GIFCT Red Team Working Group:
Executive Summary

August 2023

By Tom Thorley, in collaboration with GIFCT's Red Team Working Group

This year has been dramatic in the trust and safety field, and changes in the online and technical landscape have led to significant challenges and opportunities for those seeking to prevent terrorism and violent extremism online.

GIFCT's Red Team Working Group (RTWG) convened a panel of experts from tech, civil society, and government to look at five of the most critical areas of technology: Generative AI, Social Media, Fediverse and Distributed Web Technologies, End-to-End Encryption, and 3D Printing. They projected ahead approximately two years to identify how terrorists may use these technologies and what should be done now to manage these risks and take advantage of the technologies' opportunities.

Across all of these technologies, they found that some terrorists and violent extremists were already experimenting with their use or actively exploiting them and that upcoming technology, policy, and legislative changes are likely to have a significant impact on how future adoption manifests. The extent to which these technologies become accessible and usable at scale will also determine how much adoption occurs in terrorist groups. Hence, safety by design and human rights impact assessments early in product development cycles are critical to managing the risk of exploitation by terrorists and violent extremists.

Generative artificial intelligence has exploded in popularity over the past year. In the future, TVE actors will be able to produce higher-quality content cheaper, easier, and faster, personalize recruitment efforts, and potentially leverage code generation models for malware and technical abuse. There is the potential for risk management through technical controls such as safeguards and detection mechanisms.

Since the inception of social media, violent extremists and terrorists have been attempting to exploit it to advance their goals. Today we see a diversity of actors, aligned with ideologies spanning Islamic Extremism, Accelerationism, Incel, and White Supremacy, moving between platforms and using different online services for every aspect of their activities. As much of this production of content and radicalization will continue to occur on alt-tech platforms that will not come to the table and even willfully support hate-based ideologies, having a strategy for addressing them will be key. Such a strategy should include diversifying GIFCT membership and building sustainable solutions that companies of all sizes can access, as well as interventions to address different parts of the tech stack while respecting fundamental and universal human rights.

Given that we expect the ease of production of terrorist and violent extremist content to increase, make use of new technologies, and adapt – and that we expect the networks disseminating such material to decentralize – positive interventions that encourage disengagement from hate-based ideologies will also increase in importance, as will interventions designed to address the root causes of the problems we face as a society.

At the same time, as novel technologies begin to shape the eco-system, end-to-end encryption continues to play a significant role in terrorist use of the internet, especially in supporting direct communication relating to attack planning or other topics that terrorist groups deem sensitive. When combined with other features, such as the ability to archive and share large quantities of data or host large groups with relative anonymity, these technologies remain extremely attractive to terrorists. However, the same technology is critical for protecting freedom of expression and privacy. As governments and tech companies seek to prevent terrorism, they should adopt strategies that minimize the impact on the human rights of everyday citizens.

Finally, 3D printing is being actively exploited by violent extremists, notably as part of efforts to produce craft firearms or parts for them. Much of the content shared online relating to this activity is not illegal, but as content begins to express terrorist and extremist points of view, reducing its reach is critical. Platforms should work together with civil society to exchange experiences and expertise and develop best practices and a strategy for handling such content to reduce potential harm while respecting freedom of expression and monitoring future innovations of this technology.

In summary, terrorists and violent extremists are adaptable by nature, and use resources available to them in a frugally innovative manner. This dynamic quality means it is hard to say what the impact of a given policy change or technical innovation will be on the community. Many other significant technologies are emerging, and terrorists and violent extremists have already been experimenting with some and exploiting others. RTWG has identified that work is needed now to explore technologies such as consumer drone technology, blockchain, non-fungible tokens (NFTs), and on-demand manufacturing of biochemical compounds (especially in combination with generative AI). Just over the horizon, the group also noted the potential impact of quantum computing, terabyte internet connectivity, and quantum encryption, making all of the challenges we currently face even more complex regarding the volume and velocity of risks to manage. To address these challenges, a framework for evaluating technology, its existing exploitation and potential for harm, and what measures can be taken to manage these risks in a human-rights-based way is needed.