human moderators should be increased to advise on local cultures and contexts, including minority cultures and contexts.

Assessment/Identification of Hate Speech

54. The transparency of Internet companies with respect to their platforms should be radically improved. Internal definition, interpretation and application of hate speech standards should be elaborated and shared publicly to ensure better protection for minorities, but also enable minority activists to share evidence of human rights abuse and violations.
55. Social media companies should ensure that persons belonging to minorities, as amongst those most affected by hate speech, are involved in analysing the context and content of hate speech and identifying the most effective tools to address its harms;
56. The recommender algorithms should be subjected to human rights and minority rights impact assessments to prevent any inadvertent biased results.³
57. Social media companies should ensure that they are able to identify “dog-whistle” rhetoric, that is rhetoric implying and inciting hate without explicitly stating it, to prevent such material from spreading.

Addressing/Removing Hate Speech

58. Hate speech removal data needs to be disaggregated by protected characteristics, the type and severity of hate speech, and by categories of actors (dangerous organisations and individuals) along with key drivers for increased occurrences such as before elections and after terrorist attacks. Data on removals should be accompanied with accuracy of automated removal systems. Academic

³ An example of good practices includes Mozilla’s initiative for people to download software that will monitor YouTube’s recommendations to users.

researchers, organisations and other stakeholders should be given access to data for independent analysis and to ensure transparency.

59. Social media companies should develop and adopt alternatives to banning of accounts and removal of content, including de-amplification and de-monetization, education, counter-speech, the promotion of human rights and positive social values, as well as reporting and training which promotes alternative and positive narratives about persons belonging to minorities. This could be done in coordination with NHRIs and/or INGOs.

Appealing Hate Speech-related decisions

60. Social media companies should offer two layers of appeal with far more detailed and transparent processes before there is a right to appeal to an independent regulator or oversight mechanism. A full body of cases should be published to ensure transparency.

Session 4: Towards a safer space for minorities: positive initiatives to address online hate speech: the role of NHRIs, human rights organisations, civil society and other stakeholders (regulate what and how?)

61. Independent statutory bodies, such as NHRIs and equality bodies, human rights organisations, civil society and other stakeholders should prioritise initiatives to counter online expressions of antisemitism, Islamophobia, anti-Gypsyism, caste-based hatred and other forms of racism, and xenophobia, while ensuring participation of minorities.
62. Independent statutory bodies, such as NHRIs and equality bodies, human rights organisations, civil society and other stakeholders should build and strengthen partnerships with each other, and with minority communities. All minority groups should be provided equitable representation.
63. Independent statutory bodies, such as NHRIs and equality bodies, human rights organisations, civil society and other stakeholders should help to ensure safe spaces for minorities to discuss issues related to hate speech.
64. Independent statutory bodies, such as NHRIs and equality bodies, should have the necessary resources and be sufficiently empowered to effectively counter hate speech.
65. Independent statutory bodies, such as NHRIs and equality bodies, human rights organisations, civil society and other stakeholders should support and promote research on manifestations of hate speech, as well as academic collaboration across regions in order to foster meaningful outputs, and effective contributions in the context of applied research; and should collect and make available data on hate speech.
66. The representatives of minorities should build coalitions with each other at the international, regional, national, and local levels to effectively address and counter online hate speech and exchange experiences.
67. Minorities should monitor and counter hate speech within their own institutions, media and communities.
68. Human rights and civil society organisations, including minority organisations, should advocate that tech companies to counter growing online hate speech based on antisemitism, Islamophobia, anti-Gypsyism, caste and other grounds, and ensure that measures to address hate speech are incorporated into corporate human rights due diligence processes and requirements.

69. Human rights and civil society organisations, including minority organisations, should be included in formal and informal educational activities about the negative consequences of hate speech, including in schools and universities.
70. Media institutions should promote correct, equitable representation of minority groups and include information about human rights, diversity, non-discrimination and prejudice in their reporting.
71. Media representations of minorities should be based on principles of respect, fairness and the avoidance of stereotyping.⁴ Media should avoid referring unnecessarily to race, ethnicity, nationality, religion, language, caste, and other group characteristics in a manner that may promote intolerance.
72. Drawing on, among other, the recommendations deployed in the [United Nations Guidance Note on Addressing and Countering COVID-19 related Hate Speech](#), media institutions should provide accurate and objective information about the COVID-19 pandemic without hate speech, disinformation, stereotypes, and without unnecessarily referring to factors such as race, ethnicity, nationality, religion, language, caste, and other protected characteristics.

⁴ The role of the mass media is detailed in [Principle 6 of the Camden Principles on freedom of expression and equality](#).