

Transparency Report

July 2021





Table of Contents

GIFCT Annual Transparency Report	4
The Purpose of Transparency Reporting	4
Members & Mentorship	5
Strategic Pillars	6
Prevent : Hash-Sharing Database	7
Prevent: URL Sharing	12
Prevent: Working Groups	13
Respond : Incident Response & Content Incident Protocol (CIP)	16
Learn : Member Resource Guide	18
Learn: E-Learnings	17
Learn: Conducting & Funding Research	18
Contact	20

GIFCT Annual Transparency Report

The mission of Global Internet Forum to Counter Terrorism (GIFCT) is to prevent terrorist and violent extremists from exploiting digital platforms. We believe that by working together and sharing technological and operational elements of our individual efforts, GIFCT members can have a greater impact on decreasing the threat of terrorist and violent extremist activity online.

This is GIFCT's third annual Transparency Report and represents output and metrics from June 1, 2020, through May 31, 2021. This report also reflects the progress made in GIFCT's first full year as an independent 501(c)(3) non-governmental organization, registered in the United States with an independent Executive Director and staff. More information about the structure of GIFCT, mission, vision and values can be found here.

GIFCT is governed by an Operating Board made up of the Forum's founding companies: Facebook, Microsoft, Twitter, and YouTube. The Operating Board provides GIFCT's operational budget and ensures overall GIFCT operations align with its mission. The Operating Board chair rotates annually. Twitter is the 2021 chair of GIFCT.

GIFCT is advised by an Independent Advisory Committee (IAC) made up of representatives from civil society, government, and intergovernmental organizations. A majority of members of the IAC are drawn from civil society, including advocacy groups, human rights specialists, foundations, researchers and technical experts. The current Chair of the IAC is Bjørn Ihler, co-founder of the Khalifa Ihler Institute. More details of the GIFCT governance structure, Operating Board, and IAC can be found here.

The Purpose of Transparency Reporting

It is somewhat unique for an NGO to release an annual transparency report. However, since GIFCT manages and builds out tools that facilitate platforms, and because we require member companies to produce a transparency report on at least an annual cadence, we hold ourselves accountable to the same standard. At the July 2020 Annual Summit, GIFCT Executive Director Nick Rasmussen identified transparency as one of three primary values that guides the organization's work, particularly as it began to develop its infrastructure and operations in the first year as an independent organization.

Reporting of this nature encourages an open and inclusive internet and multi-stakeholder approach. We have also worked to highlight member companies' transparency reports in our Member Resources Guide so that they are more easily accessible to broader stakeholder groups.

Additions to this year's report in comparison to previous versions were made based



on guidance provided by the Human Rights Impact Assessment carried out by Business for Social Responsibility (BSR) between December 2020 and May 2021. The results of this assessment and GIFCT's response can be found here. The assessment encouraged a strategic and deliberate approach to transparency that will (1) enable enhanced accountability, (2) spread expertise, insight, and learning on how to prevent terrorists and violent extremists from exploiting digital platforms and (3) address various myths and misunderstandings that exist about GIFCT. Additions were also made based directly on feedback from GIFCT's Working Group on Transparency. Some of these recommendations we were able to incorporate in this year's report, while other recommendations we will continue to discuss and look to incorporate in future reports.

Members & Mentorship

There are 17 current GIFCT Members: Airbnb, Amazon, Discord, DropBox, Facebook, Instagram, JustPaste.lt, LinkedIn, Mailchimp, Mega.nz Microsoft, Pinterest, Tumblr, Twitter, WhatsApp, WorPress.com and YouTube.



Tech companies seeking to join GIFCT can apply online and need to fulfill membership criteria. For tech companies seeking membership that do not meet all of the GIFCT membership requirements, GIFCT provides guidance by connecting companies with our partner, Tech Against Terrorism, whose mentorship program facilitates developing best practices around transparency in a way that works for the specific platform. In Tech Against Terrorism's Mentorship Program, a service supported by GIFCT, Tech Against Terrorism assesses and provides recommendations on platforms' overall transparency efforts and measurement against the GIFCT Membership Criteria.

As of July 2021, three companies have published their first transparency report as a result of TAT's support, and nine companies had their content standards updated following TAT recommendations. Through TAT's Mentorship Program, eight companies were recommended and accepted as GIFCT members.

Strategic Pillars

Guided by its mission, GIFCT organizes its work to convene and enhance the collective capacity of the industry with respect to its three strategic pillars of prevent, respond, and learn. The aim of our Transparency Report is to give data and details about the outputs and impacts of these three workstreams.

Prevent: Equipping digital platforms and civil society groups with awareness, knowledge, and tools (including technology) to develop sustainable programs in their core business operations to disrupt terrorist and violent extremist activity online.

- Technical Support: The Hash-Sharing Database and URL Sharing Efforts aid tech companies in mitigating the spread of terrorist content on their platforms.
- CSO Support: Supporting CSO/NGO capacity-building through a Campaign Toolkit and CSO training.
- Working Groups: Keeping GIFCT ahead of the curve in thinking through larger issues on key topics related to the nexus between terrorism and technology.

Respond: Developing tools for and the capacity of platforms to work together and with other stakeholders to mitigate the online impact of a terrorist/violent extremist attack.

 Incident Response: GIFCT has been evolving its Incident Response and Content Incident Protocol. This includes fostering connectivity and communication between relevant government and law enforcement with the tech companies. This work is led by the Working Group on Crisis Response and dedicated internal efforts.

Learn: Empowering researchers to produce action-oriented analysis on terrorism and violent extremism and its nexus to technology, including best practices for multistakeholder cooperation and preventing abuse of digital platforms.

- Member Resource Guide: This guide links to publicly available resources and information of member companies about their counterterrorism, CVE, safety, policy and transparency efforts.
- GIFCT-Tech Against E-Learning Webinars: TAT and GIFCT hold monthly E-Learning Webinars with international experts to cover timely topics related to terrorism, extremism, and technology.
- GNET: GIFCT Funds the Global Network on Extremism and Technology (GNET), a global network of academics providing open access analysis.

In all GIFCT efforts, GIFCT members maintain their independence in enforcing their own policies, standards, and procedures. GIFCT works to provide helpful tools and resources; however, it is up to member companies how they utilize these benefits in line with how their platform operates.



Prevent: Hash-Sharing Database

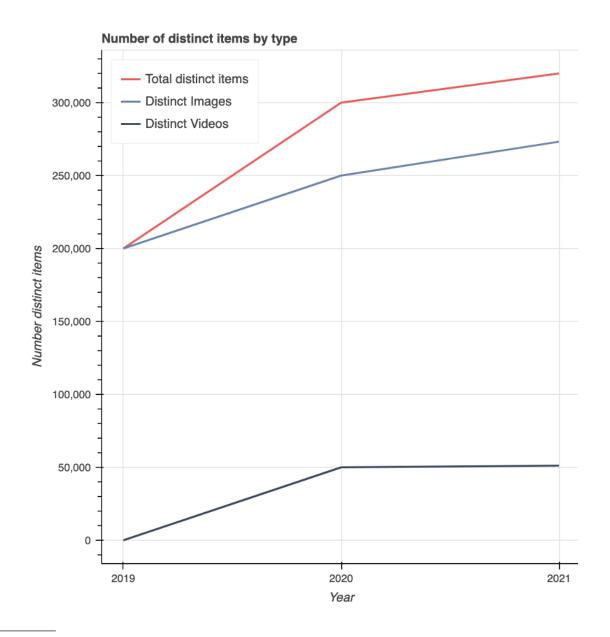
The largest cross-platform technical tool supported by GIFCT is the industry hash-sharing database. The database enables sharing of "hashes" (or "digital fingerprints") of known terrorist images and videos between GIFCT member companies. An image or video found by a member company is "hashed" in its raw form, ensuring there is no link to any source original platform or user data. Hashes appear as a numerical representation of the original content, which means they cannot be easily reverse engineered to recreate the image and/or video. Each company that is part of the hash-sharing database determines its use of and engagement with the database, depending on (among other things) their own terms of service, how their platform operates, and how they utilize technical and human resources.

GIFCT is neither a tech company nor a social media platform and we do not own or store any source data or personally identifiable information of any users associated with member platforms. GIFCT provides further explanation of how hashes and the hash-sharing database works in an explainer video.

Contents of the database

To date, the hash-sharing database contains approximately 320K unique hashes relating to 270K visually distinct images and 51K visually distinct videos. The reason this number is approximate is because "unique hashes" represent clusters of hashed content that are visually identical or near-identical to the human eye. There are 2.3 million hashes in total making up the 320K unique hashes.

Hash-sharing database access is given to: Ask.fm, Cloudinary, Dropbox, Facebook, Instagram, JustPaste.it, LinkedIn, Microsoft, Pinterest, Reddit, Snap, Twitter, Verizon Media, and YouTube.¹



¹ Not all GIFCT members are members of the hash sharing consortium. For some companies, user generated content is not how their platforms operate, others are in the process of joining, and there remain a few legacy companies that have access but are not yet officially GIFCT members. For any new access to the hash-sharing database, GIFCT membership requirements must first be met.



Taxonomy

Just like other sectors, tech companies often have slightly different operational definitions of "terrorism" and "terrorist content" that guide them in the surfacing, review, and removal of content. To find common ground, the original scope of the hash-sharing database is limited to content related to organizations on the United Nations Security Council's Consolidated Sanctions List. Hashes related to the UN Sanctions List are also labeled with a severity framing taxonomy in order to assist platforms in how they might triage and review content that matches based on a hash:

- Imminent Credible Threat (ICT): A public posting of a specific, imminent, credible threat of violence toward non-combatants and/or civilian infrastructure.
- **Graphic Violence Against Defenseless People (GVADP):** The murder, execution, rape, torture, or infliction of serious bodily harm on defenseless people (prisoner exploitation, obvious non-combatants being targeted).
- Glorification of Terrorist Acts (GTA): Content that glorifies, praises, condones, or celebrates attacks after the fact.
- **Recruitment and Instruction (R&I):** Materials that seek to recruit followers, give guidance, or instruct them operationally.

The only hashes that appear in the hash-sharing database that do not correspond to entities on the UN list are added during a declared Content Incident Protocol (CIP). The first precedent for this unique label within the database occurred during the Christchurch attacks in New Zealand. The following are CIP related labels within the database.

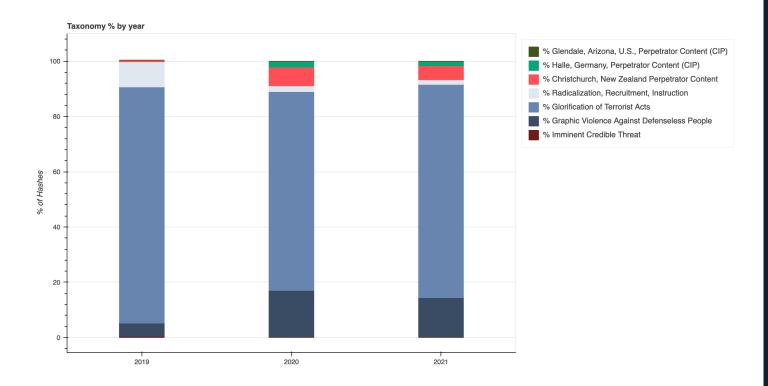
- Christchurch, New Zealand Perpetrator Hashes: On March 15, 2019, the need for a separate hash label was declared after an attacker live-streamed his attack on two mosques.2
- Halle, Germany, Perpetrator Hashes: On October 9, 2019, a CIP was declared following an attacker livestreaming his attack on a synagogue.
- **Glendale, Arizona, U.S., Perpetrator Hashes:** On May 20, 2020, a CIP was declared following an attacker livestreaming his attack on the Westgate Entertainment District.

9

² In the aftermath of the Christchurch attacks, GIFCT companies came together to formulate what is now the GIFCT Content Incident Protocol. This included testing the incident protocols during the July 14, 2019, Utica, New York murder, whereby an image of the murder was disseminated online and a video before and during the murder had also been created. However, this incident was not a terrorist or violent extremist incident. Test hashes were subsequently removed from the database.

Hash Taxonomy Category	% of total hashes
Imminent Credible Threat	0.1%
Graphic Violence Against Defenseless People	14.3%
Glorification of Terrorist Acts	77.2%
Recruitment and Instruction	1.7%
Christchurch, New Zealand, Perpetrator Hashes	5.1%
Halle, Germany, Perpetrator Hashes	1.5%
Glendale, Arizona, U.S., Perpetrator Hashes	0.1%

The figure below represents a breakdown of the UN listed hashed content as well as CIP hashed content for the last three years.



Feedback on hashes

New functionality was added to the hash-sharing database in 2019 allowing GIFCT members to give feedback on hashes shared within the database. Members can give feedback on hashes in a number of ways, including showing agreement or disagreement with a hash's inclusion in the database or a hash's categorization. This feedback is visible



for hash-sharing participants so that third parties can consider it in their own processes and reviews. To date, this feature has not been used extensively but plans are in place to accelerate the adoption and use of this part of the system.

We have so far received feedback twice or more on approximately 9000 visually distinct items representing approximately 2.9% of all visually distinct items in the database. (All items in the database have one response by definition because in order to add a hash to the database a member company must assert that the item is within GIFCT's taxonomy and label it appropriately.)

In addition to the low rate of response, feedback is not evenly sampled across all visually distinct items. Therefore any conclusions drawn from this data should be treated with caution, should not be considered statistically significant, and using these figures for extrapolation to the full database is not methodologically sound.

Hash-Sharing Database Feedback Metrics

Total Visually distinct items with two or more responses	9000
% visually distinct items with two or more responses	2.9%
% feedback that indicate agreement	69%
% feedback on CIP related items that indicate agreement	95%
% of visually distinct items with one or more responses indicating an item does not fit in current taxonomy (U.N. designated entities or CIP related)	0.63%

Law Enforcement requests

Over the last year, GIFCT received no formal requests for data or access from a government entity regarding hashed content in the hash-sharing database. Questions and requests for specific content should be directed to member companies since hashes are only numerical representations of source content and cannot practically be reverse engineered to recreate an image or video.

Broadening GIFCT's Hash-Sharing Database's Taxonomy Framework

GIFCT recognizes that the UN designation list and hashes generated from Content Incident Protocols represent a small amount of what is globally considered terrorist and violent extremist content (TVEC). While there have been calls by governments and experts to broaden the hash-sharing taxonomy, there have also been concerns from human rights activists over possible over-censorship for list-based approaches that go

beyond government-regulated lists. On this basis, GIFCT launched a multi-stakeholder effort in February 2021 to develop an expanded taxonomy framework for the hash-sharing database.

GIFCT sponsored the development of five proposals from international experts to develop approaches for the expansion of hash-sharing efforts to ensure increased parity and understanding of how TVEC manifests online. For a complete explanation regarding GIFCT's approach to expansion please see the full report. Based on this analysis and feedback from GIFCT member companies, an initial expansion will include three new categories of data in the hash-sharing database:

Attacker Manifestos: There have been numerous cases of attackers that have posted their manifestos online in advance of carrying out attacks. This material is often intended for virality and shared among sympathizers. Hashed images and hashed text extracted from PDFs of violent extremist and terrorist attacker manifestos will be included in the database.

Branded Terrorist Publications: Branded content allows tech companies to have clear indicators that online content is inherently terrorist or violent extremist. To date, the hash-sharing database has focused on image and video hashes; however, the clearer signal for branded terrorist content is in PDF form. GIFCT will add hashes of the text and images from these PDFs to the database.

TCAP URLs: In an effort to build on the utility and impact of the Terrorist Content Analytics Platform (TCAP) efforts, GIFCT will include hashes of URLs that TCAP flags to tech companies. The TCAP taxonomy includes ISIS and Al Qaeda organizations and affiliates as well as Five Eyes designated far-right extremist terrorist organizations. This process ensures a "one-to-many" signal can be shared without sharing personally identifiable information.

Prevent: URL Sharing

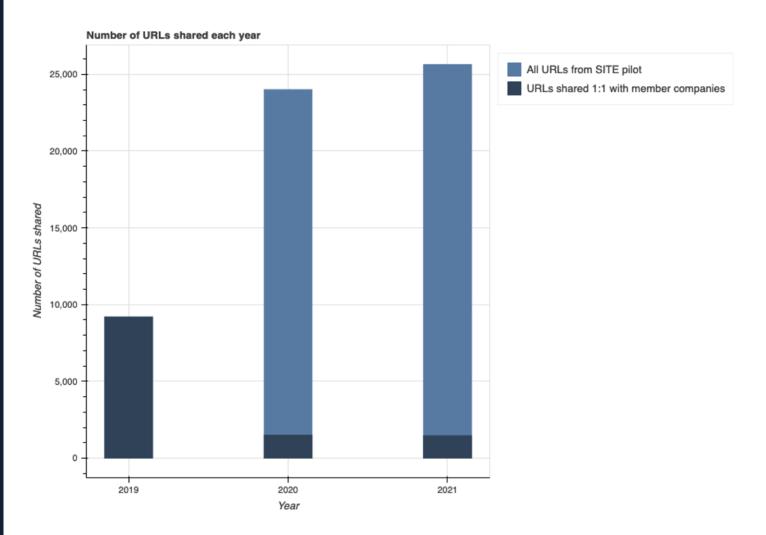
For the past 18 months, GIFCT has run a URL sharing pilot with SITE Intelligence. The pilot project gave three newer GIFCT member companies access to SITE's SourceFeed database and flagged URLs directly to the companies that had been labeled as being associated with international Islamist extremist terrorist organizations, based on the UN Consolidated List. The database also provided context around a given URL, including organizational affiliation of the terrorist content, translation of content into English, and further context support.

GIFCT has facilitated the sharing of 50k URLs through the SITE Intelligence Pilot and 9.2k URLs from internal one-to-one sharing of URL's between platforms from 2019. Of these 59k URLs approximately 20% related to the three member companies who were



participating in the trial.

Over the next 12 months, we plan to move to an alternative system of hashing URLs from TCAP and sharing the resulting hashes with all GIFCT members.



Future commitments on tech and tooling transparency

Over the next 12 months, working closely with the GIFCT Transparency Working Group, we aim to bring greater transparency to our technical and tooling efforts. As GIFCT expands the taxonomy of content that can be hashed and added to the database, it is particularly important we establish a more robust framework to assess the composition of the database. We will also focus on ways to better explain how the system operates and how it is used by our members.

Prevent: Working Groups

The GIFCT working groups convened 200 experts and practitioners from across the world; holding more than 55 meetings with representatives from 10 tech companies, 13 governments and international governing bodies, 26 civil society organizations and 41

research and academic institutions.

Working groups outputs are presented at the GIFCT Annual Summit, and output from each working group can be found on the GIFCT website. From July 25 through August 25, 2021 new members can apply to join working groups via an application link. Working group themes and their output is listed below:

- 1. Academic and Practical Research Output: The Academic and Practical Research working group focused on developing a paper that assesses the knowledge gaps and barriers that affect multiple stakeholders within the field of preventing and countering violent extremism (P/CVE) to begin moving toward holistic and coordinated solutions. Separately, the group also produced a White Paper on the landscape for new and emergent issues that are understudied and (consequently) misunderstood.
- 2. Content-Sharing Algorithms, Processes, and Positive Interventions Output: The group produced two Research Briefs. One focused on Content-Sharing Algorithms and Processes, which map the type of algorithmic processes that could be exploited by violent extremists and terrorists. The second briefing looks at Positive Interventions with a range of international case studies.
- **3. Crisis Response Output:** The Crisis Response working group produced two internal directories created for cross-sector crisis communication in linking relevant government authorities to the right contacts at GIFCT member companies. They also produced a briefing paper discussing how various crisis protocols are initiated and how to debrief relevant stakeholders and review protocols in the aftermath of a crisis.
- **4. Legal Frameworks Output:** The Legal Frameworks working group produced a Gap Analysis Report. This paper focuses on what constitutes "data" and identifies a number of policy questions and challenges that arise from the operational use of information by various actors.
- 5. Technical Approaches Output: The Technical Approaches working group's output tasked Tech Against Terrorism with producing the group's Gap Analysis Report. The report summarizes working group member feedback and discusses gaps, solutions, and recommendations for how cross-platform technical collaborations can be strengthened. The report is aimed at small tech platforms and focuses specifically on content. The report is accompanied by an Executive Summary by the co-leads of the group, which outlines the working group's output.
- **6. Transparency Output:** The Transparency Working Group produced a briefing paper that breaks down various aspects of what meaningful transparency means to different sectors and gives explicit recommendations to GIFCT directly as well as companies and relevant sectors like government in how to further transparency efforts.



Respond: Incident Response & Content Incident Protocol (CIP)

Following the attacks in Christchurch, New Zealand in March of 2019, GIFCT members established a centralized communications mechanism to share news of ongoing incidents that might result in the spread of violent content tied to the specific incident unfolding. These communications allow for widespread situational awareness and a more agile response among member companies.

To date, only two incidents have activated GIFCT's Content Incident Protocol (CIP), the highest level of incident response developed in this mechanism. A CIP is activated when all four of the following conditions are met and are discussed in the Hash-Sharing Database transparency section:

- 1. A real-world terrorist, violent extremist, or mass violence event;
- 2. Has been recorded or broadcast via livestream;
- 3. Depicting murder or attempted murder; and
- 4. Is being distributed on GIFCT member platforms or so broadly online that such distribution appears inevitable.

Recognizing that significant terrorist incidents occur that require a coordinated response from GIFCT and our members, this year we have created a more robust Incident Response Framework with two further levels:

When a Content Incident (CI) is declared, it will enable collaboration and information sharing between members. This includes ensuring that hashes relating to perpetrator produced content can be shared between members. A CI will be enacted when the following conditions are met:

- 1. A real-world, offline terrorist or mass violence event;
- 2. Content other than video or livestream (e.g. photo, audio, or text) by perpetrator or accomplice;
- 3. On a member platform (so broadly available online it is inevitable);
- 4. Depicting or advocating murder or violence from a terrorist attack;

When an Incident is declared, GIFCT tracks the incident closely to identify any additional risks that may elevate our response to a CI or CIP, and provide members with situational awareness. An Incident will be enacted when the following conditions are met:

- 1. A real-world, offline terrorist or mass violence event:
- 2. An online aspect.

Since establishing its incident response framework, member companies have initiated communications in response to 153 incidents (as defined above) following an event in as close to real-time as possible.



Fig 4. The above map indicates the location of incidents that have been communicated about across five continents as part of GIFCT's Incident Response Framework between April 21, 2019 and May 31, 2021. See details here.

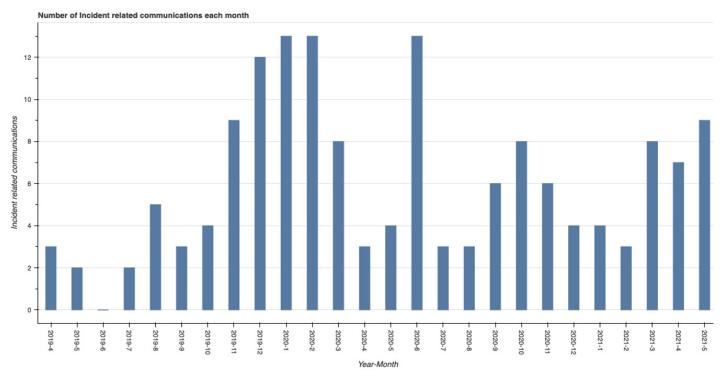


Fig 5. Showing the number of incidents that have been communicated as part of GIFCT's Incident Response Framework over time - See details here.



Learn: Member Resource Guide

In April 2021 GIFCT launched the GIFCT Member Resource Guide, offering a "one-stop-shop" for tech companies, civil society groups, academic researchers, and governments to learn how GIFCT members counter terrorist and violent extremist activity, understand the tools they've developed to combat radicalization, find links to company transparency reports, and discover the best practices they offer to help people stay safe online. This web page also includes key resources from our partners, the Global Network on Technology and Extremism and Tech Against Terrorism, and serves as a guide to how each member fulfills their GIFCT membership criteria and supports GIFCT's mission.

Learn: E-Learnings

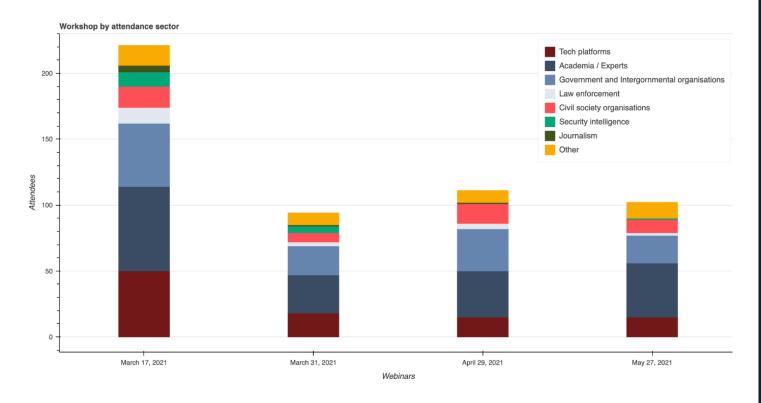
GIFCT partners with TAT for monthly E-Learning workshops that are open to global participants. Previously, GIFCT would co-host three to four workshops a year internationally with TAT.3 Workshops were put on hold since January 2020 due to the ongoing pandemic. In March 2021 GIFCT and TAT began monthly E-Learning workshops bringing global experts on key topics to the virtual stage. The four topics hosted in 2021 have been:

- The Nexus Between Violent Extremism and Conspiracy Theory Networks Online: Understanding and Challenging groups like QAnon, Oath Keepers and The Boogaloo Movement (March 17, 2021)
- Technical Approaches to Countering Terrorist Use of the Internet: URL Sharing and Collaborative Tech Sector Efforts (March 31, 2021)
- Countering Terrorist Use of Emerging Technologies: Assessing Risks of Terrorist Use of End-to-End-Encryption and Related Mitigation Strategies (April 29, 2021)
- The Nuts and Bolts of Counter Narratives: What Works and Why? (May 27, 2021)

These four E-Learnings had 528 participants in total dial-in from around the world coming from a variety of sectors, as shown in the figure on the next page.

17

³ Between August of 2017 and December of 2019 there were 13 workshops that engaged 140 tech companies, 40 NGOs, and 15 government bodies across four continents (taking place in Australia, Belgium, France, Germany, Indonesia, India, Israel, Jordan, UAE, the UK and the USA).



Learn: Conducting & Funding Research

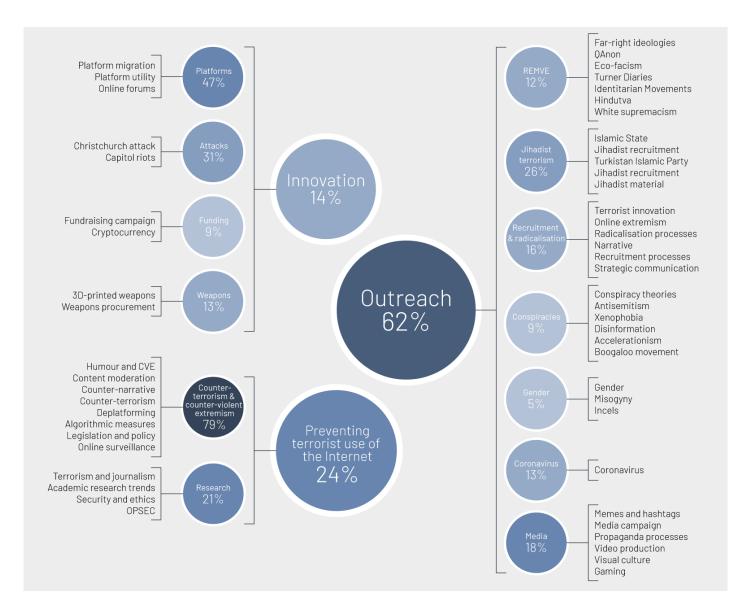
In January 2020, GIFCT began Phase Two of support for its academic research network, led by the International Centre for the Study of Radicalisation (ICSR), based at King's College London. 4 ICSR established the Global Network on Extremism and Technology (GNET), bringing together an international consortium of leading academic institutions and experts with core institutional partnerships from the US, UK, Australia, Germany and Singapore to study and share findings on terrorist and violent extremist use of digital platforms.

GNET Insights, Reports and Workshop Metrics

Between June 2020 and July 2021 **GNET published 198 insights from 245 authors based in twenty-four countries around the world**. Insights are short but concise papers, which empower experts to probe and explore contentious issues as they relate to violent extremist behaviors and technology. A breakdown of output themes of insight reports are highlighted on the next page.

⁴ In Phase 1 (2018–2019) GIFCT supported the Global Research Network on Terrorism and Technology (GRNTT), aimed at developing research and providing policy recommendations around the prevention of terrorist exploitation of technology. Thirteen papers were published in 2019 from GRNTT and can be found here.





Every year GNET also funds a selection of longer research papers focused on terrorist and violent extremist use of technology, which offer actionable findings and practical solutions to the tech industry. As of May 31, 2021, GNET produced eight reports from authors based in five different countries. Reports are available in English, French, German and Arabic:

- Decoding Hate: Using Experimental Text Analysis to Classify Terrorist Content, by Abdullah Alrhmoun, Shiraz Maher, and Charlie Winter, ICSR.
- Artificial Intelligence and Countering Violent Extremism: A Primer, by Marie Schroeter, ICSR.
- Migration Moments: Extremist Adoption of Text-Based Instant Messaging Applications, by Bennett Clifford, George Washington University.
- Researching Extremist Content on Social Media Platforms: Data Protection and Research Ethics Challenges and Opportunities, by Manjana Sold and Julian Junk, The Peace Research Institute, Frankfurt.

GIFCT Transparency Report 2021

- Conspiracy Theories, Radicalisation and Digital Media, by Daniel Allington, King's College, London.
- Polarising Narratives and Deepening Fault Lines: Social Media, Intolerance and Extremism in Four Asian Nations, by Jordan Newton, Yasmira Moner, Kyaw Nyi Nyi, Hari Prasad and The Centre of Excellence for National Security (CENS).
- Bringing Women, Peace and Security Online: Mainstreaming Gender in Responses to Online Extremism, Alexis Henshaw.
- GNET Survey on the Role of Technology in Violent Extremism and the State of Research Community-Tech Industry Engagement, Lydia Khalil, Lowy Institute.

To further facilitate multi-sector knowledge-sharing opportunities and to provide expertise to a range of stakeholders, GNET - working with GIFCT and in partnership with institutional partners in different parts of the world - also curated 21 workshops focusing on the nexus between terrorism and technology. The workshops were hosted virtually and managed by partner institutes in Australia, Singapore, the United Kingdom, the United States, Germany and Syria. Workshops were attended by a total of 353 participants.

In May 2021, GNET launched its First Annual Conference bringing together a range of diverse panels, which encouraged and facilitated discussions and dialogue between the tech sector and expert academics, civil society representatives, and government. The five different panels saw 350-400 unique visitors logging into the various sessions throughout the day.

Contact

For more information about GIFCT, please contact outreach@gifct.org or press@gifct.org.To keep up to date with GIFCT, sign up for our monthly stakeholder newsletter and follow us on Twitter and LinkedIn.

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

