

ARTICLE¹⁹

Implementing UN Human Rights Council Resolution 16/18 in the digital sphere

December 2022

ARTICLE 19**T:** +44 20 7324 2500**F:** +44 20 7490 0566**E:** info@article19.org**W:** www.article19.org**Tw:** @article19org**Fb:** facebook.com/article19org

First published by ARTICLE 19, December 2023

www.article19.org

Copyright ARTICLE 19, December 2023 (Creative Commons License 3.0)

ARTICLE 19 works for a world where all people everywhere can freely express themselves and actively engage in public life without fear of discrimination. We do this by working on two interlocking freedoms, which set the foundation for all our work. The Freedom to Speak concerns everyone's right to express and disseminate opinions, ideas, and information through any means, as well as to disagree with and question power holders. The Freedom to Know concerns the right to demand and receive information by power holders for transparency, good governance, and sustainable development. When either of these freedoms comes under threat, by the failure of power holders to adequately protect them, ARTICLE 19 speaks with one voice, through courts of law, through global and regional organisations, and through civil society wherever we are present.

About Creative Commons License 3.0: This work is provided under the Creative Commons Attribution-Non-Commercial-ShareAlike 3.0 license. 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 license identical to this one.

To access the full legal text of this license, please visit:

<http://creativecommons.org/licenses/by-nc-sa/3.0/legalcode>

Contents

Introduction	5
The problem: The rise of religious hatred and intolerance in the digital sphere	7
Regulating religious ‘hate speech’ online	7
The case of the Rohingya community	9
Role of digital companies in targeting religious ‘hate speech’ online	11
<hr/>	
International human rights framework	13
The ICCPR	13
The right to freedom of opinion and expression	13
The right to freedom of religion or belief	14
The right to equality and non-discrimination	15
Positive measures under international human rights standards	15
Human Rights Council Resolution 16/18	15
Istanbul Process	17
Rabat Plan of Action	17
Additional commitments on freedom of religion or belief and freedom of expression in the digital sphere	20
UN Strategy and Plan of Action on Hate Speech	20
<hr/>	
Implementing HRC Resolution 16/18 and related commitments in the digital sphere	21
Recommendations for states	21
Recognising and speaking out against religious intolerance and hatred	24
Fostering dialogue and public debate	24
Public education, information campaigns, and training	25
Monitoring and data collection	25
Recommendations for other stakeholders	26
Recommendations for internet companies	26
Recommendations for civil society	28

Introduction

Across the world, intolerance, discrimination, and violence against persons on the basis of their religion or belief increasingly takes place in the digital sphere.¹ While these manifestations of hatred are not new, digital technology has allowed it to be created, disseminated, and amplified at a scale not known before. The human rights impacts are diverse and can be severe, ranging from incitement to crimes against humanity and even genocide, to discriminatory harassment, abuse, and threats against individuals. The potential for content to go viral, contributing to its extensive digital footprint, and difficulties in identifying its authors present challenges from legal, regulatory, and policy perspectives.

The UN Human Rights Council (HRC) has repeatedly affirmed in its resolutions that: ‘the same rights that people have offline must also be protected online’.² It is clear that states and other stakeholders must do more to uphold the interrelated and interdependent rights of freedom of religion and belief and freedom of expression online.

A range of UN initiatives provide guidance to tackle religious intolerance and hatred – including UN HRC [Resolution 16/18](#) and the [Rabat Plan of Action](#) (RPA) – yet none explicitly addresses the importance of promoting inclusion, diversity, and pluralism in the digital sphere.

ARTICLE 19 believes that states and other actors must effectively tackle religious intolerance and hatred online as well as offline, and legislation, public policies, and programmes must be attentive to how hate is disseminated in the digital sphere.

This brief aims to provide guidance for states, digital companies, and various stakeholders to advance implementation of Resolution 16/18 and related commitments in the digital sphere. ARTICLE 19 hopes that the recommendations outlined in this brief can assist states in their identification of further positive examples of alternative policy measures in

order to more effectively address the root causes of religious hatred and increase awareness among various stakeholders of the issues around religious hatred.

The problem: The rise of religious hatred and intolerance in the digital sphere

Discrimination, intolerance, and violence based on religion or belief are not new phenomena. While digital technologies, particularly social media platforms, have enabled individuals to exercise their right to freedom of expression and participate in civic affairs, they have also allowed [widespread distribution of hateful messages](#). The UN Special Rapporteur on minorities [warned](#) that ‘hate speech’ on social media directly contributes to a dramatic increase in many countries of hate crimes targeting religious, ethnic, and other minorities, including in places of worship in Sri Lanka, New Zealand, India, or the United States. The UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (Special Rapporteur on FOE) [also raised concerns](#) that the prevalence of online hate poses challenges to everyone, first and foremost the individuals at risk of discrimination who are its principal targets.

However, states and digital companies have too often acted out of line with international human rights standards when addressing and responding to religious online ‘hate speech’,³ while also failing to address the root causes of such hatred.

Regulating religious ‘hate speech’ online

‘Hate speech’ is a broad term that has no definition under international human rights law.⁴ The expression of hatred towards an individual or group on the basis of a protected characteristic can be divided into three categories, distinguished by the response required from states under international human rights law:

- severe forms of ‘hate speech’ that international law requires states to prohibit, including through criminal, civil, and administrative measures, under both international criminal law and Article 20(2) of the International Covenant on Civil and Political Rights (ICCPR);
- other forms of ‘hate speech’ that states may prohibit to protect the rights of others under Article 19(3) of the ICCPR, such as discriminatory or bias-motivated threats or harassment; and

- ‘hate speech’ that is lawful but nevertheless raises concerns in terms of intolerance and discrimination, meriting a critical response by the state but which should be protected from restrictions under Article 19(3) of the ICCPR.

The Human Rights Committee (HR Committee) has made clear that limitations on forms of communication or expression disseminated over the internet must be justified according to the same criteria as non-electronic or ‘offline’ communications, as set out earlier.⁵

‘Hate speech’ laws and their online application are problematic in most countries of the world. The deficiencies can be [identified](#) in terms of the broad and vague definitions, their compatibility with applicable international freedom of expression standards, and the inconsistencies in their application.

Although international human rights law requires states to prohibit any advocacy of religious hatred that constitutes incitement to discrimination, hostility, or violence through criminal, civil, or administrative measures, states tend to regulate ‘hate speech’ mostly through criminal law. In addition, when assessing incitement cases, authorities fail to take into account the guidance of the RPA which should also be applied to online expressions: the content, the context, and the magnitude of the expression, with special focus on their online appearance; the position and the intent of the speaker; and the reasonable probability of discrimination, hostility, and violence as a direct consequence of the expression.⁶

At the same time, vague definitions of ‘hate speech’ laws seem to inhibit governments from addressing genuine harms such as [the kind that incites violence or discrimination against vulnerable communities or silencing of the marginalised](#) all around the world. Moreover, the rise in prejudice and intolerance based on religion or belief can in many cases also be directly linked to governments’ own policies and communication strategies or to their failures to adequately respond to ‘hate speech’ and violent threats against religious minorities. As a consequence, serious incidences of incitement to hostility, discrimination, or violence remain unpunished.

The case of the Rohingya community

Since April 2020, hateful messages targeting the Rohingya community in Malaysia have proliferated on social media platforms. Many posts included discriminatory and dehumanising language and images, as well as calls for Rohingya in Malaysia to be forcibly returned to Myanmar. Numerous online petitions calling for the expulsion of Rohingya were launched on Change.org and other platforms. Some petitions garnered thousands of signatures. Online users threatened prominent Rohingya activists, as well as their supporters, with physical attacks, murder, and sexual violence. The government failed to adequately respond to the surge in 'hate speech' and violent threats being directed at the Rohingya population. Moreover, statements by Malaysian officials, including the Minister of Home Affairs, could also have heightened tensions and incited violence and discrimination against the Rohingya.

ARTICLE 19 emphasises the key role of government officials and politicians in condemning all forms of intolerance, discrimination, and hatred directed at the Rohingya. Authorities should investigate and prosecute threats of violence and expression that constitutes incitement to violence or discrimination in a manner that is consistent with the standards set out in the RPA, as outlined earlier. In addition, the government should also step up efforts to implement positive policy measures to counter 'hate speech' and intolerance in line with Resolution 16/18 and the RPA.

Regulating content moderation by states

While international human rights law places obligations on states to protect, promote, and respect human rights, it is widely recognised that [business enterprises also have a responsibility to respect human rights](#). Importantly, the Special Rapporteur on FOE has long held that censorship measures should never be delegated to private entities,⁷ and states should not require or otherwise pressure the private sector to take steps that unnecessarily or disproportionately interfere with freedom of expression, whether through laws, policies, or extra-legal means. He further recognised that ‘private intermediaries are typically ill equipped to make determinations of content illegality’,⁸ and reiterated criticism of notice and takedown frameworks for ‘incentivizing questionable claims and for failing to provide adequate protection for the intermediaries that seek to apply fair and human rights-sensitive standards to content regulation’, i.e. the danger of ‘self- or over-removal’.⁹

There has been a growing trend of states mandating digital companies to remove ‘hate speech’, ‘extremist’ content, and ‘fake news’, or they are strongly incentivised to remove content that may be in breach of the law of the country in which they operate. [ARTICLE 19](#) is concerned that many of these regulatory models are deeply problematic because they fail to comply with the requirements of international human rights law:

- They focus on suppressing content without clearly defining what types of content are targeted. The categories generally described by national laws, such as blasphemy or defamation of religion, are too broad to enforce reliably, create uncertainty, and risk chilling speech.
- They outsource further responsibilities to platforms and undermine due process safeguards by deputising platforms to undertake law enforcement functions. Some governments informally ask or pressure the platforms to remove illegal content under their private terms of service, rather than proceeding under applicable law.
- They foster further reliance on automated and artificial intelligence (AI)-driven content moderation. Automation is unreliable for ‘hate speech’, particularly for religious ‘hate speech’ which requires nuanced and contextual assessment. Using AI to assess

potential 'hate speech' cases can lead to errors and potential over-deletion of legal content.

Role of digital companies in targeting religious 'hate speech' online

It is important to acknowledge the role of digital companies in promoting and protecting freedom of expression. These companies now play a major role in the way in which a growing proportion of the global population interacts with the world. The right of individuals to freedom of expression online is, therefore, to a large extent, determined and influenced by the actions of social media platforms. While social media platforms have had a positive effect on freedom of expression, they can easily be misused as a forum for stigmatisation, discrimination, exclusion, and, in the worst cases, incitement to violence.

First, the profit-driven business models of these companies rely on algorithms to prioritise users or news feeds. While this usually takes place on the basis of the perceived interests of their users, it is also the result of advertising or other marketing agreements. [This model](#) is fundamentally optimising for more hatred and more violence on the platforms.

Second, although social media companies are in principle free to restrict content on the basis of freedom of contract, they should respect human rights, including the rights to freedom of expression, privacy, and due process. ARTICLE 19 recognises that over the last decade, the large tech companies have proved themselves unwilling or too slow to address challenges for the protection of freedom of expression, religion or belief, and other rights on their platforms.

In addition, the way in which social media companies have dealt with content issues on their platforms, especially around 'hate speech', has also been widely criticised by many stakeholders, including the Special Rapporteur on FOE in his recent reports. [The Rapporteur argued](#) that companies in the ICT sector fail to apply the Guiding Principles on Business and Human Rights of the United Nations and that human rights are not integrated into their products by design and by default. The consequences of this might be

tragic events, as illustrated by Facebook's failure to address incitement against the Rohingya Muslim community in Myanmar.¹⁰

[ARTICLE 19](#) finds that the content moderation policies of social media companies are problematic for several reasons, including:

- **Lack of legal certainty:** Social media companies use automated processes to filter content across the range of their rules, including 'hate speech'. Besides the lack of clarity around their use, algorithms are not capable of evaluating cultural context,¹¹ detecting irony, or conducting the critical analysis necessary to accurately identify 'hate speech' content. In addition, company definitions of 'hate speech' are in general difficult to understand although companies vary on this matter. Definitions are vague and it is not clear how they define incitement, intent, result, and other relevant issues with regard to 'hate speech'. With regard to religious 'hate speech', the content moderation policies of social media companies often lack the [cultural competency](#) to recognise religious minorities who are the main targets of 'hate speech' on their platforms: the community guidelines are not translated and localised into local languages; religious 'hate speech' disaggregated by category is not tracked; and the moderation team lacks knowledge and understanding of the nuances of how religious hatred operates locally.
- **Lack of accountability or transparency over decision-making:** The majority of the decision-making processes and practices of social media platforms underpinning their content policies – including the use of automated decision-making processes – remain opaque. Existing procedures do not ensure sufficient accountability for those decisions. Although some progress has been made with regard to transparency reporting over the years, there is still too little information available about the way in which platforms apply their community guidelines in various circumstances. In addition, there is a significant barrier to external review (academic, legal, and other) of 'hate speech' policies. While the rules are public, the details of their implementation, at the aggregate and granular levels, are nearly non-existent.

International human rights framework

The ICCPR

The ICCPR is legally binding on states that have ratified it and includes the right to freedom of opinion and expression, the right to freedom of religion or belief, and the right to equality and non-discrimination.

The right to freedom of opinion and expression

The right to freedom of expression is protected by Article 19 of the ICCPR. The scope of the right to freedom of expression is broad. It 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 UN HR Committee, the treaty body of independent experts monitoring states' compliance with the ICCPR, has affirmed the scope extends to the expression of opinions and ideas that others may find deeply offensive, and this may encompass discriminatory expression.¹²

While the right to freedom of expression is fundamental, it is not absolute. A state may, exceptionally, limit the right under Article 19(3) of the ICCPR, provided that the limitation is:

- provided for by law, so any law or regulation must be formulated with sufficient precision to enable individuals to regulate their conduct accordingly;
- in pursuit of a legitimate aim, listed exhaustively as: respect of the rights or reputations of others, or the protection of national security or of public order (order public), or of public health or morals; or
- necessary in a democratic society, requiring the state to demonstrate in a specific and individualised fashion 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.¹³

Thus, any limitation imposed by the state on the right to freedom of expression, including limiting 'hate speech', must conform to the strict requirements of this three-part test.

The HR Committee has made clear that limitations on forms of communication or expression disseminated over the internet must be justified according to the same criteria as non-electronic or 'offline' communications.¹⁴

Additionally, states are obliged by law under Article 20(2) to prohibit any advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence. However, criminalisation of forms of 'hate speech' should be reserved for serious cases, to be proven beyond reasonable doubt, while less serious cases should be addressed by means other than criminal law, taking into account, inter alia, the nature and extent of the impact on targeted persons and groups. The application of criminal sanctions should be governed by principles of legality, proportionality, and necessity.¹⁵

The right to freedom of religion or belief

The right to freedom of religion or belief is protected by Article 18 of the ICCPR. The right to freedom of religion or belief provides for and guarantees the right to freedom of thought, conscience, and religion or belief broadly, and protects the profession and practice of different kinds of beliefs, whether theistic, non-theistic, or atheistic, and the freedom not to disclose one's religion or belief.¹⁶ International law also guarantees and protects the right not to have a religious confession.¹⁷

Article 18 of the ICCPR protects the rights of individuals or communities to freedom of religion or belief, including the right to manifest their religion or belief in worship, practice, and teaching.¹⁸ It does not protect religions or beliefs, religious ideas, symbols, or personalities.¹⁹ Furthermore, the scope of the protection of Article 18 of the ICCPR does not insulate any religion or belief from [criticism or ridicule](#). The provision covers the protection of individuals and groups who hold interpretations of religions which depart from or may be insulting or offensive to the feelings of many other followers. Such groups may include religious minorities or artists.

Under Article 18(3), states may, exceptionally, limit the right to manifest one's freedom of religion or belief, but must prove that the limitations prescribed by law are necessary to

protect public safety, order, health, morals, or the fundamental rights and freedoms of others.

The right to equality and non-discrimination

The right to equality and non-discrimination is given legal force in Articles 2(1) and 26 of the ICCPR. They oblige states to guarantee equality in the enjoyment of human rights, including the right to freedom of expression.

Positive measures under international human rights standards

International human rights standards highlight the importance of a range of positive policy measures states should employ in order to more effectively address the root causes of religious intolerance, discrimination, and violence.

Human Rights Council Resolution 16/18

[Resolution 16/18](#) addresses ‘combating intolerance, negative stereotyping and stigmatisation of, and discrimination, incitement to violence, and violence against persons based on religion or belief’ and was adopted by consensus in March 2011. It is widely regarded as a landmark achievement of the HRC’s first decade. A parallel set of resolutions, containing the same action points, have been adopted at the UN General Assembly since 2011, with follow-up resolutions adopted by the UN HRC since 2012.

Resolution 16/18 commits states to addressing religious intolerance through promoting the interrelated and mutually reinforcing rights to freedom of expression, freedom of religion or belief, and non-discrimination.

The agreement replaced calls to combat the deeply problematic concept of ‘defamation of religions’ with commitments to address religious intolerance through promoting the related rights to freedom of expression, freedom of religion or belief, and non-discrimination. Reiterations of the resolution have been adopted by consensus at the HRC every year since.²⁰

The resolution sets out an eight-point action plan for states to speak out against hatred, and initiatives to foster greater inclusion, diversity, and pluralism. Only one point pertains to limitations on expression.

Eight-point action plan

- 5a. *Create collaborative networks to build mutual understanding, promote dialogue and inspire constructive action in various fields.*
- 5b. *Create a mechanism within governments to identify and address potential areas of tension between members of different religious communities, and assist with conflict prevention and mediation.*
- 5c. *Train government officials in effective outreach strategies.*
- 5d. *Encourage efforts of leaders to discuss within their communities the causes of discrimination, and evolve strategies to counter them.*
- 5e. *Speak out against intolerance, including advocacy of religious hatred that constitutes incitement to discrimination, hostility or violence.*
- 5f. *Adopt measures to criminalise incitement to imminent violence based on religion or belief.*
- 5g. *Combat denigration and negative religious stereotyping of persons, as well as incitement to religious hatred, including through education and awareness-building.*
- 5h. *Recognise that the open, constructive and respectful debate of ideas plays a positive role in combating religious hatred, incitement and violence.*

Istanbul Process

The Istanbul Process is a series of meetings, initiated in 2011, to promote and guide implementation of Resolution 16/18. It was conceived of as a space for various stakeholders to exchange good practices and experiences of implementing the Resolution 16/18 action plan, outside of the sphere of multilateral politics. The Istanbul Process has enormous potential to be a strong vehicle for the implementation of Resolution 16/18, and to identify and replicate innovative and human rights-compatible approaches to promoting inclusivity, pluralism, and diversity.

ARTICLE 19 and Universal Rights Group have developed the [Istanbul Process 16/18 website](#) which provides information on the Istanbul Process and a bank of best practices and resources for implementing Resolution 16/18.

Rabat Plan of Action

The RPA on ‘the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to hostility, discrimination or violence’, developed by international experts with the support of the UN Office of the High Commissioner on Human Rights (OHCHR), provides practical legal and policy guidance to states on implementing Article 20(2) of the ICCPR, which requires states to prohibit certain severe forms of ‘hate speech’.

In particular, the RPA outlines the following:

- **Incitement:** Prohibitions should only focus on the advocacy of discriminatory hatred that constitutes incitement to hostility, discrimination, or violence, rather than the advocacy of hatred without regard to its tendency to incite action by the audience against a protected group.
- **A high threshold for limitations on expression:** The RPA proposes a six-part ‘threshold’ test for determining when limitations on ‘incitement’ are necessary under the

requirements of the ICCPR. Those factors are: (i) the social and political context; (ii) the speaker, e.g. their status and influence over their audience; (iii) the intent of the speaker; (iv) the content and form of the expression; (v) the extent of the expression; (vi) the likelihood and imminence of discrimination, hostility, or violence occurring as a direct consequence of the expression.

- **Protected characteristics:** States' obligations to protect the right to equality more broadly, with an open-ended list of protected characteristics, supports an expansive interpretation of the limited protected characteristics in Article 20(2) of the ICCPR to provide equal protection to other individuals and groups who may similarly be targeted for discrimination or violence on the basis of other recognised protected characteristics.
- **Proportionate sanctions:** The term 'prohibit by law' does not mean criminalisation. The HR Committee said it only requires states to provide 'appropriate sanction' in cases of incitement.²¹ Civil and administrative penalties will in many cases be most appropriate, with criminal sanctions an extreme measure of last resort.
- **Blasphemy laws:** The RPA expressly calls for the repeal of [blasphemy laws](#) for two main reasons. First, rather than encouraging mutual understanding, blasphemy laws fuel division by shutting down debates, often denying already marginalised groups the opportunity to speak or be heard, and are used to justify incitement to violence, as well as acts of violence by state and non-state actors, against those minority views. Second, international human rights law does not protect the ideas, religions, or beliefs themselves, as such. The right to freedom of religion or belief belongs to individuals, not religions.

The RPA also calls for a variety of positive policy measures from states, many of which are also found in the Resolution 16/18 action plan. In addition, it emphasises a number of other measures for states to take, including recommendations to:

- create 'equality bodies' or enhance the function of national human rights institutions (NHRIs) established in accordance with the Paris Principles to promote dialogue, but also in relation to accepting complaints about incidents of incitement under Article 20(2) of the ICCPR;
- create mechanisms and institutions to systematically collect data in relation to incitement under Article 20(2) of the ICCPR;
- have in place a public policy and a regulatory framework which promote pluralism and diversity of the media, including new media, and which promote universal and non-discriminatory access to and use of means of communication;
- promote and provide teacher training on human rights, and strengthen intercultural understanding as part of the school curriculum for pupils of all ages; and
- build the capacity of security forces, law enforcement agents, and those involved in the administration of justice on issues concerning the prohibition of incitement under Article 20(2) of the ICCPR.

Additional commitments on freedom of religion or belief and freedom of expression in the digital sphere

UN Strategy and Plan of Action on Hate Speech

The [UN Strategy and Plan of Action on Hate Speech](#) sets out guidance for the UN system to address ‘hate speech’ at the national and global levels. Out of the 13 commitments, two concern the issue of ‘hate speech’ in the digital sphere. First, the UN system should establish and strengthen partnerships with new and traditional media to address ‘hate speech’ narratives and promote the values of tolerance, non-discrimination, pluralism, and freedom of opinion and expression. Second, using technology, UN entities should keep up with technological innovation and encourage more research on the relationship between the misuse of the internet and social media for spreading hate speech and the factors that drive individuals towards violence.

Implementing HRC Resolution 16/18 and related commitments in the digital sphere

Recommendations for states

ARTICLE 19 recommends that the implementation of the action points and recommendations of Resolution 16/18 and related commitments be premised on the following complementary areas of action:

1. Create an enabling environment for human rights protection.

States have an obligation to guarantee to everyone an equal protection of the rights to freedom of expression, freedom of religion or belief, and equality by creating an enabling environment where opportunities to expose and counter religious intolerance and hate are maximised.

In view of the principles that ‘human rights protections in an offline context must also apply to online speech’ and that ‘there should be no special category of online hate speech for which the penalties are higher than for offline hate speech’, ARTICLE 19 recommends that states take the following actions:

- Ensure that the right to freedom of expression and information, through any medium of communication, is enshrined in domestic constitutional provisions or their equivalent, in accordance with international human rights law. In particular, domestic constitutional provisions should set out clearly the scope of permissible restrictions on the right to freedom of expression, including that such restrictions must be provided by law, be narrowly defined to serve a legitimate interest recognised in the constitution, and be necessary in a democratic society to protect that interest.
- Enact or strengthen anti-discrimination legislation in line with states’ obligations, ensuring that such laws cover online forms of discrimination and include the broadest scope of protected characteristics recognised under international human rights law.

- Establish or strengthen the role of independent equality institutions or expand the mandate of NHRIs²² to promote and protect the right to freedom of expression and religion both offline and online. Such institutions should be properly resourced with mandates to:
 - review complaints of discrimination occurring online on the grounds of the broadest scope of protected characteristics;
 - monitor and gather data on online hate based on religion or belief;
 - conduct research and thematic studies,²³ in collaboration with civil society, research institutions, and academia on freedom of expression and freedom of religion or belief; and
 - develop counter-narratives to address online hate based on religion.
- Strictly define and prohibit the advocacy of discriminatory hatred constituting incitement to hostility, discrimination, or violence in compliance with Articles 19(3) and 20(2) of the ICCPR and the guidance of the RPA. The list of prohibited grounds on incitement should be non-exhaustive and should cover grounds not mentioned in Article 20(2) of the ICCPR.
- Ensure that individuals may have access to justice and remedies when suffering cognisable harms relating to Article 20(2) of the ICCPR in the digital sphere.
- Refrain from adopting criminal approaches to broad and open-ended concepts such as online abuse or online harassment, as well as to expressions such as insults and offensive or false content, which are easily abused and can be used against the very interests of religious minorities.
- Develop guidelines for law enforcement authorities to prevent, investigate, and prosecute online hatred based on religion or belief in line with international human rights standards such as the six-part test of the RPA and with the involvement of religious minorities for greater understanding of the types of intolerance, discrimination, and violence they face in the digital sphere.²⁴

- Ensure that any state regulation of media supports [effective and independent self-regulatory mechanisms](#), with a comprehensive approach to developing and upholding media ethics, as a preferable model to press regulation.
- Ensure that any state regulation of internet companies has transparency, accountability, and the protection of human rights at its heart. In particular, refrain from imposing pressure on internet intermediaries to censor or restrict third party content, and ensure that intermediaries are not liable for any content relating to their services unless they specifically intervene in that content or refuse to obey an order adopted in accordance with due process guarantees by an independent, impartial, and authoritative oversight body.
- Publish transparency reports on all content-related requests issued to intermediaries and involve civil society organisations in all regulatory considerations.
- Ensure that human rights organisations, religious organisations, and internet companies are consulted with regard to law reform or any legislation and policies that address tackling hate online.
- Enhance efforts to promote connectivity and ensure universal access to the internet for religious minorities, including through promoting digital literacy, with special measures developed for those with disabilities or living in remote and rural areas. This includes policies that create an enabling and inclusive internet regulatory environment for small and non-profit internet operators.

2. Adopt a range of positive measures to promote human rights.

Both Resolution 16/18 and the RPA rest on the rationale that the promotion of inclusion, diversity, and pluralism is the best antidote to intolerant expression, coupled with positive measures to tackle the root cause of discrimination.

In this section, ARTICLE 19 outlines a variety of measures for the implementation of Resolution 16/18 and related commitments in order to address intolerance and hate based on religion or belief, in the digital sphere. We propose that states adopt comprehensive national implementation plans with full and effective participation and

cooperation with civil society, including religious organisations, academics, and internet companies, especially social media platforms.

Recognising and speaking out against religious intolerance and hatred

Public officials, including politicians, have a key role to play in recognising and promptly speaking out against intolerance, discrimination, and violence based on religion or belief. In this area, states should:

- adopt a communication strategy and communication plan for broader dialogue to counter religious intolerance and hatred online, including by adopting policies to ensure that public officials avoid making public statements that promote discrimination against persons based on religion or belief online.

Fostering dialogue and public debate

The most effective means of challenging religious intolerance and hatred is through informed debate and dialogue that aim to expose prejudice and build inclusive, diverse, and pluralistic societies. In order to foster dialogue and public debate, ARTICLE 19 recommends that states take the following actions:

- Ensure an environment for open, robust debate and dialogue, taking the necessary and appropriate measures to promote free, open, interoperable, reliable, and secure access to the internet.
- Encourage dialogue with and within religious communities on various different online platforms to better understand what freedom of religion or belief entails. The dialogue should cover areas in which there is – real or perceived – conflict, e.g. issues related to religious minorities, freedom of expression online, access to information, or access to justice. This can include a wide variety of approaches, including intra-faith dialogues, community councils, and social media campaigns, among other initiatives.
- Develop and introduce early warning mechanisms that monitor tensions within the society or between different religious communities both offline and online, with special

focus on social media platforms, and encourage alternative mechanisms for intercommunal interaction and dialogue where tensions occur.

Public education, information campaigns, and training

Education, information campaigns, and training are essential in order to [address and respond to online hate](#) based on religion or belief. Therefore, states should:

- promote and invest in digital, media, and information literacy, including knowledge and skills to identify and counteract messages of hatred, and knowledge on freedom of expression, and freedom of religion and belief in the online environment;²⁵
- integrate digital, media, and information literacy education into primary, secondary, and tertiary curricula;
- promote and provide training for public officials and state institutions including law enforcement agencies on freedom of expression, freedom of religion or belief, and equality in the context of the digital sphere;
- promote and provide special training for police, prosecutors, and the judiciary on the application of incitement standards under Article 20(2) of the ICCPR in the context of the digital sphere with the guidance of the RPA;
- collaborate with NGOs, equality bodies, religious institutions, and international organisations on education campaigns, awareness-raising, and equality training.

Monitoring and data collection

In order to effectively address online religious hatred through positive measures, it is important to assess and understand its characteristics. Therefore, states should take the following actions:

- Establish monitoring mechanisms and collaborate with different stakeholders to gather qualitative and quantitative data on online hate based on religion or belief. Data collection must be disaggregated, including on the basis of religious and other relevant minorities (gender, sexual orientation, and gender identity).

- Collaborate with NGOs, equality bodies, and religious institutions to monitor religious online hate.
- Support and conduct research based on the data gathered for policy responses.

Recommendations for other stakeholders

Online hate based on religion or belief cannot be addressed by states alone. A whole-society response is needed, mobilising multiple stakeholders, to tackle the root causes of discrimination, in line with states' obligations and commitments under international human rights law. The RPA also emphasises the role of non-state actors in speaking out against 'hate speech' and countering intolerance.

Recommendations for internet companies

ARTICLE 19 recommends that internet companies should:

- ensure international standards on freedom of expression, and freedom of religion and belief are embedded in their terms and conditions and community guidelines;
- conduct due diligence and impact assessments to prevent or mitigate any adverse impact on human rights resulting from their operations, products, or services;
- review their business models and advertising policies to reverse commercial incentives to amplify the spread of religious intolerance online;
- define the category of content that they consider to be 'hate speech' with reasoned explanations for users and the public, and approaches that are consistent across jurisdictions;
- ensure that any enforcement of 'hate speech' rules involves an evaluation of context and the harm that the content causes to users and the public, in compliance with the recommendations of the RPA, including by ensuring that any use of automation or AI tools includes human involvement;

- ensure that contextual analysis involves the communities – including religious minorities – most affected by content identified as ‘hate speech’ and that communities are involved in identifying the most effective tools to address harms caused on the platforms;
- promote education, counter-speech, reporting, and training on religious intolerance and ‘hate speech’ as alternatives, when appropriate, to the banning of accounts and the removal of content;
- evaluate how their products and services affect the human rights of their users and the public, in particular through periodic and publicly available human rights impact assessments.

ARTICLE 19 proposes a new model of effective self-regulation for internet companies called [Social Media Councils](#) (SMCs). SMCs would operate as voluntary mechanisms to moderate the content practices of the participating platforms. All participants (social media platforms and other stakeholders) sign up to a mechanism that does not create legal obligations for either party. Its strength and efficiency rely on voluntary compliance by platforms, whose commitment, when signing up, will be to respect and execute the SMC’s decisions (or recommendations) in good faith. These councils would be autonomous and transparent organisms in which various stakeholders have a say. Ideally, the councils would serve as representatives of civil society as a whole, composed of unions and associations of journalists, academics, human rights groups, the advertising industry, and vulnerable groups.

SMCs would be guided by a code of principles based on international law in regard to freedom of expression and would operate in two major areas:

- elaborating guidelines for social media platforms, setting international standards for content moderation; and
- functioning as an appeals mechanism, reviewing cases on an individual basis. For this task, the first approach would be algorithms and internal processes, leaving the SMC as the body for final appeals. Thus, big and difficult cases taken on by SMCs could set precedents to guide later decisions.

ARTICLE 19 believes that effective self-regulation could offer an appropriate framework through which to address the current problems with content moderation by social media companies, including religious ‘hate speech’ on their platforms, providing it also meets certain conditions of independence, openness to civil society participation, accountability, and effectiveness. Such a model could also allow for the adoption of tailored remedies without the threat of heavy legal sanctions.

Recommendations for civil society

Civil society organisations also play an important role in advancing the protection and promotion of human rights. 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 of expression and tackle inequality and discrimination. With regard to religious ‘hate speech’ online, civil society organisations are recommended to:

- map and collect data on religious ‘hate speech’ online with special focus on social media platforms – these data can be used to formulate campaigns and projects that strategically target different stakeholders;
- encourage and foster research on religious ‘hate speech’ online, its origins and causes, as well as the role of the media in spreading it;
- explore ways to promote constructive dialogues, and to mainstream debates and critical voices by allying with religious leaders;
- provide training to different stakeholders on ‘hate speech’ monitoring;
- provide online platforms to address and respond to ‘hate speech’ through debunking and providing space for unheard voices, including members of religious minorities.

Endnotes

¹ In this brief, under digital sphere, we understand the online space where people interact with each other and participate in electronic communication through the use of all segments of the technology sector that moderate content or act as gatekeepers. In the text, we will refer to these actors as digital services, intermediaries, or internet companies.

² See paragraph 1 of UN HRC (2021) [Resolution Adopted by the Human Rights Council on 13 July 2021 47/16. The Promotion, Protection and Enjoyment of Human Rights on the Internet \(A/HRC/RES/47/16\)](#). Geneva: UN.

³ As described in this brief, international standards require different responses to different types of ‘hate speech’ based on the level of severity. For these reasons, ARTICLE 19 uses the term in inverted commas.

⁴ For a full explanation of ARTICLE 19’s policy on ‘hate speech’, see page 8 of ARTICLE 19 (2015) [‘Hate Speech’ Explained: A Toolkit](#). London: ARTICLE 19.

⁵ See paragraph 43 of UN HR Committee (2011) [General Comment No. 34 – Article 19: Freedoms of Opinion and Expression \(CCPR/C/GC/34\)](#). Geneva: UN.

⁶ See the case of Konstantin Gerasimov in ARTICLE 19 (2019) [Rights in Extremis: Russia’s Anti-extremism Practices from an International Perspective](#). London: ARTICLE 19.

⁷ See paragraphs 75–76 of UN HRC (2011) [Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue \(A/HRC/17/27\)](#). Geneva: UN.

⁸ See paragraph 44 of UN HRC (2016) [Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression \(A/HRC/32/38\)](#). Geneva: UN.

⁹ See paragraphs 43–44 of [Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression \(A/HRC/32/38\)](#).

¹⁰ They have an enormous impact on human rights, particularly but not only in places where they are the predominant form of public and private expression, where a limitation of speech can amount to public silencing, or a failure to deal with incitement can facilitate offline violence and discrimination. See paragraphs 70–75 of UN HRC (2019) [Report of the Independent International Fact-finding Mission on Myanmar \(A/HRC/42/50\)](#). Geneva: UN.

¹¹ Problems also arise when platforms’ AI is poorly adapted to local languages, and companies have invested little in staff fluent in them. The consequences of ungoverned online hate can be tragic, as illustrated by Facebook’s failure to address incitement against the Rohingya Muslim community in Myanmar. See UN HCR (2018) [Report of the Independent International Fact-finding Mission on Myanmar \(A/HRC/39/64\)](#). Geneva: UN.

¹² See paragraph 11 of [General Comment No. 34 – Article 19: Freedoms of Opinion and Expression](#).

¹³ See paragraph 22 of [General Comment No. 34 – Article 19: Freedoms of Opinion and Expression](#).

¹⁴ See paragraph 43 of [General Comment No. 34 – Article 19: Freedoms of Opinion and Expression](#).

¹⁵ See paragraph 12 of UN Committee on the Elimination of Racial Discrimination (2013) [General Recommendation No. 35: Combating Racist Hate Speech](#). Geneva: UN.

¹⁶ See paragraphs 2 and 5 of UN HR Committee (1993) [General Comment No. 22: Article 18 \(Freedom of Thought, Conscience or Religion\)](#). Geneva: UN.

¹⁷ See paragraph 4 of [General Comment No. 22: Article 18 \(Freedom of Thought, Conscience or Religion\)](#).

¹⁸ See paragraph 4 of [General Comment No. 22: Article 18 \(Freedom of Thought, Conscience or Religion\)](#).

¹⁹ See paragraph 38 of UN HRC (2006) [Implementation of General Assembly Resolution 60/251 of 15 March 2006 Entitled ‘Human Rights Council’ \(A/HRC/2/3\)](#). Geneva: UN.

²⁰ See UN HRC (2017) [Resolution Adopted by the Human Rights Council on 24 March 2017: 34/32. Combating Intolerance, Negative Stereotyping and Stigmatization of, and Discrimination, Incitement to Violence and Violence Against, Persons Based on Religion or Belief. \(A/HRC/RES/34/32\)](#). Geneva: UN; and UN General Assembly (2018) [Resolution Adopted by the General Assembly on 19 December 2017: 72/176. Combating Intolerance, Negative Stereotyping, Stigmatization, Discrimination, Incitement to Violence and Violence Against Persons, Based on Religion or Belief \(A/RES/72/176\)](#). Geneva: UN.

²¹ See paragraph 2 of OHCHR (1983) [General Comment No. 11: Prohibition of Propaganda for War and Inciting National, Racial or Religious Hatred \(Art. 20\)](#). Geneva: UN.

²² All NHRIs should be established in accordance with the Principles Relating to the Status of National Institutions (Paris Principles) – UN General Assembly (1994) [Resolution Adopted by the General Assembly 48/134: National Institutions for the Promotion and Protection of Human Rights \(A/RES/48/134\)](#). Geneva: UN.

²³ As good practice, the project launched by the Hungarian Commissioner for Fundamental Rights is worth mentioning. The project, [Using Communication for Equal Dignity – Integrating Speech vs. Hate Speech](#), included several thematic studies examining different ways to prevent ‘hate speech’. This comprehensive study also included issues related to media regulation and analysis of the case law of the Media Council.

²⁴ See [Freedom of Religion or Belief and Security: Policy Guidance](#) from the Organization for Security and Co-operation in Europe Office for Democratic Institutions and Human Rights. The guidance encourages states to train law enforcement officials and the judiciary to understand and apply the six-part test set out in the RPA (context; speaker; intent; content or form; extent of the speech; and likelihood of harm occurring, including imminence) in order to determine whether the threshold of incitement to hatred is met or not.

²⁵ ‘MIL [media information literacy] is an umbrella concept that covers a package of literacies on- and offline. It includes the development of the technical skills and abilities required to use digital technologies, as well as the knowledge and abilities needed to find, analyse, evaluate and interpret specific media texts, to create media messages, and to recognise their social and political influence. Multiple and complementary literacies are seen as essential for the exercise of rights and responsibilities in regard to communications.’ UNESCO (2015) [World Trends in Freedom of Expression and Media Development: Special Digital Focus 2015](#). Paris: UNESCO, 51.