POLICY OVERVIEW: A HUMAN RIGHTS APPROACH TO PROPOSALS FOR PREVENTING OR COUNTERING VIOLENT EXTREMISM ONLINE

Governments, policymakers, and law enforcement across the world are showing increased interest in pushing for proactive monitoring, surveilling, censoring, or otherwise modifying certain types of online content, under the broad rubric of "preventing" or "countering" violent extremism (PVE or CVE).

These proposals risk targeting satire, journalism, activism and organizing, political protest, and other forms of speech and undercutting existing rule of law and human rights safeguards. Troublingly, several proposals have suggested that companies and web platforms should proactively modify online content and communications or create opaque and unaccountable channels of cooperation — **despite the clear indication that such practices would undermine fundamental rights, the rule of law, and wider trust in the internet.**

Whilst it is important that policymakers respond to concerns around terrorism and violent extremism, **ill-conceived policies and practices will undoubtedly damage the ability of the world's peoples to use a free and open internet**.

Our paper[1] makes one thing quite clear: The concept of countering "violent extremism" and "extremism" online **should not be used as the basis for restricting freedom of expression, nor violating the right to privacy**.

An initiative to counter violent extremism must be grounded in a definition of that term that focuses on specific criminality, avoiding sweeping generalizations with identifiers such as ethnic origin, political affiliation, etc.

The definition must be anchored in an accountable and independent legal system with adequate oversight in order to prevent abuse and ensure the right to appeal. Further, initiatives that employ tactics tantamount to surveillance must be conducted using the same human rights safeguards applicable to all communications surveillance.[2]

In order to address the lack of clarity and rights-invasive activities conducted in the area of CVE, Access Now has set out three high-level principles and subject-specific recommendations to protect users' rights:

PRINCIPLE ONE	**Foster dialogue and education transparently, without bias.** Efforts to counter violent extremism by promoting open dialogue or education online must be transparent and not privilege certain forms of speech.
PRINCIPLE TWO	**Respect users' privacy.** Any approach for countering violent extremism that constitutes surveillance — such as social media monitoring, algorithmic content reporting, or content referral programmes — must be subject to the same normative and legal restrictions applicable to communications surveillance in other contexts.
PRINCIPLE THREE	**Avoid coercion of private industry to undermine free expression protections.** Governments must not compel companies to conduct programs to counter violent extremism, either by advancing new legislation or by threatening to screen or censor speech outside of legal process.

The following recommendations, rooted in well-established human rights law, are specifically tailored for public officials and policymakers, companies, and civil society. While we do not seek with these recommendations to provide a complete guide for when actions taken for CVE are appropriate, they do provide a baseline for ensuring that human rights are not undermined in their pursuit.

PRINCIPLE ONE

Foster dialogue and education transparently, without bias

Efforts to counter violent extremism by promoting open dialogue or education online must be transparent and not privilege certain forms of speech.

For policymakers and public officials:

1. Adopt policy frameworks and legislative measures that favour internet-enabled independent journalism, blogging platforms, and investigative reporting; review existing legal measures and prosecution policies to prevent clamping down on this critical channel for disseminating facts and supporting dialogue.
2. Support — and prevent the chilling of — efforts to drive forward genuine academic inquiry conducted via the internet on issues connected with "violent extremism".
3. Explore ways to support efforts to create further dialogue using the internet, without preferential treatment for how content is openly disseminated. This could include methods such as helping genuine dialogue-supporting organizations and community leaders establish an online presence, funding public advertising (for example, providing publicly disclosed advertising grants to nonprofits or independent institutions that promote inter-community dialogue), or developing additional outreach and communications channels.

For companies:

1. Ensure that any efforts to provide support to groups working to counter violent extremism are transparent, sound in methodology, and do not endanger the furtherance of human rights. There is an urgent need for more transparency and understanding of impact for company programmes or pilot efforts in this regard.
2. Undertake further research and dialogue to explore how product design efforts — such as enabling direct replies — can support meaningful dialogue, discourage echo chambers, and reduce speech that directly incites violence.

For public-private partnerships:

1. Government and private sector partnerships for countering "violent extremism" should at a minimum follow transparency and disclosure norms in this space — including following regulations for national lobbying or state propaganda. There should be ongoing commitment to oversight of any such partnerships by independent government oversight agencies, civil society organisations, human rights experts, national human rights institutions, and multi-stakeholder groups. Any counter-narrative messaging paid for or otherwise supported by governments must be clearly labeled and attributed.

PRINCIPLE TWO

Respect users' privacy

Any approach for countering violent extremism that constitutes surveillance — such as social media monitoring, algorithmic content reporting, or content referral programmes — must be subject to the same normative and legal restrictions.

For policymakers and public officials:

1. Governments must not force or request online platforms to undertake actions regarding user data disclosure or other surveillance measures that are outside of rule-of-law processes that comply with international human rights law and policy (including the Necessary and Proportionate Principles).[3] For example, forcing platforms to proactively disclose "Protected Information"[4] regarding users alleged to be involved in violent extremism content is unacceptable if it does not follow the Necessary and Proportionate principles, such as those regarding legality, independent judicial consideration, due process, etc.
2. Given the potential for social media monitoring to interfere with the rights to free expression and privacy when it pertains to Protected Information, any such practices must be provided for by law in a manner compatible with the Necessary and Proportionate principles. In particular, this includes — but is not limited to — the following:

2.1. When information that is collected is Protected Information, then it should only be collected when it is both necessary and proportionate to a legitimate aim to do so. State action to authorise the collection of Protected Information must respect the Necessary and Proportionate principles, with a stepwise process regarding government application for information, judicial consideration, search, appeals and remedies, and international cooperation (if applicable)

2.2. Government agents should never seek access to Protected Information outside of legal process, and particularly not through misleading methods, such as by creating fake profiles to follow or "friend" a user.

PRINCIPLE THREE

Avoid coercion of private industry to undermine free expression protections

Governments must not compel companies to conduct programs to counter violent extremism, either by advancing new legislation or by threatening to screen or censor speech outside of legal process.

For policymakers and public officials:

1. Mass take-downs of content are often counterproductive, and should not take place until content is specifically identified as unlawful or illegal. Such efforts do not deter the cultivation of "violent extremism", and in fact may encourage it, inflaming resistance and helping "violent extremist" recruiters discredit platforms that might otherwise support online expression and debate.
2. Governments must not compel the expansion or influence Terms of Service agreements in ways that "deputise" or pressure corporations to carry out the aims of the state.
3. Government usage of platform "flagging" tools or other automated processes should not be allowed to become a channel to bypass corporate transparency or rule of law processes and human rights safeguards that normally govern governmental powers regarding restricting speech and expression.
4. Government steps targeting the removal of violent extremist content must operate within the restrictions placed by human rights standards and fundamental rights, and must not force or request platforms to remove content unless
    - it has been adjudicated to be unlawful or specifically ordered to be removed under rights respecting legal process; and
    - mechanisms for notice and redress for the accused speaker is provided for, within the relevant laws.

For companies:

1. Companies should strive to indicate the reason for removing content or banning an account, rather than merely communicating that a decision has been implemented.
2. Reporting/flagging tools, and appeal mechanisms when content or users are flagged, must meet high standards for transparency, accountability, and human rights remedy. This also extends to programmes to deputise "super-users", whether in or outside government, to report or flag allegedly violent extremist content on online platforms. Multi-stakeholder bodies — if properly constituted with the active and meaningful engagement of civil society — could oversee the development and deployment of these tools and mechanisms.
3. Companies should be wary about deploying intrusive programmes that implement proactive filtering and reporting of content using "voluntary" models that circumvent national law and international standards for interfering with free speech and expression.

For civil society and academia:

Many countries already have official policies or legal regimes in place regarding online speech. Civil society and academia should watchful of these policies and how they are interpreted and implemented, to ensure they are not used to silence speech or quell protest.

FOOTNOTES

[1] The forthcoming Access Now paper, A Human Rights Approach to Proposals for Preventing or Countering Violent Extremism Online, will be released in September.

[2] See International Principles on the Application of Human Rights to Commuications Surveillance, May 2014, https://necessaryandproportionate.org/principles (herinafter referred to as the "Necessary and Proportionate Principles").

[3] The Necessary and Proportionate Principles show how existing human rights law applies to digital surveillance. They have been widely adopted and signed by more than 200 organisations and 275,000 individuals globally including legal experts, political parties, and elected officials.

[4] Protected Information is information that includes, reflects, arises from, or is about a person's communications, and that is not readily available and easily accessible to the general public.

Access Now (accessnow.org) defends and extends the digital rights of users at risk around the world. By combining innovative policy, global advocacy, and direct technical support, we fight for open and secure communications for all.

The forthcoming full paper, *A Human Rights Approach to Proposals for Preventing or Countering Violent Extremism Online*, will be released in September. Visit our website at **accessnow.org** for more details.

For more information, please contact **Raman Jit Singh Chima** at **raman@accessnow.org**