Digital Platforms and Extremism: Are Content Controls Effective?

It’s easy to agree on the need to fight the spread of terrorism, violent extremism, and hate through digital medium. With each horrific act of terror or tragedy, the demand for better solutions becomes louder. The Christchurch Call to Action, signed on May 15, 2019, is the latest; an action plan joining governments and companies in an intensified effort to eliminate such content online. There is serious attention on the issue. But it seems that malicious actors are still one step ahead. How do we effectively stop them?

Session name: Religious Radicalism: A Combat 2.0

Panel date: 13 November 2018

The Paris Peace Forum convened stakeholders for its inaugural edition in November 2018 event to dig into this issue. The frank discussion swayed back and forth between the responsibility of the tech industry to act, the practical challenges of identifying troubling content (and fast enough), the role of governments, and the duty to civil liberties. Attention gravitated to technology platform providers, but not only. There was no easy answer except to significantly step up investment in this battle.

What’s the Problem

The internet and social media have become a vital tool in the spread of extremist views. It is now a dominant battleground for ideology and a sophisticated recruitment platform for issues ranging from religious radicalism to hardline political agendas. Extremist groups or individuals can develop high-quality content at low cost and easily reach worldwide to spread fear and ideological propaganda. Or even, as we saw in the Christchurch attack in New Zealand in March of 2019, using it to amplify terror through livestreaming. Society, governments, and technology providers know that we need to get ahead of this and neutralize the qualitative edge of the extremists and their online agendas. But success has been elusive.

The challenge is that internet platforms are open for everyone, built on the principle of creating a place for open dialogue and exchange of information. In today’s world, that openness is easily — and often — abused.
Over the last few years, efforts have been greatly intensified to block, remove, or temper abusive content ranging widely from violent extremism, hate, and child safety to regular spam mail. Tools to filter, identify, and respond are improving. But it’s not enough.

With the volume of content now transmitted over the internet and social media, it’s an uphill battle. At the Forum, Miriam Estrin from Google noted that the preceding quarter of 2018, Google removed 7.7 million pieces of content. Although the vast majority of that is spam and only a very small fraction of it relating to violent extremism, all it takes is one of these “low volume, high risk” pieces to leak through to potentially cause disastrous impact.

**Be Smart**

Digging for these needles in the haystack is not only a volume challenge. Algorithms still misfire in categorizing content. Engineers are intensely working to improve the efficacy of automated filters, but errors still occur too often.

Catching harmful content before it gets widely distributed is a race against time, as we saw with the livestreaming of the Christchurch attack in New Zealand. That attack was designed specifically to draw viral attention. At the outset, only a few hundred people were watching, which is already horrific. But over the next 24 hours, even after the original stream was cut, new versions circumvented the automatic controls of social media and the video was transmitted thousands of times, even unwittingly.

Violent extremism linked to religious radicalization is even further complicated. Hakim El Karoui from the Institut Montaigne described to the Forum that the challenge has shifted from a few years ago when the focus was the online spread of violent jihadism. Improved filtering systems have been relatively successful in shutting down these clear-cut cases that call openly for violence. (Although not fast enough as the Christchurch livestreaming debacle showed a few months later.)

Radical jihadist groups have pivoted their digital campaigns to concentrate on the ideology, maneuvering around content controls disposed towards freedom of expression and freedom of religion. But the threat remains; these groups are still recruiting, radicalizing, and inspiring violent acts through this ideological agenda.

El Karoui argued that internet and social media platforms must become more active in this fight. Staying neutral is, in reality, taking a side. Liberal content policies work in favor of the extremists, which have figured out how to use the algorithmic model to drive users to their content. With the level of resources these groups are pouring into internet sites and social media campaigns, it becomes a vicious cycle.

“Isn’t there a responsibility for moderates to match the organization and dedication of religious extremists?”
- Annette Young, France 24
Moderates don’t have the resources to compete. El Karoui called on the tech platforms not only to tighten their community standards but also to actively help moderates promote a counter-narrative. For example, adjust their algorithms in order to profile moderate speakers more prominently in searches.

Miriam Estrin discussed some ways that Google is working to amplify such voices and connect vulnerable individuals with positive content. With the “Creators for Change” program at YouTube (a subsidiary of Google), they help already popular voices, ranging from comedians to bloggers, who want to use their platform for positive counter-narratives. They also work with schools to educate children about what “us versus them” language looks like online and how to recognize it and report it. They are looking for ways to pool the unique resources of open tech platforms to fight this battle and amplify the counter-narrative. By simultaneously trying as many routes as possible — whether that be providing digital capacity-building, developing redirect campaigns, or funding civil society projects — they are trying to find creative solutions that work. While at the same time, investing heavily in cutting-edge technology to try to improve systems.

What’s at the End of the Tunnel

Two themes emerged from the Forum’s discussion. First, the need for more action on the technical side from the platform providers, whether that be detection and removal, or actively boosting moderate voices. Second, content removal strategies can only be part of the solution. The battle for ideas is much broader and requires actively disseminating a counter-narrative, beyond the digital arena.

On the technical side, platform providers are pouring money into it, recruiting intensively, and digging for technical methods to stem the problem. Priority mandates are coming from the top. The Google representative pointed out that by investing heavily in machine learning technology, they have automated the work of the equivalent of 180,000 people in this area, dramatically increasing speed and efficiency of review and removal. Over the last few years, leading companies have been strengthening their collaborative work to prevent terrorist exploitation of technology through initiatives such as the Global Internet Forum to Counter Terrorism.

Even if the implementation of filtering and review systems reaches a gold standard of reliability, there remains the challenge of finding the balance in censorship practices. Journalists in open societies know this tightrope well. The panel discussion debated this tension.

Some argued strongly in the public interest obligation for stricter standards on internet and social media platforms. Others questioned how far the private sector should go in curbing the free expression of ideas. One way in which the tech providers are trying to navigate this is by developing more options short of total removal. For example, by creating functions that can strip a video from using accessibility features (i.e., forwarding), thereby limiting the spread of borderline material.

Across the board, the panel acknowledged that removal strategies should only be part of the equation. Eiman Kheir of the African Union (AU), who leads the youth project iDove (Interfaith Dialogue on Violent Extremism), emphasized more attention on the drivers of vulnerability to recruitment online.

The iDove project works closely with individuals who have been through the radicalization path, often youth, to decipher how best to preempt it. From that perspective, she pointed out that tendencies to censor the ideas driving religious or political extremism must not go too far in cutting off information about these
topics. Educate and debunk the manipulative narrative, rather than stripping access to the debate. Otherwise, it leaves disaffected youth primed for targeting by hardliners who can tap into the simple fact that no one else is engaging them. Protecting a space for open dialogue is essential to combat the narrative of radicalization.

“We are losing the fight against extremism. Most approaches are top-down, military, border centric (city or country at most), not really involving civil society or youth. Solutions are often very traditional and don’t include technology.”

- Eiman Kheir, African Union

While leaving space for dialogue, Kheir emphasized that counternarratives need an aggressive push. She applauded efforts to amplify moderate voices and increase awareness of alternatives to the jihadist interpretation of Islam. She encouraged all actors to ramp up these efforts and added a nuance that they should also be better contextualized (language, region, social context). The ideological battle is transcending borders and so must solutions. Specifically, she warned that more attention is needed to Africa and Asia. For tech companies, that means more awareness feeding into the filtering analysis, more nuance of what they are reviewing.

Who’s on the Hot Seat

Most of the attention on this issue gravitates to the platform providers — stricter controls, unambiguous removal of dangerous content, smarter technology to catch abusive content, and much more financial investment in this battle. Rightly so, given their gateway role.

But how far can and should the private sector go? Should they be stepping into the realm of editorializing and reshaping narratives? Do they have a public interest obligation to combat radicalization by actively promoting a particular narrative? Where does the self-regulation of companies end and the social responsibility of governments take over?

More specifically, what can governments do, especially when they get into the tricky ground of censorship? In cases of direct threat, the choice is relatively clear. But in free societies it is much more delicate to define clear boundaries when it comes to the battle of ideas. The panel pointed out some areas to navigate. For example, actively fund and promote moderate counter-narratives, even if it means breaking away from neutrality and getting into a battle of values. Likewise, funding initiatives and projects like iDove that are working from the bottom-up, engaging the most vulnerable communities and learning from them. Use technology as an asset. If youth is a particular focus, then use the tools that reach them. Maximizing the opportunities of technology is as important as blocking the dangers.
What’s Next

These days, the major tech companies have defined clear standards to remove a broad swath of content that explicitly or implicitly is inciteful or dangerous. Most of them are working aggressively to tackle the challenge. Still, they need to become much more nimble and less ambiguous in their review and removal of abusive content, even while defending the freedom of expression and the foundational principles of a free and open internet. Extremist elements are ever-more adept at maneuvering the digital space and their threshold for success is much lower. But even with the most effective content controls, the ideology that drives discontent and opens the door for radicalization will remain. Strategies to combat extremism and radicalization need to look more comprehensively beyond the digital arena.

Given the urgency and the multi-faceted nature of the task, it will require a wide range of stakeholders and ideally a coordinated line of effort. There are natural tensions between the self-regulatory tendencies of the private sector and civil society or public sector responsibilities. But there are plenty of avenues for them to work together to test out ideas and scout for solutions. Governments have a particular advantage in their convening power; tech companies have significant resources for funding; civil society groups can work with sensitive communities in ways that others cannot. But solutions must be cross-cutting.

This is a fast-moving issue. All the stakeholders involved seem to recognize the danger. No one wants to see the consequences of a dangerous piece slipping through. The panel’s call for action: Do more, much more.
Panel Participants

The Forum thanks panel participants: Hakim El Karoui, Senior Fellow, Institut Montaigne; Miriam Estrin, Policy Manager for Europe, Middle East, and Africa, Google; Eiman Kheir, Diaspora Officer, African Union.

The panel was moderated by: Annette Young, Journalist France 24.

This is a publication of the Paris Peace Forum reflecting the debates at the Forum’s inaugural session in November 2018. It does not necessarily represent the conclusions of each individual participant.