GIFCT
Global Internet Forum
to Counter Terrorism

GIFCT Technical Approaches Working Group

Executive Summary

July 2021

The GIFCT Technical Approaches Working Group (TAWG) is a global multi-stakeholder group of experts and practitioners coming from tech companies, academia, government, and civil society. Tech Against Terrorism, the United Kingdom Home Office, and Facebook facilitated the group as co-leads to ensure perspectives from NGOs, government, and tech platforms.

From July 2020 through July 2021 the working group has been meeting every 4 to 6 weeks to consider strategic questions relating to technical approaches to preventing terrorist use of the internet. It has also evaluated gaps between the technical requirements of smaller tech companies and the availability of solutions in order to inform the overall strategy of the Working Group. The group's aim is to support the development and adoption of technological solutions to prevent and disrupt the spread of terrorist content online while assessing ethical and human rights concerns associated with various technical approaches.

As part of the TAWG, Tech Against Terrorism was commissioned to produce a Gap Analysis and Recommendations report for deploying technical solutions to tackle the terrorist use of the internet. The full report lays out in detail different considerations and approaches to developing a strategic framework for GIFCT to develop a portfolio of technical resources aimed at supporting smaller content sharing platforms in preventing terrorist use of their services.

From this report, the TAWG co-leads distilled some key themes and recommendations which they believe are critical to the success of any such framework and important for policymakers in government, tech companies, academia, and civil society to understand as they work together to combat this threat.

Key Themes

Ensure the economic viability of technical approaches: Establishing a fund to help finance and implement tech solutions to identify and tackle terrorist content on smaller platforms will help to ensure the advances in content moderation technology can be applied responsibly where they are most needed. Many smaller platforms, which are often the most affected by terrorist exploitation, cannot afford to develop or purchase tech solutions to deploy on their services. Further, even when having access to such solutions, smaller platforms may not have the capability to implement them. To that end, a fund that sponsors the development and deployment of such solutions on smaller platforms will be effective in tackling the dissemination of online terrorist content.

Look beyond content removal and focus on additional use cases. There is currently a range of tech solutions aimed at supporting companies in removing content. Tooling that can carry out photo and video matching, logo detection, language processing, and recidivism detection have all been used in various capacities by larger companies. While content removal is one part of the challenge faced by tech companies, it does not solve the entirety of the threats posed around the dissemination of terrorist content. Often, as large companies increase their capacities to identify, surface, and remove terrorist content, the adversarial shift is for this content to be displaced onto other less equipped platforms. Thus, smaller platforms require support in several stages of the content moderation process, not only in removing content.

Ensure that solutions are developed in accordance with appropriate and adequate transparency mechanisms which account for human rights concerns raised by civil society groups and experts to avoid risks posed by automation. A holistic approach is needed ensuring that human rights and transparency are integral to all phases of design, development, and maintenance of any technical approaches.

Focus on technology supporting collaboration between platforms to ensure shared best practices for content removal, both in terms of efficacy and transparency. An example of this includes the need for more harmonized emergency content removal processes in threat-to-life scenarios. Policymakers should first consider the technology needed to operationalize increased regulation; too often policy does not adequately consider the limited resources of smaller platforms and therefore runs the risk of stifling innovation and undermining competition. Collaboration can also come from learning best practices from other sectors and harms areas.

Support safety by design and regulatory risk assessments. Develop tools and approaches to support platforms in evaluating their features and tech stack to support improved regulatory requirements for risk assessments. In particular small platforms are likely to require support in meeting increased regulatory demands.

Recommendations

1. Devise a **success framework** before investing in technical approaches. This framework should aim to quantify the scale of the problem prior to the solution being implemented and then monitor progress against these objectives.
2. Formulate a **strategy** that encourages stakeholders to work together towards developing technical approaches to deliver success as defined in the success framework.
3. Devise a prioritized **roadmap** based on the magnitude of the threat and likely impact of investing in technical approaches, prioritizing platforms most in need and capabilities based on the greatest impact.
4. Establish a **fund** to finance and implement technical solutions to identify and mitigate the exploitation of smaller platforms by terrorists and violent extremists. The fund should be guided by the roadmap, strategy, and success framework.
5. Increase information sharing between policymakers, the intelligence community, tech platforms, and researchers looking at terrorist and violent extremist content on a systematic basis (legal and regulatory constraints permitting) with appropriate oversight and transparency.
6. Ensure technical solutions are considered alongside policy responses, legal requirements, transparency, and human rights.
7. Create better infrastructure for GIFCT and TaT to support smaller platforms in terms of membership, resources, and tooling.
8. Encourage large tech platforms to “open source” more of their content moderation technologies, including those that have been effective in combatting other high severity abuse type, and share more about how they approach terrorist and violent extremist content moderation for the benefit of smaller platforms.

Many thanks go to the working group participants. We look forward to seeing this working group evolve in response to the themes and recommendations brought forward by the group and this paper.

To learn more about the Global Internet Forum to Counter Terrorism (GIFCT), please visit our website or email outreach@gifct.org.