

# 'Hate Speech' Explained

## A Toolkit

---

2015 Edition

---

## **ARTICLE 19**

Free Word Centre  
60 Farringdon Road  
London,  
EC1R 3GA  
United Kingdom

T: +44 20 7324 2500  
F: +44 20 7490 0566  
E: [info@article19.org](mailto:info@article19.org)  
W: [www.article19.org](http://www.article19.org)  
Tw: [@article19org](https://twitter.com/article19org)  
Fb: [facebook.com/article19org](https://facebook.com/article19org)

ISBN: 978-1-910793-25-1

---

### **© ARTICLE 19, 2015**

This work is provided under the Creative Commons Attribution-Non-Commercial-ShareAlike 2.5 licence. You are free to copy, distribute and display this work and to make derivative works, provided you:

- 1) give credit to ARTICLE 19;
- 2) do not use this work for commercial purposes;
- 3) distribute any works derived from this publication under a licence identical to this one.

To access the full legal text of this licence, please visit: <http://creativecommons.org/licenses/by-nc-sa/2.5/legalcode>.

ARTICLE 19 would appreciate receiving a copy of any materials in which information from this report is used.

ARTICLE 19 appreciates the generous funding support of the Dutch Ministry of Foreign Affairs.

---

<b>What is this toolkit for?</b>	<b>3</b>
<b>Part I: Identifying ‘hate speech’</b>	<b>5</b>
What is the right to freedom of opinion and expression?	6
What is the right to equality?	8
What is ‘hate speech’?	9
Who is the target of ‘hate speech’?	13
Why use the term ‘hate speech’?	16
A proposed typology of ‘hate speech’	18
<b>The Hate Speech Pyramid</b>	<b>19</b>
‘Hate speech’ that must be prohibited	20
‘Hate speech’ that may be prohibited	22
Lawful ‘hate speech’	22
Are ‘hate speech’ and “hate crime” the same?	24
What expression is not automatically ‘hate speech’?	28
Deeply offensive expression	28
Blasphemy or “defamation of religions”	29
The denial of historical events	32
Inciting terrorist acts and violent extremism	34
Protection of “the state” and public officials	37
Defamation	39
<b>Part II: Responding to ‘hate speech’</b>	<b>40</b>
1) Creating an enabling environment for the rights to freedom of expression and equality	41
a) An enabling environment for the right to freedom expression	41
b) Ensuring the full protection of the right to equality and non-discrimination	45
2) Positive policy measures by States	49
Recognising and speaking out against intolerance	49
Equality training	51
Public policy for pluralism and equality in the media	52
Public education and information campaigns	55
Transformative justice	56
3) Voluntary initiatives by other stake-holders	57
Civil society initiatives	57
Mobilisation of influential actors and institutional alliances	58
Role of an independent and pluralistic media	58
Role of internet intermediaries	60
Role of meaningful inter-group dialogue	65

---

<b>Part III: Restricting ‘hate speech’</b>	<b>66</b>
Direct and public incitement to genocide	68
Advocacy of discriminatory hatred constituting incitement to discrimination, hostility or violence	70
Key elements of Article 20(2) ICCPR	71
The conduct of the speaker	74
The intent of the speaker	77
Severity threshold	78
Sanctions for incitement should not be limited to criminal penalties	82
Prohibitions on ‘hate speech’ under Article 19(3) of the ICCPR	84
<b>Annex I</b>	<b>86</b>

---

# What is this toolkit for?

In this toolkit, ARTICLE 19 provides a guide to identifying 'hate speech' and how effectively counter it, while protecting the rights to freedom of expression and equality. It responds to a growing demand for clear guidance on identifying "hate speech," and for responding to the challenges 'hate speech' poses within a human rights framework.

As such, it addresses three key questions:

1. How do we identify 'hate speech' that can be restricted, and distinguish it from protected speech?
2. What positive measures can States and others take to counter 'hate speech'?
3. Which types of 'hate speech' should be prohibited by States, and under which circumstances?

The toolkit is guided by the principle that coordinated and focused action taken to promote the rights to freedom of expression and equality is essential for fostering a tolerant, pluralistic and diverse democratic society in which all human rights can be realised for all people. It is informed by, and builds upon, ARTICLE 19's existing policy work in this field.

The toolkit is structured as follows:

- First, we outline that there is no uniform definition of 'hate speech' under international human rights law, rather, it is a broad concept which captures a wide range of expression. The toolkit advances a typology for identifying and distinguishing different forms of 'hate speech' according to their severity, guided by states' international human rights law obligations. (Part I)
- Second, we provide guidance on what policy measures State and non-state actors can undertake to create an enabling environment for freedom of expression and equality that addresses the underlying causes of 'hate speech' while maximising opportunities to counter it. (Part II)

- 
- Finally, we outline the exceptional circumstances in which the State is obliged by international law to prohibit the most severe forms of ‘hate speech’, and where also States may under international law place other restrictions on ‘hate speech’. This includes guidance on ensuring that such prohibitions are not abused, and to ensure that where sanctions are imposed they are appropriate and proportionate, as well as ensuring support and redress for victims. (Part III)

ARTICLE 19 believes that ensuring that responses to ‘hate speech’ comply with international human rights law is crucial. Prohibitions that censor offensive viewpoints are often counter-productive to the aim of promoting equality, as they fail to address the underlying social roots of the kinds of prejudice that drive ‘hate speech’. In most instances, equality is better-promoted through positive measures which increase understanding and tolerance, rather than through censorship.

*This toolkit is not a definitive version, and will be continuously updated to reflect the developing case law and best practices in this area.*

# Part I: Identifying 'Hate Speech'



---

To identify 'hate speech', it is essential that we first understand the importance of the mutually reinforcing human rights to freedom of expression and equality.”

In this section, ARTICLE 19 also proposes a typology for identifying 'hate speech': distinguishing different forms according to the severity of the expression, and its impact. We believe that this is critical to inform effective and nuanced responses to 'hate speech' and - in exceptional cases - prohibitions on 'hate speech' (see [Part III](#)).

## What is the right to freedom of opinion and expression?



Freedom of opinion and expression (freedom of expression) is a fundamental human right, protected in Article 19 of the Universal Declaration of Human Rights (UDHR)<sup>1</sup> and given legal force through all major international and regional human rights treaties.<sup>2</sup>

International human rights law requires States to guarantee to all people the freedom to seek, receive or impart information or ideas of any kind, regardless of frontiers, through any media of a person's choice.



**The scope of the right to freedom of expression is broad.** It includes, for example, the expression of opinions and ideas that others may find deeply offensive, and this may encompass discriminatory expression.<sup>3</sup>

It is often said that all human rights are universal, indivisible, interrelated, interdependent and mutually reinforcing. There are, furthermore, two reasons why international law grants particular importance to the right to freedom of expression as a cornerstone right:

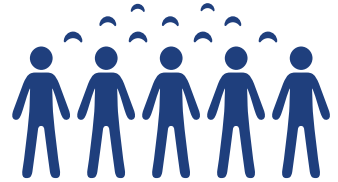
- <sup>1</sup> The UDHR is not strictly binding on states, however, many of its provisions are regarded as having acquired legal force as customary international law since its adoption; see *Filartiga v. Pena-Irala*, 630 F. 2d 876 (1980) (US Circuit Court of Appeals, 2nd circuit).
- <sup>2</sup> See Article 19 of the International Covenant on Civil and Political Rights (ICCPR), Article 9 of the African (Banjul) Charter on Human and Peoples' Rights (ACHPR); Article 13 of the American Convention on Human Rights (AmCHR), and Article 10 of the European Convention on Human Rights (ECHR).
- <sup>3</sup> Human Rights Committee (HR Committee), General Comment No. 34, CCPR/C/GC/34, 12 September 2011, para 11.





– At a personal level, freedom of expression is a key to the development, dignity and fulfilment of every person. People can gain an understanding of their surroundings and the wider world by exchanging ideas and information freely with others. People feel more secure and respected if they are able to speak their minds.

– At a state level, freedom of expression is necessary for good governance and therefore for economic and social progress. It ensures accountability by enabling people to freely debate and raise concerns with government, including for the protection and promotion of other human rights.



Having said this, **the right to freedom of expression is not an absolute right**, and the State may, under certain exceptional circumstances, restrict the right under international human rights law (see [below](#)).

---

## What is the right to equality?

International human rights law guarantees equality and non-discrimination for all people.<sup>4</sup> States are obligated to guarantee **equality in the enjoyment of human rights**, and **equal protection of the law**.

The principle of non-discrimination has three conjoined elements. It is understood as:

1. any distinction, exclusion, restriction or preference against a person
2. based on a protected characteristic recognised under international human rights law,
3. which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise, on an equal footing, of human rights and fundamental freedoms in the political, economic, social, cultural or any other field of public life.<sup>5</sup>

It is the protection of dignity for all people, without discrimination, which motivates most responses to ‘hate speech’, including restrictions on the right to freedom of expression.<sup>6</sup> Responses to, and prohibitions on, ‘hate speech’ are also often justified on the basis of protecting national security, public order, or public morals. However, where these aims are conflated with the aim of protecting individuals from discrimination, responses that limit expression can easily become overbroad and subject to abuse.

<sup>4</sup> See Article 1 of the UDHR and Articles 2(1) and Article 26 of the ICCPR. At the regional level, for example, freedom of expression is protected in Articles 2 and 19 of the ACHPR, Articles 1(1) and 24 of the AmCHR and Article 14 of ECHR and in Protocol 12 to ECHR.

<sup>5</sup> See, e.g. HR Committee, General Comment No. 18, 1989, para 6

<sup>6</sup> Many responses and prohibitions to “hate speech” are also justified on the basis of protecting national security, public order, or public morals. However, where these aims are conflated with the aim of protecting individuals from discrimination, responses that limit expression can easily become overbroad and subject to abuse.

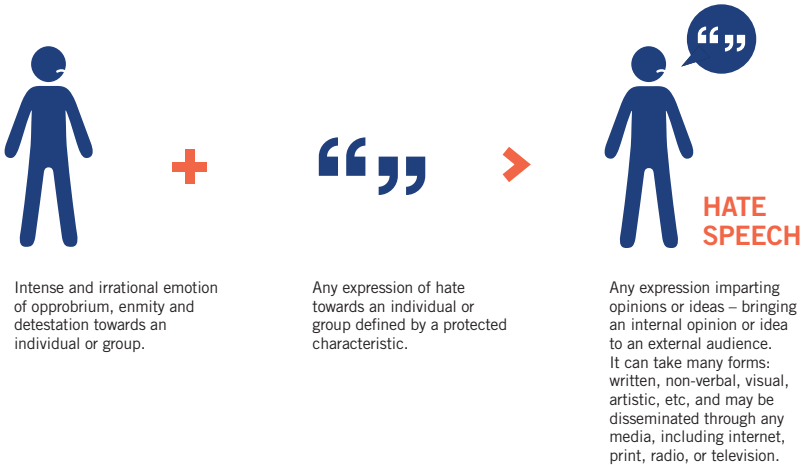
---

## What is ‘hate speech’?

‘Hate speech’ is an emotive concept, and there is no universally accepted definition of it in international human rights law. Many would claim they can identify ‘hate speech’ where they see it, but the criteria for doing so are often elusive or contradictory.

International and regional human rights instruments imply varying standards for defining and limiting ‘hate speech’: these variations is reflected in differences in domestic legislation.<sup>7</sup> In everyday settings, the use of the term and meanings attached to it vary – as do calls for regulating it. This could explain much of the confusion around the term, and what it means for human rights.

Many proposed definitions of ‘hate speech’ have been formulated in response to specific and perniciously discriminatory social phenomena or incidents. Definitions have also been adapted over time to address new situations, and to accommodate shifts in language, shifting understandings of equality, and the harms of discrimination, or developments in technology.



<sup>7</sup> See [Annex I](#).

- 
- Hate: the intense and irrational emotion of opprobrium, enmity and detestation towards an individual or group, targeted because of their having certain - actual or perceived – protected characteristics (recognised under international law).<sup>8</sup> “Hate” is more than mere bias, and must be discriminatory. Hate is an indication of an emotional state or opinion, and therefore distinct from any manifested action.
  - Speech: any expression imparting opinions or ideas – bringing an internal opinion or idea to an external audience. It can take many forms: written, non-verbal, visual or artistic, and can be disseminated through any media, including internet, print, radio, or television.

Put simply, ‘hate speech’ is any expression of discriminatory hate towards people: it does not necessarily entail a particular consequence. This lowest common denominator definition captures a very broad range of expression, including lawful expression. This definition, therefore, is too vague for use in identifying expression that may legitimately be restricted under international human rights law.

Beyond these two basic elements, the meaning of ‘hate speech’ becomes more contested; some people argue that discriminatory hate by itself isn’t enough, and that more must be shown. Opinions on what constitutes ‘hate speech’, and when it can be prohibited, vary widely, but include disagreement on the following elements:

- What constitutes a **protected characteristic** for identifying an individual or group that is the targets of ‘hate speech’;
- The degree of focus given to the **content** and **tone** of the expression;
- The degree of focus given to **harm caused**; whether the expression is considered to be **harmful in itself** for being degrading or dehumanising or is considered to have a potential or actual harmful consequence, such as:
  - inciting a manifested **action** against the target by a third person or group of people, such as violence;
  - causing an **emotional response** in the target, such as insult or distress; or

<sup>8</sup> The Camden Principles on Freedom of Expression and Equality, ARTICLE 19 (2009), Principle 12.1.

- 
- **negatively affecting societal attitudes**, by “spreading” or “stirring up” hatred;
  - The need for **causation** to be proven between the expression and the specified harm;
  - The need for any harm to be **likely** or **imminent**.
  - The need to **advocate** harm, implying that the speaker has an intent for harm to occur, and public **dissemination** of the expression.

Understandings of what ‘hate speech’ means can therefore fall anywhere between the lowest common denominator definition and one incorporating varying combinations of the above factors. At the same time, definitions are often ambiguous regarding one or more of these details, allowing flexibility for identifying ‘hate speech’ in its various manifestations, creating uncertainty and disagreement over what constitutes ‘hate speech.’

Here are just a few examples from different institutions and private actors to demonstrate the variety of approaches:

- **The European Court of Human Rights**, in a definition adopted by the **Council of Europe's** Committee of Ministers, considers 'hate speech' as: "all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility towards minorities, migrants and people of immigrant origin."<sup>9</sup>
- **YouTube**, in its community guidelines, describes 'hate speech' as: "content that promotes violence or hatred against individuals or groups based on certain attributes, such as: race or ethnic origin, religion, disability, gender, age, veteran status, or sexual orientation/ gender identity."<sup>10</sup>
- The UN's **International Committee on the Elimination of Racial Discrimination** understands 'hate speech' as "a form of other-directed speech which rejects the core human rights principles of human dignity and equality and seeks to degrade the standing of individuals and groups in the estimation of society."<sup>11</sup>
- The **Broadcasting Complaints Commission of South Africa** considers 'hate speech' to be "material which, judged within context sanctions, promotes or glamorizes violence based on race, national or ethnic origin, colour, religion, gender, sexual orientation, age, or mental or physical disability" or "propaganda for war; incitement of imminent violence; or advocacy of hatred that is based on race, ethnicity, gender or religion, and that constitutes incitement to cause harm."<sup>12</sup>

<sup>9</sup> Recommendation No. R(97)20 of the Council of Europe Committee of Ministers on "Hate Speech," 30 October 1997. See also, the European Court of Human Rights (European Court), *Gündüz v. Turkey*, App. No. 35071/97 (2004), paras 22 and 43. In Recommendation CM/ Rec (2010)5 "on measures to combat discrimination on the grounds of sexual orientation or gender identity," the Committee of Ministers has since recommended the following definition for homophobic and transphobic "hate speech": "all forms of expression, including in the media and on the Internet, which may be reasonably understood as likely to produce the effect of inciting, spreading or promoting hatred or other forms of discrimination against lesbian, gay, bisexual and transgender persons."

<sup>10</sup> [Youtube Community Guidelines](#), Hate speech.

<sup>11</sup> UN Committee on the Elimination of Racial Discrimination, General Recommendation No. 35 on combatting racist hate speech, 26 September 2013, CERD/C/GC/35, para 7

<sup>12</sup> [Code of the Broadcasting Complaints Commission of South Africa](#), applicable from 7 March 2003.

---

## Who is the target of 'hate speech'?



Put simply, 'hate speech' targets people, as individuals or groups, because of who they are.

ARTICLE 19 considers that grounds for protection against 'hate speech' should include all those protected characteristics which appear under the broader non-discrimination provisions of international human rights law. While this may seem obvious, it is sometimes contested.

Due to the lack of a universally accepted definition, obligations that permit or require limitations on certain types of 'hate speech' are pieced together from many different treaties.

It becomes complicated, as not all treaties which deal with discrimination require States to prohibit "hate speech."<sup>13</sup> Also, even in treaties that require prohibitions, the protected characteristics expressly listed are often cautiously narrow.

There are two explanations for this:

- Prohibitions on 'hate speech' are often drafted in response to discriminatory large-scale or persistent human rights violations where 'hate speech' was regarded as a causative factor. These prohibitions reflect the circumstances to which they respond, and may also be limited by the prejudices of society at the time of their drafting.<sup>14</sup>
- Requirements to prohibit 'hate speech' in international instruments have always been controversial, as several States have resisted such broad obligations, sometimes due to a belief that they unduly limit freedom of expression."

<sup>13</sup> For example, while the International Covenant on Elimination of All Forms of Racial Discrimination (ICERD) includes a broad obligation on States to prohibit types of racist "hate speech," the International Covenant on Elimination of all Forms of Discrimination Against Women (CEDAW) and the UN Convention on the Rights of Persons with Disabilities (CRPD), do not expressly require the similar prohibitions.

<sup>14</sup> E.g., while the persecution of homosexuals, political dissidents and persons with disabilities during the World War II is well-documented, Article 20(2) of the ICCPR does not expressly recognise sexual orientation, disability, or political opinion hatred as a vehicle through which persons can be incited to violence, hostility or discrimination.

---

A patchwork of overlapping international and regional instruments has resulted in divergent approaches to different forms of ‘hate speech’ in domestic law, including in relation to the protected characteristics. These international and regional instruments are surveyed in Annex I.

ARTICLE 19 has argued that the realisation of human rights should not be constrained by an overly formalistic commitment to the original wording of any instrument, or even to the intent of the drafters, if that interpretation would unnecessarily narrow the enjoyment of rights.

Also, international human rights instruments have been interpreted over time to support the principle of equality on a broad understanding of the term’, applying to protected characteristics specifically listed in treaties, as well as to grounds not expressly listed.<sup>15</sup> Numerous States recognise protected characteristics, in national laws prohibiting ‘hate speech’, which are reflective of the characteristics protected under their broader obligations to guarantee equality and non-discrimination.

With sufficient safeguards for freedom of expression, we consider that ensuring ‘hate speech’ provisions should be inclusive of the broadest range of protected characteristics. These should include but not be limited to: race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth, indigenous origin or identity, disability, migrant or refugee status, sexual orientation, gender identity or intersex status.

<sup>15</sup> Articles 2 and 26 of the ICCPR leave the list of protected grounds deliberately open ended, allowing an expansive reading of additional characteristics. See, e.g. Manfred Nowak, UN Covenant on Civil and Political Rights, CCPR Commentary, 2nd revised edition, page 628; or European Court, *Tyrer v. UK*, A 26 (1978).



---

## Protected Characteristics

ARTICLE 19 considers that grounds for discriminatory hate should include all those protected characteristics which appear under the broader non-discrimination provisions of international human rights law.



## Why use the term ‘hate speech’?

Accurately labelling certain expression as ‘hate speech’ can play an important role in advancing the values of dignity and equality which underpin international human rights law. However, too readily identifying expression as ‘hate speech’ should also be avoided, as its use can also have negative consequences.<sup>16</sup> The term is highly emotive, and can be abused to justify inappropriate restrictions on the right to freedom of expression, in particular in cases of marginalised and vulnerable groups.

'Hate Speech' as a Term	
Pros	Cons
Publicly recognising and rejecting the prejudice motivating "hate speech", and its connection to current and historical harms;	Shutting down legitimate debate on matters of public interest, in particular by people in positions of power;
Inviting a broader discussion on the implications of "hate speech" for the protection of human rights;	Overstating connection between the speech and the alleged harm - either by misjudging the influence a speaker or the likelihood of harm occurring – and/or overlooking the propensity of individuals to engage in counter-speech that has a strong and positive impact;
Showing solidarity and support to those targeted, acknowledging their dignity, and empowering them to also speak out;	Assuming that speakers intentionally advocated harm, while their intent may have either been more frivolous (e.g. an ill-judged or flippant comment on social media) or more nuanced (to satire, or provoke a debate on a challenging issue, including through art);
Exposing speakers and their supporters to counter-arguments;	Wrongly implying that all "hate speech" is unlawful and calling for criminal or other sanctions that might be inappropriate or ineffective;
Educating society, increasing understanding of the impact of "hate speech" and reduce propensity for it;	Increasing the audience of speakers, especially if they can frame themselves as "martyrs" of censorship or frame unsuccessful attempts at censorship as a vindication of their views;
Allowing for the monitoring of discrimination in society, to inform policy-making on effective responses.	Increasing policing and state or private surveillance of discourse, including online, and encouraging over-reliance on censorship instead of addressing institutional discrimination.

<sup>16</sup> Also, arguments have been made that the term "hate speech" is of limited use due to its exclusive reliance on either legal or judicial responses, see Agnes Callamard, Comments and Recommendations on ECRI General Policy Recommendation No. 15 on Combating Hate Speech, 20 May 2015.

---

For these reasons, some advocate alternative, more narrowly-defined, concepts, such as “dangerous speech”<sup>17</sup> or “fear speech,”<sup>18</sup> that focus more on the propensity of expression to cause widespread violence. In some contexts, such as in resolutions of the UN Human Rights Council, the term “hate speech” is avoided in favour of more elaborate formulations such as “intolerance, negative stereotyping and stigmatisation of, and discrimination, incitement to violence, and violence against persons based on religion or belief,”<sup>19</sup> or “the spread of discrimination and prejudice,” or “incitement of hatred.”<sup>20</sup> This perhaps demonstrates reluctance to normalise, or give legitimacy, use of the expression ‘hate speech’, given its status as a heavily contested term.

<sup>17</sup> Susan Benesch, *Dangerous Speech: A Proposal to Prevent Group Violence*, 23 February 2013.

<sup>18</sup> A. Buyse, *Words of Violence: ‘fear speech’, or How Violent Conflict Escalation Relates to the Freedom of Expression*, *Human Rights Quarterly*, Vol. 91, p. 1435 – 84.

<sup>19</sup> UN HRC Resolution 16/18 on Combating intolerance, negative stereotyping and stigmatisation of, and discrimination, incitement to violence, and violence against persons based on religion or belief (Resolution 16/18), *A/HRC/Res/16/18*, adopted without a vote on 24 March 2011.

<sup>20</sup> UN HRC Resolution 16/18 on situation of human rights of Rohingya Muslims and other minorities in Myanmar, *A/HRC/Res/29/21*, adopted without a vote on 3 July 2015.

---

## A proposed typology of ‘hate speech’

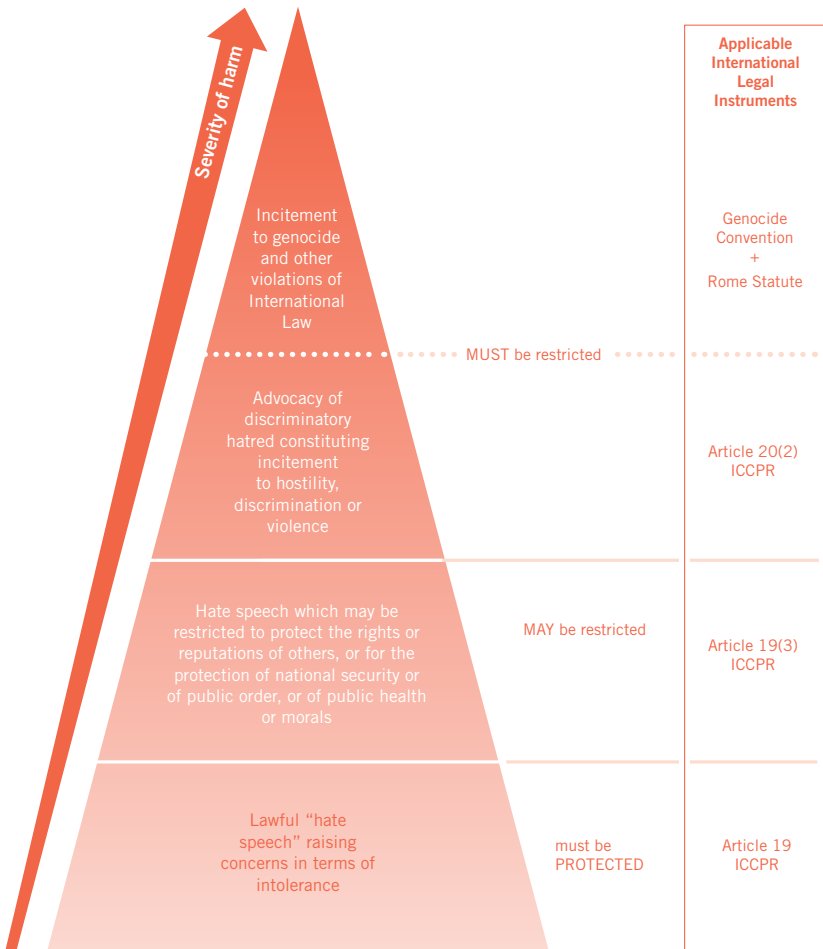
For the above reasons, we propose a typology of ‘hate speech’ - scaled according to its severity<sup>21</sup> - to give clarity to the different subcategories of expression that fit beneath the ‘hate speech’ umbrella, and to make it easier to identify appropriate and effective responses to “hate speech.” We propose to divide ‘hate speech’ into three categories:<sup>22</sup>

1. **‘Hate speech’ that must be prohibited:** international criminal law and Article 20 para 2 of the ICCPR requires States to prohibit certain severe forms of “hate speech,” including through criminal, civil, and administrative measures;
2. **‘Hate speech’ that may be prohibited:** States may prohibit other forms of “hate speech,” provided they comply with the requirements of Article 19(3) of the ICCPR;
3. **Lawful ‘hate speech’** which should be protected from restriction under Article 19(2) of the ICCPR, but nevertheless raises concerns in terms of intolerance and discrimination, and merits a critical response by the State.

<sup>21</sup> For the severity analysis for each category, see [Part III](#).

<sup>22</sup> This approach is based upon that advanced by the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (Special Rapporteur on FOE) in his annual report to the General Assembly, *A/76/357*, 7 September 2012 (the 2012 Report of the Special Rapporteur on FOE).

## The 'Hate Speech Pyramid'



---

## Hate speech' that **must** be prohibited

Under international law, States are required to prohibit the most severe forms of 'hate speech'. The prohibitions are tailored to preventing the exceptional and irreversible harms the speaker intends and is able to incite.

These are:

- **“Direct and public incitement to genocide”** - prohibited in the Convention on the Prevention and Punishment of the Crime of Genocide (1948)<sup>23</sup> (Genocide Convention) and the Rome Statute of the International Criminal Court (1998) (Rome Statute), ICC.<sup>24</sup> Prohibition of **incitement to other discriminatory violations of international criminal law**, such as the war crime of persecution, is not required by either the Genocide Convention or the Rome Statute, but should be considered within this category.
- **Any advocacy of discriminatory hatred that constitutes incitement to discrimination, hostility or violence**, as analogous to Article 20(2) of the ICCPR, but must additionally satisfy the requirements of Article 19(3) of the ICCPR.<sup>25</sup>

<sup>23</sup> The Convention on the Prevention and Punishment of the Crime of Genocide, 9 December 1948, UN Treaty Series, vol. 78, p. 277, Article 3(c).

<sup>24</sup> The Rome Statute of the International Criminal Court (last amended 2010), 17 July 1998, ISBN No. 92-9227-227-6, Article 6, Article 25(3)(e). See also Article 4(c) of the 1993 Statute of the International Criminal Tribunal for the former Yugoslavia (ICTY) and Article 2(3) (c) of the 1994 Statute of the International Criminal Tribunal for Rwanda (ICTR).

<sup>25</sup> Article 20(2) of the ICCPR only specifically requires prohibitions on the advocacy of national, racial and religious hatred. However, ARTICLE 19 argues that advocacy of hatred on the basis of all other recognised characteristics protected from discrimination under international human rights law.

- 
- The International Convention on the Elimination of all Forms of Racial Discrimination (the ICERD)<sup>26</sup> – in Article 4 – requires states to “condemn **all propaganda** and all organisations which are **based on ideas of theories of superiority of one race or group of persons of one colour or ethnic origin**, or which **attempt to justify or promote racial hatred and discrimination in any form**, and undertake to adopt immediate and positive measures designed to eradicate all **incitement to, or acts of, such discrimination** and, to this end, with due regard to the principles embodied in [the UDHR] and the rights expressly set forth in Article 5 of [the ICERD].” The Committee on the Elimination of Racial Discrimination recently adopted General Recommendation No. 35 on “combating racist hate speech,” which clarifies the scope of these provisions vis-à-vis the protection of the right to freedom of expression.<sup>27</sup> The ICERD contains much broader positive obligations on member States to prohibit particular types of speech than in Article 20(2) of the ICCPR. ARTICLE 19 has previously recommended ways in which this conflict can be reconciled, based on the Vienna Convention on the Law of Treaties.<sup>28</sup>

<sup>26</sup> International Convention on the Elimination of All Forms of Racial Discrimination, 21 December 1965, UN Treaty Series, vol. 660, p. 195.

<sup>27</sup> ICERD Committee, CERD/C/GC/35, 9 September 2013.

<sup>28</sup> [Vienna Convention on the Law of Treaties](#), 1969, Articles 31 and 32. For detailed information on reconciliation the different standards under the ICCPR and ICERD, see, ARTICLE 19, Prohibiting Incitement, op.cit., Recommendation 5.

---

## ‘Hate speech’ that **may** be prohibited

International human rights law permits States to restrict expression in limited and exceptional circumstances, complying with the three-part test under Article 19(3) of the ICCPR. These restrictions must be:

1. provided for by law
2. in pursuit of a legitimate aim, such as respect for the rights of others, and
3. must be necessary in a democratic society.

There are some forms of ‘hate speech’, which may be understood as **individually targeting an identifiable victim**. This type of ‘hate speech’ does not fit within the criteria of Article 20(2) of the ICCPR because the speaker does not seek to incite others to take an action against persons on the basis of a protected characteristic. These types of ‘hate speech’ include threats of violence, harassment and assault.

### **Lawful ‘hate speech’**

Expression may be inflammatory or offensive, but not meet any of the thresholds described above. This expression may be characterised by prejudice, and raise concerns over intolerance, but does not meet the threshold of severity (see Severity Threshold), at which restrictions on expression are justified.

This does not preclude States from taking legal and policy measures to tackle the underlying prejudices of which this category of ‘hate speech’ is symptomatic, or from maximising opportunities for all people, including public officials and institutions, to engage in counter-speech.

A (non-exhaustive) range of measures which States should be encouraged to adopt in this respect are detailed in **Part II** of this Toolkit.



### Example



A teenage boy, with a small number of followers on Twitter, tweets an offensive and sexist joke that trivialises the disappearance and likely murder of a local schoolgirl. It provokes a strong critical response against the boy online, and he eventually deletes the tweet.

Though the communication is offensive and reflects a broader problem of misogyny in society, he did not intend to incite any harmful conduct against a particular group, and in any case he does not have this kind of influence over his followers. This kind of 'hate speech' may justify soft intervention from local actors in positions of authority, such as teachers in his school or other community leaders, but it does not justify the State imposing sanctions or other restrictions.

### Useful Resources

- [Rabat Plan of Action](#) - an important document that provides authoritative guidance to States on implementing their obligations under Article 20(2) of the ICCPR to prohibit “any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence
- [Prohibiting incitement to discrimination, hostility or violence](#) – ARTICLE 19’s policy on prohibitions of incitement that elaborates on the interpretation of Article 20(2) of the ICCPR;
- [Camden Principles on Freedom of Expression and Equality](#) – a set of principles and recommendations to promote greater consensus about the proper relationship between respect for freedom of expression and the promotion of equality.

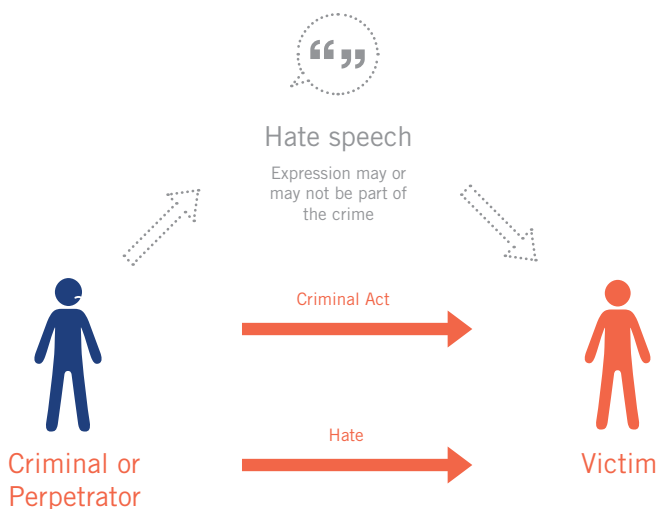
---

## Are ‘hate speech’ and ‘hate crime’ the same?

‘Hate speech’ and ‘hate crimes’ are often conflated and used interchangeably, but they should be distinguished. Both are symptomatic of intolerance and prejudice, but most ‘hate crimes’ do not involve the exercise of freedom of expression.

Although the term ‘hate crime’ is widely used, the use of the emotive term ‘hate’ may lead people to believe that any manifestation of ‘hatred,’ including ‘hate speech’, is a criminal offence. This is not the case.

While all ‘hate speech’ is a cause for concern, it will not always constitute a criminal offence, and therefore is not a “hate crime”.



---

The term “hate crime” refers to the commission of a criminal offence where the perpetrator targeted the victim in whole or in part out of a “bias motivation.” Many jurisdictions label certain criminal offences as a “hate crime” in order to acknowledge the broader prejudicial context in which a person was victimised. This acknowledgment also aims to build confidence among marginalised individuals in the criminal justice system, and allows them to feel that their full experience of the crime has been recognised.

As the Organisation for Security and Cooperation in Europe (the OSCE) has indicated, “hate crimes” are characterised by the existence of two conjoined elements:

- A “**base**” **criminal offence** (e.g. murder); and
- The crime being committed out of a “**bias motivation**” (e.g. against racial minority), which means the perpetrator chose the target of the crime based on the victim’s “protected” characteristic.<sup>29</sup>

The term “bias motivated crime” more accurately conveys that criminal responsibility is contingent on proving a criminal offence, and not on proving “hatred” only.

<sup>29</sup> [Hate Crime Laws: A practical guide](#), the OSCE, 2009.

---

### Example



A Muslim family moves into a house in a town populated mostly by Christian families. A neighbour smashes the windows of the family's new home. Asked by a passer-by why he is doing this, the perpetrator says he needs to stop "more of them arriving and taking over the town."

In this example, the crime of property damage has occurred, and the perpetrator has demonstrated a bias motivation through his communication with the passer-by. He may be prosecuted for committing property damage as a "hate crime", and receive a higher sentence to reflect the bias motivation. However, his communication to the passer-by is only relevant as evidence of his motivation for that crime; his speech wouldn't be the basis for any separate criminal offence.

In many "hate crimes," 'hate speech' will not be an element of the base criminal offence. Rather, uttered immediately prior to, during, or after the commission of a crime, it may be an indicator of a bias motivation and be introduced as evidence. In such cases, the accused would not be allowed to raise a defence based on their right to freedom of expression.

However, as established above, States are required to prohibit particularly severe forms of "hate speech," and may in exceptional circumstances do this through the criminal law. In these cases, 'hate speech' may also be the expressive act itself that is criminalised. In the typology above, the most severe types of 'hate speech' that may appropriately attract criminal sanction include "incitement to genocide", and particularly severe forms of "advocacy of discriminatory hatred that constitute incitement to violence, hostility or discrimination."

### Example



In the run-up to a heavily contested presidential election, the incumbent president makes a series of speeches to large rallies. During these rallies, he promulgates a rumour that supporters of the opposition, mostly belonging to another ethnic group, are arming themselves and are an existential threat to his supporters. As tensions increase, he uses racialised language, evoking instructions used in mass-killings in the country a few decades earlier, calling on his supporters to take urgent action to secure an election victory.

Here, the President has engaged in ‘hate speech’ which would arguably reach the threshold of **advocacy of hatred that constitutes incitement to violence**. He understands and is exploiting ethnic tensions in society, and knows that as an influential politician, his use of a particular term would be understood and likely acted on violently by individuals in the crowd against members of the ethnic group associated with the opposition. Whether or not violence actually ensues, this type of ‘hate speech’ may justifiably be criminalised as a “hate crime”.

As well as these (particularly severe) forms of ‘hate speech’ which States are required to prohibit, there are other forms of ‘hate speech’ that States may restrict. These would include individually targeted forms of bias-motivated threats, assault, or harassment (see the Typology of Hate Speech).

### Example



A same-sex couple, both women, are confronted on a train by another passenger who starts shouting sexist and homophobic abuse at them, causing the pair to reasonably fear immediate physical violence.

In many jurisdictions, this incident would, appropriately, be prosecuted as a bias-motivated crime. The abusive passenger’s expressive act falls within our broad typology of ‘hate speech’, and also amounts to the crime of an **assault**. The credible threat of violence in the expression makes it criminal conduct, and since it is characterised by bias, the content of the expression is also evidence of bias motivation.

---

## What expression is not automatically ‘hate speech’?

Given the confusion surrounding the concept, it is beneficial to achieve clarity about categories of expression which should not automatically be considered “hate speech.” In this section, we explain why certain concepts (which are often wrongly conflated with ‘hate speech’) should be distinguished and, in most cases, protected under the right to freedom of expression.



### Deeply offensive expression

International freedom of expression standards protect expression that is offensive, disturbing or shocking,<sup>30</sup> and do not permit limitations premised solely on the basis of “offence” caused to an individual or group. International human rights law provides no right to individuals to be free from offence, but it does unequivocally protect their right to counter such offence and speak out against proponents of that speech.

However, States often sanction expression labelled as “offensive,” sometimes even distinguishing degrees of offence as a basis for imposing sanctions. The nature of “offense” is inherently subjective, allowing the State to arbitrarily restrict certain views. Too often, prohibitions on “offensive” speech also lack the precision and clarity necessary to enable the public to regulate their conduct in accordance with the law. ARTICLE 19 therefore believes **offence should never be a basis for restricting expression**, even if it is discriminatory, without proof of that expression falling within the categories identified in [Part III](#).

<sup>30</sup> European Court, *Handyside v. UK*, App/ No. 5493/72, 7 December 1976; General Comment No. 34, op. cit., para 11, describes the scope of the right to freedom of expression as including “deeply offensive” speech.

---

Arguments are often made that some expression is so offensive that it holds no value and is inherently harmful. However, this argument overlooks the importance of safeguarding against abuses of power by the State, including abuses against minority groups. Furthermore, the catalyst (which offensive speech may provide) for counter-expression and debate is clearly in the public interest. Public policies and legislation that maximise opportunities for counter expression should therefore be favoured over restrictive responses in law.

### **Blasphemy or “defamation of religions”**

International human rights law protects people, and not abstract concepts, such as religions or belief systems. The right to freedom of expression cannot be limited for the purpose of protecting religions or associated ideas or symbols from criticism, or to shield the feelings of believers from offence or criticism. These are among the reasons many States give to justify retaining their blasphemy laws against international human rights standards.

Typically States’ blasphemy prohibitions fit into one or more of the following categories:

1. **Direct blasphemy:** that seek to protect a religion, its doctrines, symbols, or venerated personalities, from perceived criticism, contradiction, contempt, stigmatisation, stereotyping or “defamation;”
2. **Insult to religious feelings:** that seek to protect the feelings or sensibilities of a group of persons “insulted,” “offended,” or “outraged” by incidents of blasphemy against a religion they identify with; and
3. **Broad/vague laws limiting expression concerning religion or belief:** these are broadly or vaguely drafted laws, often concerning the protection of public morals or public order, that are applied to illegitimately limit free expression and shut down debate on religion or belief.<sup>31</sup>

<sup>31</sup> This typology is adapted from the analysis of the European Commission for Democracy through Law (the Venice Commission) in [The relationship between freedom of expression and freedom of religion: the issue of regulation and prosecution of blasphemy, religious insult and incitement to religious hatred](#), October 2008.

---

We recommend that any expression that falls within these categories **should not be considered ‘hate speech’** unless they separately meet the conditions set out for identifying ‘hate speech’ above, nor should they be **subject to any restriction** unless they exceptionally meet the **high thresholds for prohibition** (see [Severity Threshold](#) and our [6-part test](#)).

International human rights standards are unequivocal that **prohibitions on blasphemy should be repealed**. This is an explicit recommendation of the **Rabat Plan of Action**,<sup>32</sup> and strongly supported by **General Comment No. 34** of the Human Rights Committee.<sup>33</sup> Several special procedures of the UN Human Rights Council have also raised concerns regarding the impact of blasphemy laws on human rights and have recommended their repeal.<sup>34</sup> This has largely been reflected at the regional level in the Council of Europe,<sup>35</sup> European Union,<sup>36</sup> and Inter-American systems.<sup>37</sup>

There are several arguments which support the repeal of blasphemy laws, establishing that they are invalid under international human rights law, and counter-productive in both principle and practice:

- International human rights law distinguishes between the protection of ideas or beliefs from the protection of the rights of people on the basis of their religion or belief. The right to freedom of thought, conscience, religion or belief (freedom of religion or belief) under international law attaches to individuals,<sup>38</sup> and does not protect religions or beliefs per se from adverse comment or scrutiny.

<sup>32</sup> Rabat Plan of Action, op.cit.

<sup>33</sup> General Comment No. 34, op.cit., para. 48: “prohibitions of displays of lack of respect for a religion or other belief system, including blasphemy laws, are incompatible with the Covenant, except in the specific circumstances envisaged by article 20, paragraph 2, of the Covenant.” The HR Committee also underlines that it would be “impermissible” for any such prohibition to “prevent or punish criticism of religious leaders or commentary on religious doctrine and tenets of faith.”

<sup>34</sup> Report of the Special Rapporteur on minority issues, A/HRC/28/64, 2 January 2015; Report of the Special Rapporteur on freedom of religion or belief, A/HRC/28/66, 29 December 2014; the 2012 Report of the Special Rapporteur on FOE, op.cit.; UN Working Group on Arbitrary Detention, Opinion No. 35/2008 (Egypt), 6 December 2008, para 38. The HRC special procedures have also been supported by regional mandates, see e.g. [Joint Declaration on defamation of religions, and anti-terrorism, and anti-extremism legislation](#), 9 December 2008.

<sup>35</sup> Council of Europe Recommendation 1805 (2007), [Blasphemy, religious insults, and hate speech against persons on grounds of their religion](#), 29 June 2007. See also Venice Commission report, op.cit., para 89. The European Court has not considered a “blasphemy” case since



- 
- Restrictions on blasphemy are often utilised in order to prevent and punish the expression of minority or controversial views, inhibiting open and frank debate and exchange. This includes inter and intra religious dialogue, but also criticism of religious leaders and commentary on religious doctrine and tenets of faith. Individuals in positions of power often enforce these restrictions for political advantage, to target critics and avoid accountability, using “blasphemy” broadly to target any dissent.
  - Restrictions on blasphemy are often used in order to silence the expression of religious minorities and atheists or non-theists. Both the right to freedom

the report of the Venice Commission, but has previously found the application of blasphemy restrictions to be within a State's margin of appreciation, see: *Otto-Preminger-Institut v. Austria*, 20 September 1994; see also: *Wingrove v. UK*, App. no. 17419/90, 25 November 1996 (note that in 2008 the UK repealed its blasphemy law); *...A. v. Turkey*, App. no. 42571/98, 13 September 2005; and contrast *Giniewski v. France*. App. No. 64016/00, 31 January 2006; *Klein v. Slovakia*, App. No. 72208/01, 31 October 2006.

<sup>36</sup> See, e.g., the EU Guidelines on the Promotion and Protection of Freedom of Religion or Belief (2013)

<sup>37</sup> Inter-American Court of Human Rights, *The Last Temptation of Christ (Olmedo-Bustos et al.) v. Chile*, 5 February 2001.

<sup>38</sup> As protected e.g. under Article 18 of the ICCPR.

---

of expression and the right to freedom of religion or belief rely on a respect for pluralism and non-discrimination. Pluralism is essential, as one person's deeply held religious belief may be offensive to another's deeply held belief and vice versa. By privileging one belief system over another, either in law or in effect, restrictions on blasphemy inevitably discriminate against those with minority religions or beliefs.

ARTICLE 19 therefore strongly recommends that States repeal all prohibitions on “blasphemy” that fit within the above 3 categories.

In some cases, expression regarded as “blasphemy” may be used as a vehicle for expressing hatred against a particular group on the basis of a protected characteristic such as their religion or belief. Mostly, while this expression may raise concerns around intolerance, it will not necessarily reach the threshold for legitimate restrictions on expression. Where the threshold is met, State responses should be solely premised on Articles 20(2) and 19(3) of the ICCPR, as outlined in [Part III](#), and not on the protection of the religions or beliefs themselves.

### **The denial of historical events**

Various forms of “memory law” exist in numerous countries, prohibiting any expression that denies the occurrence of historical events, often connected to periods of severe persecution, genocide or other violations of international criminal law. Frequently, denialism is a direct attack on the dignity of victims and those associated with them; it often supports tenuous conspiracy theories of an atrocity being either orchestrated or fabricated by its victims, and justifies or perpetuates further discrimination. Memory laws therefore often have the stated purpose of preventing the reoccurrence of atrocities, including by combating this legacy of discrimination.

Under international human rights law, truth claims around historical events are not protected as such: importantly, international freedom of expression standards do not permit restrictions on the expression of opinions or ideas solely on the basis that those are “false” or “untrue,” even if these are deeply offensive.<sup>39</sup>

<sup>39</sup> HR Committee, General Comment No. 34, op. cit., para 49.

---

Truth claims are more reliably established through robust debate and inquiry, examining the evidence in support of competing claims and judging them on their own merit.

Where particular truth claims are elevated to dogma and the State is tasked with enforcing such claims, the value in and quality of open argument supported by evidence is underestimated in favour of permanently fixing one interpretation from one point of time. As the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (Special Rapporteur on FOE) has found:

*“by demanding that writers, journalists and citizens give only a version of events that is approved by the Government, States are enabled to subjugate freedom of expression to official versions of events.”<sup>40</sup>*

ARTICLE 19 believes that such memory laws or respective prohibitions are not necessary in a democratic society, but are in fact counterproductive. Prosecutions enable deniers to gain “martyr” or celebrity status; it provides them an opportunity, which some may deliberately seek, to market their ideas to a broader audience as simply anti-establishment or nonconformist, and misrepresent their prosecutions as evidence of the probity of their conspiracy theories.

Memory laws also raise serious concerns regarding legal certainty. Questions often arise over what constitutes an “established historical fact,” particularly in relation to laws specifying the denial of crimes that have narrow legal meanings or relate to complex factual circumstances that remain the subject of extensive historical or legal debate. This ambiguity becomes more pronounced where laws prohibit not only the denial of historical events, but also any “trivialising”, “minimising”, “justifying” or “glorifying” of those events, terms that are without definition under international human rights law and open to further abuse.

We acknowledge that the denial of historical events is often a vehicle for

<sup>40</sup> The 2012 Report of the Special Rapporteur on FOE, op.cit., para 55.

---

advocating hatred against victim-survivors and others associated with the victims of these crimes. While this raises concerns of intolerance and may legitimately be considered ‘hate speech’, prohibitions on such expression should be limited to only those acts that reach the threshold of **advocacy of discriminatory hatred that constitutes incitement to violence, hostility or discrimination**.<sup>41</sup> The protection of individual rights, as opposed to the defence of “truth claims”, must be clearly distinguished as the basis for any limitation on the right to freedom of expression.

### **Inciting terrorist acts and violent extremism**

Terrorism, and state responses to it, have raised various freedom of expression concerns in recent decades. Individuals have been targeted in terrorist acts for the exercise of their rights to freedom of expression, with attackers seeking to intimidate people into self-censorship and limit open debate, while States’ responses to terrorism have led to unjustifiable or disproportionate restrictions on fundamental rights, including freedom of expression.

In the context of States’ efforts to prevent terrorism, concepts such as “incitement to terrorism”, “violent extremism” and “radicalisation” are sometimes conflated with ‘hate speech’.

Under international law, States are obliged to prohibit **incitement to terrorist acts**.<sup>42</sup> However, being a restriction on freedom of expression to protect national security, these measures must comply with the three-part test set out in Article 19(3) of the ICCPR. The Johannesburg Principles provide that expression may be

<sup>41</sup> In its General Recommendation No. 35, the ICERD Committee brought its reasoning closer to that of the HR Committee on the issue of memory laws, by recommending “that public denials or attempts to justify crimes of genocide and crimes against humanity, as defined by international law, should be declared as offences punishable by law, provided that they clearly constitute incitement to racial violence or hatred”(para 15). General Comment No. 34, op. cit., states: “Laws that penalize the expression of opinions about historical facts are incompatible with the obligations that the Covenant imposes on States parties in relation to the respect for freedom of opinion and expression. The [ICCPR] does not permit general prohibition of expressions of an erroneous opinion or an incorrect interpretation of past events. Restrictions on the right of freedom of opinion should never be imposed and, with regard to freedom of expression, they should not go beyond what is permitted in para 3 or required under Article 20.” (para 49)

<sup>42</sup> UN Security Council Resolution 1624 (2005), para 1(a). “Terrorism” or “terrorist acts” has no universally agreed definition under international human rights law.

---

limited as a threat to national security only if the state can demonstrate that

1. the expression is intended to incite imminent violence;
2. it is likely to incite such violence; and
3. there is a direct and immediate connection between the expression and the likelihood or occurrence of such violence.<sup>43</sup>

Drawing on this test, the UN Special Rapporteur on countering terrorism has proposed a model definition for “incitement to terrorist offences”:

*It is an offence to intentionally and unlawfully distribute or otherwise make available a message to the public with the intent to incite the commission of a terrorist offence, where such conduct, whether or not expressly advocating terrorist offences, causes a danger that one or more such offences may be committed.*<sup>44</sup>

However, many states rely on “national security” justifications to restrict a much broader category of expression than this model recommends. These restrictions include prohibitions on the “justification,” “encouragement” or “glorification” of terrorist acts, or associated “extremism” and “radicalisation.” These require no proof of intent to incite violence, and do not insist on any causal connection between either the expression and the likelihood, or the expression and the

<sup>43</sup> [The Johannesburg Principles on National Security, Freedom of Expression and Access to Information](#), ARTICLE 19 (2006), Principle 6.

<sup>44</sup> Report of the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism, 22 December 2010, A/HRC/16/51, paras 30 – 31.

---

occurrence, of violence.

Such broad prohibitions are unjustified under international human rights standards.<sup>45</sup> They lack legal certainty, and may be applied arbitrarily to limit legitimate political debate; to censor minority or dissenting opinions on terrorist attacks and the efficacy or appropriateness of states' responses to them, or even commentary on broader issues in the public interest. Expression on these issues is often passionate, and frequently closely interrelated with issues around identity, and in this context 'hate speech' may also become more prevalent. However, while presenting concerns around tolerance, this speech will often not pose any real "national security" threat, and subjecting it to national security responses would not only be disproportionate but also discriminatory, stigmatising, and counter-productive.<sup>46</sup>

<sup>45</sup> See General Comment No. 34, op. cit., para. 46: "States parties should ensure that counter-terrorism measures are compatible with paragraph 3. Such offences as "encouragement of terrorism" and "extremist activity" as well as offences of "praising", "glorifying", or "justifying" terrorism, should be clearly defined to ensure that they do not lead to unnecessary or disproportionate interference with freedom of expression. Excessive restrictions on access to information must also be avoided. The media plays a crucial role in informing the public about acts of terrorism and its capacity to operate should not be unduly restricted. In this regard, journalists should not be penalized for carrying out their legitimate activities". At the regional level, the European Court in *Leroy v. France*, (App. No. 36109/03), 2 October 2008) found no violation of the right to freedom of expression where a cartoonist and magazine editor were prosecuted for depicting the 11 September attacks in New York with the caption "We have all dreamed it... Hamas did it." Given the political context in the Basque country where the magazine was published, the Court found that the reasons given by France, that the cartoon and caption were capable of stirring up violence and demonstrating a plausible impact on public order in the region, were "relevant and sufficient."

<sup>46</sup> In 2015, UN and regional freedom of expression mandates issued a joint declaration on "freedom of expression and responses to conflict situations", in which they emphasised that: "[c]ensorship is not an effective response to extremism, that open and critical debate is an important part of any strategy to address systematic attacks on freedom of expression and their underlying causes, and that overbroad criminalisation of expression can drive grievances underground and foster violence."

---

### Example



Following a bombing of a place of worship frequented by a majority religious group, a Facebook user not associated with the attack makes a public posting with discriminatory and flippant language to say the victims “deserved it.” The Facebook user has several hundred “friends”, and the post prompts heated responses on both sides; the user is not a politician or a public leader.

In several countries, overbroad counter-terrorism legislation, for example on “justifying” or “glorifying” terrorism, could be wrongly used to prosecute this Facebook user, even though the expression neither intends nor is likely to incite a terrorist attack. Though deeply offensive and ‘hate speech’, it is unlikely in consideration of the context to meet the threshold for restrictions on ‘hate speech’ either. Though Facebook may decide to remove the comment under their terms and conditions, they should not be obliged by law to do so.

ARTICLE 19 therefore recommend that states prohibit **incitement to terrorist acts**, but also that they clearly distinguish this from ‘hate speech’ more generally, and ensure that necessary elements of such a prohibition include both (i) intent to incite terrorist acts, and (ii) the likelihood or occurrence of an attack as a consequence of the expression. Laws that do not meet these requirements should be repealed.

### Protection of “the state” and public officials

States regularly exploit the label ‘hate speech’ to discredit, or even prohibit, expression that is critical of the State, the symbols of the State (such as flags and emblems), or power-holders. While some domestic laws expressly prohibit “insult” or “denigration” of abstract concepts, or particular offices of state in this context, while others, more ambiguously, prohibit “sedition” or expression that is against “national unity” or “national harmony”.

---

International standards do not permit restrictions on the right to freedom of expression which are made in order to protect “the state” or its symbols from insult or criticism. These entities cannot be the target of ‘hate speech’, because they are not people and are therefore not rights-holders.<sup>47</sup> For natural persons associated with the state, such as heads of state or other public officials, this status is not a “protected characteristic” on which discrimination claims, or the characterisation of “hate speech,” can be based. Indeed, public officials are legitimately subject to criticism and political opposition,<sup>48</sup> and are expected to display a higher degree of tolerance toward criticism than other persons.<sup>49</sup>

Though freedom of expression can be limited to protect “national security” or “public order”, these bases cannot be exploited to suppress criticism or dissent, to shield those in power from embarrassment, or to conceal wrongdoing.<sup>50</sup>

As the Johannesburg Principles state:

*No one may be punished for criticizing or insulting the nation, the state or its symbols, the government, its agencies, or public officials, or a foreign nation, state or its symbols, government, agency or public official unless the criticism or insult was intended and likely to incite imminent violence.*<sup>51</sup>

<sup>47</sup> General Comment No. 34, op. cit., para 38.

<sup>48</sup> Ibid. See also the *USA Supreme Court, US v. Eichman*, 496 U.S. 310 (1990), in which the prosecution of a person for burning a United States flag was found unconstitutional.

<sup>49</sup> See, e.g., African Commission on Human and Peoples' Rights, Resolution on the Adoption of the Declaration of Principles on Freedom of Expression, October 2002, Principle XII “protecting reputations”; European Court, *Otegi Mondragon v. Spain*, App No. 2034/07, 15 March 2011.

<sup>50</sup> Johannesburg Principles, op. cit., Principle 2(a): “a restriction sought to be justified on the ground of national security is not legitimate if its genuine purpose or demonstrable effect is to protect interests unrelated to national security, including, for example, to protect a government from embarrassment or exposure of wrongdoing, or to conceal information about the functioning of its public institutions, or to entrench a particular ideology, or to suppress industrial unrest.” See also General Comment No. 34, op. cit., para 38.

<sup>51</sup> Johannesburg Principles, op. cit., Principle 7(b).



---

### Example



During an annual celebration of the nation's armed forces, opponents of recent military campaigns stage a protest at which they deface various "national symbols", including a flag and a portrait of a historical figure associated with the nation's founding.

Though the conduct of the protesters may be offensive to some, it is political expression which is intended to convey a political message. The state's symbols, and the military as an institution, are not a person and are thus not protected as targets of 'hate speech'; also, individuals in the military, as public officials, should be expected to tolerate criticism of their conduct, and are nevertheless not personally targeted in this protest.

### Defamation

The concept of "defamation," "desacato", "slander" or "libel" is sometimes confused with 'hate speech'.

Defamation laws generally aim to protect the reputation of individuals from false statements of fact, which cause damage to their reputation. Legal actions for defamation do not require an individual to show any "advocacy of hatred", and should be distinguished from 'hate speech'. In addition, for the purposes of defamation legal actions, "groups" of individuals do not have an individual reputation, or the basis to claim legal personhood for the purpose of litigation.

ARTICLE 19 therefore generally finds laws that allow "groups" to file defamation cases problematic, and recommends their repeal. Instead, harm to a group of persons as a consequence of 'hate speech' should be litigated or prosecuted under legitimate 'hate speech' provisions.

# Part II: Responding to 'hate speech'



---

In this section, ARTICLE 19 outlines a variety of measures to respond to ‘hate speech’ and challenge the prejudice and intolerance which such expression is symptomatic of.

We propose that responses to ‘hate speech’ be premised on three complementary areas of action:

1. States must create an **enabling environment** for the exercise of the right to freedom of expression and protect the right to equality and non-discrimination;
2. States must enact a **range of positive policy measures** to promote freedom of expression and equality;
3. Other stakeholders, including civil society, the media and private businesses, should be encouraged to undertake **voluntary initiatives** to tackle the root causes of prejudice and intolerance, to contest and challenge “hate speech.”

## 1) Creating an enabling environment for the rights to freedom of expression and equality

Creating an enabling environment for the right to freedom of expression and equality is not only an obligation of States under international human rights law,<sup>52</sup> but is an essential condition for ensuring that opportunities to expose and counter ‘hate speech’ are maximised. States must ensure that they ratify international and regional human rights instruments that protect the right to freedom of expression, as well as the right to equality, and ensure that these rights are fully incorporated into their national legal frameworks.

### a) An enabling environment for the right to freedom expression

In addition to guaranteeing this right in domestic constitutions or their equivalent, States must ensure that all laws and policies comply with international freedom of expression standards. In particular, any restrictions on freedom of expression to be provided by law, narrowly defined to serve a legitimate interest, and necessary in a democratic society to protect that interest.

<sup>52</sup> See [Annex 1](#)

---

i) Any strategy to enable effective responses to ‘hate speech’ must also include **the repeal or reform of laws** that unduly limit the right to freedom of expression, in particular those that target or disproportionately affect minority or marginalised groups. ARTICLE 19 recommends that all States, where applicable, **repeal:**

- **All forms of blasphemy laws;**<sup>53</sup>
- **All laws that protect abstract notions of “nationhood” or national unity,** including provisions that protect the State or its institutions or symbols from criticism or ridicule (such as “sedition”);
- **All laws that criminalise defamation,** including laws that specifically protect public officials and heads of state from insult or ridicule (such as “desacato”);<sup>54</sup>
- **All laws that advance over-broad concepts of “public morals” or that protect so-called “traditional values”,** where those laws discriminate on the grounds of sex, gender, gender identity, or sexual orientation;<sup>55</sup>
- **Laws that require authorisation to protest, or that ban spontaneous or counter-protests,** since they inhibit individuals’ ability to effectively and collectively respond peacefully to violence and intolerance, including to incidents of ‘hate speech’;<sup>56</sup>
- **Laws that impose discriminatory, unnecessary or disproportionate barriers to the freedom of association,** since such provisions, in particular those restricting access to resources, inhibit the ability of civil society organisations to actively monitor and respond to discrimination and violence;<sup>57</sup>

<sup>53</sup> See the Rabat Plan of Action, op. cit., para 19, and the Camden Principles, op.cit., Principle 12.3.

<sup>54</sup> See, for example Defining Defamation, ARTICLE 19, London 1999.

<sup>55</sup> See [Traditional values? Attempts to censor sexuality: Homosexual Propaganda Bans, Freedom of Expression and Equality](#), ARTICLE 19, 2013.

<sup>56</sup> See ARTICLE 19, [Right to Protest Principles](#).

<sup>57</sup> See, for example, UN Special Rapporteur on Freedom of Assembly and Association, [Your Rights on One Page](#): Factsheet series.

- 
- **Broad counter-terrorism or “extremism” laws**, to ensure any restrictions on freedom of expression are narrowly tailored to protect genuine national security interests, and to guard against discrimination against or profiling of minority groups.<sup>58</sup>

ii) States must also make concerted efforts to **end impunity for attacks on independent and critical voices**. In many societies, speaking out and organising actions against intolerance and discriminatory violence can raise significant security concerns for individuals and their colleagues and families. Threats and other forms of violence against individuals exercising their rights to freedom of expression, and impunity for such crimes, is a pressing concern, and one which has a profound chilling effect for a whole society.

States must ensure that any attacks against individuals exercising their rights to freedom of expression and unequivocally condemned, and public officials must avoid statements that may encourage or indicate support for such attacks.

In particular, States should:

- Put in place special measures of protection for individuals who are likely to be targeted for what they say where this is a recurring problem;
- Ensure that crimes against freedom of expression are subject to independent, speedy and effective investigations and prosecutions; and,
- Ensure that victims of crimes against freedom of expression have access to appropriate remedies.<sup>59</sup>

<sup>58</sup> See the Johannesburg Principles, op.cit.

<sup>59</sup> See, e.g. [Joint Declaration of international freedom of expression mandates on Crimes Against Freedom of Expression](#), June 2012.

---

**iii) States must also ensure transparency in the conduct of public affairs and guarantee the right of all people to access to information.** Cultures of official secrecy, in which the right of access to information is denied, enable powerful individuals to scapegoat minorities or marginalised groups to deflect attention from their own wrongdoing or political failings. In contrast, a culture of openness and the free flow of information makes such attempts to manipulate group identities a less effective political device; it also makes publically available information that can be used to counter such attempts at division. Transparency and accountability can also build faith and trust in public institutions among all sectors of society, and therefore make individuals and groups less susceptible to calls for ‘vigilante’ forms of violence.

**iv) States must also ensure that the right to freedom of expression is fully protected in relation to digital technologies.** Digital technologies are a crucial medium for all people, but in particular for individuals belonging to minority and marginalised groups, for learning about, developing, and seeking support on identities and related issues, to build communities with others, and to speak out on issues of concern, including against intolerance and ‘hate speech’.

However, due to the volume of content, some of which is unlawful or harmful, transmitted through the Internet, there is increasing pressure on States and private companies to increase control over online content. This includes calls to prohibit or facilitate the undermining of anonymity for Internet users, and proposals for intermediaries to more proactively monitor and remove content and/or making intermediaries liable for failing to do so. These measures may threaten the nature of the Internet as an open and public forum for the robust exchange of opinions and ideas, and may also limit the Internet as a creative space for exposing, contesting, and countering ‘hate speech’. In this regard, ARTICLE 19 recommends that States:

- Protect the right to anonymity online as an essential component of the right to freedom of expression, and encourage private businesses to ensure anonymity as a real option to users of their services;<sup>60</sup>

<sup>60</sup> See, e.g. ARTICLE 19, [Right to Online Anonymity](#), June 2015.

- 
- Ensure that intermediaries are not made liable for content created by third parties, and that content is only required to be restricted on the order of a judicial authority;<sup>61</sup> and
  - Provide Internet-users with effective remedies against private parties where they have unduly interfered with users' human rights.

## **b) Ensuring the full protection of the right to equality and non-discrimination**

In addition to guaranteeing the right to equality and non-discrimination in domestic constitutions or their equivalent, States must ensure that all domestic laws guarantee equality before the law and equal protection of the law, and must guarantee against discrimination on all grounds recognised under international human rights law.<sup>62</sup>

It is often in the absence of robust and effective anti-discrimination frameworks that the individuals most impacted by 'hate speech' have limited options for redress. With few alternatives, recourse is sometimes made to the criminal law, which frequently fails to provide an effective forum for resolving incidents of discrimination and may be counter-productive.

i) There are two prongs of action which ARTICLE 19 recommends to ensure that the right to equality and non-discrimination are fully protected in practice. States should

- **Repeal all laws and policies** that formally or informally institutionalise discrimination and exclusion on any of the protected grounds recognised under international human rights law. Their mere existence creates environments in which discrimination is ignored or even tacitly encouraged.

<sup>61</sup> See [Manila Principles on Intermediary Liability](#); this does not presuppose that intermediaries may choose to intervene in the moderation of third party content, as explored below.

<sup>62</sup> For more on which protected characteristics are recognised under international human rights law, see [Annex I](#).

- 
- **Enact or strengthen anti-discrimination legislation**, and in particular ensure that such laws, at minimum:
    - Protect against **direct discrimination**, i.e. the unfavourable treatment of a person, when compared to how others would be treated in a comparable situation, due to a protected characteristic;
    - Protect against **indirect discrimination**, i.e. where a neutral rule, criterion, or practice affects a group defined by a protected characteristic in a significantly more negative way than it would others, by comparison, in a similar situation;
    - Enable temporary **special measures** to be undertaken to tackle indirect discrimination and ensure substantive equality, for as long as such measures are necessary and proportionate;<sup>63</sup>
    - Include the **broadest scope of protected characteristics** recognised under international human rights law as potential grounds for discriminatory hatred and action;
    - Apply to a **broad range of specific contexts**, including: employment; social security and access to welfare benefits; education; the provisions of goods and services; housing; access to justice; private and family life, including marriage; political participation, including freedom of expression, association and assembly; and law enforcement;
    - Include **defences** for cases in which differential treatment is based objectively upon a legitimate aim, and the means of achieving that aim are appropriate and necessary;
    - Provide for a **range of remedies** primarily in **civil and administrative law**, as well as non-legal mechanisms for redress such as mediation and alternative dispute resolution, which may be provided through equality institutions.

<sup>63</sup> See, e.g. CERD Committee General Recommendation No. 32 on The Meaning and Scope of Special Measures in the ICERD, CERD/C/GC/32, 24 September 2009, paras 21-26; and CEDAW Committee, General Recommendation No. 25 on Article 4(1) of the CEDAW, A/59/38 (supp), 18 March 2004, para 22.



---

ii) In addition, States must ensure that their **criminal law frameworks** fully recognise, and provide proportionate sanctions, for “hate crimes”, and include in these laws the broadest range of protected characteristics. The effective implementation of these laws should be monitored to ensure that victim/survivors are encouraged to report such crimes and to ensure the collection of official statistics on the number of reported incidents and successful prosecutions, disaggregated according to the type of bias at issue.

iii) States should establish or strengthen the role of **independent equality institutions**, or expand the mandate of **national human rights institutions** (NHRIs)<sup>64</sup> to promote and protect the right to equality and non-discrimination, including with respect to the right to freedom of expression.

Such institutions should be properly resourced with mandates, as appropriate, to:

- Develop data collection mechanisms on the extent and impact of discrimination in priority areas to inform the development, monitoring and evaluation of laws and policies, and to promote empirical and other research on the subject;
- Assist legislatures and the government with the development of laws and policies that comply with States’ international human rights obligations, including in relation to freedom of expression and non-discrimination, encouraging the full and effective participation of civil society in these processes;

<sup>64</sup> All NHRI should be established in accordance with the [Principles relating to the Status of National Institutions \(The Paris Principles\)](#), UN GA Resolution 48/134 of 20 December 1993.

- 
- Receive complaints regarding discrimination, and where appropriate provide alternative/voluntary dispute resolution mechanisms;
  - Complement and provide information to governmental early warning mechanisms or focal-points that monitor tensions within or between different communities; and
  - Encourage and, where appropriate, support different mechanisms for inter-communal interaction and dialogue.

It is important that NHRIs, or equality bodies, do not operate in isolation: they should be empowered to build partnerships across public sector agencies, and where appropriate with private actors and civil society, to tackle the root causes of discrimination. In this regard, they should play an integral role in **developing and implementing national action plans** to tackle the root causes of discrimination, informed by the measures outlined in this toolkit as well as in Human Rights Council resolution 16/18, and the Rabat Plan of Action.

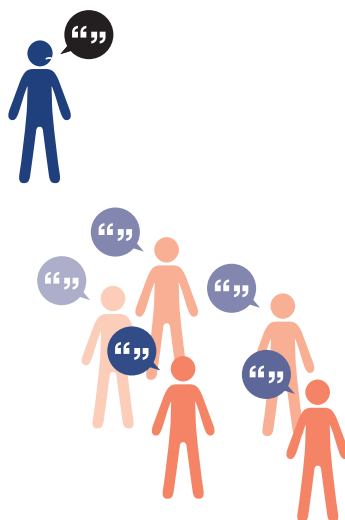
---

## 2) Positive policy measures by States

ARTICLE 19 recommends that States should, primarily, employ a range of positive non-legal measures across all aspects of public life, to tackle prejudice and discrimination, and to respond to “hate speech.” These measures must be based on, and supported by, a firm commitment by public officials to respect the human rights, and a commitment to promote dialogue and foster participation from all quarters of society.

### Recognising and speaking out against intolerance

Public officials, including politicians, have a key role to play in recognising and promptly speaking out against intolerance and discrimination, including instances of “hate speech.” This requires recognising and rejecting the conduct itself, as well as the prejudice of which it is symptomatic, expressing sympathy and support to the targeted individuals or groups, and framing such incidents as harmful to the whole of society. These interventions are particularly important where inter-communal tensions are high, or are susceptible to being escalated, and where political stakes are also high, e.g. in the run-up to elections.



Early and effective intervention from public officials can play an important preventative role to guard against tensions escalating, deterring others from engaging in similar conduct. They can also play an important role in opening space for counter-speech by other actors, in particular those targeted by “hate speech,” as well as sympathetic allies, including the “silent majority” for whom proponents of ‘hate speech’ often claim to speak. Public officials can thus play a key role in instigating or encouraging broader dialogue which might counter intolerance and discrimination.

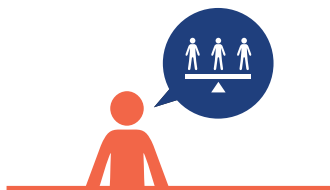
---

Further research is required to examine the circumstances in which counter-speech by public officials to instances of intolerance and discrimination is most effective. Condemnations of 'hate speech' may be insufficient if public officials fail to substantively and persuasively engage with the underlying anxieties and misperceptions that render parts of the public susceptible to such messages. Responses by public officials should therefore be nuanced, and go beyond denunciation to provide persuasive counter-narratives based on facts that appeal to and, where necessary, challenge the concerns or anxieties in the public. However, public officials should avoid responding to incidents of 'hate speech' where to do so would give undue attention to the positions of fringe individuals or groups that are not influential to public discourse.

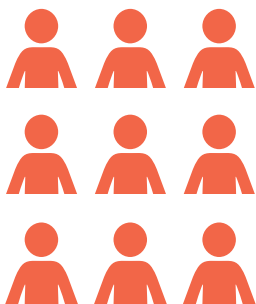
Importantly, public officials should be instructed on the importance of **avoiding statements that might promote discrimination or undermine equality**, and must understand the dangers of trivialising violence or discrimination, including in the form of 'hate speech', as well as the possibility of silence in the face of such challenges equating to tacit endorsement. In this regard, public bodies should have in place clear rules governing the conduct of individuals speaking in their capacity as public officials. Ethical codes and "no discrimination" policies adopted by political parties should also be considered as positive policy measures.

---

## Equality training



Building trust in the capacity of public institutions to tackle intolerance and discrimination requires public officials to be fully aware of the nature and impact of discrimination on different individuals and groups, and to be fully committed to promoting equality.



States should provide trainings for public officials, public figures and state institutions on the rights to equality and non-discrimination, particularly where discrimination is institutionalised, or has historically gone unchallenged. Priority contexts should include schools and other educational settings, the armed forces, the police, the judiciary, the medical profession, legal services, political associations or religious institutions.

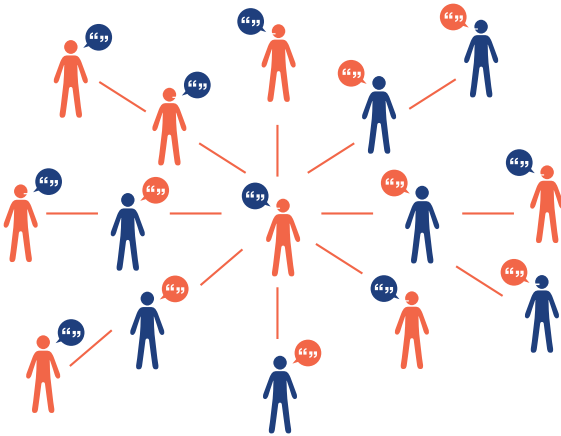
Equality training may form part of a broad range of measures designed to tackle institutionalised discrimination, and should be clearly communicated to the public to demonstrate where efforts are underway to build trust in institutions.

---

## Public policy for pluralism and equality in the media

All States should ensure that a public framework and regulatory framework for diverse **and pluralistic media** is in place, which promotes pluralism and equality, in accordance with the following:

- The framework should respect the fundamental principle that any regulation of the media should only be undertaken by bodies which are independent of the government, are publicly accountable, which operate transparently; and
- The framework should promote the right of different communities to freely access and use media and information and communications technologies for the production and circulation of their own content, as well as for the reception of content produced by others, regardless of frontiers.<sup>65</sup>



<sup>65</sup> Camden Principles, op. cit., Principle 5.

---

This framework should be implemented, among others, through the following measures:

- Promoting universal and affordable access to the means of communication and reception of media services, including telephones, the Internet and electricity;
- Eliminating discrimination in relation to the right to establish newspapers, radio and television outlets, and other communications systems;
- Allocating sufficient 'space' to broadcasting uses on different communications platforms to ensure that, as a whole, the public is able to receive a range of diverse broadcasting services;
- Making an equitable allocation of resources, including broadcasting frequencies, among public service, commercial and community media, so that together they represent the full range of cultures, communities and opinions in society.
- Requiring the governing bodies of media regulators broadly to reflect society as a whole;
- Putting in place effective measures to prevent undue concentration of media ownership;
- Providing public support, whether financial or in other forms, through an independent and transparent process, and based on objective criteria, to promote the provision of reliable, pluralist and timely information for all, and the production of content which makes an important contribution to diversity or which promote dialogue among different communities.

- 
- Repealing any restrictions on the use of minority languages that have the effect of discouraging or preventing media specifically addressed to different communities;
  - Making diversity, including in terms of media targeting different communities, one of the criteria for assessing broadcasting license applications; and
  - Ensuring that disadvantaged and excluded groups have equitable access to media resources including training opportunities.

Public service values in the media should be protected and enhanced by transforming State- or government-controlled media systems, by strengthening existing public service broadcasting networks, and by ensuring adequate funding for public service media, so as to ensure pluralism, freedom of expression, and equality in an ever-changing media landscape.



---

## Public education and information campaigns

Public information and education campaigns are essential in combating negative stereotypes of, and discrimination against individuals on the basis of their protected characteristics. Such campaigns, based on accurate information, can dispel popular myths and misconceptions, and equip individuals with greater confidence to identify and challenge manifestations of intolerance in their day-to-day interactions.



In particular, public information and education campaigns should be integrated into primary, secondary and tertiary **education**, complemented by concrete **anti-bullying policies**, including the provision of support services for victims of bullying, including peer-led initiatives. In particular, attention should be paid to ensuring diversity in school materials and the avoidance of school textbooks containing stereotypes and prejudices against particular groups.

---

## Transformative justice

In the aftermath of large scale human rights violations, including widespread and systematic discrimination, mechanisms for guaranteeing **truth, justice, reconciliation and reparations** have proven to be a positive extra-judicial means for establishing an authoritative and shared interpretation of the “truth” behind historical events, providing a basis for reconciliation in fractured societies.

By contrast, where open and inclusive discussions and critical debate on historical events are suppressed in favour of unilaterally declared or legally enforced “truths”, underlying resentment and distrust between different communities can endure and pose a danger of conflicts reoccurring.

States can play an important role in officially and publicly recognising the impact and legacy of incidents, or systemic problems of discrimination or violence, as well as **symbolically marking certain events or times** to overcome and ensure redress for respective incidents. This is often done by dedicating public sites, such as monuments, museums and in community meeting areas, and broader efforts to help people come to terms with and comprehend what has happened.



---

### 3) Voluntary initiatives by other stake-holders

Stakeholders besides the State can play an important part in promoting equality and non-discrimination, and the right to freedom of expression. Many regard this as a central part of their corporate and/or social responsibilities.

#### **Civil society initiatives**

Civil society plays a critical role in advancing the protection and promotion of human rights – even where this may not be a central part of their mandate. Their activities can be central in responding to ‘hate speech’ as they can provide the space for formal and informal interactions between people of similar or diverse backgrounds, and platforms from which individuals can exercise freedom expression and tackle inequality and discrimination. At the local, national, regional and international levels, civil society initiatives are among the most innovative and effective for monitoring and responding to incidents of intolerance and violence, as well as for countering “hate speech.”

Civil society initiatives are often designed and implemented by the individuals and communities most affected by discrimination and violence, and provide unique possibilities for communicating positive messages to and educating the public, as well as monitoring the nature and impact of discrimination. Ensuring a safe and enabling environment for civil society to operate is therefore also crucial.

---

## Mobilisation of influential actors and institutional alliances

Enhancing public understanding of discrimination and its impact requires the fostering of dialogue and engagement between government, civil society, and society at large. Key actors should attempt to forge **alliances to collaborate** on tackling manifestations of intolerance and prejudice in society – in particular seeking the support of non-government organisations, police, policymakers, equality bodies, artists, religious institutions, and international organisations to collaborate on tackling manifestations of intolerance and prejudice in society.

## Role of an independent and pluralistic media

Any policy measures to tackle ‘hate speech’ which are directed at the media should respect the fundamental principle that any form of media regulation should be undertaken by bodies independent of political influence, which are publicly accountable and operate transparently. Editorial independence and media plurality should not be compromised, as both are essential to the functioning of a democratic society.

In respect of broadcast media, any regulatory framework should promote the rights of minority and marginalised groups to freely access and use media and information and communications technologies for the production and circulation of their own content and for the reception of content produced by others, regardless of frontiers.<sup>66</sup>

All forms of mass media should recognise that they have a **moral and social responsibility** to promote equality and non-discrimination for individuals with the broadest possible range of protected characteristics. In respect of their own constitutions, mass media entities should take steps to:<sup>67</sup>

- Ensure that their workforces are diverse and representative of society as a whole;
- Address as far as possible issues of concern to all groups in society;

<sup>66</sup> Principle 5 of the Camden Principles provides more guidance on creating a public policy framework for pluralism and equality.

<sup>67</sup> Based on Principle 6 of the Camden Principles, op. cit. See also, Agnes Callamard, The Contribution of Media and Information to an Effective Strategy of Prevention to Atrocity Crimes, UN Headquarters, 11 December 2014.

- 
- Seek a multiplicity of sources and voices within different communities, rather than representing communities as homogenous entities;
  - Adhere to high standards of information provision that meet recognised professional and ethical standards;
  - Promulgate and effectively implement professional codes of conduct for the media and journalists that reflect equality principles.

To proactively combat discrimination, media entities should consider:

- Taking care to report in context, and in a factual and sensitive manner;
- Ensuring that acts of discrimination are brought to the attention of the public;
- Being alert to the danger of discrimination or negative stereotypes of individuals and groups being furthered by the media;
- Avoiding unnecessary references to race, religion, gender, sexual orientation, gender identity and other group characteristics that may promote intolerance;
- Raising awareness of the harm caused by discrimination and negative stereotyping;
- Reporting on different groups or communities and giving their members an opportunity to speak and to be heard in a way that promotes a better understanding of them, while at the same time reflecting the perspectives of those groups or communities;
- Professional development programmes that raise awareness about the role the media can play in promoting equality and the need to avoid negative stereotypes.<sup>68</sup>

Public service broadcasters should be under an obligation to avoid negative stereotypes of individuals and groups, and their mandate should require them to

<sup>68</sup> Based on Principle 9 of the Camden Principles, op. cit. See also, ARTICLE 19, [Getting the facts right: reporting ethnicity and religion](#), May 2012.

---

promote inter-group understanding and to foster a better understanding of different communities and the issues they face.

In terms of the remedies available through self-regulatory systems, a right to correction or reply should be guaranteed to protect the rights to freedom of expression and equality. This should enable individuals to demand that a mass media outlet publish or broadcast a correction in cases where that media outlet has published or broadcast incorrect information.

### **Role of internet intermediaries**

Increasingly, attention is being paid to the role of Internet intermediaries in identifying and responding to 'hate speech'.

Internet intermediaries, including web hosting companies, Internet service providers (ISPs), search engines and social media platforms, play a crucial role in enabling people to access information via the Internet. For the most part, these are privately owned companies operating across jurisdictions. Though they do not primarily engage in the creation or modification of content, but rather facilitating the communication of users, they are increasingly being called on to engage in the moderation of the content. In some cases, this involves the direct intervention of State regulation, or the adoption of civil liability regimes under which intermediaries must monitor and remove allegedly unlawful content. These factors influence how intermediaries engage in content moderation.

---

Intermediaries are also undertaking voluntary initiatives to set conditions for the use of their services, reserving to themselves in certain circumstances the role of moderator. These “terms and conditions,” sometimes framed to sound less contractual as “community standards” or “guidelines,” vary in the types of expression that they limit, though many include prohibitions on ‘hate speech’ or variations thereon. Approaches to moderation vary significantly, as do degrees of transparency around moderation processes and the availability of internal mechanisms for appealing a moderation decision.

Numerous factors seem to be incentivising a greater tendency towards removal of ‘hate speech’ content by intermediaries, including:

- Pressure to make intermediaries’ policies conform to the national laws of the several jurisdictions in which they wish to maintain or expand their operations, many of which do not comply with international freedom of expression standards, and often leading to the application of fragmented standards or a lowest common denominator standard;
- Pressure to cooperate with States and sometimes the public over ‘hate speech’ concerns, showing enthusiasm for content moderation through self-regulation to avoid the imposition of coercive and more costly forms of regulation; and
- Responses to commercial pressures from advertisers or other revenue sources who do not wish to be associated with alleged “hate speech.”

There are a number of concerns regarding the role of intermediaries in moderating content, including “hate speech,” in particular:

- **Inadequate protections for freedom of expression:** the terms and conditions of many intermediaries tend towards limiting a greater breadth of expression than States are permitted to restrict under international human rights law. The scope of so-called “private censorship” is considerable, and raises questions in respect of moral and social responsibilities to promote and protect all human rights. Initiatives to encourage intermediaries to take these responsibilities seriously often overlook concerns over the right to freedom of expression. Furthermore, there are serious doubts over the suitability of businesses, primarily motivated by profit, to objectively judge competing rights and interests;

- 
- **A lack of transparency and accountability** in the decision-making process of intermediaries when removing content, including how content is flagged and removed (e.g. if content moderation is automated, and if not what training and support exists for moderators). Many intermediaries do not publish information on removals made on their own initiative, as opposed to removals in response to requests from States or other businesses. This creates serious barriers to any analysis or evaluation of intermediaries' conduct in relation to "private censorship;"
  - **A lack of procedural safeguards, and a lack of access to an effective remedy** in the removal of content, or the imposition of other sanctions by intermediaries. There are concerns that States may take advantage of reporting mechanisms or their influence over private companies to request content removals that they cannot legally compel themselves, or to circumvent procedural safeguards that limit any powers of compulsion they have in this respect. The delegation of the policing of content from the State to intermediaries denies users any opportunity to contest or defend against the sanctions employed against them.

Though there have been numerous innovations in recent years to empower users to report 'hate speech' content for removal, either because it is perceived to be unlawful or against an intermediary's terms of service, there have not been comparable advances to empower users to guard against the unfair or unjustified removal of content. Indeed, most intermediaries do not seem to provide users with notice or reasons for the content removal. Beyond the removal, other sanctions imposed by intermediaries, such as the suspension or blocking of accounts, are rarely accompanied with notice or opportunities for appeal or remedy.



- 
- **Disproportionate impact of removals on users with minority or dissenting views:** since many intermediaries’ moderation models rely upon user reports, minority or dissenting views may attract a greater number of reports and therefore be more vulnerable to removal. The same Internet-users who find themselves on the receiving end of ‘hate speech’ may therefore also find themselves deliberately targeted through reporting tools and disproportionately impacted by content removal and sanctions against their accounts. This reflects an unfortunate reality that many Internet users that report content are unable, or are perhaps disinterested, to distinguish unlawful or harmful content, from content they simply want removed on the basis of their own prejudices.

ARTICLE 19 encourages intermediaries to take seriously their social and moral responsibility to promote and protect human rights, in accordance with the **Ruggie Principles**.<sup>69</sup>

In this regard, and as a matter of voluntary self-regulation, we encourage intermediaries to:

- Include, in their terms and conditions, a strongly-stated commitment to the promotion and protection of human rights, including the rights to freedom of expression and the right to equality and non-discrimination;
- Ensure users a right to anonymity as default, not requiring the use of “real names” or the submission of identity documents to open or maintain a social media account;
- Ensure that any content-based restrictions are specified in the terms and conditions, in a clear and accessible format so that users are able to understand the types of content that may be subject to restriction;

<sup>69</sup> International Federation for Human Rights, UN Human Rights Council adopts Guiding Principles on business conduct, yet victims still waiting for effective remedies, 17 June 2011.

- 
- Ensure that any process for applying sanctions to users, including content removal or account suspensions, are clearly detailed within terms and conditions;
  - Explore mechanisms for empowering users to respond to and contest “hate speech,” rather than the primary response being removal by the intermediary, this could include increasing awareness among users of the importance of the rights to freedom of expression, and to equality and non-discrimination;
  - Require that users given sufficient information when submitting a complaint against specific content, including: **(i)** the content at issue; **(ii)** the reasons for seeking content removal; **(iii)** contain details of the complainant; and, **(iv)** a declaration of good faith;
  - Ensure that in relation to “hate speech,” terms and conditions impose a high threshold for restrictions, which reflects as far as possible the standards set out in [Part 3](#) of this toolkit;
  - Ensure proportionality in the implementation of any sanctions against users who violate terms and conditions, taking into consideration the harm of the alleged infringement, and the user’s previous conduct on the platform. Suspension from platforms should be a measure of last resort;
  - Ensure users are given sufficiently-detailed prior-notice of complaints made against their content, with the opportunity to appeal or contest the complaint prior to the imposition of sanctions. In the absence of prior-notice, intermediaries should, at a minimum, give post-facto notice of content removal, including reasons for the content removal and availability internal mechanisms to appeal that decision.

---

## Role of meaningful inter-group dialogue

A lack of meaningful inter-group communication, and the isolation and insularity of which this is a symptom, is often identified as a significant contributing factor to inter-group tensions, where ‘hate speech’ is more prevalent, and incitement to violence, hostility or discrimination more likely.

Sustained and effective dialogue between distinct groups, in particular between communities of different religions or beliefs, can serve an effective preventative measure, by achieving the alleviation of tensions or suspicion between groups. This may be particularly effective in contexts where there is a history of inter-group tensions escalating into incitement of, or actual, incidents of violence and discrimination. However, in order to be effective, dialogue must provide the spaces for a genuine, rather than symbolic, exchange of views, and enable the discussion of differences and disagreements. Dialogue must also be inclusive, allowing for community representation beyond “traditional” leaders.

Furthermore, informal exchanges between communities, detached from intergroup dialogue, for example in the context of sport or cultural exchanges, or designed to address practical issues of common concern, can also prove to be important trust- and relationship-building exercises. The impact of inter-group dialogue and communication initiatives can be enhanced where they receive public support from government.

Outside of the context of formal or informal “dialogues”, representative of different communities, in particular religious leaders<sup>70</sup> and other community leaders, should be empowered to speak out in response to intolerance and discrimination. This is particularly important where proponents of intolerance and discrimination portray themselves as representative of, or acting on behalf of, specific communities or interest groups. Religious and community leaders are well placed not only to refute these claims of representation, but also to substantively engage with and challenge an individual’s position, and thus offer a persuasive counter-narrative.

<sup>70</sup> International Federation for Human Rights, UN Human Rights Council adopts Guiding Principles on business conduct, yet victims still waiting for effective remedies, 17 June 2011.

# Part III: Restricting 'hate speech'



---

In this section, we outline specific requirements which must be satisfied when restricting speech falling into the most 'severe' categories of 'hate speech' under the 'hate speech' triangle outlined in previous sections.

Under international human rights standards, the right to freedom of expression is not absolute, and may exceptionally be subject to restrictions provided that those meet a strict three-part test, according to Article 19 (3) of the ICCPR. Hence, all state action targeting 'hate speech' must:

1. **Be provided by law**; any law or regulation must be formulated with sufficient precision to enable individuals to regulate their conduct accordingly;
2. **Pursue a legitimate aim**, exclusively: respect of the rights or reputations of others; or the protection of national security or of public order (*ordre public*), or of public health or morals;
3. **Be necessary in a democratic society**, requiring the State to demonstrate in a specific and individualised manner the precise nature of the threat, and the necessity and proportionality of the specific action taken, in particular by establishing a direct and immediate connection between the expression and the threat.

Any prohibition on 'hate speech' must satisfy each element of this three-part test. This includes where international law separately requires a state to prohibit particular forms of "hate speech," namely **direct and public incitement to genocide**) and any advocacy of discriminatory hatred that constitutes **incitement to discrimination, hostility or violence**.

It is also important to keep in mind that not all 'hate speech' may legitimately be subject to restriction, and much will be protected under the right to freedom of expression, even if it is deeply offensive (see [Part I](#)).

---

## Direct and public incitement to genocide

The Genocide Convention, Article 3(c), requires that States prohibit and punish as a criminal offence any “direct and public incitement to genocide”, in addition to acts of genocide themselves. As noted above, this obligation is replicated in the Statute of the ICC and the statutes for the ad hoc UN international criminal tribunals.

Genocide is defined as one of five acts,<sup>71</sup> not limited to killings, “committed with intent to destroy, in whole or part, a national, ethnical, racial or religious group, as such.” The protected characteristics here are narrower than the broader range recognised under international human rights law.

Convictions for **direct and public incitement to genocide** require proof of several key elements. These are important as they distinguish incitement to genocide from other, less severe, forms of “hate speech.” At the international level, the ad hoc UN international criminal tribunals have elaborated the following elements for the offence, all of which must be satisfied in order to secure a conviction:

- **Public:** the expression inciting others to commit acts of genocide must be “public”, indicating there must be a communication in a public place, or to the public or a section of the public, for example through mass media and digital technologies;<sup>72</sup>
- **Direct:** the expression must be “direct”, i.e. the communication must be sufficiently specific as a call for action, showing a close relationship between the expression and the danger of an act of genocide occurring. However, direct does not mean explicit, as implicit expression may also directly incite genocide if in its linguistic and cultural context it is sufficiently clear to its audience;<sup>73</sup>

<sup>71</sup> Article 2 of the Genocide Convention specifies five such acts: (a) killing members of the group; (b) causing serious bodily or mental harm to members of the group; (c) deliberately inflicting on the group conditions of life, calculated to bring about its physical destruction in whole or in part; (d) imposing measures intended to prevent births within the group; (e) forcibly transferring children of the group to another group.

<sup>72</sup> ICTR, *Prosecutor v. Nahimana, Barayagwiz and Ngeze*, 3 December 2003, ICTR-99-52-T (Trial Chamber)

<sup>73</sup> ICTR, *Prosecutor v. Akayesu*, Case No. ICTR-96-4-T, Judgement (Trial Chamber), 2 September 1998, para 557.

- 
- **Intent:** the speaker must specifically intend to incite genocide, and intend for genocide to occur. This requires the speaker to specifically intend to engage in the communication calling for genocide, and either specifically intend to destroy, in whole or in part, a national, ethnical, racial or religious group, as such, or at least be “aware of the substantial likelihood that the commission of [genocide] would be a probable consequence of his acts;”<sup>74</sup>
  - Genocide need not actually occur, as **creating the potential for genocide** is sufficient to incur liability.<sup>75</sup> Where genocide does occur, the act of incitement may be considered to be an act of genocide in itself,<sup>76</sup> and charged as separately or additionally as instigation or complicity.

International law clearly requires states to **criminalise** the direct and public incitement to genocide, rather than provide for alternative and less severe forms of censure through administrative or civil law.

The Genocide Convention does not include any provisions regarding the protection of the right to freedom of expression. Nevertheless, without the presence of a declaration of a state of emergency in accordance with international human rights standards, any conviction for incitement must comply with the three-part test under Article 19(3) of the ICCPR.<sup>77</sup> This would include in the context of an armed conflict, where international humanitarian law may be concurrently applicable. Importantly, **international criminal law** does not specify “incitement” as an

<sup>74</sup> ICTY, *Prosecutor v. Naletili and Martinovi*, Case No. IT-98-34-T, Judgement (Trial Chamber), 31 March 2003, para 60

<sup>75</sup> ICTR, *Akayesu*, op. cit., para 562.

<sup>76</sup> ICTR, *Prosecutor v. Ferdinand Nahimana, Jean-Bosco Barayagwiza and Hassan Ngeze*, Case No. ICTR-99-52-T, Judgment and Sentence (Trial Chamber), 3 December 2003, para 1015.

<sup>77</sup> The **six-part test** outlined below in relation to incitement under Article 20(2) of the ICCPR would also be instructive in assessing cases of incitement to genocide, taking into account the specific intent required in incitement to genocide.

---

inchoate<sup>78</sup> form of liability in relation to other international crimes. To some offences, the concept of “instigation” applies to, though this requires any expressive act to be substantially connected to the actual commission of an offence, with specific intent on the part of the instigator for the consequential offence, rather than the mere danger or likelihood of those offences occurring. However, as a matter of domestic law, “incitement” is often recognised as a form of inchoate liability in relation to all crimes. Therefore, incitement to other forms of discriminatory international crimes, such as “persecution”, should be considered within the spectrum of “hate speech” that States may prohibit.

### **Advocacy of discriminatory hatred constituting incitement to discrimination, hostility or violence**

Article 20(2) of the ICCPR requires states to prohibit by law **“any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence”** (incitement).<sup>79</sup>

The implementation of Article 20(2) in practice has been the subject of a great deal of controversy and confusion. Reflecting this, numerous states have entered reservations to Article 20(2) of the ICCPR on the basis of concerns for the right to freedom of expression.<sup>80</sup> Domestic laws and practice diverge in their interpretation

<sup>78</sup> The word “inchoate” refers to something that has “just begun” or is “underdeveloped”, “partially completed” or “imperfectly formed.” Inchoate offenses are thus incomplete offenses, which are committed despite the fact that the substantive offense, that is, the offense whose commission they were aiming at, is not completed and the intended harm is not realized. See, e.g. A. Ashworth, *Principles of Criminal Law*, Oxford: Oxford University Press 2003 (4th ed.), p 445.

<sup>79</sup> Previously, ARTICLE 19 has used the shorthand of “incitement to hatred” to capture the obligation in Article 20(2) ICCPR. However, has at times created a misunderstanding that “hatred” alone is a proscribed outcome. This is not the case, the advocacy of hatred should only be prohibited where it constitutes incitement to a separate proscribed outcome, namely discrimination, hostility, or violence.

<sup>80</sup> See, for example Australia, Belgium, Luxemburg, Malta, New Zealand, the UK or the USA.



---

and implementation of the provisions of Article 20(2), a problem compounded by inconsistent jurisprudence at national levels and little guidance from the decisions of the HR Committee.<sup>81</sup>

There are implementations issues at both ends of the scale, with both impunity for serious cases of incitement, and overzealous enforcement of vague incitement provisions to punish legitimate expression, including the persecution of minority groups.

The **UN Rabat Plan of Action** (Rabat Plan) has identified these problems and advanced a range of conclusions and recommendations for the implementation of Article 20(2) of the ICCPR;<sup>82</sup> these correspond closely to the narrower political commitment of states in the **Resolution 16/18** to “criminalize incitement to imminent violence based on religion or belief.”<sup>83</sup> The Rabat Plan reflected to a series of recommendations which largely reflect ARTICLE 19’s input to this expert process, set out in the policy brief Prohibiting incitement to discrimination, hostility or violence (2012).

Based on the Rabat Plan and further consideration of these issues since its adoption, ARTICLE 19’s key recommendations are as follows:

### **Key elements of Article 20(2) ICCPR**

*Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.*

<sup>81</sup> The three decisions it has issued relate to two communications against Canada and one against France, each of which concerned prohibitions on anti-Semitic speech, taking a different approach in each. In General Comment No. 34, the approach of the HR Committee was not clarified ,

<sup>82</sup> This also corresponds neatly to the narrower political commitment of states in the Resolution 16/18 to “criminalize incitement to imminent violence based on religion or belief”

<sup>83</sup> The Resolution 16/18, op.cit., operational para 5(f).

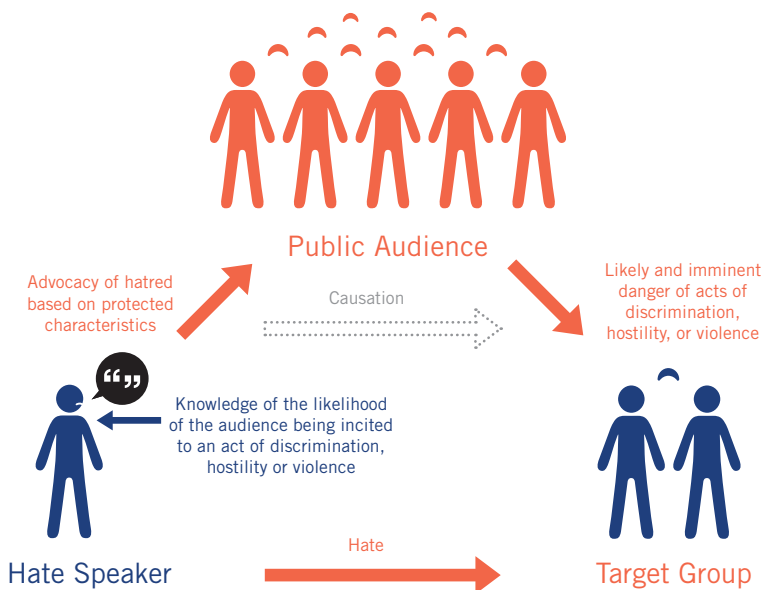
---

To address confusion around Article 20(2) of the ICCPR, it is helpful to break the prohibition into its key elements. We argue that each of these elements should be explicitly required or read into domestic provisions that implement Article 20(2) of the ICCPR. Moreover, criminal sanctions should be considered an exceptional measure and last resort for the most severe cases, with civil, administrative and alternative remedies also available.

The prohibition in Article 20(2) of the ICCPR requires:

- **Conduct of the speaker:** the speaker must address a public audience, and their expression include:
  - advocacy
  - of hatred targeting a protected group based on protected characteristics,
  - constituting incitement to discrimination, hostility or violence.
- **Intent of the speaker:** the speaker must
  - specifically intend to engage in advocacy of discriminatory hatred, and
  - intend for or have knowledge of the likelihood of the audience being incited to a discrimination, hostility or violence;
- **A likely and imminent danger of the audience actually being incited to a proscribed act**, as a consequence of the advocacy of hatred. A **six-part “severity threshold” test**, outlined below, assists in measuring whether the danger of incitement is sufficient to justify restrictions.

Incitement involves a triangular relationship between the three principal actors: the **“hate speaker”** advocating discriminatory hatred to a **public audience**; the public audience, who may engage in acts of discrimination, hostility or violence, and **the target group**, against whom these acts might be perpetrated.



---

## The conduct of the speaker

- **Advocacy** should be understood as an “intention to promote hatred publicly towards the target group.”<sup>84</sup> The idea of “promotion” is integral to advocacy; it implies more than merely stating an idea, but actually engaging in persuading others to adopt a particular viewpoint or mind-set. This may be through any medium of communication, including spoken, written or electronic.

### Example of advocacy



A prominent personality uses his social media account to disseminate to the public a series of vitriolic messages against migrants, repeating harmful stereotypes and lies about them.

Advocacy should be distinguished from **discriminatory abuse** or insults directed at a person because of their possessing a protected characteristic, where a third party audience is not present.

### Example of discriminatory abuse



A journalist posts a link to an article she has written relating to a political scandal, and as a consequence receives a series of sexist and abusive messages targeting her.

In some cases, one person’s abuse or insults, observed by others, may be understood as “advocacy of hatred” inciting a prohibited conduct. This may be the case where there is an influential “ringleader” in a situation analogous to a “mob”; one person leading discriminatory insults causes others to join with a view to inciting others to a prohibited conduct.

<sup>84</sup> Camden Principles, op. cit., Principle 12.1.; or the 2012 Report of the Special Rapporteur on FOE, op. cit., para 44(b).

---

**Hatred:** should be understood, as above, to be a **state of mind**, characterised by the “intense and irrational emotions of opprobrium, enmity and detestation” towards a target group on the basis of a protected characteristic.<sup>85</sup>

For the purpose of the prohibitions under Article 20(2), “hatred” relates to the state of mind of the speaker vis-à-vis the target group, and to the state of mind of the audience who s/he ultimately seeks to incite to the proscribed conducts. This is distinct from the feelings of insult or indignity that the target group may feel when confronted by ‘hate speech’, which is not the object of the Article 20(2).

The **type of hatred** is also relevant. Article 20(2) of the ICCPR only lists the “advocacy of **national, racial or religious** hatred” but, as noted above, we recommend that this list be interpreted expansively as “advocacy of any discriminatory hatred” against a person or group of persons. This would encompass all of the protected characteristics recognised under international human rights law, reading Article 20(2) of the ICCPR in light of the broader non-discrimination provisions in Article 2(1) and Article 26 of the ICCPR,<sup>86</sup> and other international legal instruments addressing discrimination.

<sup>85</sup> Ibid., adapted.

<sup>86</sup> Ibid. The HR Committee has not yet addressed this question directly. However, this interpretation would comply with the evolution of developments in human rights protections since the adoption of the ICCPR, considering that it was adopted before many equality movements around the world made significant progress in promoting and securing human rights for all. It has since come to be interpreted and understood as supporting the principle of equality on a larger scale, applying to other grounds not expressly included. It would also be in line with the object and purpose of international human rights law that must be interpreted “in the light of present-day conditions,” rather than being viewed as contracts with concrete terms defined by the norms that were prevailing at the moment of their drafting or ratification.

- 
- **Incitement:** Article 20(2) of the ICCPR does not prohibit advocacy of discriminatory hatred per se. It is only concerned with the advocacy of discriminatory hatred that **constitutes incitement to violence, hostility or discrimination**. While the proscribed outcome need not in fact occur, the term “incitement” strongly implies the advocacy of hatred must create “an imminent risk of discrimination, hostility or violence against persons belonging to [the target group].”<sup>87</sup>
  - It is this danger of the audience taking action against a target group on the basis of their protected characteristic which is the focus of Article 20(2), distinct from the insult or dignity the target group feel about such speech in isolation of any incitement.<sup>88</sup>

The prohibited outcomes should be understood as follows:

- **Discrimination** should be understood as “any distinction, exclusion, restriction or preference” based on any protected characteristic recognised under international human rights law,<sup>89</sup> which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise, on an equal footing, of human rights and fundamental freedoms in the political, economic, social, cultural or any other field of public life.”<sup>90</sup>
- **Violence** should be understood as “the intentional use of physical force or power against another person, or against a group or community that either results in or has a high likelihood of resulting in injury, death, psychological harm, mal-development, or deprivation.”<sup>91</sup>

<sup>87</sup> The Camden Principles, Principle 12.1.iii.

<sup>88</sup> The Rabat Plan refers to the “activation of a triangular relationship between the object and subject of the speech as well as the audience.”

<sup>89</sup> See [Annex I](#). ARTICLE 19 understands protected characteristics expansively.

<sup>90</sup> Camden Principles, op. cit., Principle 12.1

<sup>91</sup> This definition is adapted from the definition of violence by the World Health Organisation in the report World Report on Violence and Health, 2002.

- 
- **Hostility** should be understood as “a manifested action of an extreme state of mind”, rather than a mere extreme state of mind.<sup>92</sup> This may be distinguished from violence to cover threats of violence, harassment, or property damage.

Any domestic legal provision that prohibits the mere spreading of discriminatory hatred, or that concerns the effect of ‘hate speech’ on the feelings of a target group, would be significantly broader than the prohibition envisaged under the ICCPR. Similarly, laws that adapt the prohibited outcomes specified in the ICCPR to include broader concepts would be open to abuse and therefore illegitimate; for example those that seek to protect “public tranquillity” or safeguard against “unrest” or “division between religious believers.”<sup>93</sup>

### **The intent of the speaker**

The intent of the speaker to incite others to commit acts of discrimination, hostility or violence should be considered a crucial and distinguishing element of incitement as prohibited by Article 20(2) of the ICCPR. National legislation should always explicitly state that the crime of incitement is a crime of specific intent,<sup>94</sup> and not a crime that can be committed with recklessness or negligence.<sup>95</sup>

<sup>92</sup> In the 2009 Camden Principles, “hostility” was defined in the exact same terms as “hatred”. However, after further reflections and discussion, an updated definition requiring a manifested action was incorporated to our 2012 policy brief “prohibiting incitement to discrimination, hostility or violence”, *op. cit.*, at page 19.

<sup>93</sup> The 2013 Report of the Special Rapporteur on FOE, *op. cit.*, para 51.

<sup>94</sup> In some jurisdictions, also acting “wilfully” or “purposefully.”

<sup>95</sup> ARTICLE 19 notes that the legislation of many States already recognises intent as one of the defining elements of incitement, for example, the UK, Ireland, Canada, Cyprus, Ireland, Malta, and Portugal.

---

Given that there is no uniform definition of intent in international law, or across domestic laws,<sup>96</sup> ARTICLE 19 proposes that any definition of intent should include:

- Intent to engage in advocacy to hatred;
- Intent to target a protected group on the basis of a protected characteristic;
- Having knowledge that, in the given context at the time, the expression will likely cause a proscribed outcome: discrimination, hostility or violence.

### **Severity threshold**

ARTICLE 19's six-part test, incorporated to the Rabat Plan, was designed to assist in determining in which situations the danger of violence, hostility or discrimination is sufficiently present to justify prohibitions on the expression. It consists of the following criteria:

- 1 Context of the expression:** the expression should be considered within the political, economic, and social context in which it was communicated,<sup>97</sup> as this will have a bearing directly on both intent and/or causation. The contextual analysis should take into account, inter alia:
  - the existence of conflict in society, for example, recent incidents of violence against the targeted group;
  - the existence and history of institutionalised discrimination, for example in law enforcement and the judiciary;

<sup>96</sup> No international treaty body or mechanism has adopted a definition of "intent" for the purpose of prohibition on incitement. A common approach within international criminal law and domestic criminal law is to ensure liability for offences where the speaker acted with knowledge and with the intention of causing the objective elements of a crime.

<sup>97</sup> As noted by Toby Mendel, "it is extremely difficult to draw any general conclusions from the case law about what sorts of contexts are more likely to promote the proscribed result, although common sense may supply some useful conclusions. Indeed, it sometimes seems as though international courts rely on a sample of contextual factors to support their decisions rather than applying a form of objective reasoning to deduce their decisions from the context. Perhaps the impossibly broad set of factors that constitute context make this inevitable." Toby Mendel, Study on International Standards Relating to Incitement to Genocide or Racial Hatred (2006).



- 
- the legal framework, including the recognition of the targeted group's protected characteristic in any anti-discrimination provisions or lack thereof;
  - the media landscape, for example regular and negative media reports about the targeted group with a lack of alternative sources of information; and
  - the political landscape, in particular the proximity of elections and the role of identity politics in that context, as well as the degree to which the views of the targeted group are represented in formal political processes.

**2 The speaker:** the position of the speaker, and their authority or influence over their audience is crucial. Special considerations should be made when the speaker is a politician or a prominent member of a political party and public officials or persons of similar status (e.g. teachers or religious leaders) due to the stronger attention and influence they exert over the others. This analysis should also examine the relationship of the audience to the speaker, and issues such as the degree of vulnerability and fear among the various communities, including those targeted by the speaker, or whether the audience has high levels of respect or obedience of authority voices.<sup>98</sup>

**3 Intent:** as outlined above, there must be (i) intent to engage in advocacy to hatred; (ii) intent to target a group on the basis of a protected characteristic, and (iii) having knowledge of the consequences of their action and knowing that the consequences will occur or might occur in the ordinary course of events (i.e. in which no unforeseeable change or event has occurred). This should be judged on the facts of the case and its circumstances as a whole,<sup>99</sup> taking into account, inter alia, the language used, the scale and repetition of the expression, and any stated objectives of the speaker. Recklessness and negligence are not sufficient as a standard of intent; thus consideration should be giving to protecting communications that are simply ill-judged or flippant (such as a bad joke), or where the intent is more nuanced (to satire, provoke thought or challenge the status quo, including through art).

<sup>98</sup> See: [Susan Benesh, \*Dangerous Speech: A Proposal To Tackle Violence\*](#), 2011.

<sup>99</sup> See, e.g. European Court, *Jersild v. Denmark*, App. No. 15890/89, 23 September 1994, para. 31.

- 
- 4 **Content of the expression:** what was said is relevant, including the form and the style of the expression, whether the expression contained direct or indirect calls for discrimination, hostility or violence, and the nature of the arguments deployed and the balance struck between arguments. The audience's understanding of the content of the expression is particularly important, in particular where incitement may be indirect. International standards have recognised that certain forms of expression provide "little scope for restrictions",<sup>100</sup> in particular artistic expression, public interest discourse, academic discourse and research, statements of facts and value judgements.
- 5 **Extent and magnitude of the expression:** the analysis should examine the public nature of the expression,<sup>101</sup> the means of the expression and the intensity or magnitude of the expression in terms of its frequency or volume (e.g. one leaflet as opposed to broadcasting in the mainstream media, or singular dissemination as opposed to repeated dissemination). If the expression was disseminated through the mass media, consideration should be given to media freedom, in compliance with international standards.<sup>102</sup>

<sup>100</sup> European Court, *Erbakan v. Turkey*, App. No. 59405/00, 6 June 2006, para 68.

<sup>101</sup> ARTICLE 19 suggest that this should include looking at issues such as whether the statement or communication was circulated in a restricted environment or whether it was widely accessible to the general public; whether it was made in a closed place accessible by ticket or in an exposed and public area; whether the communication was directed at a non-specific audience (the general public); or whether the speech was directed to a number of individuals in a public place, and whether the speech was directed to members of the general public.

<sup>102</sup> As the European Court noted, "while the press must not overstep the boundaries set, inter alia, for the protection of the vital interests of the State, [...] it is nevertheless incumbent on it to impart information and ideas on political issues, including divisive ones. Not only does the press have the task of imparting such information and ideas: the public also has a right to receive them. The freedom to receive information or ideas provides the public with one of the best means of discovering and forming an opinion on the ideas and attitudes of their leaders;" see *Halis Doan v. Turkey*. App. No. 71984/01 (2006).

- 
- 6 Likelihood of harm occurring, including its imminence:** there must be a reasonable probability of discrimination, hostility or violence occurring as a direct consequence of the expression, but the proscribed outcome itself need not actually occur; actual occurrence of harm may be considered an aggravating circumstance in criminal cases. This should include:
- Whether the audience understand the advocacy of hatred as a call to acts of discrimination, hostility or violence;
  - Whether the speaker was in the position to influence the audience;
  - Whether the audience had the means to resort to acts of discrimination, hostility or violence;
  - Recent incidences of the targeted group suffering discrimination, hostility or violence as the result of incitement.
  - The length of time that passes between the speech and the time when discrimination, hostility or violence could take place is not so long to bring into doubt the causative impact of the expression.<sup>103</sup>

<sup>103</sup> C.f. Susan Benesh, *op.cit.*

---

## Sanctions for incitement should not be limited to criminal penalties

ARTICLE 19 believes that the sanctions for incitement should principally be contained within civil and administrative law, and criminal penalties should be imposed only as a last resort, and in the most severe cases.

- **Civil law sanctions**<sup>104</sup> provide a more victim-centred approach to providing redress in cases of advocacy of discriminatory hatred constituting incitement to violence, hostility or discrimination. Remedies for the victims of any advocacy of discriminatory hatred constituting incitement to discrimination, hostility, or violence should include compensation in the form of pecuniary and non-pecuniary damages,<sup>105</sup> and the right of correction and reply if the incitement occurred through mass media.<sup>106</sup> States should also allow NGOs to bring civil claims in relevant cases and should provide for the possibility of bringing class actions in discrimination cases. This should be part of a comprehensive anti-discrimination framework that should include all of the measures identified in **Part II** of the toolkit detailing non-discrimination laws.

<sup>104</sup> The proposal is based on requirements set by the [Racial Equality Directive](#) (2000/43/EC) against discrimination on grounds of race and ethnic origin, restricting it.

<sup>105</sup> Council of Europe Committee of Ministers Recommendation No. R 97(20) on ‘hate speech’, Principle 2. The awards of damages should be proportional and carefully and strictly justified and motivated so that they do not have a collateral chilling effect on the freedom of expression.

<sup>106</sup> *Ibid.*, Principle 2. Also, Camden Principles, *op.cit.*, Principle 7. The right of reply gives any person the right to have a mass media outlet disseminate his or her response where the publication or broadcast by that media outlet of incorrect or misleading facts has infringed a recognised right of that person, and where a correction cannot reasonably be expected to redress the wrong. This remedy also has the result of encouraging further dialogue, rather than restricting it.

- 
- **Administrative sanctions** should also be considered, in particular to enforce rules established by communication, media and press councils, consumer protection authorities, or any other regulatory bodies. Consideration should also be given to establishing formal codes of conduct and employment rules for certain actors, including for politicians, public officials, and civil servants (such as teachers). These measures should support the principle that public officials at all levels should, as far as possible, avoid making statements that promote discrimination or undermine equality and intercultural understanding. Sanctions may be in the form of an order to: issue a public apology (although this should not automatically preclude culpability or other sanctions). In relation to public service broadcasters, a framework for administrative sanctions may support the obligation to avoid communicating negative stereotypes of individuals and groups. These may include the obligation to issue a correction; provide a right of reply; allocate broadcasting time to advertise the outcome of an administrative decision, or the imposition of fines.
  - Recourse to **criminal law** should be avoided if less severe sanctions would achieve the intended effect; the experience of many jurisdictions demonstrates that civil and administrative sanctions are better suited as a response to the harm caused by incitement.<sup>107</sup> It is also important that courts, law enforcement authorities and public bodies should consider the perspective of victims in criminal law proceedings to strengthen the pursuit of justice. They should also attempt to involve them in the proceedings through other channels, for example, by inviting third party interventions in the form of amicus briefs by representatives of groups concerned by the case.

<sup>107</sup> For example, in Brazil, it has been documented that, criminal law has not been efficient due to institutional bias among law enforcement agencies, while sanctions have been levied effectively in civil proceedings. See, Tanya Hernandez, *Hate Speech and the Language of Racism in Latin America*, 32 U. Pa. J. Int'l L. 805 2010-2011.

---

## Prohibitions on ‘hate speech’ under Article 19(3) of the ICCPR

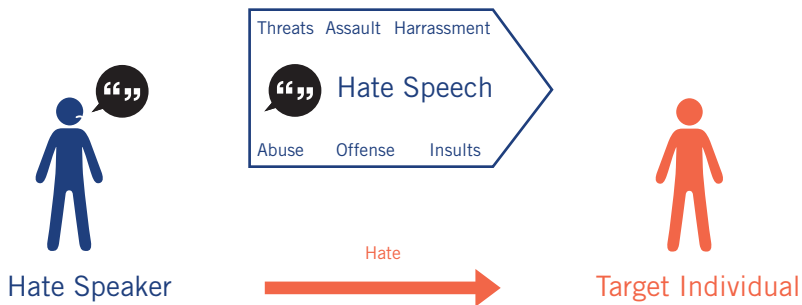
In some situations, ARTICLE 19 acknowledges that expression may not need to reach the threshold of Article 20(2) “incitement” before it constitutes a sufficiently severe problem as to require specific state responses to limit the expression.

Under international and regional human rights standards, such ‘hate speech’ may still be restricted on a number of different grounds, listed in Article 19(3) of the ICCPR, in particular the rights of others or public order:

- **‘Hate speech’ targeted at protected group:** This applies, in particular, to specific settings – such as ‘hate speech’ in broadcast media, educational institutions or before the elections;
- **Individually targeted ‘hate speech’:** In some instances, ‘hate speech’ may be targeting specific individuals, without any “advocacy of hatred.” Such expression often directly targets women or other marginalised or minority groups; it is frequently vehemently discriminatory, and can be deeply offensive and harmful to those whom it targets. In some instances, it can take on a form of assault, threatening violence, or harassment, and cause physical or psychological harm to the individual targeted.

Where a risk of physical and/or significant psychological harm exists from such conduct, restrictions may be justified where necessary to protect individuals’ rights to life, to be free from inhuman or degrading treatment, to privacy, and/or to equality.

- It is worth visualising how this kind of ‘hate speech’ is distinct from the triangulation involved between the three parties in “incitement” under Article 20(2) of the ICCPR. The six-part threshold test for “incitement” outlined above is therefore inappropriate as the expressive act involves only a hate speaker and the targeted individual:



- Though threats or harassment disproportionately impact those vulnerable to discrimination and “hate speech,” not all threats or harassment will necessarily be bias-motivated or characterised by hatred. Existing but generic prohibitions on threats or harassment may therefore be adequate to deal with these incidents, though the absence of any mechanism for recognition of bias-motivation in this conduct would be a deficiency that should be addressed.

ARTICLE 19 is concerned that restrictions on freedom of expression in particular settings and prohibitions on threats or harassment often are too broad and subject to abuse; they may be enforced to merely protect people from offence or expression they disagree with, or shield people in positions of power from legitimate criticism.

There is no consensus at the international level in relation to these specific contexts. Any such restrictions on account of other grounds should be considered legitimate in as much as they comply with the three part test outlined above.

ARTICLE 19 encourages extreme caution in any of these approaches being expanded beyond the particular circumstances they address, and the unique and tailored regulatory approaches that may be appropriate in each of these circumstances.

## Annex 1: Legal Instruments

Instrument	Recognised protected characteristics in general provisions	Recognised characteristics in 'hate speech' prohibitions
<b>Convention on the Prevention and Punishment of the Crime of Genocide</b>	Article 2 prohibits “ <b>genocide</b> ” as the commission of one of five specified acts “committed with intent to destroy, in whole or in part”, a <b>national, ethnical, racial or religious</b> group, as such.	Article 3(c) prohibits “ <b>public and direct incitement to Genocide</b> ” on the same protected characteristics as Article 2.
<b>Rome Statute of the International Criminal Court</b>	The broadest range of protected characteristics is found in the offence of “ <b>persecution</b> ”, which can be committed against any group on the basis of their “ <b>political opinion, race, nationality, ethnicity, culture, religion, gender</b> ” or “ <b>other grounds that are universally recognized as impermissible under international law</b> ” (Articles 7(1)(h) and 7(2)(g)).	Article 25(3)(e) of the Rome Statute only prohibits “incitement” in relation to Genocide, and does expressly prohibit incitement to other international crimes.
<b>The International Covenant on the Elimination of all forms of Racial Discrimination</b>	Article 1 recognises discrimination on the bases of <b>race, colour, descent, or national or ethnic origin</b> .	Article 4(a) requires states to “condemn all propaganda and all organisations which are based on ideas of theories of superiority of one <b>race</b> or group of persons of one colour or <b>ethnic origin</b> , or which attempt to justify or promote <b>racial hatred and discrimination in any form</b> , and undertake to adopt immediate and positive measures designed to eradicate all incitement to, or acts of, such discrimination and, to this end, with due regard to the principles embodied in [the UDHR] and the rights expressly set forth in Article 5 of [the ICERD].”  However, the CERD Committee has addressed ‘hate speech’ on all grounds listed in Article 1, and recognised intersectional discrimination on the grounds of <b>religion, gender, and indigenous origin</b> . <sup>108</sup> The CERD Committee has also welcomed prohibitions on “hate speech” on grounds of <b>sexual orientation and gender</b> . <sup>109</sup>

<sup>108</sup> For example, in Brazil, it has been documented that, criminal law has not been efficient due to institutional bias among law enforcement agencies, while sanctions have been levied effectively in civil proceedings. See, Tanya Hernandez, Hate Speech and the Language of Racism in Latin America, 32 U. Pa. J. Int'l L. 805 2010-2011.

<sup>109</sup> ICERD Committee, Concluding Observations on Romania, CERD/C/ROU/CO/16-19, 13, September 2010, para 4.



Instrument	Recognised protected characteristics in general provisions	Recognised characteristics in 'hate speech' prohibitions
<p><b>The Convention of the Elimination of all forms of Discrimination Against Women 1979</b></p>	<p>Article 2 provides that States Parties “condemn discrimination against <b>women</b> in all its forms” and commits them to “pursue by all appropriate means and without delay a policy of eliminating discrimination against women”. Article 1 defines “discrimination against women” as “any distinction, exclusion or restriction <b>made on the basis of sex</b> which has the effect or purpose of impairing or nullifying the recognition, enjoyment or exercise by women, irrespective of their marital status, on a basis of equality of men and women, of human rights and fundamental freedoms in the political, economic, social, cultural, civil or any other field.”</p>	<p>There is no provision explicitly requiring prohibitions on ‘hate speech’ against women. However:</p> <ul style="list-style-type: none"> <li>– Article 2(b) requires States “to adopt appropriate legislative and other measures, including sanctions where appropriate, prohibiting all discrimination against women”</li> <li>– Article 2(f) requires States “to take all appropriate measures, including legislation, to modify or abolish existing laws, regulations, customs and practices which constitute discrimination against women.”</li> <li>– Article 5 requires States “to modify the social and cultural patterns of conduct of men and women, with a view to achieving the elimination of prejudices and customary and all other practices which are based on the idea of the inferiority or the superiority of either of the sexes or on stereotyped roles for men and women.” <sup>110</sup></li> </ul>

<sup>110</sup> In its General Recommendation No. 23 on “political and public life” (1992), the CEDAW Committee identified “stereotyping, including that perpetuated by the media”, as a barrier to full participation in public life. In its General Recommendation No. 19 on “violence against women”, the Committee has recommended that: “effective measures should be taken to ensure that the media respect and promote respect for women”. In their reports to the Committee, States are also asked to “identify the nature and extent of attitudes, customs and practices that perpetuate violence against women and the kinds of violence that result”. In their Concluding Observations on State Reports, the Committee has called on States to “engage with the media with the aim of eliminating stereotypical imaging of women especially when it affects women’s health”, and “strengthen measures to address hate speech against ethnic minority and other women and girls in the media, including on internet discussion boards and in social media” (see CEDAW Committee, Concluding Observations on the 7th Periodic Report of Finland, CEDAW/C/FIN/CO/7, 28 February 2014, para 15). While there has been attention to intersecting discrimination in “hate speech” against ethnic minority women, the CEDAW Committee has not addressed the same against bisexual, lesbian or transgender women, for example.

Instrument	Recognised protected characteristics in general provisions	Recognised characteristics in 'hate speech' prohibitions
<b>Convention Against Torture</b>	Article 1 provides that one of the recognised purposes of torture is “discrimination of any kind”, without listing protected characteristics. The Committee Against Torture has adopted the following listing, noting in particular the intersection between each ground and gender: race, colour, ethnicity, age, religious belief or affiliation, political or other opinion, national or social origin, gender, sexual orientation, transgender identity, mental or other disability, health status, economic or indigenous status, reason for which the person is detained, including persons accused of political offences or terrorist acts, asylum seekers, refugees or others under international protection, or any other status or adverse distinction. <sup>111</sup>	There is no provision explicitly requiring prohibitions on 'hate speech'. However, Article 16 requires States to “prevent” “other acts of cruel, inhuman or degrading treatment or punishment which do not amount to torture.” The Committee Against Torture has recommended to individual States to sanction hate crimes as acts of “intolerance and incitement to hatred and violence” as part of the obligation under Article 16 of the CAT. <sup>112</sup>
<b>The Convention on Persons with Disabilities</b>	Recognises a broad obligation to eliminate discrimination on the basis of disability.	There is no provision explicitly requiring prohibitions on 'hate speech'. However, States are required to adopt “immediate, effective and appropriate measures” “to combat stereotypes, prejudices and harmful practices relating to persons with disabilities, including those based on sex and age, in all areas of life.”
<b>UN Declaration on the Rights of Indigenous Peoples</b>	Article 2 guarantees that “indigenous peoples and individuals are free and equal to all other peoples and individuals and have the right to be free from any kind of discrimination, in the exercise of their rights, in particular that based on their indigenous origin or identity.”	Article 8(2)(e) requires States to prevent and redress “any form of propaganda designed to promote or incite racial or ethnic discrimination directed against [indigenous peoples and individuals].”

<sup>111</sup> Committee against Torture, General Comment No.2, CAT/C/GC/2, 24 January 2008, Part V

<sup>112</sup> Concluding observations on Moldova, CAT/C/MDA/CO/2, para 27; Concluding observations on Poland, CAT/C/POL/CO/4, para. 20; and Concluding observations on Mongolia, CAT/C/MNG/CO/1, para 25. ARTICLE 19 notes that the Committee has not defined “hate crimes,” “intolerance,” “incitement to hatred” or “hate speech,” or offered guidance on the appropriate balance between the rights, leaving the nature of the legal obligation ambiguous.



## DEFENDING FREEDOM OF EXPRESSION AND INFORMATION

---

ARTICLE 19 Free Word Centre 60 Farringdon Road London EC1R 3GA

T +44 20 7324 2500 F +44 20 7490 0566

E [info@article19.org](mailto:info@article19.org) W [www.article19.org](http://www.article19.org) Tw [@article19org](https://twitter.com/article19org) [facebook.com/article19org](https://facebook.com/article19org)