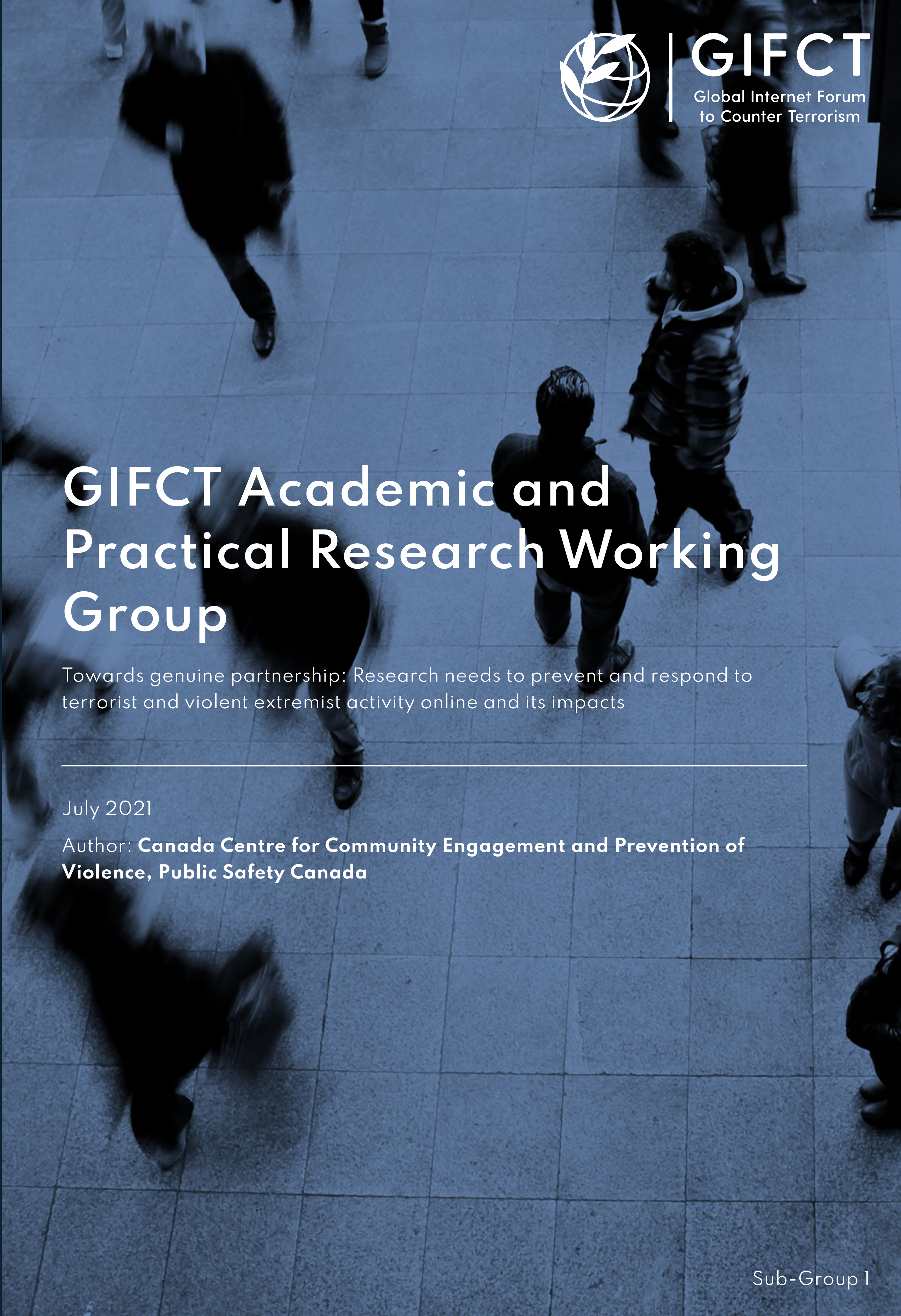

GIFCT
Global Internet Forum
to Counter Terrorism

GIFCT Academic and Practical Research Working Group

Towards genuine partnership: Research needs to prevent and respond to terrorist and violent extremist activity online and its impacts

July 2021

Author: **Canada Centre for Community Engagement and Prevention of Violence, Public Safety Canada**

Sub-Group 1

GIFCT

Global Internet Forum
to Counter Terrorism

Executive Summary

While access to data is often held up as the key challenge to better understand and address the activities and impacts of violent extremist and terrorist activity online, there are a number of other fundamental and interconnected barriers.

These include lack of standards and guidance for how to gather and use data that is or could be available in ways that are effective and appropriate; limited understanding and even fear among institutions that govern and fund relevant fields of academic research; narrow use of tools and methods, including owing to limited partnership across disciplines; and the still-early development of fields of research working on the difficult empirical challenges – such as studying small numbers of fringe actors in difficult-to-reach milieus – where more creativity in methodology is necessary for advancement. The connections across such barriers are evident, for example, when efforts to apply novel methods quickly run into challenges around ethics and identifying the right kinds of standards to ensure research subjects and researchers are protected.

There are a number of existing efforts to better understand and address these kinds of challenges within fields of research directly relevant to the mission and mandate of the Global Internet Forum to Counter Terrorism (GIFCT) and its working groups. As well, there are parallel efforts in nearby fields of research on platform governance, such as on the nature and effects of other forms of harmful and violent content or activity online.

Through early discussions among members of the GIFCT Academic and Practical Research Working Group's Sub-Group on barriers to research ('Sub-Group 1'), a priority identified for the near term is to develop a position paper to better bring together the kinds of structural, disciplinary challenges for advancing the research and evidence necessary to address the shared priorities of initiatives like GIFCT, as well as the Christchurch Call to Action, and of the stakeholders they are aiming to serve.

Aims would include supporting a shared understanding of the challenges, with data access needs showcased alongside those for data use and analysis; demonstrate that questions around areas like methodology, standards and ethics are already commonly considered; and help establish where more needs to be done to build a genuine, interdisciplinary partnership – a more established community of science – to take on the kinds of questions central to progress for the GIFCT.

Background

The first aim for the GIFCT Academic and Practical Research Working Group (APRWG) Sub-Group 1 was to produce a written assessment of the key barriers to research and knowledge transfer, along with opportunities and models for addressing them. The six GIFCT Working Groups – Academic and Practical Research; Content-Sharing Algorithms, Processes, and Positive Interventions; Crisis Response; Legal Frameworks; Technical Approaches; and Transparency – depend on reliable research and practice. As a result, the APRWG inhabits the unique position of identifying research needs that may ostensibly impact these working groups at a foundational level. This paper aims to assess the knowledge gaps and barriers that affect multiple stakeholders within the field of preventing and countering violent extremism (P/CVE) to begin moving towards holistic, coordinated solutions.

Preliminary discussions of the Sub-Group 1 members identified three main areas that could serve as a basis for this assessment:

1. Key questions central to the field's development including basic research about relevant phenomena, but also areas like evaluating impacts of policies and programs,
2. Limits to the research methods, practices, data, and access to information commonly used to try to answer those questions, and
3. Initiatives and ideas for helping to address those barriers, including in areas where funders, governments, and private companies could play a larger role

The following paper outlines, in greater detail, how Sub-Group 1 members explored each of these areas, and what options they proposed, ranging from where further work could begin immediately, to where greater planning and partnership would be required.

Best Practices

1. What best or good practice recommendations exist with regards to policies and programs for preventing radicalization online? Are these applicable to advanced digital platforms, and where do gaps still exist?
2. What evidence exists surrounding the impact of efforts to preventing/ countering online-related harms?
3. Have certain cultures, milieus or geographical areas demonstrated more or less resistance to online radicalization? If so, what are the protective/ vulnerability factors operating in these spaces?

Processes

1. What are the mechanisms by which online activity translates to offline violence?
2. What are the online dynamics of ideologically motivated violent extremist milieus? (I.e. xenophobic, neo-nazi, ethno-nationalist, anti-government, anti-law enforcement, violent misogynist, etc.)

Risk Mitigation

1. Should GIFCT's efforts focus primarily on a) countering the comparatively rare & complex phenomenon of offline terrorist violence; b) more common extremist online dynamics; c) the myriad of activities adjacent to violent extremism online; or all?
2. How do tech platforms counter misuse of services and disrupt the movement from online harm to offline violence?

Algorithms and Consumption

1. What are the links between dis/mis/mal-information and violent extremist dynamics online?
2. How does online extremism/harm support environments conducive to physical violence or offline manifestations of extremism?

In Addition

In addition to the questions identified here, other GIFCT Working Groups are motivated by additional core research questions, such as:

- How can tech organizations and technical solutions support the disruption of violent extremist content online?
- What are the ethical and human rights considerations when developing counter-violent extremism technology?
- How can transparency of resources and data be maximized while still respecting privacy and human rights?

More specifically, for example, the **Content-Sharing Algorithms, Processes, and Positive Interventions Working Group** aims to identify positive interventions and risk mitigation opportunities, which can minimize users' exposure and/or consumption of violent extremist and terrorist content. As such, there is alignment with several of the above research questions such as the role of platforms in processes of radicalization or mobilization to violence, as well as about evidence of impact of efforts at prevention. Similarly, the **Technical Approaches Working Group** aims to support the development and adoption of effective technological solutions to prevent and disrupt the spread of terrorist content online, while the **Crisis Response Working Group** aims to support multi-sector collaboration to minimize the spread of violent extremist and terrorist content online. By helping identify and address barriers to applied research relevant for the GIFCT, the Academic and Practical Research Working Group can support a number of the objectives of other working groups.

Barriers

In this context of significant and overlapping research questions, notable is how Sub-Group 1 members focused on the barriers that preclude answering them in an effective and substantive manner. Barriers discussed by members included limits on the availability and quality of data and research methods, such as guidelines and standards for their effective and appropriate use. These themes are explored in greater detail below.

The lack of standards and accessible methods used to monitor violent extremism online, along with a specific focus on ethical dimensions, was raised by several members as in need of significant improvement. Additionally, a significant barrier was the inability for researchers to access relevant data or information due to its classified nature, such as material found on seized media devices or material held by governments and tech companies, which could significantly assist the study of violent extremists and terrorists online. In the event that researchers and practitioners are able to access or gather data, participants conversely raised the issue of a lack of guidelines surrounding its use. Ambiguity around data use becomes particularly problematic when dealing with small n data, as is often the case for researchers gathering personal information about individuals who have committed a relatively rare act of offline violence.

Members of Sub-Group 1 also noted that addressing the online space is limited by a lack of knowledge and experience of emerging digital issues. These included limited bridging between the behavioral and computer sciences; insufficient use of offline and qualitative data to support richer understanding of the online; limited use and capabilities of linguistic analysis; limited training/awareness in some disciplines of tools, ethics, and guidelines around data gathering and usage; and underdeveloped knowledge on what methods work well to answer what kinds of questions.

Discussion repeatedly returned to research ethics in various ways. Protection of researchers themselves has been garnering significant attention of late, and was further discussed within the Sub-Group. Another common area of consideration was the challenge of getting ethics approval for various methods and approaches relevant for addressing some of the questions above, including to understand impact on vulnerable audiences such as children and youth, as well as studying fringe platforms and milieus that are difficult to access. Given the limited history of the field, one of the barriers is the lack of generally accepted guidance for such questions, both to inform research design, but also to inform the research ethics review process. Some members noted that some of the standards for human subject research and for data management used for ethics review and guidance are not always the right fit for this field of study, and that there are other available standards which could be a better fit.

Access to data remains a core concern for group members, with an example being the difficulty in studying the impacts and unintended consequences of content moderation, particularly where casting a wide net in removing terrorist/violent extremist content adversely affects activists and those working in the P/CVE space. Currently, little research exists to track or address the policy changes issued by online platforms and what recourse is available to those unintentionally affected, particularly where more access to various kinds of data – including about policy and decision-making processes – would be relevant.

Another common topic was the limitations of method and assumptions/ theory guiding it. This kind of barrier is exemplified by a number of studies cautioning against the use of automated hate speech detection. For over a decade, researchers have wrestled with the possibility of language detection online as a predictor of violent extremism[1]. One 2017 research study found, however, that certain types of offensive rhetoric are more likely to be mis-categorized by computer programs, leading to the potential for harmful language to go unnoticed[2]. As noted by Fernandez and Harith, the inability of automated language detection to correct for context poses both ethical and practical risks[3].

[1] Swati Agarwal, "Applying Social Media Intelligence for Predicting and Identifying On-Line Radicalization and Civil Unrest Oriented Threats" Link

[2] Thomas Davidson et al., "Automated Hate Speech Detection and the Problem of Offensive Language" Link

[3] Miriam Fernandez and Harith Alani, "Artificial Intelligence and Online Extremism: Challenges and Opportunities" Link ; Ghayda Hassan, Sébastien Brouillette-Alarie, Alava Séraphin et al., "Exposure to Extremist Online Content Could Lead to Violent Radicalization: A Systematic Review of Empirical Evidence," International Journal of Developmental Science 12, no.1-2 (2018):71-88

Working group members also emphasized that the existing limitations in the P/CVE field are often amplified by a lack of coordination across researchers. Members suggested that this could be improved by sharing lessons or guidelines among relevant actors, including technology companies, governments, or other academics.

Existing literature adds to the barriers raised by the Sub-Group. For example, in a 2018 systematic review, Hassan et al. found that the majority of existing studies of online extremism failed to demonstrate conclusive evidence of a relationship between exposure to extremist content online and violent radicalization offline[4]. The authors argue, as noted by working group participants above, that the nascent field of online extremist research has not yet developed reliable models of behavior or methodologies to support its claims.

Others note how the relationship between the online and offline space, while commonly raised as an urgent topic of discussion and research, is one where assumptions about causality, mechanisms, and choice of methods, can themselves be a barrier.

Other studies suggest that the online-offline worlds are interdependent and tend to respond to catalytic events in tandem[5]. However, the relationship between online extremism and offline violence in the absence of catalysts is less clear. Gallacher and Heerdink suggest that abusive online activity tends to precede offline violence by far-right groups[6], a finding that is supported by studies in the adjacent field of online hate speech[7] and interviews with former far-right extremists[8]. As suggested by Hassan et al, however, where such studies are based on unreliable methodologies, such findings should be treated with caution.

[4] Ghayda Hassan, Sébastien Brouillette-Alarie, Alava Séraphin et al., “Exposure to Extremist Online Content Could Lead to Violent Radicalization: A Systematic Review of Empirical Evidence,” *International Journal of Developmental Science 12, no.1-2 (2018):71-88*

[5] John D. Gallacher, Marc W. Heerdink, and Miles Hewstone, “Online Engagement Between Opposing Political Protest Groups via Social Media is Linked to Physical Violence of Offline Encounters”, Link

[6] John D. Gallacher and Marc W. Heerdink, “Mutual radicalization of opposing extremist groups via the Internet”, Link

[7] Matthew L. Williams, Pete Burnap, Amir Javed, Han Liu, and Sefa Ozalp, “Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime”, Link

[8] Tiana Gaudette and Ryan Scrivens, “The Role of the Internet in Facilitating Violent Extremism: Insights from Former Right-Wing Extremists”, Link

It is significant that many of these barriers are deeply interconnected and span across much of the research process as a whole. A lack of methods impede effective research design at its nascency, while methodologies for undertaking any such design may be stalled by difficulty accessing data or clear ethical guidelines for its use. The further lack of communication and knowledge-sharing among actors in the P/CVE space not only prevents the sharing of lessons learned, but also creates an – arguably artificial – barrier to the field's empirical development.

Ideas and Initiatives

The challenges identified by Sub-Group 1 point to an interconnected network of barriers facing researchers and practitioners alike. While this presents the need for multiple points of change, it also implies that each barrier addressed improves the process of P/CVE research and practice as a whole. Critical to this process is a joint effort across stakeholders.

Sub-Group 1 members identified a number of initiatives and ideas for helping address these limits and gaps, including in areas where funders, governments and companies could play a larger role. Members have proposed mapping funding streams to better coordinate research (subject of another deliverable for Sub-Group 1); host engagement events that include tech representatives to speak to the inner workings of their organizations; develop and publish short, op-ed styled research products; and, establish a knowledge hub and/or a forum to help address gaps, resources, guidance and methods for research on violent extremist and terrorist use of the internet. In addition, it was noted that coordination amongst the relevant GIFCT Working Groups' initiatives and/or objectives would be helpful, particularly in facilitating access to data necessary for fundamental research.

Externally of GIFCT, a number of initiatives have been undertaken to tackle the spread and dissemination of extremist content online. VOX-Pol has done significant work on the topic of developing and gathering resources to support research. In a 2019 report, the organization presented a collection of essays examining the practical efficacies and limitations of a variety of strategies, including automated language detection, counter-narratives, and crowdsourced flagging[9]. Additionally, publications such as those written by the Dangerous Speech Project share lessons learned about their efforts to use data science methods to identify dangerous speech as well as forms of counter-speech, and under what circumstances the latter can be effective[10].

Foundational to all of these initiatives is the need to create solutions that are applicable and relevant to all actors within the P/CVE space. Working group members emphasized the need to work towards building a community of science, in which both barriers and solutions are understood from a collective framework. In this way, data-driven research and practice can act as a supportive infrastructure through which to collaborate and improve P/CVE.

[9] Bharath Ganesh and Jonathan Bright, eds., "Extreme Digital Speech: Contexts, Responses, and Solutions"

[10] Link

Next Steps

A position paper on 'genuine partnership'

Given the kinds of research questions at the core of GIFCT efforts, the common barriers to achieving reliable results in ethical ways, and some of the solutions available and proposed, there are some potential next steps for Sub-Group 1 to take, including through work funded by GIFCT:

A. Solidify the assessment

Further desk review and engagement with additional experts, practitioners, and researchers in the field of digital counter-violent extremism to solidify the three areas identified in this paper: namely, the priority research questions; the main common barriers/limitations; and promising approaches to addressing them.

A more thorough treatment of these three areas can help inform a longer-term work program.

This exercise can/should include targeted outreach to and assistance from adjacent fields, such as those working on platform governance research more generally, as well as areas focusing on connected forms of harm that may not fall clearly under the banner of terrorism or violent extremism; hate; and/or grievance-fueled/targeted violence. That is, areas where there are closely similar research questions, limits to data/methods/practice, and development of solutions.

B. Deeper dives

Among the various topics identified above, in the meeting minutes, and through further discussion among Sub-Group members, targeted literature reviews could be undertaken to centralize readily available evidence on some of the key issue areas, particularly those relevant to GIFCT WGs. These deeper dives may also help identify commonalities across these groups and facilitate connections on issues where collaboration could be useful.

C. Solution development

Work towards one or more of the proposed solutions could begin in the near term, including though outreach to existing initiatives, as well as collaboration with other GIFCT WGs. Potential areas could include a knowledge sharing forum or resource hub, or even a data access initiative.

At the most recent meeting of Sub-group 1, members brought these ideas together to propose the drafting of a position paper, to address the needs, barriers, and proposed solutions from a multidisciplinary perspective.

Final Thoughts

In addition to expanding on aforementioned topics, the paper could demonstrate the creation of collaborative solutions that benefit and strengthen all actors within the P/CVE space. It could serve as a public call signaling the state of the discipline, how seriously questions around methods, standards, guidelines and multi-disciplinarity are being addressed, and as a resource for guiding work over the near term to more clearly establish and build the field of research relevant to supporting the mission and mandate of the GIFCT. A position paper also puts into action the concept of transparency, to which the GIFCT has dedicated a working group and uses as a guiding principle in all its efforts. By helping foreground the connected range of challenges and possibilities for applied research, the APRWG can help contribute to the various work streams for GIFCT and related initiatives such as the Christchurch Call to Action, along with the communities they are aiming to serve.

Critically, this paper could serve as a first step towards a genuine partnership of actors within the P/CVE space, and create an environment in which they can begin working towards solutions for the most far-reaching barriers to effective violence prevention.

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