

Introducing 2022 GIFCT Working Group Outputs

GIFCT WORKING GROUPS OUTPUT 2022



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
Automatic	Australian Government - Department of Home Affairs	American University	Mnemonic
Checkstep Ltd.	BMI Germany	ARTICLE 19	Moonshot
Dailymotion	Canadian Government	Australian Muslim Advocacy Network (AMAN)	ModusIzad - 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 (I267 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	

Good Practices, Tools, and Safety Measures for Researchers

GIFCT Positive Interventions and
Strategic Communications Working
Group



Introduction

When collecting data on actors such as terrorists and extremists, researchers (along with content moderators and investigators) are required to put themselves in harm's way for the greater good of society. A primary researcher investigating extremist activity online has one of the most vital and dangerous positions at a research organization, academic institution, or tech platform. The researchers in this area serve a vital need because they are engaging with actors online, which involves putting themselves at risk.¹ Similarly, content moderators (along with trust and safety specialists) tasked with reviewing dangerous content face similar (and in some cases even more severe) exposure to harmful data in the course of their work.

While the safeguarding of researchers should be a consideration of any research project, the sensitivity of extremism and terrorism research, along with the potential harms that researchers face, place a special onus on organizations to ensure researcher wellbeing. Conducting research for academic and organizational purposes has its own set of unique risks that lack a comprehensive response that would alleviate potential harms, including exposure to upsetting or potentially traumatic content, impacts on mental health, or in extreme cases risks to personal security.² Due consideration should be made of potential impacts on researcher welfare and sufficient mitigating actions should be enacted. Although a researcher may not be in a physical conflict zone, advanced security measures cannot always defend against bad actors who are constantly discovering new and innovative methods to harm individuals. This paper aims to help mitigate risks researchers may encounter by offering solutions and resources to assist them.

.....

1 Maura Conway, "Online Extremism and Terrorism Researcher Security and Privacy: Some Practical Advice," Global Network on Extremism and Technology (blog), February 19, 2021, <https://gnet-research.org/2021/02/19/online-extremism-and-terrorism-researcher-security-and-privacy-some-practical-advice/>.

2 Miriah Steiger et al., "The Psychological Well-Being of Content Moderators: The Emotional Labor of Commercial Moderation and Avenues for Improving Support," in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21) (New York, NY: Association for Computing Machinery), Article 341: 1–14, <https://doi.org/10.1145/3411764.3445092>.

Types of Threats

Threats of any kind should be taken seriously because it is impossible to recognize whether a threat may result in actual violence or other harm. The most common threat categories that researchers and at-risk audiences (individuals such as journalists, activists, and politicians who are frequently targeted with harassment) are exposed to include direct, indirect, veiled, and secondary:³



- **Direct:** A specific target and victim are explicitly identified by name. All threats of this nature should be taken particularly seriously as the named victim may be at risk of potential violence or attack.
- **Indirect:** Denoted by vagueness, lack of clarity, and non-specifics of a violent attack which do not provide enough information to understand the author's intentions.
- **Veiled:** The threat does not specifically call for violence because it is vaguely implied. The rhetoric used in this category of threats is often observed in online behaviors such as shitposting and trolling.⁴
- **Secondary:** These threats are often not from a specific actor and may include exposure to harmful content, the experience of vicarious or secondary trauma due to the nature of the content, interviews, or narratives in the research, and other potential threats posed by (for example) failing to protect from exposure the Personally Identifiable Information (PII) of the researcher or organization.

These threat categories represent only a few examples of potential techniques actors can use against researchers and at-risk groups.

Risk Management Frameworks for Researchers

The principles below provide a general risk management framework designed to be flexible for individual researchers, institutions, and specific populations and identities who may be at additional risk.

Do No Harm

A core principle of applied social research, especially on sensitive topics such as extremism, terrorism,

.....
 3 University of Arkansas Little Rock, "Types of Threats," University Police, n.d., <https://ualr.edu/safety/home/emergency-management-plan/threat-assessment-team/types-of-threats/>.

4 Note: Shitposting refers to off-topic, aggressive, and frequently ironic online posts often made in a public or semi-public group for the purpose of derailing or disrupting conversation or to upset others. Similarly, trolling is an online practice through which the instigator (the 'troll') engages in arguments of an illogical, spurious, or directly debasing nature. Trolls and shitposters typically use anonymous or pseudonymous handles to mask their identity. The actual intention of the troll or shitposter may be a genuine if misguided attempt at humor or credibility-building (doing something for 'the lolz').

and conflict is that of “do no harm.”⁵ This fundamental tenet is derived from the humanitarian and medical sectors and requires that researchers prevent and mitigate against harming study participants and their wider community (whether intentionally or out of negligence).⁶ Additionally, researchers and the institutions backing them should ensure that staff is not harmed through their work and that risks are mitigated to the highest extent possible. This principle should inform all research efforts in the terrorism and extremism research sphere.

Two corollary principles, also from humanitarian work, are that staff both have the right to cease their work at any time if they are in danger or are at a personal limit and that they have no “right to remain” after the organization ceases a project to protect its staff. In research, this implies that researchers ought never to be coerced to continue research when facing personal threats or having reached their limit⁷ and that individual researchers inside an organization must not continue working on a project if (after a risk management assessment by the institution) the risks are deemed too high to pursue completion.

Research Institutions: Duty of Care and Risk Management

Research institutions have a primary responsibility to protect themselves and their researchers from a variety of harms, including traumatization, harassment, doxing threats, and intimidation. Due to researchers investigating and studying a variety of topics that have real-world dangers, research institutions are responsible for protecting the well-being of their researchers from potential harm.⁸ The level of funding, institution staff size, and other resources can limit the amount of protection they are able to provide. However, this should not be an excuse for not providing a minimal level of risk management that can be facilitated regardless of capacity constraints. If an institution is unable or unwilling to commit to a duty of care to its staff on a project, a serious ethical and pragmatic evaluation of undertaking that project should be held.

A key element of the duty of care approach is ensuring that prior to the commencement and initiation of any fieldwork (online or offline), a thorough and holistic risk assessment and management plan are created.⁹ The risk assessment should include the likelihood of certain outcomes occurring that may cause harm, damage, or threat to the researcher, institutions, and (where relevant) the research subjects. Once this assessment is completed, a standardized numeric

.....

5 The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, “The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects for Research,” April 18, 1979, https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf.

6 Won Oak Kim, “Institutional Review Board (IRB) and Ethical Issues in Clinical Research,” *Korean Journal of Anesthesiology* 62, no. 1 (January 25, 2012): 3 – 12, <https://doi.org/10.4097/kjae.2012.62.1.3>.

7 The Sphere Project, “Humanitarian Charter and Minimum Standards in Disaster Response,” United Nations High Commissioner for Refugees, August 29, 2001, <https://www.unhcr.org/en-us/partners/guides/3b9cc1144/humanitarian-charter-minimum-standards-disaster-response-courtesy-sphere.html>.

8 Alice Marwick, Lindsay Blackwell, & Katherine Lo, “Best Practices for Conducting Risky Research and Protecting Yourself from Online Harassment (Data and Society Guide),” (New York: Data & Society Research Institute), 2016, https://datasociety.net/wp-content/uploads/2016/10/Best_Practices_for_Conducting_Risky_Research-Oct-2016.pdf.

9 Paul Hopkin, *Fundamentals of Risk Management: Understanding, Evaluating and Implementing Effective Risk Management* (London: Kogan Page Limited).

or threat level indicator system (for example, low-medium-high) should be instituted. This risk indicator system (the complexity of which again depends on project size, scale, and scope) should then be further developed by clear and process-driven mitigation and management-based actions that clearly outline the identification, actioning of decisions, and ongoing monitoring of risks (should they materialize).¹⁰ This relates to the chain of command, practical steps, and naming those responsible for administering each step and process (accepting that external agencies may also be part of this stage of risk management). A risk management plan works best when it is part of an overall project management system that includes distinct stages, controls, and processes for every aspect of research, communication, and day-to-day decisions. The creation of a risk log allows for a named person to manage and monitor duty of care issues.

It is vital for institutions to be aware of the dangers researchers face to develop precautionary measures. When harm against researchers and institutions shows signs of potential violence and hampers individual wellbeing, legal action should be taken and law enforcement should be informed of the threat. All threats should be archived to monitor the actions of the perpetrator because online risks can grow into real-life actions. However, falling short of these most extreme examples, a methodological and institutional policy should be implemented to safeguard researchers.

At a minimum, duty of care should include understanding the risks and developing prevention mechanisms for:

- Exposure to harmful and toxic content, including the potential for vicarious or secondary trauma responses among researchers;
- Virtual harassment or threats from target research populations, including violent extremist organizations and individuals;¹¹
- Physical threats or attacks against researchers or institutions; and
- Legal obligations and threats in the specific context of the researcher, as data collection may expose researchers to downloading or receiving terrorist or extremist content for which legal penalties may apply depending on the jurisdiction.

While no institutional approach is perfect in safeguarding staff all the time, key principles used by research organizations in this space include:

- Providing non-coercive, “opt-in” policies for research staff to ensure they are not forced to work on efforts they may find particularly upsetting or traumatizing. This could include briefing analysts on the potential risks to their mental wellbeing at the outset of a research project and the signing of consent forms for analysts;
- Providing accessible and effective mental health and wellbeing support (ideally in tandem with certified medical professionals) through providing resources and access to compensated time-

.....
 10 The United Nations Office for Disaster Risk Reduction, “Towards the Post-2015 Framework for Disaster Risk Reduction Indicators of Success: A New System of Indicators to Measure Progress in Disaster Risk Management,” Prevention Web, November 21, 2013, https://www.preventionweb.net/files/35716_newsystemofprogressindicatorsfordrr.pdf.

11 “So You’ve Been Doxed,” Crash Course Override Network, n.d., <http://www.crashoverridenetwork.com/soyouvebeendoxed.html>.

off for staff;¹²

- Directly de-briefing with researchers on harmful content and exposure to challenging material;
- The training of project managers on best practices for discussing mental health issues with staff;¹³
- Ensuring efforts are made to ensure that individuals affiliated with a research institution but not directly working on harmful content are not accidentally exposed to traumatic material;
- Ensuring there is a basic level of data and information security (infosec) at the individual and organizational level to protect against data breaches, network infiltration, and inadvertent exposure at the device to network levels;
- The use of encrypted communications whenever possible, including through free tools like Signal, ProtonMail, VeraCrypt, and vetted Virtual Private Network (VPN) services;
- Reviewing legal guidelines in the jurisdiction of the researchers and ensuring compliance or safeguarding against potential infringement; and
- Developing and deploying sufficient physical and virtual staff safety protocols, such as reporting lines for threats, pre-established communication links with law enforcement, and mitigation tactics such as identity monitoring, PII removal services, and physical security hardening through surveillance systems, locked and secure offices, and (if needed) security for researchers.

Malicious actors can target an institution and researcher at the same time by leaving messages on an institution's answering machine regarding the targeted researcher. These types of messages put the researcher and the institution at risk because individuals can increase their level of action by showing up to the institution in person creating a threatening environment.

Women and Minority Researchers

If they encounter extremists, researchers who identify as women will likely experience power dynamics, sexual harassment, and dehumanization aimed at belittling them. Extremists expressing misogynistic views may attempt to commit physical or virtual violence against female researchers. Similarly, individuals identifying with LGBTQ+ communities, especially those who express their sexual orientation and/or gender expression in particularly identifiable ways, may find themselves facing additional risks than other researchers. Mitigation measures, especially in research carried out

.....
 12 "The Importance of Providing Mental Health Days: They're a Valid Reason to Stay Home," Wellmark, n.d., <https://www.wellmark.com/blue-at-work/healthy-employees/provide-mental-health-days>.

13 "Why You Need to Talk About Mental Health in The Workplace: Five Tips for Addressing It," Wellmark, May 2020, <https://www.wellmark.com/blue-at-work/healthy-employees/mental-health-in-the-workplace>.

virtually, can be effective for protecting both women and LGBTQ+ identifying people, whose diverse views and analytical intelligence are necessary to include from a research perspective.

The threats against researchers and at-risk audiences are often the result of a publication, media appearances, or random acts of trolling. Many extremist individuals, organizations, and groups have a history of using violence and symbolism to reach their goals, which demonstrates the threat they can pose offline in addition to their virtual threats. As researchers become more accessible and build a public profile with their work – such as by publishing or appearing in media interviews – the opportunity for malicious actors to target them increases. The risks researchers face may hinder their ability to safely interview certain populations and individuals, implying a need to mitigate interviewing risks based on the danger posed by the target interviewee based and the identity expression(s) of the researcher.¹⁴

African Americans are a prime target for white supremacists because according to a narrative of “white genocide,” the white race in America is at risk of being wiped out as minorities are entering the country at an alarming rate. To white supremacists, being a minority and an outsider to their community automatically makes that researcher a threat, as they are supposedly contributing to wiping out the white race.¹⁵

The dangers of conducting research on extremism, terrorism, and violence can be reduced by taking measures to protect oneself online and offline. Personal protective measures can include using a secure virtual private network (VPN), paid or self-led services to remove personally identifiable information online, maintaining additional vigilance when publishing or publicly discussing research, and robust online hygiene and infosec practices (links to useful toolkits in this regard are provided in the annex to this report).

In-person and virtual interviews can also benefit from the added security and changed power dynamics provided by conducting interviews with a colleague, especially one who may be able to mitigate social biases from the interviewee. One helpful practice can be paired interviews with male and female colleagues, younger and older, or LGBTQ+ groupings (evaluated on a case-by-case basis). Utilizing safe risk mitigation practices decreases female and minority vulnerability to malicious actors.

14 Irène Bahati, “The Challenges Facing Female Researchers in Conflict Settings,” Governance in Conflict Network, June 10, 2019, <https://www.gicnetwork.be/bukavu-series-the-challenges-facing-female-researchers-in-conflict-settings/>.

15 “White Genocide,” Anti-Defamation League, April 5, 2017, <https://www.adl.org/resources/glossary-terms/white-genocide>; Milan Obaidi et al., “The ‘Great Replacement’ Conspiracy: How the Perceived Ousting of Whites Can Evoke Violent Extremism and Islamophobia,” *Group Processes & Intergroup Relations* (August 2021), <https://doi.org/10.1177/13684302211028293>; Manuel Castro e Almeida & Alistair Harris, “The Conflict Sensitivity Principle: Can Best Practice in Conflict Research Fill the Ethics Gap in Terrorism and Counterterrorism Research Practice?,” *Terrorism and Political Violence* 33, no. 2 (March 24, 2021): 381–396, <https://doi-org.proxyau.wrlc.org/10.1080/09546553.2021.1880159>.

Directional Harms

The Obligation to Report

In this space, researchers will likely come across and may directly be analyzing, harmful content which violates platform-specific 'terms of service' (TOS) policies, along with content that may be outright illegal. Furthermore, researchers may also unearth specific credible threats to others, which carry an additional set of risks.

In general, researchers must use their best judgment when deciding whether to report problematic content encountered in the course of their work.¹⁶ Researchers often assume the content being reported will be removed from the platform, which could hinder their ability to access valuable information. However, content that may indicate a user or actor likely poses a direct threat to themselves or others should be treated with extreme caution and in many cases ought to be reported to the platform hosting the content, law enforcement, and other relevant authorities. Researchers and their institutions should be familiar with local legal requirements as these vary: in some countries, researchers are protected when downloading terrorist or potentially illegal content in the course of their work. However, in other contexts, even accessing such content for the purposes of empirical research may be illegal. Similarly, the requirements to report content to government actors may vary depending on location.

TOS and content policies vary substantially from platform to platform. For example, content that goes against Twitter's policies may be acceptable behavior on Gab, so a researcher must use their best and informed judgment when researching across platforms. Most platforms, and Global Internet Forum to Counter Terrorism (GIFCT) members especially, have some extent of public-facing community or content standards for what is and is not allowed on that platform. Additionally, most platforms have a way to report problematic or harmful content. Ultimately, most platforms employ their own trust and safety teams responsible for investigating and reporting online behaviors, meaning that researchers do not have the sole responsibility to report content, nor will every piece of reported content ultimately be removed or acted on based on platform policies.

In general, ethical practice in social science and social work dictates that the anonymity of research participants may be lifted in cases where the participant poses a credible threat to themselves or others. A similar approach can be taken in extremism and terrorism research. Violent extremists have established an active presence on social media prior to their attacks to post content and interact with material related to their ideology. If a researcher encounters a threat of imminent, credible action that would cause extreme harm (e.g., a planned terrorist attack or school shooting or an identified user credibly threatening suicide), reporting to authorities should be strongly considered.

.....
 16 J.M. Berger, "Researching Violent Extremism: The State of Play", Resolve Network, June 2019, https://resolvenet.org/system/files/2019-09/RSVE_RVESeries_ResearchingViolentExtremism-TheStateofPlay_JMBerger_June2019.pdf.

While a researcher is examining extremist content online, they noticed that some of the content posted contained racist rhetoric. However, it did not necessarily need to be reported to the social media platform per content standards at the time. While racist material is hurtful, reporting and focusing on every instance of racism can cause moderators or researchers to overlook more serious calls for offline violence.

Protection from Platforms

Social media platforms have a limited duty to protect researchers and other users from online harm that occurs on their websites. The policies and regulations enforced by companies use a variety of tactics to address violations on their platforms. For example, Twitter offers “account-level enforcement” and “direct message-level enforcement” depending on a user’s violation.¹⁷ The enforcement can include stopping a user’s messages or placing an account in “read-only” mode which prevents a user from tweeting. Some platforms may also (if asked) ensure researcher and user information is protected and cannot be accessed without the user’s permission. In general, firms have the same responsibility to protect researchers as they do with protecting their other users.

However, as research findings in this space can be highly sensitive, researchers may also experience blowback from platforms or publishers after releasing their work. Without any intention of incurring bias in research, as a mitigating step the investigating team or individual may wish to engage the platform in conversation about the research aims. Most platforms are interested in reducing harmful content hosted on their servers: asking for input on research questions may be mutually beneficial during the research design phase. Similarly, providing an opportunity for platforms to comment on findings may provide additional context prior to public release. Still, the profile of the platform should be considered when determining if such an engagement would be beneficial or fraught, especially when considering approaching platforms that are specifically created for use by extremists, have ties to authoritarian regimes, or have previously proved to be unwilling to engage on harmful content.

Law Enforcement

When conducting research into extremism and terrorism online, there is the very real and likely possibility that analysts encounter illegal activity. This could include networks of individuals sharing content produced by proscribed terrorist groups, evidence of individuals recruiting for terrorist movements, or even credible threats of a terrorist attack or hate crime. Additionally, researchers themselves might find themselves targeted by individuals as the result of publishing insights on extremist and terrorist content online through (for example) doxing and online harassment. Taking these possibilities into account, researchers should consider scoping out appropriate mechanisms for

.....
 17 “Our Range of Enforcement Options,” Twitter Help Center, n.d., <https://help.twitter.com/en/rules-and-policies/enforcement-options>.

referring to law enforcement (such as online portals) at the outset of a research project.¹⁸ However, the capacity, human rights standing, and types of response by authorities may, depending on the context, also be considered in the ethical call of whether and to whom a report should be filed. To facilitate this, in advance of commencing active fieldwork, research organizations should develop clear “bright lines” for reporting harms.

External/Audience Harms

Informed Consent

Informed consent is the groundwork of ethical human-subject-centered research. In short, if data collection is not carried out via 100 percent open, publicly accessible information and views, and researchers intend to speak with or observe participants in non-public spaces, informed consent must be obtained. For consent to be obtained, it must fulfill three criteria by being free (voluntary and able to be withdrawn at any time), specific (directly related to the research aims), and informed (whereby the participant is fully aware of potential consequences of their participation). Study participants should be given full disclosure of the potential risks of their involvement and care should be taken that they understand those risks. Their anonymity must be ensured or the circumstances under which they may be named fully elucidated. While a full discussion of informed consent is beyond the scope of this document (free and comprehensive guidance is readily available), researchers of online extremist and terrorist content must still adhere to these guiding principles in their work.¹⁹

However, gaining informed consent in online research is not clear-cut and perceptions of privacy online are complex. In some cases, online spaces are clearly public – such as a public Twitter feed – or clearly private, such as direct messages. However, in some cases, the privacy of some spaces is more ambiguous, such as in large open groups on Facebook, or in community forums where membership is required. Additionally, perceptions of what social media spaces are public and private may vary from the legal reality or the terms of service of a platform. These discrepancies between reality and perceptions of privacy are important, and accordingly, where information is gathered through a platform’s Application Programming Interface (API) using (for example) a social listening tool, it is important to consider whether a research subject could reasonably perceive this information to be private.

Archiving Content

The online ecosystem is one of the most important sources of information for researchers to study communication, tactics, narratives, and other data on extremism. Websites and social media provide some of the richest data for researchers to analyze, which is why it is important for it to be properly archived for future research. At the same time, safely archiving content protects research and

.....

¹⁸ “Tell Us About Possible Terrorist Activity,” UK Metropolitan Police, n.d., <https://www.met.police.uk/tua/tell-us-about/ath/possible-terrorist-activity/>.

¹⁹ “Unicef Procedure for Ethical Standards in Research, Evaluation, Data Collection and Analysis,” Unicef for Children (UNICEF), April 1, 2015, <https://www.unicef.org/media/54796/file#:~:text=In%20order%20to%20ensure%20the%20protection%20of%2C%20and,for%20ethical%20research%2C%20evaluation%20and%20data%20collection%20and>.

research integrity while minimizing researcher exposure to risk. Additionally, extremist content on social media and other platforms is often removed – arguably good for the public but challenging for future researchers. Internal databases of deleted content may be maintained by platforms but are rarely (if ever) accessible externally.

In an ideal operating environment, trusted researchers would be able to work with platforms and/or the government to maintain archives of harmful online content in a safe environment away from public exposure. Barring such a situation, the need to safely archive online content falls on individual researchers and their institutions. Open-Source Intelligence (OSINT) practices have vastly improved in recent years, which has led to many guides on carrying out online open-source research and safely archiving content. For example, the open-source investigative outfit Bellingcat provides a comprehensive toolkit (see Annex), while a quick search reveals multiple similar free and paid toolkits.

In general, archiving content is one of the best methods researchers have for analyzing and comparing information as the material is removed by platforms or by the user. This process should follow several guiding principles:

- Data in the form of images, screengrabs, videos, screen recordings, or documents should be downloaded from the digital source. OSINT recording tools can help in this process to record the entire investigative process automatically, while also allowing the analyst or researcher to streamline the archival process.
- If intended for wider use beyond the researcher or their immediate team, archived data should not contain any identifiable information such as the username or channel/page name, as this will prevent individuals outside of the research field from accessing the data's origin point. Despite extremists promoting hateful content, they still have a right to privacy, and removing identifiers prevents other individuals from gaining access to the data's location. If identifiers are not removed, it provides the opportunity for individuals from gaining access to the havens where extremism and hate thrive.
- Researchers should document for their own records the method used to locate the information for future research. Keeping detailed records of any archived content and any remarks a researcher may have upon first encountering the information online can help maintain the archive's organization. The value of information is not always obvious, so detailing the justification for archiving it is vital for recognizing its significance.
- Once the data has been collected and identifiable information is removed, the content should be archived. A secure location – safe online cloud services, encrypted Universal Serial Bus (USB) or external hard drives, or hard copies in a locked location may all be options. In general, if sensitive digital content should be encrypted prior to storing using tools like VeraCrypt, encrypted .7Z storage files with SHA256, or other alternative encryption techniques. The secure location should contain an organizational system to easily find content when needed for research purposes. Researchers should use a file-saving system that includes the date, location of data, and an identifiable characteristic of the content.

Considering exploitation

Extremism and terrorism research may have security or intelligence applications. Additionally, these subjects are inherently political. Political groups, states, and actors (both benign and malicious) may have an interest in promoting a position with regards to terrorism or extremism, having groups considered in a certain way, or exaggerating or minimizing a threat or social problem. Authoritarian regimes also use the threat of extremism or terrorism to change laws or erode civil liberties or use the language of 'terrorism' and 'extremism' to justify human rights violations. Any exploitation of analysis potentially opens a researcher or research institution (as well as the subjects of their research) to risk. Accordingly, when conducting analysis, it is important that researchers consider the ways their research may be misrepresented, taken out of context, or used to justify unethical actions or positions. While the risk of exploitation can not be eliminated, it should be considered at the inception and the publishing of research projects.

Trauma & Mental Health and Mitigating Vicarious / Secondary Trauma

Researchers are exposed to hateful content which can harm their mental health. Trauma occurs on a spectrum that varies by individual. Many researchers become desensitized to online violence because they see it so often. Such high levels of exposure to violence have been shown to often lead to emotional numbness and reduced response to real-world violence. It is often difficult to ask for help or express mental health issues, which is why institutions should be proactive in their promotion of mental health resources.

Handling emotionally laden and disturbing content on a consistent basis raises the likelihood of experiencing several negative effects. The risk of vicarious or secondary trauma and burnout among researchers in this field is particularly high. Vicarious trauma is defined as a "cognitive change through empathetic engagement with trauma survivors."²⁰ This has been widely shown to impact interviewers and those who assist trauma survivors.²¹ Secondary trauma stress is characterized as "a state of physical, emotional and mental exhaustion caused by long-term involvement in emotionally demanding situations."²²

Viewing racism, hate, and violence can take a toll on a researcher's well-being, so it is vital for institutions to have resources in place. As discussed above, research institutions have a duty of care to protect staff from trauma and provide a space for discussions, opinions, and processing of the

20 Jason M. Newell & Gordon A. MacNeil, "Professional Burnout, Vicarious Trauma, Secondary Traumatic Stress, and Compassion Fatigue: A Review of Theoretical Terms, Risk Factors, and Preventative Methods for Clinicians and Researchers," *Best Practices in Mental Health: An International Journal* 6, no. 2 (July 14, 2010): 57–68, <https://calio.org/wp-content/uploads/2020/04/Professional-burnout-vicarious-trauma-secondary-stress-and-compassion-fatigue-A-review-of-theoretical-terms-risk-factors-and-preventive-methods-for-clinicians-and-researchers.pdf>.

21 Philip A. Sandick, "Speechless and Trauma: Why the International Criminal Court Needs a Public Interviewing Guide," *Northwestern Journal of International Human Rights* 11, no. 1 (2012): 105–125, <https://scholarlycommons.law.northwestern.edu/njihr/vol11/iss1/4>.

22 "Staff Well-Being and Mental Health in UNHCR," The United Nations High Commissioner for Refugees, 2016, <https://www.unhcr.org/56e2d-fa09.pdf>; Andreas Seidler et al., "The Role of Psychosocial Working Conditions on Burnout and Its Core Component Emotional Exhaustion – a Systematic Review," *National Library of Medicine: National Center for Biotechnology Information* 9, no. 10 (March 14, 2014): 9–10, <https://doi.org/10.1186/1745-6673-9-10>.

content seen online. Psychological stressors and traumatic events online can translate to physical health problems.²³ Institutions often provide mental health days for their employees outside of taking time off to allow time to rest. All research institutions and organizations can provide a variety of resources that can directly impact the mental health support an individual receives:

- Allowing researchers more mental health days to process graphic content they may have encountered that week. Mental health days should not count as paid time off or a sick day;
- Implementing hard limits on the amount of time a researcher can spend online; and,
- Provide weekly opportunities for researchers to openly discuss trauma, graphic material, and stressors with licensed professionals and/or among their peers.

There is research that suggests that accidental or unexpected exposure to harmful images can lead to greater psychological harm than that which is expected and undertaken with a clear sense of purpose. Accordingly, institutions should consider policies that prevent accidental exposure to this kind of content by other researchers or third parties. This could include the use of screen obscurers if an institution has an open office and the storing of potentially traumatic material on protected drives that other staff cannot access.

In addition to institutions implementing guidelines, researchers should find their own methods for coping with their work by choosing to engage in healthy behaviors that are non-work related. In a 2020 study, participants found peer support to be the most helpful for processing trauma because it provides immediate responses and colleagues can advise one another.²⁴ Additionally, it may be important for colleagues to engage in healthy dialogue regarding their research because some workplaces require employees to sign non-disclosure agreements that prevent them from talking about their research outside of the workplace.²⁵ Particularly compartmentalized or classified work that prevents colleagues from speaking to each other inside the workplace should be proactively addressed by making time for specific units or teams to debrief together. Other mitigation opportunities include:

- Encouraging researchers to engage in healthy activities that are non-work related to prevent work from “following” them home;
- Seeking out a licensed therapist to provide more individualized support and providing referrals to other licensed professionals if necessary; and,
- Ensuring that the research institution will provide mental health days (and proactively offer them) if the researcher needs to rest.

.....
 23 “Mental Health Disorders and Stress Affect Working-Age Americans,” Center for Disease Control, July 2018, <https://www.cdc.gov/workplace-healthpromotion/tools-resources/pdfs/WHRC-Mental-Health-and-Stress-in-the-Workplace-Issue-Brief-H.pdf>.

24 Lara Bakes-Denman, Yolanda Mansfield, & Tom Meehan, “Supporting Mental Health Staff Following Exposure to Occupational Violence – Staff Perceptions of ‘Peer’ Support,” *International Journal of Mental Health Nursing* 30, no. 1 (February 2021): 158–166, <https://doi.org/10.1111/inm.12767>.

25 Cristina Criddle, “Facebook Moderator: ‘Everyday was a nightmare,’” BBC News, May 12, 2021, <https://www.bbc.com/news/technology-57088382>.

ANNEX: Researcher Safety Resources

Institutions

1. Ethical Research Practices

- a. [Does the Institution Have a Plan for That? Researcher Safety and the Ethics of Institutional Responsibility](#)
- b. [The Development of the Framework for Research Ethics in Terrorism Studies \(FRETS\)](#)
- c. [Responsible Conduct of Research: Not Just for Researchers](#)
- d. [Ethics](#)
- e. [Do Researchers have an obligation to report dangerous actors?](#)
- f. [General Data Protection Regulation \(GDPR\)](#)

2. Protecting Researchers

- a. [Best Practices for Protecting Researchers and Research](#)
- b. [Being a woman on the internet is a nightmare. How can we fix it?](#)

3. Mental Health

- a. [The Center for Workplace Mental Health](#)
- b. [Supporting Mental Health in the Post-Pandemic Workplaces](#)

4. Archiving Content

- a. [Bellingcat's Online Investigation Toolkit](#)
- b. [Conifer](#)

Individuals

1. Safety

a. General Online Safety

- i. [Researcher Welfare 1: Privacy and Security](#)
- ii. [Digital protection guides comparison](#)
- iii. [Digital protection guides - A survey of community resources](#)
- iv. [Crash Override Network: Resource Center](#)
- v. [Tactical Tech Data Detox](#)
- vi. [Surveillance Self-Defense: Tips, Tools and How-Tos for Safer Online Communications](#)
- vii. [Best VPN Service of 2022](#)
- viii. [NSA: Keeping Safe on Social Media](#)

b. Online Harassment Protection

- i. [Speak Up & Stay Safe\(r\): A Guide to Protecting Yourself From Online Harassment](#)
- ii. [PUBLIC: PII Manager Opt-Out Template](#)
- iii. [Doxing: What it is and how to protect yourself](#)
- iv. [Report Illegal Content on the Internet: Europol](#)

c. Other

- i. [UNESCO: Safety guide for journalists: a handbook for reporters in high-risk environments](#)

2. Mental Health

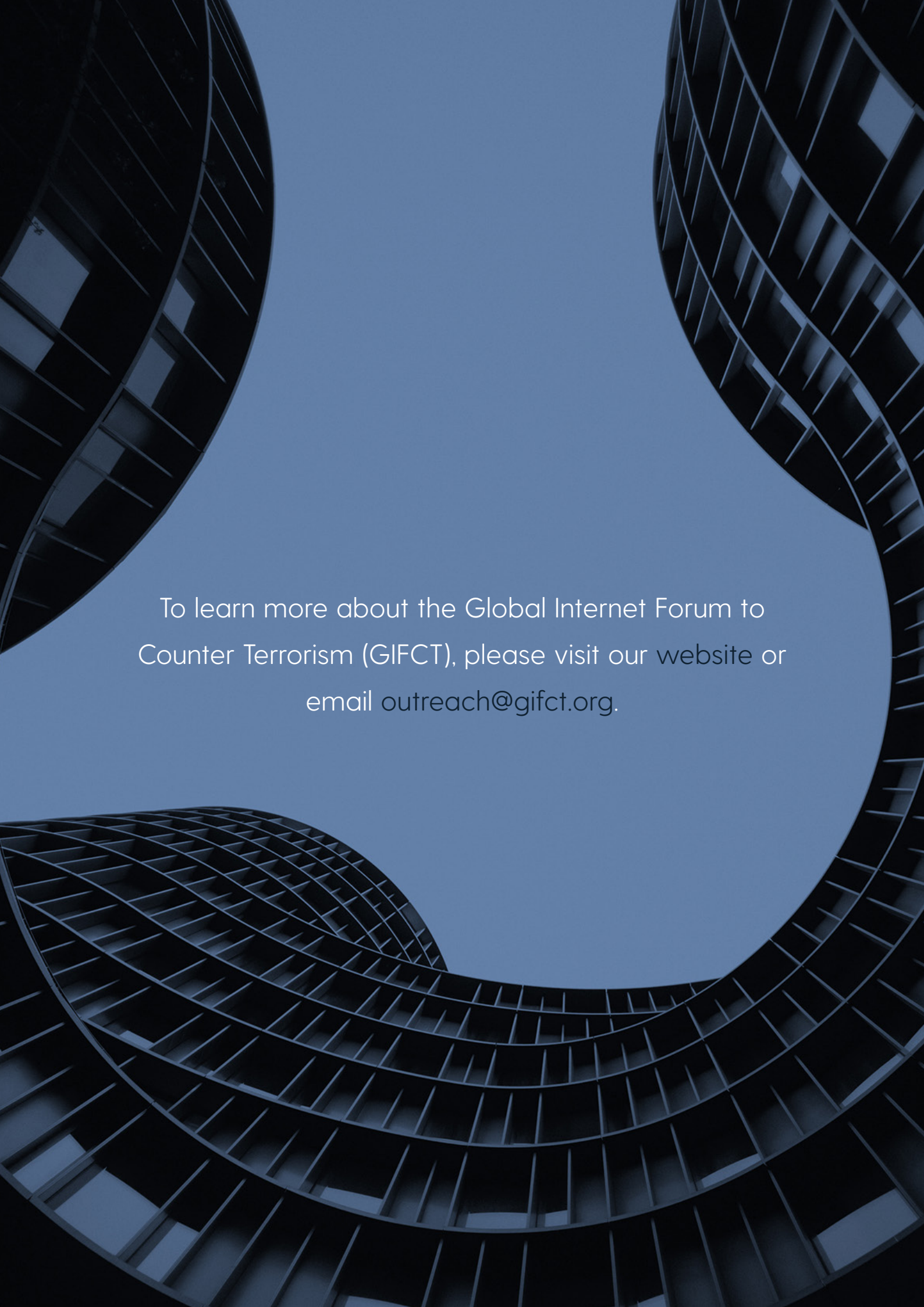
- a. [Researcher Welfare 2: Mental and Emotional Well-being and Self Care - VOX - Pol](#)

(voxpath.eu)

- b. [The Body Keeps the Score: Brain, Mind, and Body in the Healing](#)
- c. [Headington Institute](#)
 - i. [How Stressed You Are](#)
 - ii. [You Are Showing Signs of Burnout](#)
 - iii. [Understanding & Coping with Traumatic Stress](#)
 - iv. [Preventing Burnout](#)
 - v. [Understanding and Addressing Vicarious Trauma](#)
- 3. [The Center for Victims of Torture](#)
 - i. [Self-Care Guide](#)
 - ii. [Trauma Stewardship: An Everyday Guide to Caring for Self While Caring for Others](#)
 - iii. [What About You? A Workbook for Those Who Care for Others](#)
- 4. Social Media**
 - a. [Introducing the Researcher Platform: Empowering independent research analyzing large-scale data from Meta](#)
 - b. [Digital Void](#)
 - c. [Center for Countering Digital Hate](#)

Links to the Annex Resources can be found on the GIFCT Working Groups Webpage:

<https://gifct.org/working-groups/>



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