



- 0 00

- /

# A Guide for policy makers and practitioners

July 2023



society to develop context specific and national action plans on addressing and countering hate speech.

One of the commitments of the Strategy is to use new technologies and engage with social media to address online hate speech. In this regard, my Office has a longstanding engagement with technology and social media platforms to promote policies and practices on addressing online hate speech, in line with the UN Strategy. We have organized annually, since 2020, roundtable discussions with tech and social media companies on their role and responsibilities in addressing hate speech on their platforms, in line with international human rights norms and standards. These roundtables organized in collaboration with the ESRC Human Rights, Big Data and Technology Project at the University of Essex, also included the members of the United Nations Working Group on Hate Speech as well as Special Rapporteurs and civil society organizations working on this topic. The roundtables provide a platform for dialogue engagement and action on combating online hate speech. They aimed at looking beyond the limited scope of content removal to address hate speech online holistically, including through promoting positive narratives, warnings on problematic content, as well as reducing virality of posts and countering

also builds on the 2022 Policy Paper on \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

It is my firm hope that the recommendations in this document will help accelerate action to tackle online

hate speech and in doing so promoting and upholding our fundamental rights, in particular freedom of opinion and expression as well as to non-discrimination and equality.



**Alice Wairimu Nderitu**

*Under-Secretary General and Special Adviser on  
Prevention of Genocide to the United Nations Secretary  
General*

**0 /**



! 0 - / -  
- - / 0 ! 0 - - - 0  
⊠ - / - / -  
! 0 - /  
- / -00 .- / - -  
/ - / 0 - - - /

must also be protected online, in particular freedom of expression.<sup>1</sup>

International law. Under the UN Strategy and Plan of Action, hate speech is understood as any kind of









States must also be formulated to counter and address hate speech that does not reach this threshold through similar participatory efforts.

Measures taken to address hate speech must aim at . . . - / -00 . /- , offline and online. Such measures should not encourage mass surveillance<sup>7</sup>, criminalization of the exercise of freedom of expression as guaranteed under international law, undermine trust or attempt to regulate each and every piece of content.

Ensure that there is no impunity for advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence (ICCPR, article 20.2) as well as public and direct incitement to genocide as set out under the Convention on the Prevention and Punishment of the Crime of Genocide

Refrain from imposing internet shutdowns<sup>8</sup>, including as a response to hate speech. Internet shutdowns is damaging to a range of human rights. Keeping the internet on, and expanding connectivity for those on the other side of the digital divide are essential.

Ensure that official requests for - 0 - 0 - content follow existing

guidelines and are compliant with human rights norms and standards and enhance transparency of such requests, in line with the Rabat Plan of Action.

States institutions and public authorities should not weaponize social media to spread hate speech.

Refrain from imposing obligations on social media platforms to monitor online content generally, and require them to establish mechanisms to address specific content such as hate speech, especially when it amounts to incitement to discrimination, hostility or violence.

/ - 0 / - 0 - -

Conduct human rights due diligence to identify the risks the use of their services may pose to people, and take all reasonable steps to prevent or mitigate such risks, in accordance with the UN Guiding Principles on Business and Human Rights. Effective due di

---

<sup>7</sup> A/HRC/51/17

<sup>8</sup> A/HRC/50/55



Invest in improving the capacity and quality of / 0 - in all languages in which their platforms can be used.

Dedicate more resources to moderating content that could constitute incitement to discrimination, hostility or violence, or is particularly harmful, in countries or contexts that are fragile or high-risk, with an increased focus on dedicating resources to languages other than English.

Engage with the relevant entities at the United Nations, including the UN Office on Genocide Prevention and the Responsibility

*Enhance Transparency of Content Moderation,*

-

Engage in promoting **- - 0**  
/ - - to address and combat  
hate speech, including by promoting diverse and  
pluralistic journalism.

Develop / - **0** / -

- <sup>17</sup> and promote holistic equality measures  
to address root causes of hate.

Promote media and information literacy for  
the entire population and increase **0 - /** ,  
as a means to enhance resilience and empower  
individuals to identify and counter online hate  
speech.

/ - **0** / - **0** - -

Develop and implement **0- / -/**

- / in times of crisis such as conflict  
and humanitarian crisis, health emergencies where  
impact of hate speech is higher and may have more  
serious consequences on victims and groups  
affected.

Promote explicit and consistent policies to  
address online /- - **0** / **0 0** - as  
well as denial of other

on the engagement with individuals and groups, they should further promote conversations with technology and social media companies and State authorities on the shaping of protective policies, through knowledge from working closely with populations.

*Ensure Accountability, Strengthen Judicial Mechanisms and Enhance Independent Oversight Mechanisms*

! 0 - /

Strengt

Abide by human rights commitments,  
including human rights







Strengthen efforts to understand and tackle online hate speech in /- / , including language moderation, and enhance cooperation with UN on the ground and with national and local CSOs.

Tackle online hate speech in a way that is commensurate to the challenges posed by its volume, speed, and diversity. Pay increasing attention to - 0 0 0 - - 0 , diverse contexts and community standards, efforts and resources in place (including access to tools and channels to assist) where they are most needed to protect those individuals and groups mainly targeted by hate speech in social media.

Continue and strengthen emphasis on countering online hate speech against - / - - : minorities, migrants, refugees and internally displaced persons, women, sexual orientation and gender identity minorities, dissenters, human rights defenders, journalists, and other civil society voices representing targeted groups, as well as other groups in vulnerable positions (e.g. children).

Provide 0 - - 0 0- - to independent researchers, journalists and relevant bodies,

while respecting the privacy of users, to better understand how much content is managed automatically and how successfully/unsuccessfully the automated tools address key concerns on incitement to hatred, in order to enable further understanding of hate speech to inform 0 / 0.- 0 /

Amplify and support voices of /- 0 0 - - 0 / active in addressing and countering online and offline hateful rhetoric, and work with journalists to produce /

Conduct comprehensive research on the . /- - 0 - 0 - - - - as well as between the online hate and the offline harm and identify good practices in order to design multifaceted and grounded responses to hate speech.

Enhance research on existing / - 0 - / of measures adopted by technology and social media companies and online platforms to combat hate speech.

" /



Online hate speech remains a critical challenge to advancing the objectives set out in the UN Strategy and Plan of Action on Hate Speech and the pillars of the United Nations work, namely Peace and Security, Human Rights and Sustainable Development. The recommendations outlined in this policy paper, provide a framework for countering hate speech, in line with international human rights norms and standards. The recommendations are based on three years of consultations and dialogue, including with the technology and social media companies, experts, the UN Working Group on Hate Speech and civil society. Their implementation should be part of broader efforts to address hate speech globally, including its root causes and impact offline, in line with the UN Strategy and Plan of Action. The pursuit of these recommendations should also prioritize