## A Guide to GIFCT Member Company Tools and Resources

# DRAFT



Date: March 2021

**Descriptor:** This is a first draft of GIFCT's Member Resources Document for NGOs - to sit on GIFCT website. The aim is to make this a publicly available resource to highlight all the existing efforts, tools and resources member companies have in relation to preventing terrorist and violent extremist exploitation. All links included can be seen in Appendix B and C based on:

(External) GIFCT Member Open-Source Resources

## Table of Contents

Introduction & How to Use This Document	3
Community Standards	4
Transparency	7
Safety Hubs	8
Reporting Mechanisms	10
Counterspeech and Counter-Narrative Facilitation	11
Digital Literacy	13
Advertising and Marketing	14
Appendix A: GIFCT Membership Criteria and Benefits	16
Appendix B: GIFCT and Partner Resources Index	17
Appendix C: Member Resources Index	17

#### Introduction & How to Use This Document

This document serves as a guide for civil society groups, academic researchers, governments, and other tech companies wanting to know more about the open-source resources and information that GIFCT member companies have available. Specifically, under the remit of the <u>Global Internet Forum to Counter Terrorism</u> (GIFCT), these resources highlight counterterrorism and counter-extremism efforts as well as tools that can be used in countering hate-based processes of radicalization and violent extremist narratives online.

In each section we have linked directly to the existing resources on each topic developed by GIFCT member companies and GIFCT itself where relevant. Resources are listed either thematically or in alphabetical order. Resources also include those developed by GIFCT and its partners at the <u>Global Network on Technology and Extremism</u> (GNET) and <u>Tech Against Terrorism</u> where applicable.

To begin, this document maps companies' efforts fulfilling some of the <u>GIFCT membership criteria</u>. These are criteria that all tech companies must have publicly in place before being considered for GIFCT membership. First, companies must have <u>public platform standards</u> that explicitly prohibit the promotion of terrorism and/or violent extremism in their terms of service, community guidelines, or other publicly available content policies. This first section explores how companies discuss their frameworks for prohibiting terrorist and/or violent extremist use or content distribution on their platforms.

The second section of this document links to member companies' Transparency Reports. All GIFCT member companies must have regular, public data transparency reports published at least annually. While we have seen an increase in robust and progressively granular transparency reporting from many of the larger GIFCT member companies, developing an initial transparency report can be the most daunting aspect of GIFCT membership criteria for many smaller companies.

The third section covers platform safety portals and wider safety information that platforms make available. Safety information is particularly crucial for activists, academics, journalists, and users that might be vulnerable to violent extremist groups. Safety information on some platforms includes information around how to report hacked accounts, how to better ensure privacy on your account set-up, and how to flag abuse and harassment.

The fourth section reviews the ways that GIFCT members ensure they have the ability to receive and act on reports of illegal activity or activity violating their terms of service. This is also a GIFCT membership requirement, though the

format for reporting may take different forms depending on the platform. Reporting capabilities include internal flagging tools, reporting portals, and outreach emails. Whatever form it takes, it is important to ensure that the community using a platform can highlight abuses to the platform that might be missed by any proactive detection programs.

While community standards, transparency reports, safety outlets, and reporting guidance work to define and disallow certain types of content and activities on a given platform, we know that content removal alone addresses a symptom and not the root causes of radicalization leading to violence. The fifth section of this document highlights resources related to GIFCT's membership criteria to support the capacity of civil society organizations to challenge violent extremism. Under this remit, some of the larger social media companies have developed international counterspeech/counter-narrative programs and tooling to help activists and NGOs upscale and optimize their voices online in their efforts to push back on hate speech and extremism.

The final sections of this document supports preventative measures to counter extremism online, specifically, digital literacy training and tooling. We have included, where available, open-access resources related to digital literacy. In deployment of counterspeech and digital literacy we have also included relevant resources that guide activists and NGOs on how to engage and utilize advertising and marketing tools to ensure their campaigns and voices reach the relevant target audiences and can be effectively measured.

For ease of use, the appendix gives all the resource links in one comprehensive table by company, theme and URL link. GIFCT hopes that this document will serve as a useful resource and guide for better understanding and joint efforts in challenging terrorism and violent extremism online.

## Community Standards

Most platforms require users to agree to set terms of service before accessing a platform's tools. While terms of service give legal parameters for usage, community standards explain in a less legally formatted way what is and is not allowed on a given tech platform. These guidelines are put in place to ensure users can use the platforms freely and safely, while understanding where they might cross the line and have their content or engagement on the platform limited or removed. These guidelines are generally global, since it is difficult for platforms to operate on different standards in different countries. Most companies design, review and update these standards based on feedback from a range of stakeholders that might include their online communities, government bodies, and global experts in fields such as technology, public safety and human rights. GIFCT members enforce their own policies and standards, and engage in other

standard operating procedures associated with content removals, account closures, and similar enforcement actions related to violations of terms of service or community standards.

A key <u>GIFCT Membership</u> criteria revolves around content standards that explicitly prohibit the promotion of terrorism and/or violent extremism in their community guidelines or other publicly available content policies. Just like governments, intergovernmental institutions, civil society organizations, and academics, tech companies often have slightly different definitions of "terrorism," "terrorist content" and "violent extremism." While there is no one agreed upon definition of terrorism, violent extremism or hate speech, most tech companies, in their independent capacity, have developed definitions and approaches based on existing resources and in consideration of what will work best based on how their platforms operate.

For example, <u>Microsoft</u> considers terrorist content to be material posted by or in support of organizations included on the Consolidated United Nations Security Council Sanctions List that depicts graphic violence, encourages violent action, endorses a terrorist organization or its acts, or encourages people to join such groups. Given this definition, Microsoft has been amending their <u>guidelines</u>, which already prohibit hate speech and advocacy of violence against others to specifically <u>prohibit the posting of terrorist content</u> on their platforms.

In a similar vein, companies such as Facebook have <u>community standards</u> that also address a range of harmful content and abuse within their violence and criminal behavior policies, which range from violence and incitement policies to fraud and deception. More specifically, Facebook's <u>dangerous individuals and organizations policies</u> explicitly do not allow any organizations or individuals that proclaim a violent mission or are engaged in violence to have a presence on their platform. This includes organizations or individuals involved in terrorist activity, organized hate, mass murder (including attempted), human trafficking, organized violence and criminal activity. <u>Instagram community guidelines</u> also have a clear set of policies for what is and isn't allowed on their platform. These guidelines refer explicitly to the illegality of any material that supports or praises terrorism, organized crime, or hate groups.

Similarly, <u>LinkedIn</u> is also committed to ensuring their platforms remain safe for all their users and indicate that they do not tolerate any content that depicts terrorist activity, that is intended to recruit for terrorist organizations, or threatens, promotes, or supports terrorism. It also provides clarity on publishing <u>guidelines</u>. Also explicit in their guidelines, <u>Twitter</u> and <u>YouTube</u> do not allow violent organizations, including terrorist organizations, violent extremist groups, or individuals who affiliate with and promote the illicit activities of these groups to use their platforms for any purposes, including recruitment.

Dropbox has put in place an <u>Acceptable Use Policy</u> page which explicitly condemns the use of their platform to publish or share materials that contain extreme acts of violence or terrorist activity, including terrorist propaganda. Similarly, <u>MailChimp's Acceptable Use Policy</u> is clear about suspending or terminating accounts which contain behavior, content, and in particular any campaigns promoting or supporting terrorism or violent extremism.

WhatsApp's responsible use guide includes a video, a list of best practices, and a list of practices to avoid on the messaging platform. Though WhatsApp uses end-to-end encryption - meaning that the company does not have access to users' communications - it has clear reporting mechanisms and explicitly condemns illegal, obscene, defamatory, threatening, intimidating, harassing, hateful, racially, or ethnically offensive content. While MEGA's platform also features end-to-end encryption, its <u>Terms of Service</u> are explicit in not condoning, authorizing, supporting or facilitating the storage of any materials relating to violent extremism.

Platforms not typically considered to be social media platforms also recognize the importance of providing clear guidelines and rules to keep their users safe. For example, <u>Amazon</u> explicitly prohibits any activity that encourages or supports behavior that is illegal, including violence or content encouraging physical or financial harm, including terrorism. In <u>Airbnb</u>'s trust and safety hub, its <u>community standards</u> refer to five pillars to guide its users: safety, security, fairness, authenticity and reliability.

While the public tends to focus on the four to five larger GIFCT member companies, this criterion is necessary for all GIFCT members, large or small. Airbnb, Discord and Pinterest have benefited from reviewing and refreshing their Terms of Service and Community Guidelines through their membership process. One of the benefits a company receives while going through the GIFCT membership process is having access to the Tech Against Terrorism mentorship program, sponsored by GIFCT. This mentorship program supports tech companies who want to further develop policy language references to terrorism and/or violent extremism in their Terms or Service or Community Guidelines to better combat terrorist use of their services.

In addition to Tech Against Terrorism resources offered to members, GIFCT also ensures there is a large body of academic and expert insights that a company can review as they further develop their definitional language and internal understanding on how terrorist and violent extremist organizations exploit online services. GIFCT works in close partnership with the <u>Global Network on Extremism and Technology (GNET)</u>, an independent but GIFCT-funded global academic network. GNET is a highly valued resource not only for tech companies but for a wide range of stakeholders including academics, civil society, and policy

practitioners. Topical <u>insights</u>, <u>reports</u>, and workshops with global experts are designed to better understand, and counteract, terrorist use of technology.

### Transparency

In addition to explicit platform policies that prohibit terrorist and violent extremist content or activities, every GIFCT member is required to publish, at a minimum, an annual transparency report that reviews elements of how standards are upheld and enforced. Though exact format and content presentation vary from company to company, these reports tend to provide data on basics around government data request processing and wider removal compliance. As a company has more resources and ability to build out more nuanced metrics within a transparency report, these reports might also include data on the appearance and removal of specific types of violating content and the amount of content on which the platform took action. In the most advanced examples of transparency reporting to date these reports sometimes include rates of appeals to report on false/positive removal rates, and even the prevalence of violating content on the platform over time.

NGOs, academics, and governments can utilize this information to better understand how platforms enforce community standards, respond to violating content, and what the landscape of violating content looks like on a platform. This can inform how NGOs communicate about violent extremism on the platform.

Some of GIFCT's larger members include specific categories in their transparency data for terrorism and violent extremism, while others might have this information incorporated into other areas of violating or removed content. Member companies that include specific data on content flagged and/or removed for promoting terrorism are <u>Facebook</u>, <u>Microsoft</u>, <u>Twitter</u> and <u>YouTube</u>. These members are able to provide more granular data due to greater resources and additional personnel that are able to support scaled up efforts to detect, label, and review content. It is also important to note that these companies' reports rely on slightly different definitions of terrorism and violent extremism.

Transparency reports also include data on the enforcement of community standards, requests for information by governments and law enforcement, as well as how content is restricted based on local laws. <u>LinkedIn</u>, <u>MEGA</u> and <u>Mailchimp's</u> transparency reports all feature a combination of these elements.

Transparency reports from GIFCT members <u>Airbnb</u>, <u>Amazon</u>, <u>Pinterest</u> and <u>Dropbox</u> primarily provide data focused on government and law enforcement requests due to how their platforms are set up and how on-platform abuse takes place. <u>Discord's most recent transparency report</u> focuses primarily on the company's responses to reports made by users, as well as efforts by the company

to proactively enforce community standards by removing racist and extremist groups organizing on its servers.

Recognizing that tech companies, governments, and NGOs can always improve how they convey transparency and what is added to transparency data, GIFCT hosts a <u>Transparency Working Group</u> that brings together international, multisector stakeholders to foster better collaboration on how best to approach transparency in the context of tech platforms. This multi-sector effort ensures that GIFCT can best support its mission to facilitate dialogue between tech companies and other stakeholders on key issues.

For tech companies seeking membership that do not currently produce regular transparency reports, 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. As a reminder, all platforms will naturally have different approaches to transparency reporting both in terms of frequency and data included. Some larger companies give transparency updates as often as quarterly, while that sort of frequency would put a massive strain on the communication, policy, data and legal teams at smaller companies. It should also be noted that depending on how a platform is set up, there might naturally be less data available to report on.

Finally, GIFCT has its own commitment to transparency and produces an <u>annual Transparency Report</u>, which documents our organization's growth, year's activities and progress in joint-tech innovations, and knowledge-sharing initiatives.

## Safety Hubs

Often, activists, academics, journalists, and practitioners in the counterterrorism and counter-extremism space have to risk-mitigate in the online space more than the average user. By using online platforms to research or challenge hate-based extremism and terrorism, an individual can often put themselves or the sensitive communities they engage with online at risk. Safety concerns also tend to arise for many activists and NGOs who often have public facing interactions with different communities and as a result tend to face several risks online. Platforms' safety guidelines and resources are, therefore, increasingly important to ensure that users know how to flag abuse, manage privacy settings, and report things such as potentially hacked accounts or nefarious activities of dangerous organizations.

Companies such as <u>Discord</u>, <u>Facebook</u>, <u>Microsoft</u>, <u>YouTube</u> and <u>Google</u> have extensive safety hubs that cover a wide range of abuse topics and risk mitigation

work streams. For example, Discord provides account tips, server management and how to report problems directly to their Trust and Safety team as well as a <u>Parents and Educators</u> section. Facebook also ensures that its safety hubs cater to a range of different users, thereby addressing the needs of <u>law enforcement</u>, <u>parents</u>, and <u>young people</u>. Similarly, YouTube has safety resources focusing on <u>online harms</u> as well as <u>parent</u>, <u>teens</u> and <u>educator</u> resources.

Sometimes safety resources include third party facilitation and external experts to ensure safety measures develop as adversarial shifts occur. For example, platforms such as Facebook have a <u>Safety Advisory Board</u>, composed of leading internet safety organizations from around the world, who provide expertise, perspective, and insights that inform their approach to safety. The tools provided address a range of safety concerns, such as privacy measures and account security. Critically, it also engages substantially with issues of abuse and crisis response by providing a range of resources and mechanisms to address and mitigate those risks for their users. Meanwhile, Twitter has a dedicated <u>Twitter account</u> with the latest safety tools, resources, and updates to support its community. Twitter also provides tools geared towards <u>brand safety</u>.

Larger social media platforms tend to have more capacity to develop comprehensive safety tools. For example, <u>Microsoft</u>'s safety hub provides a <u>Digital Safety Content Report</u> that explicitly details their mechanisms for prohibiting certain content and conduct. In particular, it dedicates a significant section on <u>what they do to help to prevent terrorists and violent extremists from exploiting digital platforms</u>, including by addressing images and videos on their hosted services that include terrorist or violent extremist content.

<u>Instagram's Safety Hub</u> provides comprehensive safety tools and a <u>help center</u> that gives safety guidelines. The center provides resources around privacy and security, reporting abuse, and digital wellbeing resources for its community.

<u>WhatsApp</u> aims to ensure that all users remain safe while using its platform through tools including privacy controls and reporting mechanisms. It has also published a <u>white paper</u> on how WhatsApp prevents abuse of automated and bulk messaging. Similarly, <u>MEGA</u> provides users with a range of tools to remain safe across their different streams.

<u>Dropbox</u> and <u>Pinterest</u>, have developed principles and platform-specific approaches to tackle terrorist exploitation on their platforms with a focus on keeping users safe. Additionally, <u>Airbnb</u> provides safety guidance for both online and offline components of their service, providing users with proper information throughout their experience. On <u>Airbnb's Trust and Safety page</u> and in the <u>Airbnb Host Resource Center</u>, both hosts and travelers can learn more about the wider safety resources and information offered by the platform.

The evolution of safety tools and risk mitigation strategies also comes from external dialogues and international commitments with bodies such as the United Nations Counterterrorism Executive Directorate, the European Union Internet Forum, and the Christchurch Call to Action. GIFCT members Amazon, Facebook, Microsoft, Twitter and YouTube signed onto the <u>Christchurch Call to Action</u> in 2019 as part of their commitment to address the exploitation of technology by users looking to spread terrorist and violent extremist agendas online. These companies and GIFCT <u>committed to a nine-step plan</u> for individual and collective action to prevent and respond to terrorist or violent extremist content.

In the GIFCT-funded <u>Campaign Toolkit</u>, there is also a <u>Digital Security Checklist</u>, which lays out the basics for digital security on computers and phones, security practices for staff involved in operating campaigns, and steps to take if a campaign is threatened, harassed, or targeted by bad actors.

Overall, safety across digital platforms is a shared industry responsibility and efforts are furthered by a range of stakeholders including private sector, academic researchers, civil society, governmental and intergovernmental actors. Multi-stakeholder engagement to further this is one of <u>GIFCT's foundational goals</u>. When different stakeholders work together to address this ever evolving and complex challenge, we see progress that would be difficult for any one entity to tackle alone.

## Reporting Mechanisms

Related to broader safety concerns, another <u>GIFCT membership requirement</u> states that all members must have a functional way to "receive and act on reports of illegal activity or activity violating terms of service." While members approach this in a variety of ways, all member companies provide public resources and tools to assist users in reporting illegal and prohibited activity and content.

At the very least a publicly available email address or direct contact portal is available, by which users may contact the company and report content or activity in violation of the platform's guidelines. <u>Mailchimp</u>, <u>MEGA</u>, and <u>WhatsApp</u> all provide basic outreach portals of this kind. For companies like MEGA and WhatsApp, which feature end-to-end encryption, this is the most efficient way to report abuse because the company does not otherwise have access to the content and activity of users.

In addition to outreach portals, other member companies use a combination of tools including content-adjacent reporting mechanisms and interactive online forms. Some larger companies have tools for giving more granular feedback on the type of abuse violating content might fall under. Many of these reporting tools

also provide guidance on how to report content using the aforementioned content-adjacent mechanisms, ensuring that users are able to more easily report content in the future.

<u>Discord</u>, <u>Dropbox</u>, and <u>Pinterest</u> make use of general-purpose forms for reporting violating content. Discord requires users to share links to reported messages in their online form and provides instructions for users to do so successfully. In addition to their online forms, <u>Dropbox</u> and <u>Pinterest</u> also provide instructions for users to report content where it appears on the platform. Amazon allows customers to report violating content via email and content adjacent reporting mechanisms. Information on both of these tools is located at the base of <u>Amazon's Community Guidelines Page</u>.

Airbnb provides guidance for <u>reporting and blocking other users</u> and <u>reporting discrimination on their platform</u>. Additionally, Airbnb provides users with a <u>neighborhood support portal</u> to call or chat with a representative about safety and broader neighborhood concerns.

<u>Facebook</u>, <u>Instagram</u>, <u>LinkedIn</u> and <u>Twitter</u> rely on similarly structured pages in their help centers to provide information about reporting mechanisms and instructions for how to report content, profiles, and other platform aspects through content-adjacent mechanisms. In addition to on-platform reporting resources, each of these companies also provides direct contact reporting options and guidance for reporting content without having an active user account.

Microsoft has a variety of reporting pages focused on different community services and types of violating content. Microsoft has portals for reporting both Terrorist Content and Hate Speech across all of the company's consumer services, as well as a specific reporting portal for reporting concerns related to Bing.

## Counterspeech and Counter-Narrative Facilitation

While safety tools allow users to flag harm and on-platform violations that lead to possible removal of content, we also know that removal alone will never solve for the root causes of radicalization and hate-based extremism. Larger social media platforms have realized that their platforms also house crucial tools for activists and practitioners to develop content and online activities for prevention purposes. Some GIFCT members have developed tools that can optimize community-level voices that challenge hate speech and extremism. Supporting counterspeech and counter-narrative efforts emerges out of a recognition that deplatforming and removing content only addresses symptoms of radicalization, rather than causes.

Counterspeech efforts supported by tech companies often focus on helping activists and organizations produce, launch and evaluate localized content to challenge engaging narratives created by violent extremists. Other forms of counterspeech aim to create friction between individuals searching for violent extremist content by proactively surfacing alternative voices or local resources for disengagement. These efforts rely on strong partnerships between the tech companies and local civil society organizations around the world, reinforcing GIFCT's belief that multi-sector collaboration is a critical element of effectively combating extremism in online spaces.

<u>Facebook's Counterspeech</u> Hub houses a range of tips, tools, research and examples for activists and organizations to better understand Facebook's current initiatives and resources in the counterspeech space. The site's Initiatives page showcases the range of global projects, partnerships, and positive intervention approaches already underway. These include programs they run, such as the <u>Redirect Initiative</u>, as well as programs they have in partnership with bodies like the United Nations, such as <u>Extreme Lives</u>. The <u>Counterspeech Resources</u> section offers broader guidance for nonprofit organizations on how to take advantage of Facebook's tools and products, as well as more specific information on crafting counterspeech projects.

YouTube's Creators for Change program is a global initiative that partners with content creators from over 20 countries to promote positive messages, give space to community voices, and empower youth voices as drivers of social good. Creators - known as Ambassadors and Fellows - receive mentorship and promotional assistance to work on Impact Projects, which engage viewers on a variety of challenging topics, including combating extremism, hate, and intolerance. YouTube was also an early pioneer of The Redirect Method, a tool created by Jigsaw and Moonshot CVE to "redirect" Google Search and YouTube users away from violent extremist recruitment content to YouTube videos that credibly refuted extremist content and recruitment materials. Modified versions of the initial pilot are being used across Bing, Facebook, and Instagram to address a variety of challenges.

To further foster this, GIFCT worked to gather resources that serve activists and organizations in their counterspeech efforts, working with ISD Global to produce the online <u>Campaign Toolkit</u>, a digital resource trove for anyone interested in producing and promoting counterspeech campaigns to global audiences. The Campaign Toolkit provides users with guidance as they develop campaigns and counter-narrative projects, as well as resources from many of our member companies to optimize the use of their platforms. Additionally, GIFCT hosts a <u>Content-Sharing Algorithms</u>, <u>Processes</u>, and <u>Positive Interventions Working Group</u> to ensure we further dialogues around best practices, monitoring and evaluation, and innovations.

To <u>learn more about counterspeech and counter-narrative</u> deployment, you can also look to our academic partners at the Global Network on Extremism and Technology (GNET) for published work by academics and researchers on this important and ever-evolving subject.

## Digital Literacy

While counterspeech and counter-narrative campaigns and initiatives look to disrupt a cycle or pathway of hate-based radicalization, many practitioners and researchers have highlighted that preventing violent extremism needs to start with better comprehensive digital literacy for online audiences (young and old). Digital Literacy is a broad term that refers to a range of skills and information that allow individuals and organizations to safely and effectively operate in digital environments by questioning online information and having the tools to discern opinion from fact. While some of the material shared here is targeted at younger people, it is equally useful to those unfamiliar with operating on digital platforms and those who wish to engage younger audiences (activists, NGOs, teachers) by providing the basics in clear, straightforward language that can be easily communicated to others.

Some of our companies have built out substantial educational resources for digital literacy on their platforms and the internet at-large. Within <u>Discord's Safety Center</u>, users can find sections on "Controlling Your Experience" and a section for "Parents and Educators," which provide in-depth information for users to make their own experience safer. This also includes resources for those that might have responsibility for the safety of others on the platform.

Facebook hosts a <u>Digital Literacy Library</u>, which offers thorough modules for comprehension on several topics including; Security, Privacy and Reputation, and Community Engagement. Likewise, <u>Instagram's Community page</u> provides users with information about how the platform aims to protect its users, as well as a broad range of safety tools and programs to ensure safe, positive user experiences.

LinkedIn's <u>Staying Safe</u> page highlights best practices for LinkedIn users to maximize their use of the platform's products safely and effectively. The page also offers guidance for protecting personal information and resources for parents, educators and teenagers using the platform.

<u>Microsoft's digital literacy resources</u> include two LinkedIn Learning courses, as well as a series of courses presented in a series of downloadable videos. Microsoft also has an <u>Online Safety page</u>, with sections dedicated to <u>Digital Civility</u> and further <u>Resources</u> and <u>Research</u>.

Twitter, in collaboration with the United Nations Educational, Scientific and Cultural Organization (UNESCO), produced a document titled <u>Teaching and Learning with Twitter: Media and Information Literacy</u>. Additionally, <u>A Safer Twitter</u>, located within Twitter's Help Center, hosts a series of short videos that explain how to use some of the basic mechanisms offered on the platform to enhance the user's safety and experience.

Finally, YouTube provides a suite of <u>Safety Resources</u>, which cover a broad list of digital literacy subjects with specific information catering to <u>parents</u>, <u>educators</u> and <u>teenagers</u>. Additionally, YouTube also provides general-purpose resources for <u>staying safe online</u> and <u>staying safe on YouTube</u>, which educate users phishing, account safety and basic mechanisms such as reporting and blocking.

## Advertising and Marketing

Some GIFCT member companies house advertising and marketing functions based on how they function and monetize. While these tools are originally developed for more traditional commerce and marketplace use cases, we also know that these tools can be optimized by activists, NGOs and practitioners to better launch their messaging to target audiences and further counterspeech goals. Some companies have advertising and marketing tools to ensure that a range of organizations and/or activist's campaigns and voices reach the relevant target audiences and can be effectively measured.

For example, <u>Facebook for Business</u> is a free resource that allows small and medium businesses to expand their internet presence for greater visibility and reach. Supported by <u>Facebook Analytics</u>, these tools allow a user to understand basic metrics of reach and engagement with their content. Meanwhile, the <u>Facebook Social Impact</u> site provides resources around a range of impact aimed streams: charitable causes, crisis response to vulnerable and affected people, health and mentorship.

The <u>YouTube Advertising</u> team helps organizations connect with their target audiences by supporting development content. YouTube Social Impact also helps users harness YouTube's scale, technology and talent to ignite and sustain movements that drive measurable social change. In this way YouTube also partners with others such as Google.org, Google for Nonprofits, Jigsaw and YouTube Creators For Change.

Google has a specific <u>Google for Nonprofits</u> program which allows organizations, NGOs and other grassroot communities to use their platform to elevate their mission, engage with new supporters and fundraise online effectively to reach and impact maximum audiences. Furthermore, in terms of making sure a platform has a greater social impact, Google's <u>Al for Social Good</u>,

is a program that focuses Google's AI expertise on solving humanitarian and environmental challenges. With a global network of researchers in Japan, India, Israel, and Ghana it works closely with nonprofits, academia and the public sector to create systemic solutions to global issues.

Similarly, Microsoft provides a range of different open-source tools to support individuals, organizations, activist and business growth. For organizations seeking to make use of advertising tools to expand their reach, Microsoft Advertising (BING) provides insights, technology and proven expertise that empower organizations to deliver better-performing marketing campaigns. Microsoft Advertising Learning Lab offers free, on-demand, online learning courses to help users better understand the Microsoft Advertising interface. Tools and best practices for maximizing campaigns and investment with Microsoft Advertising are also available. Finally, Microsoft Nonprofit ensures the benefits of world-class cloud technology is accessible and affordable for nonprofits. They provide nonprofit grants and discounts as well as industry-leading solutions help accelerate their impact. Microsoft provides step-by-step instructions to make this process as easy as possible.

<u>Twitter</u> for business provides a range of free resources around campaign types, takeover products, advertising best practices and analytics. Especially for <u>small businesses</u> these are valuable tools on how to navigate their platform with recommendations, ideas and educational resources.

Finally, above and beyond GIFCT member resources, additional resources include a <u>Campaign Toolkit</u>, which act as an information hub, providing activists and organizations with resources on a range of useful information, in particular for advertising, audience targeting and social good.

## Appendix A: GIFCT Membership Criteria and Benefits

#### Membership Criteria

GIFCT member companies are listed on our website <u>here</u>. In order to join GIFCT, companies must meet the following requirements:

- Content standards that explicitly prohibit the promotion of terrorism in their terms of service, community guidelines, or other publicly available content policies
- ❖ The ability to receive and act on reports of illegal activity or activity violating terms of service
- A desire to explore new technical solutions to content and conduct challenges
- Regular, public data transparency
- A public commitment to respecting human rights, particularly free expression and privacy, when implementing content removal policies
- Support for expanding the capacity of civil society organizations to challenge violent extremism

If a company looking to join the GIFCT does not meet certain requirements, GIFCT offers that company mentorship through our partnership with  $\underline{\text{Tech Against Terrorism}}$ .

#### Membership Benefits

- Potential access to GIFCT's <u>hash-sharing</u> database and URL sharing program
- Participation in crisis response communications around international terrorist and violent extremist events with online implications
- Briefings from scholars associated with the <u>Global Network on Extremism</u> and <u>Technology (GNET)</u>, the academic research arm of GIFCT
- Briefings on technological approaches and solutions
- Priority participation in topical workshops, e-learnings and webinars with global experts

Contact: for any questions contact us at <a href="mailto:outreach@gifct.org">outreach@gifct.org</a>.

Sign up to receive email updates from the Global Internet Forum to Counter Terrorism

## Appendix B: GIFCT and Partner Resources Index

Partner	Webpage Title	Link
GIFCT	Membership Pillars	https://gifct.org/membership/
GIFCT	Transparency	https://gifct.org/transparency/
GIFCT	Working Groups	https://gifct.org/working-groups/
GIFCT	News Page and Newsletter Subscription	https://gifct.org/news/
GIFCT	Campaign Toolkit	https://www.campaigntoolkit.org/
GIFCT	Digital Security Checklist	http://campaigntoolkit.org/resources/campaign- toolkit-digital-security-check-list/
TAT	Mentorship	https://www.techagainstterrorism.org/membership/tech-against-terrorism-mentorship/
TAT	Workshops and Events	https://www.techagainstterrorism.org/events/project-events/
TAT	Knowledge Sharing Platform (KSP)	https://ksp.techagainstterrorism.org/
GNET	Homepage: Insights and Reports	https://gnet-research.org/

## Appendix C: Member Resources Index

Member	Webpage Title	Link
Airbnb	Community Standards	https://www.airbnb.com/trust/standar ds
Airbnb	Transparency	https://news.airbnb.com/transparency
Airbnb	Safety and accessibility	https://www.airbnb.com/help/topic/13 98/safety-and-accessibility
Airbnb	Contact Us (Safety and Discrimination)	https://www.airbnb.com/help/contact -us/topic-search
Airbnb	How do I report a message or block someone on Airbnb?	https://www.airbnb.com/help/article/ 2020/how-do-i-report-a-message- or-block-someone-on-airbnb
Airbnb	How do I report discrimination to Airbnb?	https://www.airbnb.com/help/article/l433/how-do-i-report-discrimination-

		<u>to-airbnb</u>
Amazon	Community Guidelines	https://www.amazon.com/gp/help/customer/display.html?nodeld=GLHXEX85MENUE4XF
Amazon	Security and Privacy - Law Enforcement Information Requests	https://www.amazon.com/gp/help/customer/display.html?nodeld=GYSDRGWQ2C2CRYEF
Amazon	Security and Privacy	https://www.amazon.com/gp/help/customer/display.html/ref=hp bc nav?ie=UTF8&nodeId=GRFTMVHP4HXMESSP
Discord	Community Guidelines	https://discord.com/guidelines
Discord	Transparency Report	https://discord.com/safety/360043712 232-Discord's-Transparency-Report
Discord	Safety Center	https://discord.com/safety
Discord	Trust and Safety - How to Properly Report Issues to Trust and Safety	https://support.discord.com/hc/en- us/articles/360000291932-How-to- Properly-Report-Issues-to-Trust-Safety
Dropbox	Acceptable Use Policy	https://www.dropbox.com/acceptable use
Dropbox	Transparency Overview	https://www.dropbox.com/transparen
Dropbox	Security and Privacy - Report Abuse on Dropbox	https://help.dropbox.com/accounts- billing/security/report-abuse
Dropbox	Blog, 2019: Protecting our users and society: guarding against terrorist content	https://blog.dropbox.com/topics/comp any/protecting-our-users-and-society- -guarding-against-terrorist-con
Facebook	Community Standards	https://www.facebook.com/communitystandards/
Facebook	Transparency Report	https://transparency.facebook.com/
Facebook	Digital Literacy Library	https://www.facebook.com/safety/educators
Facebook	Community Standards - Dangerous Individuals and Organizations	https://www.facebook.com/community standards/dangerous_individuals_orga nizations
Facebook	Combating Hate and Extremism (FB News, 2019)	https://about.fb.com/news/2019/09/combating-hate-and-extremism/
Facebook	What Are We Doing to Stay Ahead of Terrorists? (FB News, 2018)	https://about.fb.com/news/2018/11/sta ying-ahead-of-terrorists/
Facebook	Analytics	https://analytics.facebook.com/

Facebook	Business	https://www.facebook.com/business
Facebook	Facebook Counterspeech	https://counterspeech.fb.com/en/
Facebook	Social Impact	https://socialimpact.facebook.com/
Facebook	Safety Center	https://www.facebook.com/safety
Facebook	Policies and Reporting - How to Report Things	https://www.facebook.com/help/1380 418588640631/how-to-report- things/?helpref=hc_fnav
Instagram	Community Guidelines	https://www.facebook.com/help/instagram/477434105621119
Instagram	Community Guidelines FAQs (Instagram Blog, 2018)	https://about.instagram.com/blog/announcements/instagram-community-guidelines-faqs
Instagram	Transparency Report	https://transparency.facebook.com/
Instagram	Help Center - Report Something	https://help.instagram.com/165828726 894770/?helpref=hc_fnav&bc[0]=Inst agram%20Help&bc[1]=Privacy%20an d%20Safety%20Center&bc[2]=Report %20Something
Instagram	Community	https://about.instagram.com/community/
LinkedIn	Professional Community Policies	https://www.linkedin.com/legal/professional-community-policies
LinkedIn	Transparency	https://about.linkedin.com/transparenc
LinkedIn	Report Inappropriate Content, Messages, or Safety Concerns	https://www.linkedin.com/help/linkedin/answer/146/report-inappropriate-content-messages-or-safety-concerns?lang=en
LinkedIn	Safety Center - Identifying Abuse	https://safety.linkedin.com/identifying- abuse
LinkedIn	Safety Center - Staying Safe	https://safety.linkedin.com/staying- safe
Mailchimp	Acceptable Use Policy	https://mailchimp.com/legal/acceptable_use/
Mailchimp	Transparency Reports	https://mailchimp.com/transparency- report/
Mailchimp	Abuse Desk	https://mailchimp.com/contact/abuse/
MEGA	Terms of Service	https://mega.nz/terms
MEGA	Transparency Report 2020	https://mega.nz/Mega_Transparency_

		Report 202009.pdf
MEGA	Contact Us - Report Objectionable Material	https://mega.io/contact
Microsoft	Community Code of Conduct	https://answers.microsoft.com/en- us/page/codeofconduct
Microsoft	Report Hate Speech Content Posted to a Microsoft Hosted Consumer Service	https://www.microsoft.com/en- us/concern/hatespeech
Microsoft	Corporate Social Responsibility - Reports Hub	https://www.microsoft.com/en- us/corporate-responsibility/reports- hub
Microsoft	Digital Literacy	https://www.microsoft.com/en- us/digitalliteracy
Microsoft	Report Terrorist Content Posted to a Microsoft Consumer Service	https://www.microsoft.com/en- us/concern/terroristcontent
Microsoft	Register with Microsoft Nonprofits	https://nonprofit.microsoft.com/en- us/getting-started
Microsoft	Advertising	https://about.ads.microsoft.com/en- gb/h/a/microsoft-advertising
Microsoft	About Microsoft Advertising	https://about.ads.microsoft.com/en- gb/get-started/about-microsoft- advertising
Microsoft	Al For Good	https://www.microsoft.com/en- us/ai/ai-for-good
Microsoft	Partner Network - Sales and Marketing	https://partner.microsoft.com/en- cy/marketing
Microsoft	Online Safety	https://www.microsoft.com/en- us/digital-skills/online- safety?activetab=protect-whats- important%3aprimaryr3
Microsoft	Online Safety - Resources and Research	https://www.microsoft.com/en- us/digital-skills/online-safety-resources
Microsoft	Security	https://www.microsoft.com/en- us/security
Microsoft	Report a Concern to Bing	https://www.microsoft.com/en- us/concern/bing
Microsoft	Advertising Learning Lab	https://learninglab.about.ads.microsoft .com/
Microsoft	"Stay alert to online grooming: Online, people aren't always who they say they are" (Document)	https://query.prod.cms.rt.microsoft.co m/cms/api/am/binary/RE1If3n
Microsoft	Microsoft partners with Institute for Strategic Dialogue and NGOs to discourage online	https://blogs.microsoft.com/on-the-issues/2017/04/18/microsoft-partners-

	radicalization to violence (blog post, Microsoft on the Issues)	institute-strategic-dialogue-ngos- discourage-online-radicalization- violence/
Microsoft	Corporate Citizenship - Support for NGOs	https://www.microsoft.com/en- hk/sparkhk/support-for-ngos
Microsoft (Bing)	Advertising Training Courses	https://about.ads.microsoft.com/en- gb/resources/training/courses
Microsoft (Bing)	Advertising - Ad products: Solutions for online advertising	https://about.ads.microsoft.com/en- gb/solutions/ad-products
Microsoft	Online Safety - Promoting Digital Civility	https://www.microsoft.com/en- us/digital-skills/digital- civility?activetab=dci_reports:primaryr 4
Microsoft	Request to Reinstate Disabled Content	https://www.microsoft.com/en- us/concern/reinstatecontent
Pinterest	Community Guidelines	https://policy.pinterest.com/en/commu nity-guidelines
Pinterest	Transparency Report	https://help.pinterest.com/en/article/transparency-report
Pinterest	Help Center - Safety - Report Something on Pinterest	https://help.pinterest.com/en/article/report-something-on-pinterest
Pinterest	Help Center - Safety - Get More Help - Report a Policy Violation	https://help.pinterest.com/en/contact
Twitter	The Twitter Rules	https://help.twitter.com/en/rules-and- policies/twitter-rules
Twitter	NGO Handbook	https://about.twitter.com/content/dam/about- twitter/en/tfg/download/campaigning-on-twitter-handbook-2019.pdf
Twitter	Transparency	https://transparency.twitter.com/
Twitter	Safety and Security	https://help.twitter.com/en/safety- and-security
Twitter	A Safer Twitter	https://help.twitter.com/en/a-safer- twitter
Twitter	Violent Organizations Policy	https://help.twitter.com/en/rules-and- policies/violent-groups
Twitter	Advertising - Build Campaigns	https://business.twitter.com/en/advertising/campaign-types.html
Twitter	How Twitter Ads Work	https://business.twitter.com/en/help/troubleshooting/how-twitter-adswork.html

Twitter	Advertising - Targeting	https://business.twitter.com/en/advertising/targeting.html
Twitter	Advertising - Campaign Types	https://business.twitter.com/en/advertising/campaign-types.html
Twitter	Analytics	https://analytics.twitter.com/about
Twitter	Business Resources and Guides	https://business.twitter.com/en/resources.html
Twitter	Corporate Philanthropy	https://about.twitter.com/en/who-we- are/twitter-for-good
Twitter	Hateful Conduct Policy	https://help.twitter.com/en/rules-and- policies/hateful-conduct-policy
Twitter	Q&A: The marketing team behind @SimonBooks shares tips on creating impactful organic and paid content (Blog Post, Twitter Business)	https://business.twitter.com/en/blog/marketing-team-behind-simonbooks.html
Twitter	@TwitterSafety (Twitter Account)	https://twitter.com/twittersafety
Twitter	Help - Safety and Security - Sensitive Content	https://help.twitter.com/en/safety- and-security#sensitive-content
Twitter	Business - Twitter Video Resources	https://business.twitter.com/en/resources/video.html
Twitter	Rules and Policies - Report Violations	https://help.twitter.com/en/rules-and- policies/twitter-report-violation
Twitter	Teaching and Learning with Twitter Media and Information Literacy	https://about.twitter.com/content/dam/about- twitter/en/tfg/download/teaching- learning-with-twitter-unesco.pdf
WhatsApp	How to Use WhatsApp Responsibly	https://faq.whatsapp.com/general/sec urity-and-privacy/how-to-use- whatsapp-responsibly/?lang=en
WhatsApp	Safety Tips	https://www.whatsapp.com/safety
WhatsApp	Security	https://www.whatsapp.com/security
WhatsApp	Help Center - Advanced Safety and Security Features - Report Issues to WhatsApp	https://faq.whatsapp.com/general/sec urity-and-privacy/staying-safe-on- whatsapp/?lang=en
YouTube	Community Guidelines	https://www.youtube.com/howyoutubeworks/policies/community-guidelines/#community-guidelines
YouTube	Transparency - Community Guidelines Enforcement	https://transparencyreport.google.com/youtube-policy/removals?hl=en

YouTube	Safety Resources	https://support.google.com/youtube/topic/9386941?hl=en&ref_topic=2803240
YouTube	Policies - Violent Criminal Organizations	https://support.google.com/youtube/answer/9229472?hl=en&ref_topic=9282436
YouTube	Help - Hate Speech Policy	https://support.google.com/youtube/answer/2801939?hl=en-GB
YouTube	Advertising	https://www.youtube.com/intl/en- GB/ads/
YouTube	Creators for Change	https://www.youtube.com/creators- for-change/
YouTube	Google for Nonprofits - YouTube Nonprofit Program	https://www.google.com/nonprofits/of ferings/youtube-nonprofit-program/
YouTube	Social Impact	https://socialimpact.youtube.com/
YouTube	Help Center - Report Inappropriate Content	https://support.google.com/youtube/answer/2802027?hl=en&ref_topic=9387085
YouTube	Help Center - Other Reporting Options	https://support.google.com/youtube/a nswer/2802057?hl=en&ref_topic=93 87085
YouTube	Help Center - Report a YouTube search prediction	https://support.google.com/youtube/answer/7626105?hl=en&ref_topic=9387085
YouTube	Help Center - Trusted Flagger Program	https://support.google.com/youtube/answer/7554338?hl=en&ref_topic=9387085
YouTube	Ever wonder how YouTube works?	https://www.youtube.com/intl/ALL_uk_/howyoutubeworks/
YouTube/Go ogle	Request Google Nonprofit Account	https://www.google.com/nonprofits/a ccount/u/O/orgs
YouTube/Go ogle	Google for Nonprofits	https://www.google.com/nonprofits/
YouTube/Go ogle	Google Al for Social Good	https://ai.google/social-good/
YouTube/Go ogle	Google Digital Garage	https://learndigital.withgoogle.com/digitalgarage
YouTube/Go ogle	Google Marketing Platform	https://marketingplatform.google.com/about/
YouTube/Go ogle	Google Safety Center	https://safety.google/

	Think with Google - Discover Marketing Research and Digital Trends	https://www.thinkwithgoogle.com/
YouTube/Go ogle	Think with Google - Build a Content Plan	https://www.thinkwithgoogle.com/mar keting-strategies/video/build-a- content-plan/