GIFCT WORKING GROUPS OUTPUT 2022

# Introducing 2022 GIFCT Working Group Outputs

GIFCT
Global Internet Forum to Counter Terrorism

Dr. Erin Saltman
Director of Programming,
GIFCT

In July 2020, GIFCT launched a series of Working Groups to bring together experts from across sectors, geographies, and disciplines to offer advice in specific thematic areas and deliver on targeted, substantive projects to enhance and evolve counterterrorism and counter-extremism efforts online. Participation in Working Groups is voluntary and individuals or NGOs leading Working Group projects and outputs receive funding from GIFCT to help further their group's aims. Participants work with GIFCT to prepare strategic work plans, outline objectives, set goals, identify strategies, produce deliverables, and meet timelines. Working Group outputs are made public on the GIFCT website to benefit the widest community. Each year, after GIFCT's Annual Summit in July, groups are refreshed to update themes, focus areas, and participants.

From August 2021 to July 2022, GIFCT Working Groups focused on the following themes:

- Crisis Response & Incident Protocols
- Positive Interventions & Strategic Communications
- Technical Approaches: Tooling, Algorithms & Artificial Intelligence
- Transparency: Best Practices & Implementation
- Legal Frameworks

A total of 178 participants from 35 countries across six continents were picked to participate in this year's Working Groups. Applications to join groups are open to the public and participants are chosen based on ensuring each group is populated with subject matter experts from across different sectors and geographies, with a range of perspectives to address the topic. Working Group participants in 2021–2022 came from civil society (57%), national and international government bodies (26%), and technology companies (17%).

Participant diversity does not mean that everyone always agrees on approaches. In many cases, the aim is not to force group unanimity, but to find value in highlighting differences of opinion and develop empathy and greater understanding about the various ways that each sector identifies problems and looks to build solutions. At the end of the day, everyone involved in addressing violent extremist exploitation of digital platforms is working toward the same goal: countering terrorism while respecting human rights. The projects presented from this year's Working Groups highlight the many perspectives and approaches necessary to understand and effectively address the ever-evolving counterterrorism and violent extremism efforts in the online space. The following summarizes the thirteen outputs produced by the five Working Groups.

## Crisis Response Working Group (CRWG):

The GIFCT Working Group on Crisis Response feeds directly into improving and refining GIFCT's own Incident Response Framework, as well as posing broader questions about the role of law enforcement, tech companies, and wider civil society groups during and in the aftermath of a terrorist or violent extremist attack. CRWG produced three outputs. The largest of the three was an immersive virtual series of Crisis Response Tabletop Exercises, hosted by GIFCT's Director of Technology, Tom Thorley. The aim of the Tabletops was to build on previous Europol and Christchurch Call-led Crisis Response events, with a focus on human rights, internal communications, and external strategic communications in and around crisis scenarios. To share lessons learned and areas for

improvement and refinement, a summary of these cross-sector immersive events is included in the 2022 collection of Working Group papers.

The second output from the CRWG is a paper on the Human Rights Lifecycle of a Terrorist Incident, led by Dr. Farzaneh Badii. This paper discusses how best GIFCT and relevant stakeholders can apply human rights indicators and parameters into crisis response work based on the 2021 GIFCT Human Rights Impact Assessment and UN frameworks. To help practitioners integrate a human rights approach, the output highlights which and whose human rights are impacted during a terrorist incident and the ramifications involved.

The final CRWG output is on Crisis Response Protocols: Mapping & Gap Analysis , led by the New Zealand government in coordination with the wider Christchurch Call to Action. The paper maps crisis response protocols of GIFCT and partnered governments and outlines the role of tech companies and civil society within those protocols. Overall, the output identifies and analyzes the gaps and overlaps of protocols, and provides a set of recommendations for moving forward.

## Positive Interventions & Strategic Communications (PIWG):

The Positive Interventions and Strategic Communications Working Group developed two outputs to focus on advancing the prevention and counter-extremism activist space. The first is a paper led by Munir Zamir on Active Strategic Communications: Measuring Impact and Audience Engagement. This analysis highlights tactics and methodologies for turning passive content consumption of campaigns into active engagement online. The analysis tracks a variety of methodologies for yielding more impact-focused measurement and evaluation.

The second paper, led by Kesa White, is on Good Practices, Tools, and Safety Measures for Researchers. This paper discusses approaches and safeguarding mechanisms to ensure best practices online for online researchers and activists in the counterterrorism and counter-extremism sector. Recognizing that researchers and practitioners often put themselves or their target audiences at risk, the paper discusses do-no-harm principles and online tools for safety-by-design methodologies within personal, research, and practitioner online habits.

## Technical Approaches Working Group (TAWG):

As the dialogue on algorithms and the nexus with violent extremism has increased in recent years, the Technical Approaches Working Group worked to produce a longer report on Methodologies to Evaluate Content Sharing Algorithms & Processes led by GIFCT's Director of Technology Tom Thorley in collaboration with Emma Llanso and Dr. Chris Meserole. While Year 1 of Working Groups produced a paper identifying the types of algorithms that pose major concerns to the CVE and counterterrorism sector, Year 2 output explores research questions at the intersection of algorithms, users and TVEC, the feasibility of various methodologies and the challenges and debates facing research in this area.

To further this technical work into Year 3, TAWG has worked with GIFCT to release a Research Call

for Proposals funded by GIFCT. This Call for Proposals is on Machine Translation. Specifically, it will allow third parties to develop tooling based on the gap analysis from last year's TAWG Gap Analysis. Specifically, it seeks to develop a multilingual machine learning system addressing violent extremist contexts.

## Transparency Working Group (TWG):

The Transparency Working Group produced two outputs to guide and evolve the conversation about transparency in relation to practitioners, governments, and tech companies. The first output, led by Dr. Joe Whittaker, focuses on researcher transparency in analyzing algorithmic systems. The paper on Recommendation Algorithms and Extremist Content: A Review of Empirical Evidence reviews how researchers have attempted to analyze content-sharing algorithms and indicates suggested best practices for researchers in terms of framing, methodologies, and transparency. It also contains recommendations for sustainable and replicable research.

The second output, led by Dr. Courtney Radsch, reports on Transparency Reporting: Good Practices and Lessons from Global Assessment Frameworks. The paper highlights broader framing for the questions around transparency reporting, the needs of various sectors for transparency, and questions around what meaningful transparency looks like.

## The Legal Frameworks Working Group (LFWG):

The Legal Frameworks Working Group produced two complementary outputs.

The first LFWG output is about Privacy and Data Protection/Access led by Dia Kayyali. This White Paper reviews the implications and applications of the EU's Digital Services Act (DSA) and the General Data Protection Regulation (GDPR). This includes case studies on Yemen and Ukraine, a data taxonomy, and legal research on the Stored Communications Act.

The second LFWG output focuses on terrorist definitions and compliments GIFCT's wider Definitional Frameworks and Principles work. This output, led by Dr. Katy Vaughan, is on The Interoperability of Terrorism Definitions. This paper focuses on the interoperability, consistency, and coherence of terrorism definitions across a number of countries, international organizations, and tech platforms. Notably, it highlights legal issues around defining terrorism based largely on government lists and how they are applied online.

## Research on Algorithmic Amplification:

Finally, due to the increased concern from governments and human rights networks about the potential link between algorithmic amplification and violent extremist radicalization, GIFCT commissioned Dr. Jazz Rowa to sit across three of GIFCT's Working Groups to develop an extensive paper providing an analytical framework through the lens of human security to better understand the relation between algorithms and processes of radicalization. Dr. Rowa participated in the Transparency, Technical Approaches, and Legal Frameworks Working Groups to gain insight into

the real and perceived threat from algorithmic amplification. This research looks at the contextuality of algorithms, the current public policy environment, and human rights as a cross-cutting issue. In reviewing technical and human processes, she also looks at the potential agency played by algorithms, governments, users, and platforms more broadly to better understand causality.

We at GIFCT hope that these fourteen outputs are of utility to the widest range of international stakeholders possible. While we are an organization that was founded by technology companies to aid the wider tech landscape in preventing terrorist and violent extremist exploitation online, we believe it is only through this multistakeholder approach that we can yield meaningful and long-lasting progress against a constantly evolving adversarial threat.

We look forward to the refreshed Working Groups commencing in September 2022 and remain grateful for all the time and energy given to these efforts by our Working Group participants.

## Participant Affiliations in the August 2021 - July 2022 Working Groups:

| Tech Sector | Government Sector | Civil Society / Academia / Practitioners | Civil Society / Academia / Practitioners |
|---|---|---|---|
| ActiveFence | Aqaba Process | Access Now | Lowy Institute |
| Amazon | Association Rwandaise de Défense des Droits de l'Homme | Anti-Defamation League (ADL) | M&C Saatchi World Services Partner |
| Automattic | Australian Government - Department of Home Affairs | American University | Mnemonic |
| Checkstep Ltd. | BMI Germany | ARTICLE 19 | Moonshot |
| Dailymotion | Canadian Government | Australian Muslim Advocacy Network (AMAN) | Moduslzad - Centre for applied research on deradicalisation |
| Discord | Classification Office, New Zealand | Biodiversity Hub International | New America's Open Technology Institute |
| Dropbox, Inc. | Commonwealth Secretariat | Bonding Beyond Borders | Oxford Internet Institute |
| ExTrac | Council of Europe, Committee on Counter-Terrorism | Brookings Institution | Partnership for Countering Influence Operations, Carnegie Endowment for International Peace |
| Facebook | Department of Justice - Ireland | Business for Social Responsibility | Peace Research Institute Frankfurt (PRIF); Germany |
| JustPaste.it | Department of State - Ireland | Centre for Analysis of the Radical Right (CARR) | PeaceGeeks |
| Mailchimp | Department of State - USA | Center for Democracy & Technology | Point72.com |
| MEGA | Department of the Prime Minister and Cabinet (DPMC), New Zealand Government | Center for Media, Data and Society | Polarization and Extremism Research and Innovation Lab (PERIL) |
| Microsoft | DHS Center for Prevention Programs and Partnerships (CP3) | Centre for Human Rights | Policy Center for the New South (senior fellow) |
| Pex | European Commission | Centre for International Governance Innovation | Public Safety Canada & Carleton University |
| Snap Inc. | Europol/EU IRU | Centre for Youth and Criminal Justice (CYCJ) at the University of Strathclyde, Scotland. | Queen's University |
| Tik Tok | Federal Bureau of Investigation (FBI) | Cognitive Security Information Sharing & Analysis Center | Sada Award, Athar NGO, International Youth Foundation |
| Tremau | HRH Prince Ghazi Bin Muhammad's Office | Cornell University | Shout Out UK |
| Twitter | Ministry of Culture, DGMIC - France | CyberPeace Institute | Strategic News Global |
| You Tube | Ministry of Foreign Affairs - France | Dare to be Grey | S. Rajaratnam School of International Studies, Singapore (RSIS) |
| | Ministry of Home Affairs (MHA) - Indian Government | Dept of Computer Science, University of Otago | Swansea University |
| | Ministry of Justice and Security, the Netherlands | Digital Medusa | Tech Against Terrorism |
| | National Counter Terrorism Authority (NACTA) Pakistan | Edinburgh Law School, The University of Edinburgh | The Alan Turing Institute |

| | | | |
|---|---|---|---|
| | Organisation for Economic Co-operation and Development (OECD) | European Center for Not-for-Profit Law (ECNL) | The Electronic Frontier Foundation |
| | Office of the Australian eSafety Commissioner (eSafety) | Gillberg Neuropsychiatry Centre, Gothenburg University, Sweden, | The National Consortium for the Study of Terrorism and Responses to Terrorism (START) / University of Maryland |
| | Organization for Security and Co-operation in Europe (OSCE RFoM) | George Washington University, Program on Extremism | Unity is Strength |
| | Pôle d'Expertise de la Régulation Numérique (French Government) | Georgetown University | Université de Bretagne occidentale (France) |
| | North Atlantic Treaty Organization, also called the North Atlantic Alliance (NATO) | Georgia State University | University of Auckland |
| | Secrétaire général du Comité Interministériel de prévention de la délinquance et de la radicalisation | Global Network on Extremism and Technology (GNET) | University of Groningen |
| | State Security Service of Georgia | Global Disinformation Index | University of Massachusetts Lowell |
| | The Royal Hashemite Court/ Jordanian Government | Global Network Initiative (GNI) | University of Oxford |
| | The Office of Communications (Ofcom), UK | Global Partners Digital | University of Queensland |
| | UK Home Office | Global Project Against Hate and Extremism | University of Salford, Manchester, England, |
| | United Nations Counter-terrorism Committee Executive Directorate (CTED) | Groundscout/Resonant Voices Initiative | University of South Wales |
| | UN, Analytical Support and Sanctions Monitoring Team (1267 Monitoring Team) | Hedayah | University of the West of Scotland |
| | United Nations Major Group for Children and Youth (UNMGCY) | Human Cognition | Violence Prevention Network |
| | United States Agency for International Development (USAID) | Institute for Strategic Dialogue | WeCan Africa Initiative & Inspire Africa For Global Impact |
| | | International Centre for Counter-Terrorism | Wikimedia Foundation |
| | | Internet Governance Project, Georgia Institute of Technology | World Jewish Congress |
| | | Islamic Women's Council of New Zealand | XCyber Group |
| | | JOS Project | Yale University, Jackson Institute |
| | | JustPeace Labs | Zinc Network |
| | | Khalifa Ihler Institute | |
| | | KizBasina (Just-a-Girl) | |
| | | Love Frankie | |

GIFCT WORKING GROUPS OUTPUT 2022

# GIFCT Executive Summary and Discussion of Dr. Jazz Rowa's Algorithms Research

GIFCT
Global Internet Forum to Counter Terrorism

Dr. Erin Saltman
Director of Programming,
GIFCT

GIFCT recognizes the increasing concern from governments, researchers, technologists, and human rights advocates about the potential link between algorithmic amplification and processes of radicalization towards violence. Increased legislative language around the world has turned to 'algorithmic transparency' and one of the primary themes of the Christchurch Call to Action's Second Anniversary Summit in 2021 was to support methods to better understand user journeys online and the role algorithms may play in processes of radicalization. There is a fear that the nature of online environments may amplify hatred and glorify terrorism and violent extremism in a way that drives others towards violence. To effectively counter terrorism and violent extremism online, GIFCT aims to support research, analysis, and tools to better understand the true nature of the problem so that action can be taken. On the topic of understanding algorithmic processes there remain large knowledge gaps. GIFCT commissioned an extensive research effort by Dr. Jazz Rowa to assist in framing and better understanding the role of algorithms as part of GIFCT's 2022 Working Group outputs. This executive summary of her longer research paper, The Contextuality of Algorithms: An Examination of (Non)Violent Extremism in the Cyber-Physical Space, serves as a briefing document and reflection from GIFCT about some of Dr. Rowa's key findings. As of September 2022, her longer report can also be found on the GIFCT website under Working Group output and under our highlighted resources.

## Background

In the first year of GIFCT Working Groups, held September 2020 through July 2021, GIFCT convened a group of global experts focussed on Content-Sharing Algorithms, Processes, and Positive Interventions, with participants from across tech companies, government, and civil society. Since an algorithm can be almost any input online with an output, the group adopted the shared goal of mapping which content-sharing algorithms and processes used by industry had the potential of facilitating consumption of content that may amplify terrorist and violent extremist content, or user interest in such content. The group also mapped and considered positive interventions and risk mitigation points for safety-by-design. The results of this paper honed in on the algorithmically optimized surfaces and tools that could potentially be exploited by bad actors, such as terrorists or violent extremists. This allowed the conversation on algorithms to focus more specifically on three online surfaces: search functions, recommendation features, and ad targeting algorithms.

In Year 2 of Working Groups, held September 2021 through July 2022, GIFCT commissioned Dr. Jazz Rowa to take this conversation and analysis further. GIFCT Working Groups had sub questions related to algorithms and the nexus with extremism in 3 of our 5 groups and asked Dr. Rowa to sit across these groups to develop this extensive paper. She has provided an analytical framework through the lens of human security to better understand the relation between algorithms and processes of radicalization. Dr. Rowa participated in the Transparency, Technical Approaches, and Legal Frameworks Working Groups to gain insight into the real and perceived threat from algorithmic amplification. This participation was supplemented with empirical research and a range of first-person interviews. This research looks at the contextuality of algorithms, the current public policy environment, and human rights as a cross-cutting issue. In reviewing technical and human processes, she also looks at the potential agency played by algorithms, governments, users, and platforms more broadly to better understand causality.

## Findings

While this paper presents a myriad of findings and poses further questions, identifying gaps for further research, there are some key takeaways that stuck out to our teams at GIFCT, which we will be processing and looking to build further work around in the future. The first takes us back to the age-old questions of definitions. In group discussions and interviews it remains clear that **there is no overarching agreement between different sectors or geographies on what online terrorist content is, what violent extremism is, what algorithms are, and what "extremist" or "borderline" content is.** If it can't be well defined, or if legislative language is vague on these points, we are still left with too much ambiguity to apply technical solutions or to ensure rigorous oversight or accountability mechanisms. Specifically for online spaces, the better you can define harm parameters the more you can measure, evaluate and risk mitigate. Vague or ambiguous terminology can lead to over censorship, under censorship, or the inability to measure and understand the nature of the problem in the first place.

While pressure escalates for tech companies to "do more", the analysis notes that **the current guidance on human rights in national, regional, and international legal frameworks is technologically suboptimal.** The pressure to expand technical solution-building is not equally matched with practical guidance of what human rights applications for technological ecosystems should look like. The paper also found that even some government representatives were wary that the term "algorithm" had become the latest buzzword and hot topic in the international debates on preventing and countering terrorism and violent extremism online, without enough clarity on the concept or the scope.

Dr. Rowa addresses the multiple reasons **why understanding algorithms, and attempts to provide meaningful algorithmic transparency, remains difficult. There is a notable difference between algorithmic explicability, interpretability, and auditability.** However, approaching algorithmic systems and its "black box" effect for analyzing input and output variables is compounded for a number of reasons; very few people understand the technical side of digital technologies, there remains a system of self-regulation for the technical evolution and review of technologies, there are methodological limitations for external researchers reviewing algorithmic systems, all combined with a trend of reactionary government regulation. The disclosure of an algorithmic formula or source code is viewed by some as useful and many as irrelevant in understanding a program's predictive behavior. Meanwhile there is a multi-dimensional and ever-changing landscape for both terrorist and violent extremist actors online and technical dynamism of platforms themselves. This conceptualisation of audits and the design of mechanisms for algorithmic oversight must therefore acknowledge the complexity of such an undertaking. To work towards greater algorithmic transparency, more work will need to be done to fully understand what "meaningful" data and algorithmic transparency means to policy makers and relevant stakeholders. Data and information sharing from tech companies can take many forms and alignment on understanding what data is useful and meaningful is crucial.

**The current discourse on the role of algorithms in (non)violent extremism has for the most part**

**created a false dichotomy between the online and offline spaces**. The discussion around user, platform, and government furthers the complexity in trying to interpret causality in processes of radicalization and agency. User agency and lived experiences particularize contextual phenomena and inform the integration of the online and offline dynamics of extremism. Dr. Rowa points out that the interplay between the user and how an algorithm operates is intrinsically tied. Algorithmic systems are representations of human decisions and worldviews. What happens in the online realm cannot be detached from real life actions. This interplay needs to also inform legislative thinking.

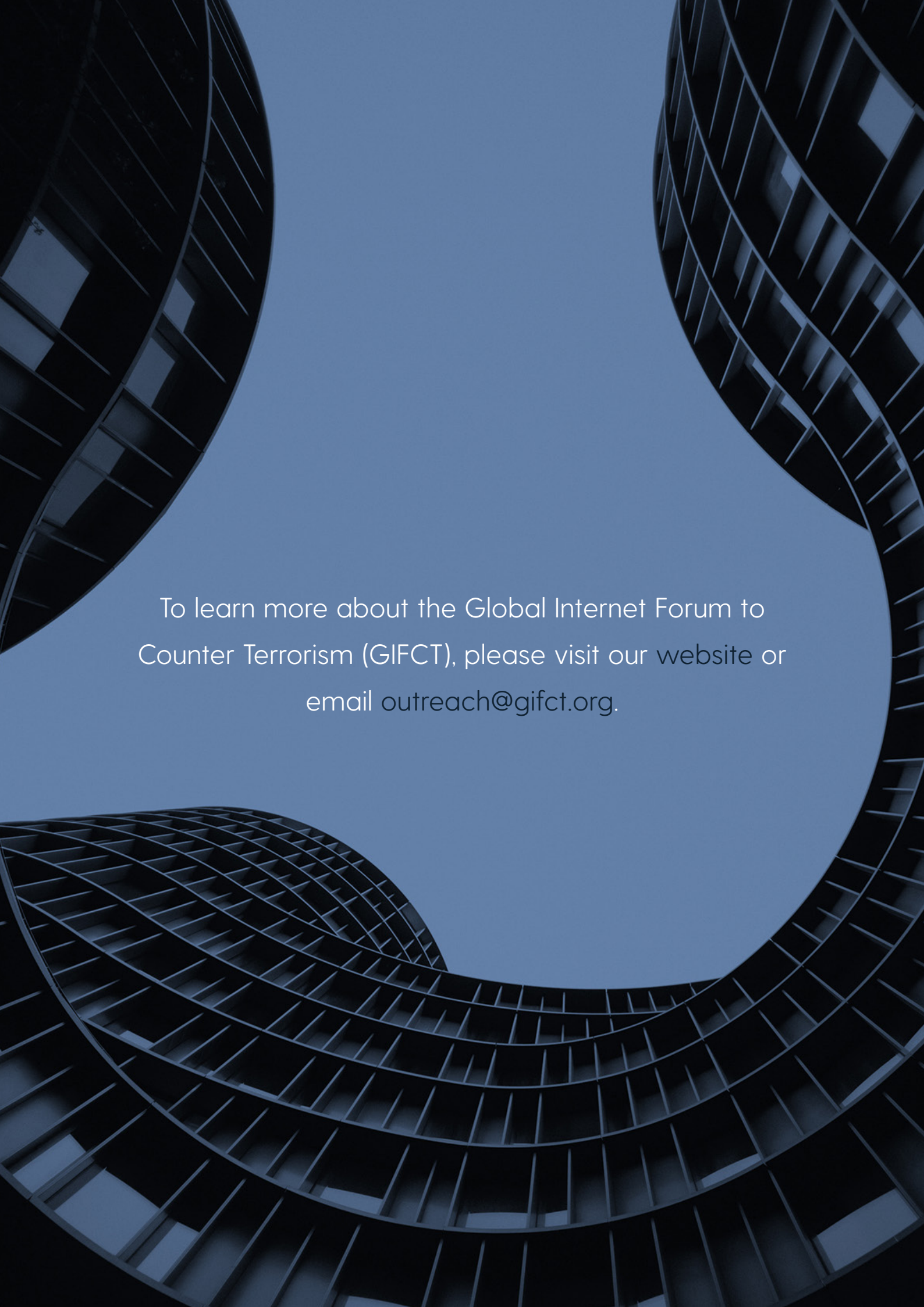
Related to the discourse around user and platform accountability and responsibility, the interviews highlighted the **continued discomfort with non-violent and non-violating extremist content in what might be determined "gray area" content, and what, if anything, tech companies should do about it.** If users create legal, non-violating content and other users actively search and engage with the content, should private technology companies exert absolute control over the curation and restriction of legal but 'extreme' content? The concerns over borderline content are tied to the overarching debate on the definition of extremist content, liability for content creation, and the dispersal of content across digital publics (within hybridized or algorithmically amplified systems).

**While some algorithm/user interplay could potentially amplify extremist content, there remain many spaces online that are beacons for violent extremist and terrorist sympathizers, yet have no algorithmic optimization associated with content surfacing or group recommendation features.** These platforms remain a beacon to hate-based groups simply because they lack proactive moderation of content. The analysis notes that the recent lone actor attack in Buffalo, New York is seen as a case of "radicalization on 4chan" by other users giving social constructive information, documents, and social feedback. The attacker was also previously known to police, meaning there were offline signals that could have been used to provide support or have led to PVE/CVE interventions.

The overall research creates many avenues for further dialogues and multistakeholder work. However, it is important to recognize where positive opportunities for future work lie. **The research concludes that algorithmic processes, while being the core scrutiny of this paper, are equally where solutions can be found.** Despite the initial research question for the paper, Dr. Rowa points out that, paradoxically, algorithmic systems are conceived as automated problem solvers. In concert with other agencies, algorithms can act as conduits for the reconciliation, remediation, and reconstitution of an increasingly dysfunctional cyber-physical order. Whereas algorithms pose (un) known challenges for extremism, the opportunities they present in the mitigation and resolution of this and other societal challenges is equally consequential.

We at GIFCT hope that this research is of utility to the broadest range of stakeholders working to counter terrorism and violent extremism online and are grateful to Dr. Jazz Rowa for the time and energy she put into this extensive research over the last year.

Dr. Erin Saltman
Director of Programming
GIFCT



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