



# General Assembly

Distr.: General  
3 March 2021

Original: English

---

## Human Rights Council

### Forty-sixth session

22 February–19 March 2021

Agenda item 3

**Promotion and protection of all human rights, civil,  
political, economic, social and cultural rights,  
including the right to development**

## Minority issues

### Report of the Special Rapporteur on minority issues, Fernand de Varenes\*, \*\*

#### *Summary*

In the present report, the Special Rapporteur on minority issues, Fernand de Varenes, provides an overview of his activities since his previous report (A/HRC/43/47), and a thematic report in which he addresses the widespread targeting of minorities through hate speech in social media. He describes related phenomena, including the widespread denial or failure of State authorities to recognize or effectively protect minorities against prohibited forms of hate speech. He concludes by emphasizing the responsibility of States, civil society and social media platforms to acknowledge that hate speech is mainly a minority issue and, as a matter of urgency, their duty to take further steps towards the full and effective implementation of the human rights obligations involved.

---

\* The present report was submitted after the deadline so as to reflect the most recent information.

\*\* The annexes to the present report are circulated as received, in the language of submission only.



## I. Introduction

1. The mandate of the Special Rapporteur on minority issues was established by the Commission on Human Rights in its resolution 2005/79 of 21 April 2005, and subsequently extended by the Human Rights Council in successive resolutions.
2. The current Special Rapporteur, Fernand de Varennes, was appointed by the Human Rights Council in June 2017 and assumed his functions on 1 August 2017. The mandate was extended in 2020 for a three-year period by the Council in its resolution 43/8.
3. The Special Rapporteur is honoured to be entrusted with the mandate and thanks the Human Rights Council for its trust in him. He also wishes to thank the Office of the United Nations High Commissioner for Human Rights (OHCHR) for its support in the implementation of the mandate.

## II. Activities of the Special Rapporteur

4. The Special Rapporteur wishes to draw the attention of the Human Rights Council to the mandate's web page, where general information is provided on the activities associated with the mandate, including communications, press statements, public appearances, country visits and thematic reports.<sup>1</sup> An overview of the activities of the mandate holder between 1 January and 1 July 2020 is provided in the report presented to the General Assembly at its seventy-fifth session (A/75/211).
5. The Special Rapporteur has continued to focus on increasing the visibility and raising awareness of minority issues – particularly related to his thematic priorities (statelessness; education, language and the human rights of minorities; hate speech targeting minorities in social media; and the prevention of ethnic conflicts) – among United Nations institutions and Member States, and more generally among the general public and other regional and international organizations. He has also focused on new approaches in order to improve the accessibility of activities under the mandate such as the Forum on Minority Issues.
6. This work has included two main initiatives:
  - (a) In cooperation with the Tom Lantos Institute and numerous regional minority and human rights organizations, consolidating, the continued organization of regional forums each year on the same themes as the Forum on Minority Issues;
  - (b) Clarifying, for the purposes of the mandate, a working definition of the concept of a minority, and the significance and scope of the four categories of minorities recognized in United Nations instruments (national or ethnic, religious and linguistic minorities).
7. The Special Rapporteur has signalled as a priority the thirtieth anniversary of the adoption of the Declaration on the Rights of Persons Belonging to National or Ethnic, Religious and Linguistic Minorities, expected to be marked in 2022 with a high-level event during the seventy-seventh session of the General Assembly in New York and associated activities.

### A. Country visit

8. The Special Rapporteur looks forward to continuing dialogue with Cameroon, India, Jordan, Kenya, Nepal, the Russian Federation, South Africa, South Sudan, the Syrian Arab Republic, Timor-Leste, the United States of America and Vanuatu, to which he has made requests to visit. He is appreciative to the Government of Paraguay, which has indicated its willingness for the Special Rapporteur to conduct a visit in 2021.
9. In his visits, the Special Rapporteur focuses on the importance of addressing discrimination, exclusion and other violations of human rights involving particularly vulnerable minorities, such as the Bidoon, Dalits and Roma, and doubly or even triply

---

<sup>1</sup> [www.ohchr.org/EN/Issues/Minorities/SRMinorities/Pages/SRminorityissuesIndex.aspx](http://www.ohchr.org/EN/Issues/Minorities/SRMinorities/Pages/SRminorityissuesIndex.aspx).

marginalized minority women, and issues pertaining to deaf and hard-of-hearing persons who, as users of sign language, are members of linguistic minorities.

10. The Special Rapporteur conducted an official visit to Kyrgyzstan from 6 to 17 December 2019 at the invitation of the Government (see A/HRC/46/57/Add.1).

## **B. Communications**

11. The Special Rapporteur sent communications and urgent action letters to the Member States concerned based on information received from diverse sources about human rights violations perpetrated against national or ethnic, religious and linguistic minorities. Those communications and the responses thereto are publicly available.

12. A total of 78 communications have been sent to Governments since January 2020. This was a considerable increase over the previous year's total number of communications. All of the communications were sent jointly with other special procedure mandate holders: 16 were urgent appeals, 53 were letters of allegation, and 9 were other letters expressing concerns regarding legislation and policy.

13. The largest number of communications were sent to States in Asia and the Pacific (43), followed by Europe and Central Asia (16), the Middle East and North Africa (9), the Americas (5) and sub-Saharan Africa (4). One communication was sent to a private company.

## **C. Conferences and awareness-raising activities**

14. Raising awareness and increasing the visibility of the human rights of minorities has been repeatedly highlighted as an important dimension of the Special Rapporteur's work since his election by the Human Rights Council in June 2017. Accordingly, the Special Rapporteur has frequently participated in and contributed to conferences, seminars and meetings at the international, regional and national levels throughout the world and with a variety of governmental and non-governmental organizations. He also has given frequent media interviews on issues involving the human rights of minorities.

15. A summary of the main activities undertaken by the Special Rapporteur from July 2020 to December 2020 is contained in annex I to the present report. The activities undertaken prior to this period are summarized in the Special Rapporteur's report to the General Assembly at its seventy-fourth session (A/74/160).

## **III. Update on 2020 forums on minority issues**

16. In his first report to the Human Rights Council, in 2018, the Special Rapporteur identified the need for a more regional approach with respect to the Forum on Minority Issues in order to make the Forum more accessible to minorities in different parts of the world and more receptive to regional concerns and contexts (A/HRC/37/66, para. 64). The first steps towards implementing such an approach were taken in 2019, when three regional forums were held. Because of the uncertainties in 2020 owing to the coronavirus disease (COVID-19) pandemic, only two of the four regional forums envisaged were able to proceed, on the thematic priority of tackling hate speech and incitement to hatred against persons belonging to minorities through social media. Still, more than 400 participants were involved in 2020 in the regional forums, held in Europe and Asia and the Pacific. The regional forums were possible thanks to the coordination of the Tom Lantos Institute and the contributions and assistance of numerous non-governmental organizations, State representatives, and regional and international organizations such as the United Nations Educational, Scientific and Cultural Organization, the Organization for Security and Cooperation in Europe (OSCE), the European Union, the Council of Europe and others.

17. The Forum on Minority Issues itself was established in 2007 by the Human Rights Council in its resolution 6/15 and its important role reaffirmed in 2012 in Council resolution 19/23. The Forum is mandated to provide a platform for promoting dialogue and cooperation on issues pertaining to national or ethnic, religious and linguistic minorities, and to provide

thematic contributions and expertise to the work of the Special Rapporteur. The Special Rapporteur is tasked with guiding the work of the Forum, preparing its annual meetings and reporting to the Council on the Forum's thematic recommendations. The Forum meets annually in Geneva for two working days allocated to thematic discussions. In recent years, the Forum is usually attended by more than 600 participants.

18. The thirteenth session of the Forum was held online on 19 and 20 November 2020, on the theme "Hate speech, social media and minorities". The number of participants, at fewer than 400, was greatly reduced as compared to previous years because of uncertainty over the holding of the Forum in the light of the pandemic, and the absence for a significant period of time of an OHCHR consultant to organize the Forum. The absence of a full-time consultant meant that OHCHR staff who worked with the Special Rapporteur but had never previously been involved in the Forum were assigned to ensure the Forum's success. The limited availability of staff before the Forum in November, the absence of a consultant with experience of the Forum, and the untried online format of the Forum led to the greatly reduced level of participation.

19. The complete report on the Forum and its recommendations is being presented to the Human Rights Council separately (A/HRC/46/58).

#### **IV. Thematic report: hate speech, social media and minorities**

20. The Special Rapporteur wishes to express his regret that, perhaps ironically, a computer failure at a critical moment led to a significant loss of data for and delays in the preparation of the thematic report.

##### **A. Introduction**

21. The year 2020 has seen the rapid spread of two pandemics: one a disease of the body, and the other a disease of minds. The ravages of the COVID-19 pandemic have at times overshadowed the illness of minds in social media, which has too often led to atrocities, massacres and even calls to genocide. While the scourge of COVID-19 is being addressed and may be under control within a year, the disease of minds shows no signs of weakening. On the contrary, hate speech in social media is spreading and strengthening. Minorities are overwhelmingly the main victims of hate and incitement to violence and discrimination. Where disaggregated data is available on hate speech in social media or on hate crimes, approximately 70 per cent or more of those targeted tend to belong to minorities.

22. The menace of hate speech affects minorities first and foremost. Whether by omission or not, many actors in the field fail to systematically acknowledge and nominally admit who the main targets are of racism, prejudice, scapegoating and even incitement to violence in social media. By not specifically mentioning minorities, the extent and brutality of hate speech is ignored, even camouflaged in a fog of generalities. In a sense, everyone becomes an accomplice to hate when the main victims remain unnamed. The result is fertile ground to feed intolerance and exclusion, the godparents of hate towards minorities. To add insult to injury, hate has become extremely profitable for some, and useful for others.

23. In the present thematic report, some of the gaps in efforts to combat hate speech in social media are examined, covering such aspects as the extent to which minorities are the main targets of hate speech in social media and the extent to which owners of social media platforms largely still ignore or are not taking sufficient or effective measures to curtail the spread of forms of hate speech that involve human rights obligations in international law. To add another layer of complexity, hate speech in social media actually involves a non-State, private third party in the actual owners of these platforms, which are among the largest and most profitable of global corporations.

24. The theme of hate speech, social media and minorities pits many of the most vulnerable communities – for example, ethnic minorities such as persons of African descent, Asian communities, Dalits, Rohingya and Roma, and religious or belief minorities such as Baha'i, Muslims, Jews and Christians – in countries in Africa, Asia, the Americas and Europe

against the interests of some of the most powerful corporate entities, with States sitting uncomfortably between the two.

25. Minorities are particularly vulnerable or disadvantaged in their use of and portrayal in online services such as social media. In addition to being the main targets of hate speech in social media, they can also be disproportionately affected by restrictions and removal measures in content moderation systems, as a result of bias in the notification systems themselves among users and third parties, and bias inherent in algorithms used in the content moderation tools, since behind every algorithm, there is a human.

26. This situation was not the one expected some 20 years ago when, perhaps naively, it was thought the Internet – and its offspring, social media – would simply provide humanity with amazing opportunities for fast, inexpensive means to communicate and exchange, and to support the transfer of information and education across borders almost instantaneously. The dichotomy of good and evil is as old as the story of the world itself, but it has been amplified and, because of the business model of some owners of social media platforms, intentionally directed so as to be monetized and more profitable.

27. The dark side of social media is now impossible to ignore, as they have become a breeding ground and propaganda vehicle for dangerous tropes, conspiracy theories, misinformation, disinformation, racism and scapegoating of minorities. Examples include a resurgence of some of the crudest forms of antisemitic propaganda, even all the way back to the *Protocols of the Elders of Zion* of more than a century ago. The events in the United States Capitol in January 2021 show that social media can be used to directly undermine democracy and stable societies, and as a propaganda tool for violence.

28. In the present report, some of these dimensions are considered with a view to better understanding, highlighting and providing guidance on the human rights obligations of States, corporations and other actors in relation to addressing the most prevalent instances of hate speech in social media, namely those targeting national or ethnic, religious and linguistic minorities.

## **B. International human rights law: freedom of expression as the starting point**

29. Even before the modern formulation of human rights, which emerged from the ashes of the Second World War and its horrors, the freedoms of opinion and expression were already presented as one of the pillars of a free, stable and democratic society. They offer a safe space in which individuals can be themselves and freely express their differences and diversity. The freedoms of opinion and expression are, in a sense, the oxygen of democracy. As the Human Rights Committee has pointed out, they are indispensable conditions for the full development of the person, and constitute the foundation stone for every free and democratic society.<sup>2</sup>

30. The freedoms of opinion and expression are widely recognized in universal and regional instruments as fundamental rights of all humans.<sup>3</sup> As the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression makes clear in his 2019 thematic report to the General Assembly, on online hate speech (A/74/486), the freedoms of opinion and expression should be viewed as a default starting point with only very strictly constructed restrictions.

31. The present report does not repeat the detailed and clear exposition of the central prominence and role of the freedoms of opinion and expression that the Special Rapporteur

<sup>2</sup> General comment No. 34 (2011) on the freedoms of opinion and expression, para. 2.

<sup>3</sup> International Convention on the Elimination of All Forms of Racial Discrimination, art. 5; Convention on the Rights of the Child, art. 13; Convention on the Rights of Persons with Disabilities, art. 21; International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families, art. 13; American Convention on Human Rights, art. 13; African Charter on Human and Peoples' Rights, art. 9; African Charter on the Rights and Welfare of the Child, art. 7; Arab Charter on Human Rights, art. 32; and Convention for the Protection of Human Rights and Fundamental Freedoms, art. 10.

on the promotion and protection of the right to freedom of opinion and freedom of expression has already strongly and eloquently made in his report. Rather, the present report on social media focuses on elements that have been largely omitted or unaddressed by this and other colleagues and United Nations entities: specifically, the most prevalent forms of hate speech in social media, such as those targeting minorities, and whether, in what contexts and to what extent freedom of expression may be restricted.

32. It should be made clear that even in a report of 2013, the United Nations High Commissioner for Human Rights expressed alarm at the often extraordinarily negative portrayal in many countries of migrants, but also of minority groups, by the media, politicians and other actors in the society, and she called for measures to curb growing xenophobic attitudes and underlined the need to swiftly denounce hate speech and prosecute those suspected of inciting racial violence and those who had perpetrated racist and violent actions (A/HRC/22/17/Add.4, annex, para. 3). Her report included the recommendations and conclusions of a series of expert workshops, which led to the adoption by the experts of the Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence (*ibid.*, annex, appendix).

33. One point that emerged during this process still needs to be emphasized much more strongly: the overwhelming majority of cases of hate speech in social media involve minorities as victims and targets. Clarifications, priorities and efforts to combat hate speech should therefore logically focus on minorities. However, this has largely not been the case, with little if any references to minorities – those most often affected by hate speech – in most United Nations and other initiatives. The idea that it is possible to tackle the root causes of some three quarters of cases of hate speech around the world without mentioning that these cases involve mainly minorities, and therefore to develop strategies that are relevant and adapted to this phenomenon, is perplexing to say the least.

34. In any event, the starting point for addressing the scourge of hate speech in social media must be full respect of freedom of expression. It follows that whatever the nature or content of hate speech, international law provides for only a limited number of permissible restrictions on freedom of expression. As outlined in the present report, it is therefore essential to clearly set out the human rights obligations and responsibilities of the parties most directly involved, particularly States and owners of social media platforms, as they relate to the most often occurring instances of hate speech in social media, namely those against minorities.

## **C. A contextualization: a pandemic of hate and the obligation to take action against prohibited forms of hate speech in social media**

### **1. A spreading poison of minds**

35. Freedom of expression should not be restricted, except where absolutely necessary. Unfortunately, until now, States and owners of social media platforms have largely been unable or unwilling to acknowledge and address the most vile and dangerous forms of hate circulating in the world targeting minorities and not only poisoning the minds of individuals, but also threatening the peace and stability of entire countries.

36. The Secretary-General did recognize that the pandemic had given rise to a new wave of hate, xenophobia, scapegoating and scaremongering, and that anti-foreigner sentiment had surged online and in the streets, antisemitic conspiracy theories had spread, and anti-Muslim attacks related to COVID-19 had occurred.<sup>4</sup> More precisely, this hatred mainly targets minorities. The pandemic has been weaponized by some and monetized by others, but this has been increasingly the case for the past decade. Minorities and other vulnerable groups constitute the majority of victims of incitement to hatred.

37. In addition to the groups mentioned by the Secretary-General, Roma in many parts of Europe have been the targets of choice for the poison of hate and violence propagated by

<sup>4</sup> See “United Nations guidance note on addressing and countering COVID-19 related hate speech”, 11 May 2020.

social media platforms. Dalits in South Asia have been also at the receiving end, as have migrants and refugees, who are also, in any event, members of minorities. Antisemitism and Islamophobia have grown and expanded rapidly online.<sup>5</sup>

38. What the world is experiencing through social media includes calls to genocide, against the Igbo minority in northern Nigeria in 2017 and against the Rohingya minority in Myanmar in 2018. There are calls to violence, murder, rape and other atrocities targeting Baha'i, Muslim and Christian minorities in some countries, Dalits in others, and incitement to violence against Roma minorities in Europe, against persons of African descent in many parts of the Americas, and so on. Migrants, who are also members of minorities, have continued to be scapegoated as sources of crime and disease, or portrayed as threats or corruptive influences. Hate speech, hate crimes and atrocities are surging against minorities, "facilitated by a handful of Internet companies that amount to the greatest propaganda machine in history".<sup>6</sup>

39. Social media have also portrayed minorities as a threat or danger, and contribute to violence towards them. For example, in 2018, a shooter who killed 11 worshippers at a synagogue in Pittsburgh, United States, had used the social media network Gab to circulate a conspiracy theory that Jews were bringing immigrants in to the United States to submerge white people in the country. Jews are the objects of a "great replacement" trope, exemplified by the Pittsburgh shooting, which seems to have originated in recent years with the far right in France and according to which certain non-white or non-Christian minorities would eventually demographically overcome and threaten "European" culture through immigration and high birth rates.

40. In Assam, India, where an exercise was under way to update the National Register of Citizens, Avaaz recently conducted a study of 800 posts on Facebook, and found a preponderance of hate speech against Bengali immigrants, who are openly referred to as "criminals", "rapists", "terrorists", "pigs" and other dehumanizing terms. These posts had been shared almost 100,000 times, adding up to around 5.4 million views for violent hate speech. Another similar, India-wide, study of hate content on Facebook, conducted by Equality Labs, provided a breakdown of hate content on the platform: 37 per cent of the posts were Islamophobic (including anti-Rohingya material, posts referring to "love jihad", glorification of earlier instances of violence against Muslims, and Islamophobic profanities), 16 per cent constituted fake news, 13 per cent targeted gender or sexuality, 13 per cent targeted caste minorities (40 per cent of these posts targeted caste-based reservation, while the rest consisted of caste slurs, anti-Ambedkar messages, posts decrying inter-caste personal relationships, and so on) and 9 per cent targeted other religious minorities.

41. These examples are the rule and not the exception worldwide in terms of the shape and content of hate speech. The OSCE hate crime data for 2018, covering European and other OSCE countries, indicate that more than 76 per cent of hate crimes involve Jewish, Muslim, Roma and other ethnic and religious minorities.

42. According to the information provided to the Special Rapporteur, minorities are increasingly and to a heartbreaking extent becoming victims of vitriol and hate, horrific in their brutality and animosity, in social media.

43. Hate speech begets hate crimes, as can misinformation and disinformation. As pointed out in one of the submissions to the Special Rapporteur, the Holocaust started not with the gas chambers, but with hate speech against a minority. False information may end up being harmful for minorities, even fatal, and result in so-called social media vigilante violence or "WhatsApp lynchings". One well-known case in France led to physical attacks against a number of members of the Roma minority in 2014 after disinformation in social media that Roma had kidnapped a child in a white van. In a more deadly case, in Sri Lanka in 2018, rumours of a Muslim minority plot to sterilize the Sinhalese majority, circulating mainly on

<sup>5</sup> Savvas Zannettou and others, "A quantitative approach to understanding online antisemitism", 24 November 2019. Available at [www.arxiv.org/pdf/1809.01644.pdf](http://www.arxiv.org/pdf/1809.01644.pdf).

<sup>6</sup> Anti-Defamation League, "Sacha Baron Cohen's keynote address at ADL's 2019 Never Is Now summit on anti-semitism and hate", 21 November 2019.

Facebook, led to deaths, with mobs in several towns burning mosques, shops and homes owned by Muslims minorities. Examples such as these against minorities are legion.

44. The Special Rapporteur shares the concern expressed in one submission that dehumanizing language, often reducing minority groups to animals or insects, normalizes violence against such groups and makes their persecution and eventual elimination acceptable, and that, when committed with a discriminatory or biased intent, these violations become a pathway of demonization and dehumanization that can lead to genocide. Individuals can find themselves drawn by social media into dehumanizing language and hate environments and end up surrounded by people with similar viewpoints. They can thus become enmeshed in confirmation bias in social media, an incubating environment that has become particularly conducive to the expression of – and indeed strengthened and confirmed – racist, intolerant and even violent viewpoints against certain scapegoated minorities.

## **2. Legal and regulatory context**

45. An invaluable compendium on the regulation of online hate worldwide shows that there is a wide range of approaches to the legal prohibition of hate speech, definitions, if any, of the concept, the type of restrictions or limitations imposed, and the availability of remedies, if any.<sup>7</sup> While it is impossible to generalize, the compendium does suggest that there is no single approach to hate speech, and quite often none specially adapted to the particular nature and challenges of hate speech in social media.

46. Some submissions to the Special Rapporteur highlight the fact that there is often a lack of enforcement of restrictions on hate speech in social media, particularly those intended to protect minorities. It has been suggested that in some countries, there are no data on hate speech cases in social media, and situations whereby existing legislation against hate crimes has never been used or is too onerous or vague to be invoked successfully for prosecution. In some of the submissions, it is suggested that minorities hesitate to bring cases of hate speech to the relevant authorities because their own experiences suggest that public authorities will not intervene, that there will be no consequences for those who breach the legislation, or that the use of moderation or complaint mechanisms for social media was unlikely to remedy the situation. The consequences of the lack of effective legal and other responses by public authorities and owners of social media platforms can be tragic to the point of being lethal and lead to massive atrocities and violations of human rights, as shown by Facebook's failure in Myanmar to address incitement against the Rohingya minority. Social media were used to demonize the Rohingya minority ahead of and during a campaign of ethnic violence, and the independent international fact-finding mission on Myanmar confirmed that Facebook had been a useful instrument for those seeking to spread hate (A/HRC/39/64, para. 74). The consequence was a foreseeable and planned human rights catastrophe of gang rapes, thousands of killings, and burning of schools, marketplaces, homes and mosques as part of ethnic cleansing and possible genocide attempt, all resulting in a horrific humanitarian crisis involving hundreds of thousands of men, women and children belonging to the Rohingya minority.

47. From the point of view of international law, companies such as social media platforms do not have the same obligations as Governments. Governments, however, have direct obligations, at a minimum, to prohibit incitement to genocide and advocacy that constitutes incitement to discrimination, hostility or violence. Given the impact of social media today in the propagation of hate speech, constituting grave violations of the rights of millions of people and even threats to their life and safety, Governments have the obligation to take measures to ensure that incitement to genocide or advocacy of hatred that constitutes incitement to discrimination, hostility or violence are prohibited.

48. A recurring issue is whether social media platforms should be subject to consequences and penalties, as is the case for mainstream traditional media, when they are allowed to become vehicles or a fertile ecosystem for incitement to genocide or advocacy that constitutes

---

<sup>7</sup> Hogan Lovells and PeaceTech Lab, *The Global Regulation of Online Hate: A Survey of Applicable Laws* (December 2020). Available at [www.hoganlovells.com/~media/hogan-lovells/pdf/2020-pdfs/2020\\_12\\_15\\_the-global-regulation-of-online-hate\\_a-survey-of-applicable-laws\\_special-report\\_december-2020.pdf?la=en](http://www.hoganlovells.com/~media/hogan-lovells/pdf/2020-pdfs/2020_12_15_the-global-regulation-of-online-hate_a-survey-of-applicable-laws_special-report_december-2020.pdf?la=en).

incitement to discrimination, hostility or violence. The above-mentioned compendium on the global regulation of online hate suggests that social media platforms are often not subject to such consequences and penalties.

49. The Special Rapporteur has been advised of the more recent steps of some platform owners to respond to messages that may constitute incitement to hate or harm, such as the recent suspension by Amazon of the Parler platform from its web hosting services because of Parler's failure to moderate and remove violent content.

50. The Special Rapporteur also notes and welcomes the setting up of an oversight board of independent members, which will make final and binding decisions on the content that Facebook and Instagram should allow or remove, on the basis of respect for freedom of expression and the human rights obligations to remove any incitement to genocide or advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence.

51. However, a point which is repeatedly raised in submissions is that much of the legislation and other measures to address hate speech in social media are too vague or weak, or are insufficient. The scourge of hate speech and the harm and violence that it fuels are not only continuing, but intensifying and threatening to destabilise societies, as illustrated by recent attacks on the United States Capitol and the German Reichstag.

#### **D. Permissible restrictions on freedom of expression to prevent harm**

52. As the Special Rapporteur on freedom of expression emphasized in his 2019 report (A/74/486, para. 26), and is widely acknowledged in the jurisprudence of international human rights bodies, content might for many be offensive, controversial, unpopular or disturbing, but freedom of expression protects the right to express such information and ideas. Minorities, in particular, can be the object of extremely widespread expression that may be offensive, intolerant and even racist: whether Governments should prohibit any such forms of expression, or at what threshold of severity or ugliness a Government would be entitled to restrict freedom of expression, are matters for debate. Nevertheless, as a matter of fundamental principle, restrictions on any form of expression must remain an exception, allowable only in cases that fall into one of three categories acknowledged in international human rights law, as outlined below.

53. On the basis of previous reports – such as the 2019 report of the Special Rapporteur on freedom of expression, the 2015 report of the preceding Special Rapporteur on minority issues (A/HRC/28/64), and others too numerous to enumerate – it is possible to identify the exceptional contexts in which freedom of expression may be restricted. In international human rights law, as enshrined in articles 19 and 20 of the International Covenant on Civil and Political Rights (and repeated, broadly speaking, in other human rights treaties) and in the Convention on the Prevention and Punishment of the Crime of Genocide, there are three strictly delimited contexts in which freedom of expression may be restricted:

- (a) States must criminalize expression that constitutes incitement to genocide;<sup>8</sup>
- (b) States have an obligation to prohibit by law, though not necessarily to criminalize, any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence;<sup>9</sup>
- (c) States may further restrict the exercise of the right to freedom of expression only as provided by law and as necessary for respect of the rights or reputations of others or for the protection of national security, public order, or public health or morals.<sup>10</sup>

54. It is noteworthy, although again largely unacknowledged, that the first two contexts of permissible restrictions on freedom of expression under international law are connected to

<sup>8</sup> Under art. III (c) of the Convention on the Prevention and Punishment of the Crime of Genocide, “direct and public incitement to commit genocide” is punishable.

<sup>9</sup> Covenant, art. 20 (2).

<sup>10</sup> *Ibid.*, art. 19 (3).

minorities being more susceptible to hate and violence from majority populations and national Governments. Most targets of genocide in modern history, for example, have been minority communities, such as Jews, Roma, Bosniaks and Tutsi. The reference to advocacy of “national, racial or religious” hatred that constitutes incitement again points to the main characteristics of minority communities around the world that are most likely to be vulnerable to targeted efforts of discrimination, hostility and violence.

55. While the three contexts of permissible restrictions on freedom of expression may appear to be unrelated, there is in fact an unappreciated connecting thread: harm, both tangible and potential. All three categories deal with harmful speech of greater or lesser gravity, moving from the obviously most harmful threat of incitement to genocide, the physical destruction of a national, ethnic, racial or religious – and usually minority – group. This worst form of harm entails the greatest obligation: criminalization of such forms of expression.

56. The second most severe form of harm is acknowledged to be any advocacy or expression of hate on the basis of national origin, race or religion. This restriction does not merely target intolerance, bias or uncomfortable prejudices. It is very narrowly constructed. States have an obligation to prohibit by law forms of expression that reach the threshold of constituting incitement to discrimination, hostility or violence; that is, encouraging another to commit an act of harm, in the form of discrimination, hostility or violence against others. Only the most severe forms of incitement to discrimination, hostility or violence would merit criminalization. In less severe instances, as noted in the Rabat Plan of Action (A/HRC/22/17/Add.4, annex, appendix, para. 34), civil sanctions and remedies should be considered, including pecuniary and non-pecuniary damages, along with the right of correction and the right of reply. Administrative sanctions and remedies should also be considered, including those identified and put in force by various professional and regulatory bodies.

57. Whereas the first category entails the criminalization of certain forms of expression and the second entails prohibition by States, the third category involves situations in which it is permissible, but not mandatory, for States to regulate and restrict speech on a limited number of grounds. In such situations, there is such potential for harm that States may be permitted under human rights law to impose limited restrictions on the exercise of the right to freedom of expression.

58. The conditions under which such restrictions are permissible are strict and demanding, as the Special Rapporteur on freedom of expression pointed out in his 2019 report (A/74/486, para. 6): “States may restrict expression ... only where provided by law and necessary to respect the rights or reputations of others or protect national security, public order, public health or morals. These are narrowly defined exceptions ... and the burden falls on the authority restricting speech to justify the restriction, not on the speakers to demonstrate that they have the right to such speech.” He went on to specify that any limitations must meet three conditions: legality, legitimacy, and necessity and proportionality.

59. Hate speech in social media may thus involve one of the three contexts in which freedom of expression may be subject to permissible restrictions under international human rights law. No other speech that might be reprehensible, uncomfortable or objectionable may be restricted under the fundamental pillar of freedom of expression in international law.

60. There are a number of specific issues relating to the recognition and protection of the human rights of minorities that are often insufficiently addressed when considering hate speech in its various forms and social media:

(a) The extent to which States comply with their legal obligation to criminalize incitement to genocide in social and other media;

(b) The extent to which States comply with their legal obligation to prohibit advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence in social and other media;

(c) The legislation in place to ensure that States comply with these international legal obligations. In addition to the general conditions of legality, legitimacy and necessity and proportionality, legislation to protect against all forms of hate speech must be carefully

and narrowly tailored in order to restrict freedom of expression only to the extent permissible. Much existing hate speech laws are vague in ways that can be used, and indeed are used, to restrict the freedom of expression of minorities and human rights defenders, including by preventing them from expressing legitimate grievances. Vagueness can render legislation useless or even threaten freedom of expression unduly, leading to impermissible restrictions on freedom of expression by motivating social media platforms to remove content rather than risk a fine (a phenomenon known as “overblocking”);

(d) The extent to which States prosecute breaches of permissible restrictions on freedom of expression in social and other media;

(e) The legislation that States have in place to comply with their legal obligations to criminalize or prohibit impermissible harmful speech in social media, setting out the obligations, including potential financial liabilities, of owners of social media platforms. Such platforms constitute favourable ecosystems for the propagation of the most harmful forms of hate speech, including incitement to genocide and advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence;

(f) The legislation in place requiring owners of social media platforms to adopt policies and protocols acknowledging and incorporating the fact that States have an obligation to prohibit the described illegal forms of harmful speech, and may legislate on others in accordance with international law. While social media platforms are private entities, such legislation is key as social media have become the main carriers of hate in the twenty-first century, particularly against minorities;

(g) The legislation in place requiring social media companies to manage hate speech on their platforms with reference to the human rights implications of their products. They must therefore have in place human rights review processes that refer to and focus specifically on the groups most susceptible to hate speech in the States concerned, and in particular minorities, women and lesbian, gay, bisexual, transgender and intersex communities;

(h) The transparency requirements in place, which must allow public oversight. Such requirements operate on two main levels: (i) in procedures and mechanisms within State agencies and social media platforms for the mandatory collection of data on incitement to genocide and advocacy that constitutes incitement to discrimination, hostility or violence. States are required to ensure that such data are disaggregated according to the basis of the hatred, whether national, ethnic, racial or religious hatred, which all relate to vulnerable minorities, or hatred affecting other vulnerable groups, on the basis of such factors as gender and sexual orientation; and (ii) in measures to ensure the actual enforcement of legislation in place to deal with hate speech.

61. States may restrict the exercise of the right to freedom of expression under the third category only as provided by law – that is, where each of the above-mentioned specific issues is addressed – and only as necessary to address sufficient threats of harm to respect of the rights or reputations of others, or to protect national security, public order, or public health or morals. Restrictions on freedom of expression outside these types of harm is not permitted.

62. While it is often claimed that what constitutes impermissible is too uncertain or contextual, this is not altogether accurate. The Rabat Plan of Action includes a six-part threshold test for assessing the severity of hatred in expressions that could be judged to constitute illegal hate speech under international law. Similarly, principle 12 of the Camden Principles on Freedom of Expression and Equality provides clear and succinct legal definitions of the main relevant terms, including “hatred”, “advocacy” and “incitement”.<sup>11</sup>

63. Any limitations to freedom of expression must remain within the strictly defined parameters flowing from international human rights, in particular the Convention on the Prevention and Punishment of the Crime of Genocide, the International Convention on the Elimination of All Forms of Racial Discrimination and the International Covenant on Civil and Political Rights. The implication of freedom of expression is that it must be possible to express views and openly debate and criticize opinions and institutions, even if offensive or

<sup>11</sup> ARTICLE 19, *The Camden Principles on Freedom of Expression and Equality*, (London, 2009).

disturbing. The key is that expression that constitutes direct and public incitement to genocide must be criminalized, and advocacy of hatred that constitutes incitement to discrimination, hostility or violence must be prohibited.

64. As discussed, under international law, a number of other forms of expression, some of which may be understood to constitute less severe hate speech, may be restricted exceptionally only as provided by law and as necessary for respect of the rights or reputations of others or for the protection of national security, public order, or public health or morals.

65. Focusing on the forms of hate speech that States must prohibit and in some cases criminalize (incitement to genocide and advocacy of hatred that constitutes incitement to violence, hostility or discrimination), it should be made clear that States are required to ensure that their legal obligations are respected by social media platforms, since the latter constitute the main ecosystems in which prohibited forms of hate speech have found particularly fertile ground.

66. At the other end of the spectrum, Poland has recently proposed legislation that owners of social media platforms should be prohibited from deleting content or banning users that do not break Polish legislation. While this measure has been presented and can be perceived as intended to protect freedom of expression, the question remains as to whether Polish law conforms to the State's international legal obligations to criminalize incitement to genocide and prohibit advocacy of hatred that constitutes incitement to discrimination, hostility or violence. In other words, legislation such as that proposed could breach human rights law if it restricts social media owners and protects freedom of speech in a way that is inconsistent with international law.

## **E. Role and responsibilities of owners of social media**

67. As private corporations, social media platforms currently enjoy a large degree of financial and other immunity in some States for being the carriers of hate and violence and even calls of genocide against minorities and others. Nevertheless, while private owners of property and platforms such as social media have the right to decide whom they will serve and accommodate, they are not above the law if their action results in forms of harm that must be prohibited under international human rights law.

68. A small number of the world's social media platforms, such as Facebook, Google, YouTube and Twitter, reach billions of people almost instantaneously. They are among the richest companies in the world. Their business models rely on engagement, and on algorithms that intentionally amplify content to keep users engaged, but they are also echo chambers that are too often narrow sources of information and concentrate bias and prejudices.

69. The online platforms' business models depend on maximizing profit by maximizing reading or viewing times. They make their money by enabling advertisers to target audiences with precision, so it is in their interests to design algorithms that will funnel users towards communities and themes on which they will spend the most time. Studies confirm that this has had an unexpected consequence of "rabbit-holing" individuals, of diverting them towards extreme or obsessive content, such as videos that promote conspiracy theories or are otherwise divisive, misleading or false – and hateful. In June 2019, for example, YouTube changed one algorithm to halve the number of views of videos deemed "borderline content" for spreading misinformation, as part of efforts to remove neo-Nazi and white supremacist videos.

70. These algorithms feed, concentrate and funnel hate and intolerance: almost two out of three people who join an extremist group (most of which again propagate hate speech against minorities) do so because of the recommendations pushed forward by the algorithms in social media.<sup>12</sup> The results have been the driving force behind an explosion of hate, radicalization, dehumanization, scapegoating, incitement to genocide and advocacy of hatred that constitutes incitement to violence, hostility or discrimination against minorities in social

---

<sup>12</sup> Guillaume Guichard, "Facebook a minimisé des initiatives internes visant à affaiblir les contenus extrémistes", *Le Figaro*, 27 May 2020.

media, leading to alarming increases in hate crimes and atrocities.<sup>13</sup> Reports widely confirm that hate speech online has been linked to a global increase in violence toward minorities, including mass shootings, lynchings and ethnic cleansing.<sup>14</sup> Hate pays, minorities suffer: social media platforms are hugely profitable, while minorities increasingly experience hate and incitement to violence through the platforms.

71. Changes have begun, though belatedly, following a recognition of the scale of hate speech targeting minorities in social media. As pointed out in submissions to the Special Rapporteur, even when social media platforms have vetting policies in place, they may be too slow, ineffective or broadly formulated.<sup>15</sup> Implementation often fails to protect the most vulnerable from harm. Additionally, content posted by minorities, who are particularly subjected to hate speech, demonized or scapegoated, is more frequently removed than content posted by the majority containing discriminatory or racist speech. Numerous studies on hate speech have noted that incidents against minorities are treated by the Government and law enforcement as unimportant or as pranks, and casually dismissed.<sup>16</sup> Ironically, perpetrators of incidents that reach the threshold of incitement to genocide or advocacy of hatred that constitutes incitement to discrimination, hostility or violence are often not prosecuted or punished in some countries, while, as noted in the Rabat Plan of Action (para. 11):

At the same time, members of minorities are de facto persecuted, with a chilling effect on others, through the abuse of vague domestic legislation, jurisprudence and policies. This dichotomy of (1) non-prosecution of “real” incitement cases and (2) persecution of minorities under the guise of domestic incitement laws seems to be pervasive. Anti-incitement laws in countries worldwide can be qualified as heterogeneous, at times excessively narrow or vague. ... [W]hile several States have adopted ... policies, most of them are too general, not systematically followed up, lacking focus and deprived of proper impact assessments.

72. One recent example of change reflecting the recognition of the scale of hate speech targeting minorities, and the need to adopt an approach that reflects the particular vulnerability of minorities to hate speech and the greater harm that they experience, is the indication by Facebook that its broad definition of hate speech needs to take into account those that are particularly targeted and subjected to harm. In 2020, Facebook was in the process of altering its algorithms to prioritize the flagging of hate speech targeting minorities such as persons of African descent, Muslims and Jews.

73. Other examples abound of the dangers of algorithms being developed and the use of artificial intelligence, which, in the absence of protocols and human rights impact assessments that duly take into account the vulnerability and targeting of minorities in social media, are prone to contribute to and accentuate the hate and harm experienced by minorities. In January 2021, a South Korean chatbot driven by artificial intelligence called Lee Luda, which had almost a million users, was taken down just a few weeks after its launch after spewing hate speech against minorities.<sup>17</sup> This was not the first such occurrence. Another artificial intelligence bot, Microsoft’s Tay, caused controversy in 2016 and was shut down just 16 hours after its launch, as it had been manipulated into using Islamophobic and white supremacist slurs against minorities. Artificial intelligence used in interactive platforms face

<sup>13</sup> According to the Commissioner for Human Rights of the Council of Europe, Dunja Mijatović, in her annual activity report for 2019 (21 April 2020), “[a]ntisemitism, Islamophobia and anti-Gypsyism have reached alarming levels. ... Hate speech and crimes against Roma also [remain] widespread.”

<sup>14</sup> Zachary Laub, “Hate speech on social media: global comparisons”, Council on Foreign Relations, 11 April 2019.

<sup>15</sup> The European Commission, in its fourth round of monitoring of implementation of the European Union code of conduct on countering illegal hate speech online, emphasized that the code “complements [national] legislation fighting racism and xenophobia, which requires authors of illegal hate speech offences – whether online or offline – to be effectively prosecuted.” See European Commission, “How the Code of Conduct helped countering illegal hate speech online”, February 2019. Available at [https://ec.europa.eu/info/sites/info/files/hatespeech\\_infographic3\\_web.pdf](https://ec.europa.eu/info/sites/info/files/hatespeech_infographic3_web.pdf).

<sup>16</sup> Mari J. Matsuda, “Public response to racist speech: considering the victim’s story”, *Michigan Law Review*, vol. 87, No. 8 (August 1989).

<sup>17</sup> Justin McCurry, “South Korean AI chatbot pulled from Facebook after hate speech towards minorities”, *Guardian*, 14 January 2021.

the same flaw: their algorithms allow bots to learn from past chat records or their interactions with current users. Given the large-scale prevalence of hate speech against minorities in social media, and the “teaching” or manipulation of artificial intelligence bots by racist and intolerant users, the bots are easily funnelled towards forms of intolerance and discrimination and themselves begin to spout hate speech against minorities.

74. There are other artificial intelligence issues of grave concern for the protection of the human rights of minorities that are outside the scope of the present report. One such issue is the use of facial recognition in artificial intelligence and digital technology by police and security forces and how it can interfere with the prohibition of discrimination and freedom of movement, expression and association, particularly the rights of specific groups, such as minorities and indigenous peoples.<sup>18</sup> Another is the targeting of specific minorities by police forces through facial recognition technology that can conduct racial profiling and specifically pick out members of a minority.<sup>19</sup> These and other issues could be considered in a future thematic report.

75. Overall, there is a sense that the explosion of hate, misinformation and disinformation, and the harm, radicalization, polarization and dangers that they are increasingly creating, must somehow be addressed, leading to increased recommendations, proposals, regulatory pressure and measures.<sup>20</sup>

76. Countries such as Germany have passed laws on online hate speech obliging social media platforms to remove hate speech within strict time limits or face heavy fines.<sup>21</sup> However, some aspects of the legislation have led to grave concerns regarding freedom of expression, including the prohibition of the dissemination of information on the basis of vague and ambiguous criteria, such as “insult” or “defamation”.<sup>22</sup>

77. In 2016, the European Union introduced a voluntary code of conduct on hate speech.<sup>23</sup> At the time of writing, it was also drafting a digital services act,<sup>24</sup> which may impose legal obligations and liabilities in recognition of the fact that, while social media and other online platforms already moderate illegal and harmful content, there remains little transparency, accountability or, in many cases, effectiveness. To date, Facebook, Microsoft, Twitter, YouTube, Instagram, Snapchat, Dailymotion, Jeuxvideo.com and TikTok have joined the

<sup>18</sup> Studies confirm the existence of bias against ethnic minorities in many facial recognition systems. One study showed that Asian and African American minorities in the United States were up to 100 times more likely to be misidentified than white males, depending on the particular algorithm and type of search (Drew Harwell, “Federal study confirms racial bias of many facial-recognition systems, casts doubt on their expanding use”, *Washington Post*, 20 December 2019). Such bias in artificial intelligence, and the resulting false positives and false negatives, have obvious serious consequences for minorities.

<sup>19</sup> Several technology companies in China filed patents for various devices and systems that included identification of Uyghur and other ethnic minorities. See IPVM, “Patenting Uyghur tracking – Huawei, Megvii, more”, 12 January 2021.

<sup>20</sup> See, for example, Committee on the Elimination of Racial Discrimination, general recommendation No. 35 (2013) on combating racist hate speech; Council of Europe, recommendation No. R (97) 20 of the Committee of Ministers to member States on “hate speech”, 30 October 1997; European Commission against Racism and Intolerance, general policy recommendation No. 15 on combating hate speech, 8 December 2015; and OSCE High Commissioner on National Minorities, *The Tallinn Guidelines on National Minorities and the Media in the Digital Age & Explanatory Note* (The Hague, 2019).

<sup>21</sup> See the Network Enforcement Act, 1 September 2017. Available at [www.gesetze-im-internet.de/netzdg/BJNR335210017.html](http://www.gesetze-im-internet.de/netzdg/BJNR335210017.html) (in German only).

<sup>22</sup> Letter dated 1 June 2017 from the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression addressed to the Permanent Representative of Germany to the United Nations Office and other international organizations in Geneva. Available at [www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-DEU-1-2017.pdf](http://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-DEU-1-2017.pdf).

<sup>23</sup> Available at [https://ec.europa.eu/newsroom/just/document.cfm?doc\\_id=42985](https://ec.europa.eu/newsroom/just/document.cfm?doc_id=42985). The code is very brief and quite general, does not define hate speech and makes no mention of minorities.

<sup>24</sup> European Commission, proposal for a regulation of the European Parliament and of the Council on a single market for digital services (Digital Services Act) and amending Directive 2000/31/EC, 15 December 2020. Available at <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en>.

code of conduct. The fifth – and latest – round of monitoring of the code’s implementation, conducted in 2020, suggested that on average the companies were now assessing 90 per cent of flagged content within 24 hours and that 71 per cent of the content deemed illegal hate speech was removed.<sup>25</sup> However, again, there remain concerns that the current process does not faithfully capture many forms of hate speech against minorities.

78. While it can be suggested that some two thirds of hate speech in social media targets minorities, the community standards or content moderation guidelines of most social media platforms pay little direct attention to minorities, or even fail to mention them specifically. TikTok’s Community Guidelines, for example, refer to matters such as “content that depicts or promotes the poaching or illegal trade of wildlife”, and defines hate speech or behaviour as “content that attacks, threatens, incites violence against, or otherwise dehumanizes an individual or a group” on the basis of attributes such as race, ethnicity and religion.<sup>26</sup> However, the Guidelines do not contain a single reference to the word “minority”. While “wildlife” legitimately gets a mention, minorities, unfortunately and counter-intuitively, do not, despite the fact that minorities are overwhelmingly the targets and victims of most hate speech in social media.

79. The above is not to suggest there have been no positive developments. Most social networks developed their content moderation rules on an ad hoc basis and only in recent years. Facebook first set down comprehensive internal moderation guidelines in 2009, but issued public guidelines only in 2018. There are increasingly indications that most owners of social media platforms are moving towards improving transparency and collaboration with civil society organizations on content moderation, including perhaps human rights impact assessments in some areas, and these efforts are to be commended.<sup>27</sup> However, the most glaring gaps remain the relative silence on the issue of greater attention to minorities as the main targets of hate speech, and few or no policy steps for their protection. One notable exception is the above-mentioned signal from Facebook that it was in the process of altering its algorithms to prioritize the flagging of hate speech targeting minorities. However, it is not known whether this change will be applied worldwide or on a more limited basis.

## F. Minorities and civil society organizations

80. Civil society initiatives, presence and involvement are essential in the modern world of communication and information through social media, and particularly since hate speech is mainly the scourge of minorities. Indeed, the United Nations also clarifies that the focus must be on those groups in situations of vulnerability due to entrenched or historic stigmatization, discrimination, long-standing conflicts, and exclusion and marginalization from the political, economic and social life of the society.<sup>28</sup> Most of these groups are minorities, and social media platforms must therefore focus more on minorities as the most at risk.

81. For this reason, it is increasingly acknowledged there is a need to introduce or strengthen the structured representation of minorities in social media. Platform owners therefore need to engage with and ensure the presence and participation of minorities in their advisory, consultative and other bodies, particularly those related to human rights impact assessments and hate speech issues. Direct and transparent inclusion of minorities is more likely to enable the owners of social media platforms to better understand and take account of the underlying root causes of hate, racism and intolerance in social media, and to more effectively manage and operate their platforms and products in order to avoid profiting from hate.

<sup>25</sup> See [https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combating-discrimination/racism-and-xenophobia/eu-code-conduct-counteracting-illegal-hate-speech-online\\_en](https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combating-discrimination/racism-and-xenophobia/eu-code-conduct-counteracting-illegal-hate-speech-online_en).

<sup>26</sup> See [www.tiktok.com/community-guidelines?lang=en#38](http://www.tiktok.com/community-guidelines?lang=en#38).

<sup>27</sup> See, for example, OHCHR, “Public comment by UN Special Rapporteur on minority issues relating to cases on hate speech and minorities”, 23 December 2020.

<sup>28</sup> *United Nations Strategy and Plan of Action on Hate Speech: Detailed Guidance on Implementation for United Nations Field Presences* (2020), p. 11.

## V. Recommendations

82. The Special Rapporteur invites States, regional organizations and non-governmental organizations to recall the recommendations of the Special Rapporteur on freedom of expression in his 2019 report (A/74/486), and not to impose on social media platforms any restrictions that fail to meet the standards set out in international human rights law.

83. The Special Rapporteur on minority issues reminds States that restrictions on forms of expression in social media are permissible only as provided by law and as necessary for respect of the rights or reputations of others or for the protection of national security, public order, or public health or morals. Such restrictions must also be clearly and narrowly defined, respond to a pressing social need, be the least intrusive measure available to be effective and the least restrictive of freedom of expression, not be overbroad or vague, and be proportionate in that the benefit to the protected interest outweighs the harm to freedom of expression.

84. The Special Rapporteur also invites States and other interested parties to refer to existing guidance on what constitutes forms of expression that may be restricted under international human rights law in the area of impermissible hate speech, in particular the six-part threshold test in the Rabat Plan of Action and principle 12 of the Camden Principles on Freedom of Expression and Equality.

85. The Special Rapporteur reminds States that in order to comply with their international obligations, they must adopt legislation against certain forms of hate speech, including the prohibition of any advocacy of national, ethnic, racial or religious hatred that constitutes incitement to discrimination, hostility or violence, and the criminalization of incitement to genocide in social media. These forms of prohibited hate speech must be interpreted narrowly, consistent with the six-part threshold test in the Rabat Plan of Action and the detailed guidance on the United Nations Strategy and Plan of Action on Hate Speech. To comply with these international obligations, States are invited to study and detail the effective enforcement of these prohibited forms of hate speech, and to consider the potential liability of social media platforms for failure to comply.

86. States should review their legal framework to ensure that it incorporates a list of protected characteristics or identifies the minorities (national or ethnic, religious and linguistic) that are recognized in international human rights law, particularly since most incidents of impermissible hate speech worldwide involve minorities.

87. The Special Rapporteur invites State agencies and social media platforms to put in place in procedures and mechanisms for the mandatory collection of data on hate speech, and at a minimum on incitement to genocide and advocacy that constitutes incitement to discrimination, hostility or violence. Such data should be disaggregated according to the basis of the hatred, whether national, ethnic, racial or religious hatred, which all relate to vulnerable minorities, or hatred affecting other vulnerable groups, on the basis of such factors as gender and sexual orientation.

88. The Special Rapporteur recommends that national legislation be in place that requires social media platforms to adopt policies and protocols to identify hate speech on the grounds of national origin, ethnicity, religion and language, as well as other factors such as gender and sexual orientation, while respecting the right to privacy of individuals.

89. The Special Rapporteur encourages social media companies to manage hate speech on their platforms with reference to the human rights implications of their products, including algorithms and artificial intelligence programs such as chatbots. They must therefore have in place human rights review processes that refer to and focus specifically on the groups most susceptible to hate speech in the States concerned, and in particular minorities, women and lesbian, gay, bisexual, transgender and intersex communities.

90. The Special Rapporteur invites States to act decisively, quickly and effectively to address and counter hate speech against minorities in online communications, including by swiftly and effectively investigating and prosecuting those responsible, holding them accountable, and ensuring that victims have effective access to justice and remedy.
91. The Special Rapporteur invites States, the United Nations and its entities, and in particular OHCHR, to initiate a process to develop a global voluntary code of conduct for social media platforms to combat hate speech. He also requests that they support his efforts to draft guidelines on combating hate speech targeting minorities in social media, as a matter of urgency, in 2021–2022.
92. The Special Rapporteur emphasizes that social media’s content moderation systems and community standards and any oversight or appeal entity should clearly commit to protecting vulnerable and marginalised minorities and other groups. Minorities should specifically be identified as priorities for social media platforms.
93. The Special Rapporteur invites owners of social media platforms to apply the Guiding Principles on Business and Human Rights in their own operations. Human rights standards should be systematically integrated into the content policies and decision mechanisms of social media platforms, so that, as specified in the Guiding Principles, their owners “comply with all applicable laws and respect internationally recognized human rights wherever they operate”, and “treat the risk of causing or contributing to gross human rights abuses as a legal compliance issue wherever they operate”, for which they could be liable.
94. The Special Rapporteur urges States to establish independent and authoritative specialized institutions that meet international standards to carry out work to counter hate speech, and accessible mechanisms for civil society organizations to report hate speech online.
95. The Special Rapporteur recommends that States provide adequate and specialized training to law enforcement agencies and the judiciary on minority rights, in particular with regard to hate speech online against minorities.
96. The Special Rapporteur encourages States and social media platforms to comprehensively address distortion and systemic bias against Jews and Muslims, as evidence suggests that antisemitism and Islamophobia are pressing challenges for minority rights.
97. The Special Rapporteur urges States to adopt human rights education initiatives on minority rights, including in school curricula; to promote diversity and pluralism; and to combat discrimination, stereotyping, xenophobia, racism and hate speech by disseminating positive, alternative and inclusive narratives that invite dialogue.
98. In order to improve mechanisms and content vetting policies for the handling of hateful content, and to ensure incorporation of the concerns of the main targets of hate speech in social media, the Special Rapporteur urges that minorities, as the most targeted and vulnerable groups, be represented in advisory and other relevant boards.
99. The Special Rapporteur invites States, national human rights institutions and civil society representatives to use the United Nations human rights mechanisms, including the special procedures of the Human Rights Council, the universal periodic review and the human rights treaty bodies, and regional human rights mechanisms and other appropriate forums to combat online hate speech directed against minorities.
100. The Special Rapporteur encourages all stakeholders to look into innovative, educational and preventive strategies that focus on protection of and respect for diverse communities in order to counter hate speech.

## Annex I

1. On 9 July 2020, the Special Rapporteur took part in a webinar entitled ‘Debating Challenges on Minority Protection’, organised by the Tom Lantos Institute, in cooperation with the Human Rights Consortium at the University of London and other organisations. The webinar focussed on the interrelationship between conflicts, minority rights and the promotion of inclusiveness and stability.
2. On 21 July 2020, he participated in a webinar on “Keeping the Faith in Times of Hate: The Practical Utility of Human Rights” co-hosted by Religions for Peace, UN Human Rights and the Convention on the Elimination of all Forms of Discrimination Against Women, raising the implications of COVID-19 for minorities who simultaneously face an increase in hate speech in social media. On 21 July 2020, the Special Rapporteur also addressed a webinar organised by the OSCE Transnational Threats Department’s Strategic Police Matters Unit in collaboration with the OSCE High Commissioner on National Minorities entitled “Towards resilient and cohesive societies: community policing to prevent and combat human trafficking” where he spoke on the particular vulnerability to trafficking of minority women and girls.
3. On 31 July 2020, the UN Special Rapporteur gave the keynote addresses at the plenary session of the Unrepresented Peoples Organisation (UNPO) General Assembly on the impact of COVID-19 on minorities and indigenous peoples worldwide.
4. On 11 August 2020, the Special Rapporteur in the webinar celebrating International Youth Day organised by the National Youth Council of Moldova, in cooperation with the Youth Platform of Interethnic Solidarity of Moldova and former UN Minority Fellows Programme.
5. On 31 August 2020, he gave an online lecture to students in the Human Rights programme at Vesalius College in Brussels on his mandate on minority issues and the protection of minority rights in international law.
6. On 10 September 2020, the Special Rapporteur spoke on ‘Inequalities, disenfranchisement and frailty: fault lines in global human rights protection and challenges faced by minorities’ during a webinar session of the online international Conference on Global Justice, Rule of Law and Human Rights on Pandemic Perspectives, organised by the Research Group Culture, Law and Society of the State University of Maranhão, Brazil.
7. On 14 September 2020, the Special Rapporteur participated in a webinar on “Situation of Human Rights Violation of Dalits in Nepal and Role of the International Community”.
8. On 21 and 22 September 2020, the Special Rapporteur participated in his mandate’s two-day online regional forum for Europe on hate speech, social media and minorities. More than 200 experts, state, human rights and minority representatives participated in the event during four panels looking more closely to the vastly disproportionate, and growing, targeting and scapegoating of minorities in social media for hate and incitement to violence and discrimination.
9. On 23 September 2020, Dr de Varennes gave the plenary presentation for the World Federation of the Deaf webinar and annual meeting for the International Day of Sign Languages. The Special Rapporteur addressed the language rights of users of sign languages as members of linguistic minorities.
10. On 25 September 2020, the Special Rapporteur gave the keynote speech for the 2020 Symposium on French and the Law organised by Massey College at the University of Toronto, on equity and access to education and the legal system.
11. On 8 October 2020, the Special Rapporteur participated online with discussions with and a presentation to students in the Global Minority Rights Summer School 2020 organised by the Tom Lantos Institute in Budapest.
12. On 19 and 20 October 2020, the Special Rapporteur participated in his mandate’s two-day online regional forum for Asia-Pacific on hate speech, social media and minorities. Some 200 experts, state, human rights and minority representatives participated in the event during

four panels looking more closely to the vastly disproportionate, and growing, targeting and scapegoating of minorities in social media for hate and incitement to violence and discrimination.

13. On 23 October 2020, the Special Rapporteur participated in the annual meeting of the European Language Equality Network Steering Committee Plus and discussed on the pandemic and minority languages, as well as the conclusions and recommendations of his 2019 mission report to Spain.

14. On 26 October 2020, the Special Rapporteur made introductory remarks and participated in the online conference “Minority Protection and Ethnic Groups’ Rights in Central and Central Eastern Europe” organised by the Federal Union of European Nationalities and the Cultural Foundation of German Expellees for Science and Research (Kulturstiftung der deutschen Vertriebenen für Wissenschaft und Forschung).

15. On 12 November 2020, the Special Rapporteur participated in the online consultation on anti-Muslim hatred online and Islamophobia organised by his colleague, the Special Rapporteur on freedom of religion or beliefs.

16. On 18 November 2020, the Special Rapporteur participated in Mexico’s National Transparency Week in an international conference on “Public Health and Transparency, the Importance of Public Information to Face Sanitary Crises” organised by the Mexican National Institute for Transparency, Access to Information and Personal Data Protection.

17. On 19 and 20 November 2020, the Special Rapporteur guided the 13th Forum on Minority Issues, held virtually this year exceptionally because of the pandemic on the theme of Hate Speech, Social Media and Minorities. Some 300 participants discussed during four panels the most significant areas of concern in relation to hate speech, social media and minorities, including on the causes, scale and impact of hate speech targeting minorities in social media, the international legal and institutional framework, the regulation of online hate speech, the role and responsibility of intergovernmental organisations, States, internet companies and social media platforms; and positive initiatives to address online hate speech.

18. On 8 December 2020, the Special Rapporteur made introductory remarks at the presentation of the results of the Justice and Social Equality Programme of the OHCHR Regional Office for Central Asia in collaboration with the Government of the Republic of Kyrgyzstan as well as the support of the European Union and the United Nations Peace Fund. He also on the same day participated online in an expert panel on discrimination in the justice system during the regional dialogue for the Latin America and Caribbean region on Promoting and Protecting the Rights of People of African Descent and their Communities in the Context of COVID-19 and Implementing the International Decade for People of African Descent, the Montevideo Consensus on Population and Development and the 2030 Agenda.

19. On 15 December 2020, the Special Rapporteur made the closing remarks and was presented the recommendations for the advocacy training workshop held online and in Baghdad on the preservation and protection of minority languages in Iraq organised by the JSSOR Youth Organization with the support of the United Nations Assistance Mission in Iraq and the OHCHR.

20. On 23 December 2020, the Special Rapporteur participated in the webinar entitled “Covid-19: I am minority do I matter?”, organised by the Swedish Institute Network of Future Global Leaders Malmö.

21. The Special Rapporteur also intervened in various capacities, including in an intervention to the Facebook Oversight Board in relation to its first 6 cases, three of which concerned hate speech and had some relation to minority issues. He drew special attention to the need to incorporate a minority rights approach and retain focus on the overarching aim of protecting vulnerable and marginalised groups which are in most cases minorities.

22. The Special Rapporteur also conducted numerous media interviews concerning his mandate and minority rights issues globally.

## Annex II

### Hate Speech, Social Media and Minorities: sample questionnaire and list of contributors

#### A. Sample questionnaire

##### Call for Submissions by 30 November 2020

In accordance with his mandate pursuant to Human Rights Council resolution 34/6, the Special Rapporteur on minority issues, Dr. Fernand de Varennes, will present a thematic report at the 46th Session of the UN Human Rights Council, which will provide a detailed analysis and highlight the issue of “Hate speech, social media and minorities”. Previously, in his first report to the UN Human Rights Council in March 2018, the Special Rapporteur identified the issue of ‘hate speech’, xenophobic rhetoric and incitement to hatred against minorities as one of the thematic priorities of his mandate. In this regard, his report to the Human Rights Council will build and further expand upon the work carried out by his predecessors on this particular area (see report [A/HRC/28/64](#)), as well as the contributions by other human rights mechanisms and the work done by civil society in this field.

Inspired by the [Rabat Plan of Action](#), the [Beirut Declaration](#) and its [18 commitments](#) on “Faith for Rights” and the 2019 launch by the Secretary-General of the [UN Strategy and Plan of Action on Hate Speech](#), the present report will refer to the global discussions on ‘hate speech’ and incitement to discrimination, hostility or violence through social media and will highlight the significance of the promotion and protection of the human rights of minorities, whilst respecting the right of freedom of expression.

##### Context

Across the world, discriminatory practices and hateful messages serve to stigmatize 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. Disaggregated data in some countries suggest that perhaps three-quarters or more of hate speech, for example, target ethnic, religious or linguistic minorities. ‘An ethnic, religious or linguistic minority is any group of persons which constitutes less than half of the population in the entire territory of a State whose members share common characteristics of culture, religion or language, or a combination of any of these. A person can freely belong to an ethnic, religious or linguistic minority without any requirement of citizenship, residence, official recognition or any other status’ according to the concept adopted in the UN Special Rapporteur’s [2019 annual report](#) to the UN General Assembly.

Expressions of hate and discrimination are increasingly present in certain political agendas and discourses and can contribute to a climate of fear amongst individuals and communities. They can at times also create a climate of rejection, exclusion and even scapegoating, threatening societal values and undermining the respect of human dignity and the protection of human rights. A state’s international obligations are thus central in any effort to address the direct challenges to and even rejection of the fundamental values at the core of the United Nations’ human rights’ architecture.

In the digital age, internet companies and social media platforms present opportunities for the online expression, association, participation and empowerment for members of ethnic, religious and linguistic minorities around the world, as well as for advocacy for the protection and promotion of their human rights. However, it has become increasingly obvious that there have also been numerous and flagrant examples of social media platforms being exploited to spread hatred, unfounded and stigmatizing rumours, fostering a climate of insecurity and animosity, and in the most extreme cases, leading to violence against members of minorities. Thus, while platforms offer members of minorities expanding opportunities for generating, framing and sharing information and reporting on issues that affect these individuals and groups, the same platforms can also host online ‘hate speech’, including incitement to discrimination, hostility or violence against those perceived as “others”, mainly members of

minorities. In extreme cases, they can even be used to propagate calls for genocide against minorities. Thus, unregulated online expressions of hate can increase the chances of human rights violations taking place offline against some of the most marginalised segments of society.

At the same time, some internet companies are responding to pressure to remove online content that is deemed harmful. This can occur through the imposition of intermediary liability, the application of filters, as well as automated tools. Some companies have also set their own content standards in this regard. Frequently, however, these measures have the purpose or effect of unlawfully, illegitimately and unnecessarily restricting the exercise of human rights – especially freedom of opinion and expression – online and operate in the absence of any meaningful oversight mechanisms. Furthermore, such measures may have a disproportionate impact upon or even target individuals from marginalised groups, particularly persons belonging to minorities (as well as political opponents, critics, and human rights defenders), while limiting the possibilities for information-sharing, awareness-raising and advocacy for civil society organisations, human rights defenders and representatives of persons belonging to minorities. Moreover, companies' and social media platforms' online content moderation policies can lack transparency and any precise and meaningful basis in international human rights standards, raising the possibilities that the decisions made by these actors undermine the human rights of individuals, including those belonging to minority groups.

Thus, the fundamental, two-pronged concern first raised by the Rabat Plan of Action in October 2012 (A/HRC/22/17/Add.4) – that members of minorities are effectively persecuted through the abuse of vague domestic legislation, jurisprudence and policies on 'hate speech', whereas actual incidents which meet the threshold of incitement to discrimination, hostility or violence under international human rights law are not addressed – has become an even more pressing issue at the beginning of the new decade, one that requires effective and urgent responses from States, social media platform owners, and other stakeholders which are based on international human rights law and standards.

Freedom of expression and the essential communication tools and services provided by or dependent on an open and accessible internet must be protected, as minorities and others who are marginalised or vulnerable must be protected from hate speech, incitement to discrimination, hostility or violence, and even calls to genocide.

### **Call for submissions**

In accordance with the established practice of thematic mandate-holders, the Special Rapporteur welcomes inputs by States, UN agencies, regional and international organizations, national human rights institutions, civil society, scholars and research institutions, private companies including those from the ICT sector, and others who may wish to submit for this purpose. Such submissions may include, for instance, recommendations, evidence and case studies. The following questions are intended to guide submissions:

1. Please provide annual disaggregated data since 2017 if possible on hate speech in social media, and in particular hate speech targeting minorities (national or ethnic, religious and linguistic minorities). Please additionally indicate whether there are future plans to include specifically disaggregated data on hate speech targeting minorities, considering that in most countries, the victims of hate speech on social media are usually members of minorities.
2. Please identify the mechanisms and processes in place to remove, penalise or address hate speech in social media targeting minorities. Please also specify and include any studies or reports assessing their implementation and effectiveness.
3. Please provide (legal and non-legal) examples of good practices of appropriate responses developed by States, internet companies, civil society and other relevant stakeholders to address online 'hate speech', including incitement to discrimination, hostility or violence, against persons belonging to minorities. Please include assessments, if any, on the effectiveness of these examples.

4. Please identify legal, institutional and policy challenges to address online ‘hate speech’, particularly incitement to discrimination, hostility or violence, on social media against minorities, as well as how specifically is it being addressed so as not to infringe freedom of expression.

5. Please identify good practices to strengthen the participation of persons belonging to minorities and their representatives in the development of laws and policies dealing with hate speech and social media, including their representation in advisory or other similar bodies, as well as their participation in national and other forums on online ‘hate speech’ and the promotion of diversity.

6. Please provide examples of trainings or other initiatives to strengthen partnerships among various stakeholders and to build their capacity to address ‘hate speech’ against minorities specifically, especially incitement to discrimination, hostility or violence on social media platforms.

## **B. List of contributors**

### **States:**

Azerbaijan, Croatia, Colombia, Cyprus, Greece, Guatemala, Hungary, India, Italy, Kenya, Kyrgyzstan, Latvia, Lebanon, Mexico, Poland, Portugal, Qatar, Singapore, Slovenia, Switzerland and the European Union

### **International and Regional Organizations:**

UNESCO

### **National Human Rights Institutions:**

Australian Human Rights Commission

Comisión de Derechos Humanos de la Ciudad de México

Commission nationale des droits de l’homme Togo

Czech Public Defender of Rights

Defensoría del Pueblo de la Nación de Argentina

Ombudsperson of Azerbaijan

Office of the Croatian Ombudsman

Slovak National Centre for Human Rights

### **Civil society organizations and Academia:**

Anhad

Baha’i International Community

Citizens against Hate

National Christian Evangelical Alliance of Sri Lanka (NCEASL)

Rights for Peace

RyangOk Ku and other lawyers in Japan

University of Minnesota