

Social Media and its Impact on Terrorism and Violent Extremism in the Next 2-5 Years

GIFCT Red Team Working Group

September 20, 2023



About GIFCT Year 3 Working Group Outputs

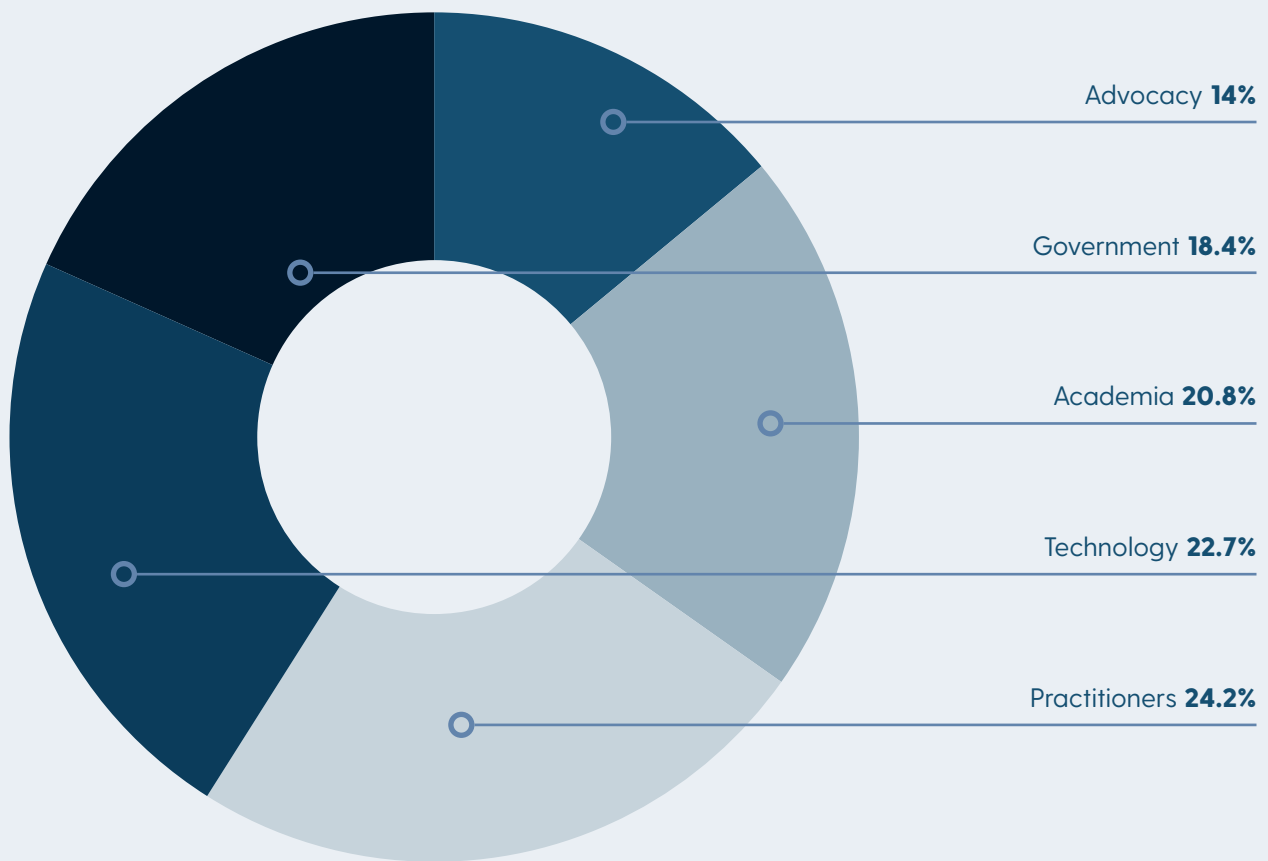
By Dr. Nagham El Karhili, Programming and Partnerships Lead, GIFCT

In November 2022, GIFCT launched its Year 3 Working Groups to facilitate dialogue, foster understanding, and produce outputs to directly support our mission of preventing terrorists and violent extremists from exploiting digital platforms across a range of sectors, geographies, and disciplines. Started in 2020, GIFCT Working Groups contribute to growing our organizational capacity to deliver guidance and solutions to technology companies and practitioners working to counter terrorism and violent extremism.

Overall, this year's five thematic Working Groups convened 207 participants from 43 countries across six continents with 59% drawn from civil society (14% advocacy organizations, 20.8% academia, and 24.2% practitioners), 18.4% representing governments, and 22.7% in tech.

WG Participants

Sectoral Breakdown



Beginning in November 2022, GIFCT Year 3 Working Groups focused on the following themes and outputs:

- 1. Refining Incident Response: Building Nuance and Evaluation Frameworks:** This Working Group explored incident response processes and protocols of tech companies and the GIFCT resulting in a handbook. The handbook provides guidance on how to better measure and evaluate incident response around questions of transparency, communication, evaluation metrics, and human rights considerations.
- 2. Blue Teaming: Alternative Platforms for Positive Intervention:** After recognizing a gap in the online intervention space, this GIFCT Working Group focused on highlighting alternative platforms through a tailored playbook of approaches to further PVE/CVE efforts on a wider diversity of platforms. This included reviewing intervention tactics for approaching alternative social media platforms, gaming spaces, online marketplaces, and adversarial platforms.
- 3. Red Teaming: Assessing Threat and Safety by Design:** Looking at how the tech landscape is evolving in the next two to five years, this GIFCT Working Group worked to identify, and scrutinizes risk mitigation aspects of newer parts of the tech stack through a number of short blog posts, highlighting where safety-by-design efforts should evolve.
- 4. Legal Frameworks: Animated Explainers on Definitions of Terrorism and Violent Extremism:** This Working Group tackled questions around definitions of terrorism along with the impact that they have on minority communities through the production of two complementary animated videos. The videos are aimed to support the global counterterrorism and counter violent extremism community in understanding, developing, and considering how they may apply definitions of terrorism and violent extremism.
- 5. Frameworks for Meaningful Transparency:** In an effort to further the tech industry's continued commitment to transparency, this Working Group composed a report outlining the current state of play, various perspectives on barriers and risks around transparency reporting. While acknowledging the challenges, the Working Group provided cross sectoral views on what an ideal end state of meaningful transparency would be, along with guidance on ways to reach it.

We at GIFCT are grateful for all of the participants' hard work, time, and energy given to this year's Working Groups and look forward to what our next iteration will bring.

To see how Working Groups have evolved you can access Year One themes and outputs [HERE](#) and Year Two [HERE](#).

Social Media and its Impact on Terrorism and Violent Extremism in the Next 2-5 Years

Introduction

Since its inception, violent extremists and terrorists have been attempting to exploit social media to advance their goals, with propaganda and radicalization central to terrorist groups' engagement with these platforms. In 1985, Aryan Nations established *Aryan Liberty Net*, one of the first social media platforms.¹ Two decades later in the mid-2000s, the Islamic State exploited social media on an industrial scale, flooding platforms with propaganda. Since 2008, the Taliban have used emerging social media platforms alongside their own web presence, and Al Qaeda first circulated *Inspire Magazine* online in 2010. Today we see a diversity of actors aligned with ideologies including Islamic Extremism, Accelerationism, Incel, and White Supremacy, moving between platforms and using different online services for every aspect of their activities.

Terrorism and violent extremism continue to evolve dynamically, as do the technologies and platforms that make up social media. This paper aims to lay out the interplay that exists in different hypothetical situations based on parallel trends in the technology landscape and the violent extremist community while looking ahead to consider the impact on terrorism and violent extremism. In line with our mission, GIFCT is spearheading this theoretical investigation to ensure that we can prevent future exploitation of social media platforms by terrorists and violent extremists.

Scenario 1

Increased Regulation of and Liability for Social Media

Governments around the world are seeking to update regulations and legislation governing how social media operates and how content is moderated, but such changes carry significant implications. Examples of this include laws such as House Bill 20 in Texas, which seeks to prohibit social media companies from banning users' posts based on political viewpoints. Other legislation could conflict with such laws, such as regulations under the EU's Digital Services Act and Terrorism Content Online regulations, which aim to limit the spread of illegal content online and "establish a new set of obligations for private actors with the aim to create a secure and safe online environment for all."² There are also significant implications of a U.S. company like Meta being sued in a country like Kenya³ or Telegram

.....

¹ Wayne King, "COMPUTER NETWORK LINKS RIHTIST GROUPS AND OFFERS 'ENEMY' LIST," The New York Times, February 15, 1985. <https://www.nytimes.com/1985/02/15/us/computer-network-links-rihtist-groups-and-offers-enemy-list.html>.

² Eliska Pirkova, "The Digital Services Act: Your Guide to the EU's New Content Moderation Rules," Access Now, July 6, 2022, <https://www.accessnow.org/digital-services-act-eu-content-moderation-rules-guide/>.

³ Reuters, "Meta Can Be Sued in Kenya by Ex-Content Moderator, the Country's Court Rules," February 6, 2023, <https://www.reuters.com/technology/meta-can-be-sued-kenya-by-ex-content-moderator-countrys-court-rules-2023-02-06/>.

accepting a fine in Brazil rather than removing content that they felt did not breach terms of service.⁴ While in the in the U.S., the Supreme Court ruled against the family of a 2017 ISIS attack victim who sought to hold tech companies liable for allowing ISIS to use their platforms in its terrorism efforts⁵ the court left the question of the scope of the immunity that Section 230 grants to social media companies, declining to weigh in on this issue in the case of *Gonzalez vs Google*.⁶

We can easily imagine a future in which social media companies face increased liability for the content on their platforms and the recommendations that they make. With increased pressure to respond quickly to removal requests from law enforcement entities and a complex patchwork of different regulations applied in different jurisdictions, responses to these pressures could manifest in a multitude of different ways.

Mitigating the legal risks associated with changes in the regulatory landscape may lead to stricter content moderation practices – which necessarily err on the side of caution – in the form of removing, deranking, or otherwise avoiding recommending content that even comes close to breaching these regulations. Historically, stricter moderation has led to a greater impact on freedom of expression, especially for vulnerable groups where more nuanced decision-making is required. But the impact of stricter moderation on terrorism and violent extremism is unclear. More moderation is not necessarily better moderation when it comes to preventing terrorists and violent extremists from exploiting the internet, especially if increased moderation happens only on bigger social media platforms with the greatest global user bases, leading to further balkanization of the online landscape and driving more people to less regulated spaces. While an over-moderated space is far from ideal, a completely unmoderated space would be equally undesirable; as Aaron Rabinowitz has argued, “discourse really only works if it is properly moderated and everyone is committed to the system.”⁷

An alternative response to mitigating legal risks may be moving to federated technologies and networks such as [BlueSky](#) and [Nostr](#), which (alongside [Mastodon](#)) make use of decentralized technology. In a tech landscape dominated by a “fediverse,” the large social media companies would no longer “host” the content that users post, but instead the content would exist in a distributed network that is not owned by any one company. Social media companies would each provide a lens into this network, allowing users to view and interact with a distributed network of posts.

This technology could have a profound impact on the current approach to content moderation. While

.....
 4 Consultor Jurídico, “Plataformas São Ameaça à Democracia e Devem Ser Reguladas, Afirma Dino,” n.d., <https://www.conjur.com.br/2023-fev-24/plataformas-digitais-sao-ameaca-democracia-flavio-dino>; Henrique Lessa, “STF Multa Telegram em R\$ 1,2 milhão por não bloquear perfil de Nikolas Ferreira,” *Política*, <https://www.correiobraziliense.com.br/politica/2023/01/5068918-telegram-descumpre-decisao-e-mo-raes-multa-aplicativo-em-rs-12-milhao.html>.

5 Howe, A., & Amy-Howe. (2023). Supreme Court rules Twitter not liable for ISIS content. SCOTUSblog, <https://www.scotusblog.com/2023/05/supreme-court-rules-twitter-not-liable-for-isis-content/>

6 SCOTUSblog, *Gonzalez v. Google LLC*, March 15, 2023, <https://www.scotusblog.com/case-files/cases/gonzalez-v-google-llc/>.

7 Aaron Rabinowitz, “The Curse of Monster Island: A Four Year Experiment in Unmoderated Free Speech,” *The Skeptic*, October 15, 2020, <https://www.skeptic.org.uk/2020/10/monster-island-free-speech-experiment/>.

human rights can be positively impacted by federated networks by ensuring access and availability for all and giving individuals the power to express themselves freely, the very concept of content moderation would have to undergo revision. Control over what content is available online would change significantly, moving away from social media and content-sharing platforms. Moderation in a federated paradigm would become more about what can be accessed through a particular lens and who can post and interact through a particular portal. Some developers of these networks have stated that hosting providers will still be legally required to “remove illegal content according to their local laws,”⁸ assuming that hosting providers can be identified, jurisdiction can be determined, and enforcement is in any way viable.

Governments might react to this loss of control by seeking to further control the internet with full or partial shutdowns. Figures from Access Now show internet outages increased globally by 15 percent in 2021, leading to a less open internet and further degrading human rights.⁹ Many individuals, organizations, and businesses around the world depend on internet services for essential functions such as data storage and processing and financial transactions based in various countries. Aside from the significant human rights impacts, the interruption of access to these services inevitably reduces productivity – it disrupts the availability of service platforms (including e-government services) and payment systems, resulting in significant economic losses. An internet shutdown affects long-term investments in a country and presents major risks for a whole set of companies and investors, in particular those who develop infrastructure and/or services. Internet shutdowns can undermine trust and highlight that a country’s internet infrastructure is neither resilient nor reliable, making investment unattractive. Internet shutdowns erode the trust people place in the internet infrastructure to be available and reliable when needed to build and sustain their economic activities.

As regulations are developed and interpreted, we must be vigilant and careful in designing online spaces, the regulations themselves, and the appropriate trust and safety responses. At the forefront of our strategic thinking must be the online harms that we seek to mitigate, as well as the fundamental and universal human rights and vulnerable communities that may be disproportionately impacted. Safety by Design¹⁰ encourages technology companies to anticipate, detect, and eliminate online risks to make our digital environments safer and more inclusive, especially for those most at risk. Industry, government, and civil society must commit to greater collaboration to develop innovative content moderation approaches and regulatory frameworks that effectively balance rights in decentralized services.

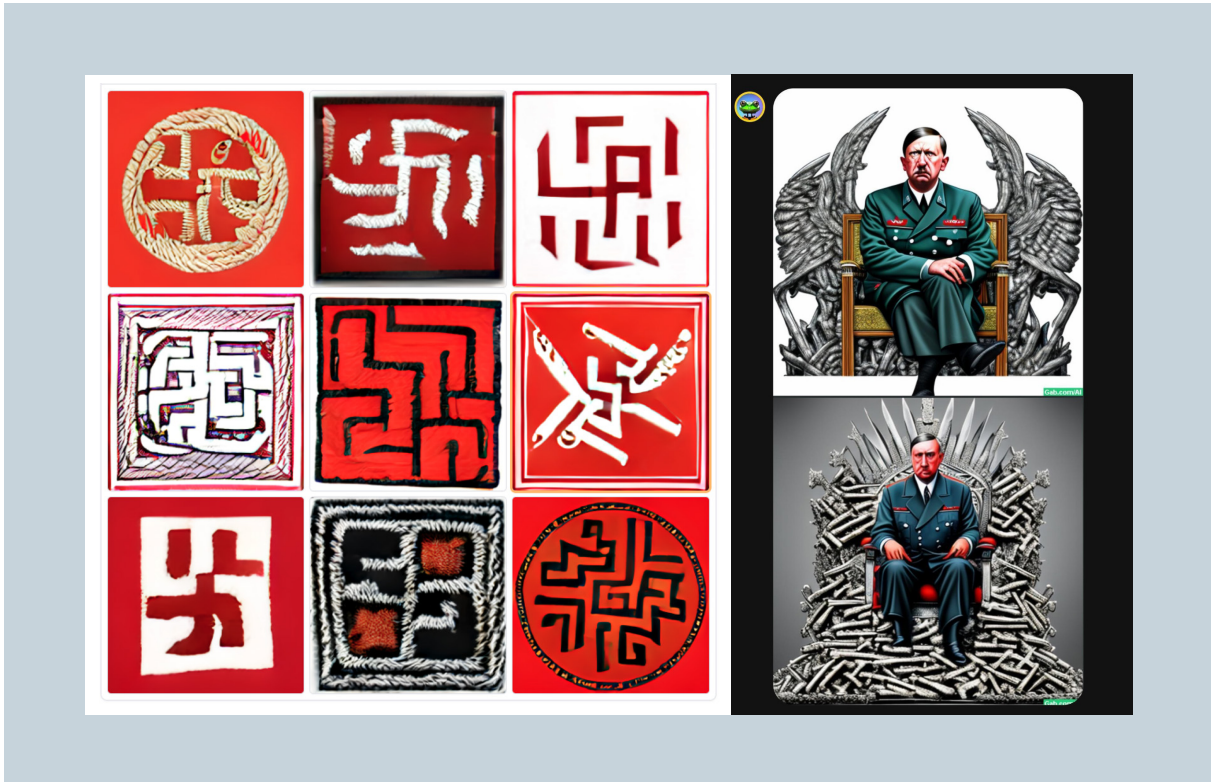
.....
 8 Brian Quarmby, “Decentralized Twitter’ Bluesky Releases Code, Outlines Content Moderation,” Cointelegraph, May 6, 2022, <https://cointelegraph.com/news/decentralized-twitter-bluesky-releases-code-outlines-content-moderation>.

9 Pirkova, “The Digital Services Act.”

10 Safety by Design (SbD). (n.d.). World Economic Forum. <https://www.weforum.org/projects/safety-by-design-sbd>

Scenario 2

Disruption of Social Media via Rapid Growth and Content Moderation Challenges



AI-Generated Images of "Day of the Rope Swastika" using Dalle-2 and "Hitler on Throne" using Gabby

Regulation comes and goes, but the kinds of policies that platforms are developing and building are aligned with broader strategies that do not only shift with the introduction of regulation. As we have seen recently with Chat-GPT, the entry of new technologies can have a significant impact on the market.

Generative AI is currently disrupting the search market and reigniting the so-called "search wars," and its potential impact on social media and content moderation could be equally significant. As users seek to generate content in ever more complex ways, there is a drive to provide new ways to create, manipulate, and recycle content. These technologies have already been disruptive in countering child sexual abuse material, because synthetic material is now mixed in with real child exploitation materials, making it harder to distinguish real from fake content, identify and protect victims, and prevent further abuse. When it comes to terrorism, the definitional challenges currently faced become more acute when encountering "synthetic material" and assessing harm. Generative technologies also offer the capability to alter videos and produce different versions of content at scale, exacerbating the moderation challenge. We have already seen examples of this with audio being generated that

appeared to be Emma Watson reading *Mein Kampf*,¹¹ as well as Gab releasing their own DALLE-2 style image generator and it immediately being used to create Pepe the Frog memes.¹² As large language models continue to improve, we can expect that they will increasingly be relied upon for content moderation purposes. However, we can also expect AI to be capable of navigating the edges of acceptable use policies in ways that undermine companies trying to tackle complex decisions on issues like incitement of violence.

We can imagine a future where social media apps make generative AI significantly more accessible, available in much the same way as filters are today. This accessibility and the ability to quickly scale content generation – creating endless permutations of material that requires highly complex subject matter expertise and decision-making to moderate – will require significant development and investment in evolving content moderation and other trust and safety interventions for identifying, moderating, and removing violative content. Technology needs to be developed to support moderators and trust and safety professionals, augmenting their expertise and helping to scale their work in an increasingly complex environment.

Scenario 3

Significant Growth in Violent Extremist Movements Leveraging New Technology

What governments do and what technology is developed are not the only drivers of change. As is often said by U.S. military commanders, “The Enemy Gets a Vote.” When thinking about the activity of individual terrorists and violent extremists online and its impact, we can identify two types of activity. First is their terrorist and violent extremist activity, which tends to occur more on “alt-tech” sites such as 4chan and Rumble. Second is normal activities like engaging with friends and family while conducting their everyday lives, just like any other internet user. This type of activity typically happens on mainstream platforms. While the content may allude to fairly extreme views, it is typically carefully kept within community guidelines, creating challenging edge cases for tech companies tasked with enforcing policies in a consistent manner without always having broader context about the specific user’s intent that could enable a clearer decision about whether the content violated their respective policies.

One recent example of this bifurcated use of social media can be seen in the case of Luke Kenna and Michael J. Brown. When they were charged with conspiracy to commit an armed bank robbery, they were investigated by the FBI Albany Field Office’s Joint Terrorism Task Force.¹³ The two men allegedly

.....
 11 Joseph Cox, “AI-Generated Voice Firm Clamps Down After 4chan Makes Celebrity Voices for Abuse,” Vice, January 30, 2023, <https://www.vice.com/en/article/dy7mww/ai-voice-firm-4chan-celebrity-voices-emma-watson-joe-rogan-elevenlabs>.

12 David Gilbert, “White Supremacist Networks Gab and 8kun Are Training Their Own AI Now,” Vice, February 22, 2023, <https://www.vice.com/en/article/epzjpn/ai-chatbot-white-supremacist-gab>.

13 U.S. Attorney’s Office, Northern District of New York, “Two Men Charged in Bank Robbery Conspiracy,” December 21, 2022, <https://www.justice.gov/usao-ndny/pr/two-men-charged-bank-robbery-conspiracy>.

ran an openly militant Neo-Nazi Telegram channel called “Aryan Compartmented Elements” (ACE).¹⁴ Meanwhile, both conducted their personal and business activities on Instagram and Facebook. Brown flirted with crossing the line regarding what was acceptable on the latter platform, offering services including “AK-47 & AR-15 Modifications,” and acted similarly on YouTube, posting “instructional videos” (since removed) on knife fighting. Both instances stayed within each platform’s acceptable use policies while hinting at more extreme views.

Terrorists and violent extremists are adaptable by nature; they use resources available to them in a frugal yet innovative manner, following the unofficial military motto “Semper Gumby.” This dynamic quality means that it is hard to say what the impact of a given policy change or technical innovation will be on the terrorist and violent extremist community. The pace of change regarding their tactics online is very different from the legislative pace of change; significant changes in how terrorists use the internet can and do occur between events prompting government action and regulations actually being passed. For example, regardless of efforts to constraint the use of the internet to recruit new terrorists, some degree of terrorist messaging is always going to take place on mainstream social media platforms, either by only persisting for a short time, masquerading as something more benign (e.g., legitimate news coverage), or carefully staying within acceptable use policies.

As we continue to operate in a fluid and dynamic terrorist landscape, we can expect the trends of decentralization to continue, moving from highly structured groups with clear ideologies to loosely connected movements with intersecting and overlapping belief systems. This also means that the risk of individuals carrying out attacks due to being inspired by others (but without direct coordination) will continue and escalate, as seen with both the attacks in Buffalo and Bratislava in 2022. This decentralization trend is complex, and when it comes to the production of content designed to radicalize, spread hate-based ideologies, and dehumanize certain groups, informal networks and collaborations seem likely to continue to flourish, especially on alt-tech platforms.

For the tech sector, this means that as new technology and new policies are developed, a reactive adversary is to be expected. In terms of preventing attacks, targeting propaganda and inspirational material will remain important, as well as instructional and attack planning material. However, disrupting networks that are driving radicalization will be increasingly important, and the community countering and preventing terrorism and violent extremism must develop more agility in order to keep pace with the dynamic nature of the threat while protecting human rights to access and utilize these platforms.

Conclusion

The multitude and diversity of on-going developments mean that the tech sector cannot merely plan for just one possible future. As noted above, definitions of both terrorism and violent extremism


.....
 14 Mack Lamoureux, “2 Men With Neo-Nazi Ties Arrested in Armed Bank Robbery Scheme,” Vice, December 21, 2022, <https://www.vice.com/en/article/4axeab/brown-kenna-bank-robbery-operation-werewolf>.

as well as technical concepts have impacts on the development and implementation of policy and technology.¹⁵ These definitions and designations lists will continue to be a challenge for the tech sector. The lack of clarity and agreement in the international community – as well as the pace of designations by governments – will not keep pace with the threat. This necessitates behavioral-based approaches to trust and safety on tech platforms, with sufficient investment in detection, moderation, enforcement, risk assessment, and policy development. Given the current economic climate in the tech sector, the fight for resources to achieve the increasingly complex trust and safety work required will only intensify.

At the same time, as much of this production of content and radicalization will continue to occur on alt-tech platforms, having a strategy for circumventing the negative effects of platforms that will not come to the table – and that willfully support hate-based ideologies – 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 the production of terrorist and violent extremist content will likely adapt and increase given the use of new technologies while the networks disseminating such material will likely 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.

.....
¹⁵ See GIFCT's Definitions and Principles Framework Site: <https://def-frameworks.gifct.org/>.



Copyright © Global Internet Forum to Counter Terrorism 2023

Recommended citation: Tom Thorley and GIFCT's Red Team Working Group, *Social Media and its Impact on Terrorism and Violent Extremism in the Next 2-5 Years* (Washington, D.C.: Global Internet Forum to Counter Terrorism, 2023), *Year 3 Working Groups*.

GIFCT is a 501(c)(3) non-profit organization and tech-led initiative with over 20 member tech companies offering unique settings for diverse stakeholders to identify and solve the most complex global challenges at the intersection of terrorism and technology. GIFCT's mission is to prevent terrorists and violent extremists from exploiting digital platforms through our vision of a world in which the technology sector marshals its collective creativity and capacity to render terrorists and violent extremists ineffective online. In every aspect of our work, we aim to be transparent, inclusive, and respectful of the fundamental and universal human rights that terrorists and violent extremists seek to undermine.



www.gifct.org



outreach@gifct.org