GIFCT YEAR 4 WORKING GROUP OUTPUT

# Hash-Sharing Database Review: Challenges and Opportunities

## GIFCT Year 4 Working Group

December 2024

START
NATIONAL CONSORTIUM FOR THE
STUDY OF TERRORISM AND RESPONSES TO TERRORISM

GIFCT
Global Internet Forum
to Counter Terrorism

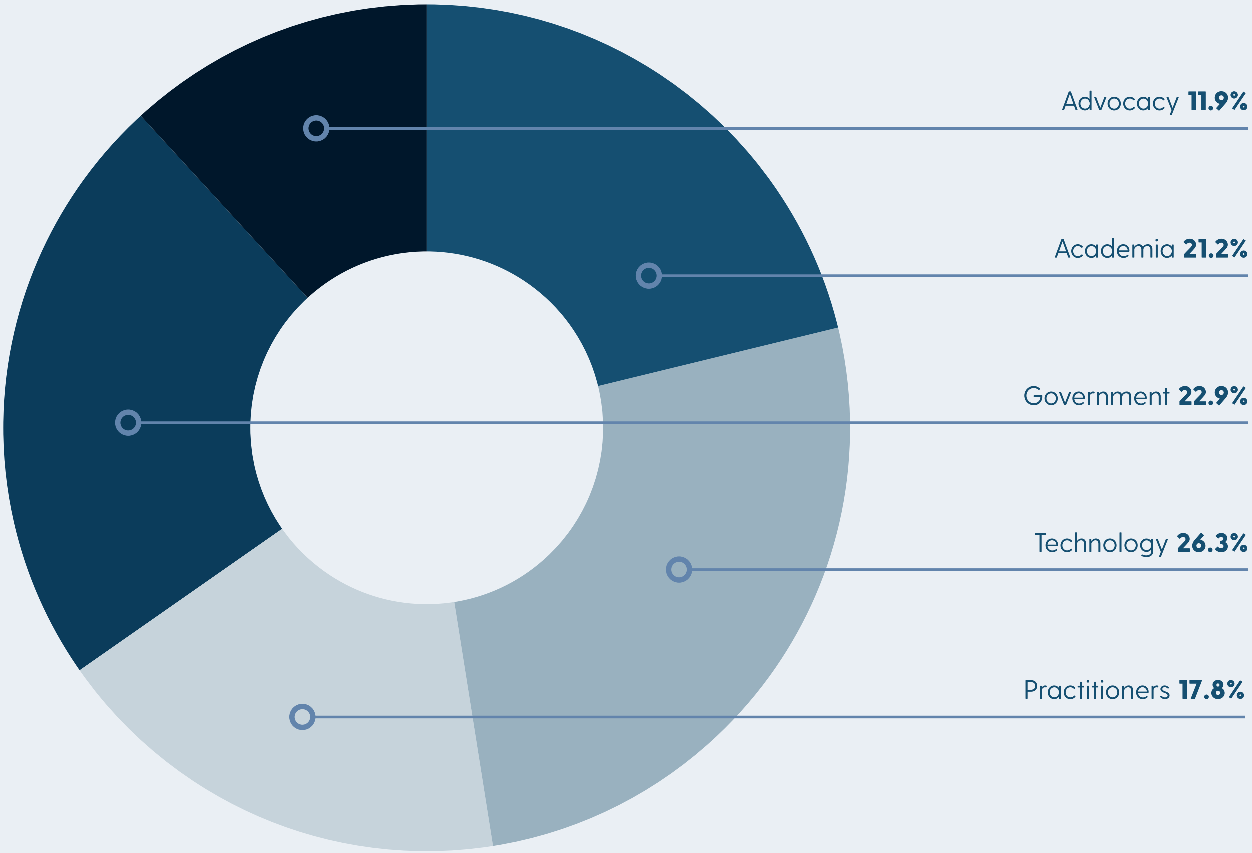
# Table of Contents

## Introducing GIFCT Year 4 Working Groups

In May 2024, GIFCT launched its Year 4 Working Groups to facilitate dialogue, foster understanding, and produce outputs to directly support our mission of preventing terrorists and violent extremists from exploiting digital platforms across a range of sectors, geographies, and disciplines. Started in 2020, GIFCT Working Groups contribute to growing our organizational capacity to deliver guidance and solutions to technology companies and practitioners working to counter terrorism and violent extremism, and offer multi-stakeholder perspectives on critical challenges and opportunities.

Overall, this year's three thematic Working Groups convened **145** participants from **32** countries across **6** continents with **51%** drawn from civil society (**12%** advocacy, **21%** academia, and **18%** practitioners), **23%** representing governments, and **26%** in tech.

### Sectoral Breakdown of Working Group Participants



The 2024 GIFCT Working Groups focused on the following three topics:

## Hash Sharing Working Group: Evolving Inclusion Parameters

GIFCT has managed and continually enhanced its Hash-Sharing Database (HSDB), which contains perceptual hashes of terrorist and violent extremist content, since 2017. The current inclusion parameters for the database have evolved through close consultations with global experts. As technologies, content, and types of violent extremist and terrorist groups change, GIFCT aims to continuously review its definitions and parameters to evolve with the times.

In order to enhance the transparency and accuracy of GIFCT's HSDB, this Working Group reviewed the existing inclusion criteria, identified potential gaps, and put forward suggestions to enhance its use. Meetings included consultations with current GIFCT member companies and feedback sessions with global experts. The Working Group resulted in a final report mapping out recommendations and expectations on the future trajectory of GIFCT's HSDB taxonomy.

## Incident Response Working Group: Future-proofing GIFCT's Incident Response Framework

GIFCT has continuously evolved its Incident Response Framework (IRF) since it launched in 2019 following the attacks in Christchurch, New Zealand. The IRF provides a centralized communications mechanism to share news of ongoing incidents that might result in the spread of violent content online, enabling widespread situational awareness and a more agile response among GIFCT member companies. Activations of the IRF allow GIFCT to heighten member awareness of ongoing incidents, circulate critical information regarding related online content, respond to member needs and requests regarding substantive or contextual information, and facilitate related uploads to the HSDB.

This Working Group reviewed and provided suggestions to future proof GIFCT's IRF. To do so, the Working Group evaluated the societal harms around terrorist and violent extremist attacks and mass violent events, examined case studies across different regions, and assessed different types of content, including AI-generated and synthetic materials, and their implications. Meetings included consultations with current GIFCT member companies and feedback sessions with global experts. The Working Group resulted in a set of recommendations regarding GIFCT's IRF. These inputs will inform GIFCT's ongoing efforts to assess lessons learned and good practices in strengthening the IRF and engagement with key stakeholders.

## Gaming Community of Practice: Supporting Gaming Tech Safety

GIFCT established its Gaming Community of Practice (GCoP) to foster collaboration, knowledge sharing, and innovation among practitioners in the gaming industry and to enhance the development of best practices to prevent terrorists and violent extremists (TVE) from exploiting games, gaming-adjacent services, and the gaming community.

This Working Group invited researchers, policy makers, and subject matter experts to support the GCoP by sharing their insights and feedback on the ways in which game-play spaces should evolve their safety work, review safety policies, tools, and practices, and anticipate evolving safety risks. Participants joined GCoP meetings in 2024 to contribute to specific themed discussions to help inform the Community of Practice's themes and goals such as positive intervention potentials across game-play services and sessions with international law enforcement bodies to understand threat signals. Outputs from this year's GCoP include Safety-By-Design one-pagers on best practices on specific gaming surfaces; a review of interventions approaches and research; and early concept work for expanding how terrorist and violent extremist signals can be shared across GIFCT platforms.

# Foreword

**Dr. Nagham El Karhili and Skip Gilmour**
**The Global Internet Forum to Counter Terrorism**

## The GIFCT Hash Sharing Data Base 2017 - Now

Since 2017, the Global Internet Forum to Counter Terrorism (GIFCT) has maintained a Hash Sharing Database (HSDB) of hashes of terrorist and violent extremist content (TVEC) available to GIFCT Member companies. The HSDB enables GIFCT member companies to collectively share hashes (or "digital fingerprints") that correspond to TVEC identified and removed from their respective platforms. Content found by member companies is hashed into a numerical representation of the original content using hashing algorithms that are impossible to reverse engineer, allowing for the circulation of signals about the content without sharing the original TVEC or associated user information. To govern the sharing of these hashes in accordance with privacy and human rights, GIFCT established a taxonomy and inclusion criteria stipulating categories of TVEC that can be hashed and included for submission to the database, agreed upon by member companies and in consultation with the broader multistakeholder community.

The HSDB taxonomy has shifted with the terrorist and violent extremist (TVE) threat landscape. Initially, GIFCT's HSDB was limited to material associated with designated terrorist organizations on the United Nations Security Council's Consolidated Sanctions List associated with  Security Council Resolution 1267. This decision was made to triangulate areas of consensus between founding member's diverse "TVE content" policies, and to reflect global consensus among stakeholders at the time. As the threat landscape continued to change, particularly in the context of lone attackers using online tactics to amplify their terrorist aims (as was seen in the 2019 attacks on Muslim worshippers in Christchurch, New Zealand), non-designated actors and groups became a prominent source of TVE threats. In response, GIFCT created its Incident Response Framework in 2019 and published its 2021 taxonomy report which resulted in an updated expanded taxonomy covering a new set of content, In addition to TVEC related to the UN Sanctions List, GIFCT's updated taxonomy now covers incident response framework activations, and GIFCT's behavioral inclusion parameters. These parameters enable GIFCT members to use the HSDB to share content related to TVE threats from non-UN-designated actors or organizations. Currently GIFCT has 33 members with 21 accessing the HSDB in various stages of onboarding and integration.

## 2024 Hash Sharing Perspective

As part of GIFCT's 2024 Hash Sharing Working Group, two researchers from the National Consortium for the Study of Terrorism and Responses to Terrorism (START) led the Hash Sharing Working Group (HSWG), which included multiple sessions with member company representatives and experts from government and civil society organizations. As a result, the START team led on the authorship of a report reviewing the current status of the HSDB, its use by current GIFCT members, and overall the efficacy of GIFCT's HSDB and taxonomy. In response to this report's findings, GIFCT plans to take the following actions:

### *Reshape Inclusion Criteria*

The START team found that some GIFCT members reported uncertainty about the operability of some elements of the HSDB taxonomy. Some of this uncertainty surrounds how best to interpret key nomenclature like "hate-based ideology" or what qualifies as advocating for violence. To remedy this, GIFCT plans to reshape the inclusion criteria and labeling guidelines to clarify the existing categories and nomenclature.

### *Follow-Up Engagements*

Per the START team's finding discussed further in this report, GIFCT's taxonomy is not the primary barrier to a company's non-adoption of the HSDB. Companies that have not adopted the HSDB or do not regularly contribute have frequently cited logistical, onboarding, or resource constraints associated with their at-scale moderation data pipelines. In the coming months, GIFCT will schedule follow-up engagements with member companies to assess barriers to adopting the expanded taxonomy and gather information on their hashing workflows and labeling systems.

### *Product Strategy for Non-Hashed Resources*

The START team noted that members are also interested in non-hashed resources to complement the HSDB, in order to supplement what hashes might surface and give further context for moderation efforts. GIFCT recognizes the value of this recommendation and will be prioritizing a product strategy in 2025 to develop effective programs for deploying non-hashed resources.

### *New Opportunities for Member Collaboration*

One of the START team's key recommendations is that GIFCT should find ways to incentivize more active participation in GIFCT information sharing. To accomplish this, GIFCT will seek to engage member companies on multiple levels. Firstly, GIFCT will create an Investigator's Community of Practice (ICOP), and engage members via focus groups to assess additional use cases and user types for GIFCT information-sharing initiatives. Additionally, GIFCT will encourage members to join industry-wide hash-sharing engagements to facilitate better onboarding and promote participation within the broader GIFCT membership.

# Hash-Sharing Database Review: Challenges and Opportunities

Dr. Sean Doody and Dr. Michael Jensen

**The National Consortium for the Study of Terrorism and Responses to Terrorism**

## Introduction and Executive Summary

The Global Internet Forum to Counter Terrorism's (GIFCT) Hash-Sharing Database (HSDB) is a significant counter terrorism achievement. Constituted by millions of discrete hashes, the HSDB enables members to address the shared problem of terrorist and violent extremist content (TVEC) on their platforms. The HSDB has been especially useful in the domains of video and image hashing, and in 2021, evolved to include other content types that expand the possibilities for hashing and information sharing among member companies. While originally designed for sharing hashes pertaining to TVEC produced by designated entities on the United Nations Consolidated Sanctions List, the HSDB expanded in 2019 with content related to the Incident Response Frameworks, and in 2021 to include TVEC produced by non-designated actors.

Whenever GIFCT expands or modifies the HSDB, the goal is to make the tool more representative of the existing threat landscape, thus maximizing its utility to members. To this end, the Year 4 Hash-Sharing Working Group (HSWG) was convened to evaluate the current state of the HSDB in light of the 2021 taxonomy expansion. The goals of the Year 4 HSWG were to:

1. Review the types of TVEC in the HSDB that fall under the 2021 taxonomy expansion;
2. Identify any challenges pertaining to hashing newly-allowed TVEC since the 2021 expansion; and
3. Make recommendations for GIFCT to consider as it seeks to further evolve and refine the HSDB.

As part of GIFCT's 2024 Hash Sharing Working Group, two researchers from The National Consortium for the Study of Terrorism and Responses to Terrorism (START) held multiple focus groups with GIFCT members, evaluated internal GIFCT documents and resources, and received feedback from the multi-stakeholder counter terrorism community (e.g., academia, government, and civil society).

### The START team's core findings are:

- **The 2021 HSDB taxonomy expansion provides a flexible framework for hashing a large variety of TVEC.**
    - GIFCT's behavioral inclusion parameters provide a reasonable definitional framework for hashing content (e.g., images, videos, PDF/text, URLs) associated with non-designated

violent extremist actors, moving beyond the previous requirement for UN Sanctions List association.

  - GIFCT's labeling guide has developed a suite of labels for describing hashes that cover the most relevant aspects of TVEC (e.g., ideology, organization or movement, reason for inclusion, etc.).

- **Despite the expanded taxonomy, members primarily share TVEC from designated entities.**
  - Members noted that they do not have the internal pipelines and processes needed to hash and send non-designated yet internally moderated content to the database. As a result, they cannot share TVEC that meets the behavioral parameters of the HSDB at scale.
  - Members also noted that they lack pathways to work with each other or GIFCT to identify ways to overcome the technical and workflow challenges preventing them from sharing more varied TVEC to the HSDB.
- **Members desire a standardized and more systematic use of labels when sharing hashes to the HSDB.**
  - While GIFCT provides members with a comprehensive labeling guide, companies are not required to apply content labels beyond basic rationale for inclusion when hashing TVEC.
  - Members support establishing mandatory core labels for all hashes, along with a modular labeling system for additional content features.
  - While potentially beneficial, a modular design would require engineering resources from member companies to develop and implement.
- **A small number of the largest GIFCT member companies are responsible for most of the activity in the HSDB.**
  - Platforms with the largest user bases are most frequently targeted by violent extremist and terrorist actors, and so contribute the most to the HSDB.
  - However, as a collective endeavor, the HSDB requires broader member participation to better reflect the full TVEC threat landscape.
- **Members desire additional non-hashed resources to complement the HSDB.**
  - This would include a queryable database of unhashed logos and other symbols used by extremist groups and movements.
  - Members supported a plaintext URL database of links to known TVEC that would eliminate the need for hashing URLs.

Based on these findings, the START team offers four recommendations for GIFCT to consider:

1. **Work with members to understand their updated processes for classifying TVEC and resource constraints, focusing on developing scaled pipelines for hashing.**
   - This is a long-term goal that will require GIFCT to work directly with members to understand what is preventing them from developing scaled pipelines for hashing TVEC that meets the HSDB's behavioral inclusion parameters.
2. **Work with members to understand and address barriers to HSDB participation, as the database's value depends on broad engagement.**
   - The HSDB is a cooperative enterprise that depends on the collective sharing and using of hashes in the database. It is important for GIFCT to better understand what is preventing members from participating more actively in the HSDB and to find ways to encourage more engagement.
3. **Work with members on standardizing and improving the use of labels within the HSDB.**
   - While GIFCT's current HSDB taxonomy adequately captures TVEC varieties and should be maintained, GIFCT should collaborate with members to establish mandatory core labels for ideological and/or movement affiliations in addition to the rationale for inclusion in the HSDB.
   - GIFCT should work with members on developing mappings between their internal labeling systems and the labels the HSDB uses to represent the taxonomy.
   - To address conceptual challenges in operationalizing behavioral inclusion parameters, GIFCT should develop interactive case studies featuring both "obvious" and "edge" cases, with detailed annotations showing how classification decisions link to specific content features.
   - Additionally, the HSDB labeling system needs stronger quality control mechanisms, including more robust validation processes and auditing controls, with GIFCT receiving formal authorization to investigate and rectify labeling errors.
4. **Work with members to develop non-hashed resources, including a symbology database of logos, imagery, co-opted symbols, and phraseology with contextual explanations, as well as a plaintext URL database that would better serve member needs than hashed URLs.**
   - The START team's recommendations see GIFCT continuing to play a proactive role by working closely with members to ensure the continued success of the HSDB. This will require commitments of time and resources, and the START team encourages members and GIFCT to determine the investments that will be required to address these challenges. The following sections review key findings and expand upon the recommendations.

## FINDINGS

The 2021 HSDB taxonomy expansion provides a flexible framework for hashing a large variety of TVEC.

Despite the expanded taxonomy, members primarily share TVEC from designated entities.

Members desire a standardized and more systematic use of labels when sharing hashes to the HSDB.

A small number of the largest GIFCT member companies are responsible for most of the activity in the HSDB.

Members desire additional non-hashed resources to complement the HSDB.

## RECOMMENDATIONS

Work with members to understand their updated processes for classifying TVEC and resource constraints, focusing on developing scaled pipelines for hashing.

Work with members to understand and address barriers to HSDB participation, as the database's value depends on broad engagement.

Work with members on standardizing and improving the use of labels within the HSDB.

Work with members to develop non-hashed resources.

## Section 1: A Review of the Hash-Sharing Database Taxonomy

GIFCT's HSDB is a unique tool for combating terrorist and extremist content online. The original initiative to establish the HSDB began in 2017, led by the founding members of the Forum.[1] The HSDB was designed to facilitate the sharing of hashed data to allow for a collective response to TVEC on members' platforms. The HSDB is not meant to be the final authoritative source on what constitutes TVEC, and tech companies are always free to remove content according to their own moderation policies. However, the HSDB allows members to work together to address a shared challenge and set a collective cross-sector standard. This can be especially important in times of crisis.

At the time of its founding, the inclusion criteria for the HSDB were at their most narrow, only allowing hashed content produced by designated entities on the United Nations Security Council's Consolidated Sanctions List. Additionally, the only types of content originally hashed by the HSDB were images and videos. These narrow inclusion criteria built consensus around what counts as terrorist content among tech companies—a difficult task given widespread disagreements about the definition of TVEC.[2] They also operated efficiently for the context in which they emerged, which was characterized by a large volume of TVEC being produced and shared by ISIS and its sympathizers.

The limitations of the original HSDB taxonomy were made apparent during the March 2019 mosque attacks in Christchurch, New Zealand, when a white supremacist live-streamed the shootings on social media. The video of his attacks circulated widely when copies of the live-stream were re-uploaded to members' platforms. Prior to his attacks, the perpetrator also released a manifesto laden with white supremacist vitriol. As an unaffiliated lone actor, the assailant was not a member of a terrorist organization on the UN Sanctions List, and thus the content related to his attacks fell outside the parameters of the HSDB inclusion criteria. In addition, because his manifesto was widely distributed as a PDF document, it did not meet the original hashing criteria for included content types.

### Strategic Expansion of the Taxonomy

In light of the Christchurch attacks, GIFCT took immediate steps to begin reconciling the HSDB's content gaps, leading to the formation of the Content Incident Protocol (CIP) as part of GIFCT's Incident Response Framework (IRF).[3] The goal of the CIP is to enable GIFCT member companies to quickly

[1] GIFCT Staff, "Background: A Guide to the Taxonomy of GIFCT's Hash-Sharing Database," in Broadening the GIFCT Hash-Sharing Database Taxonomy: An Assessment and Recommended Next Steps (Global Internet Forum to Counter Terrorism, 2021), 9–24.

[2] Nicholas J. Rasmussen and Johnathan Lowin, "Introduction," in Broadening the GIFCT Hash-Sharing Database Taxonomy: An Assessment and Recommended Next Steps (Global Internet Forum to Counter Terrorism, 2021), 4–8.

[3] GIFCT subsequently expanded the Content Incident Protocol into a robust Incident Response Framework (IRF) with three levels: Incident, Content Incident (CI), and Content Incident Protocol (CIP). Content created by a perpetrator or accomplice of an attack that leads to the activation of the CIP or CI levels of GIFCT's IRF are eligible for inclusion in HSDB.

identify and address TVEC that is circulating on their platforms as a result of an offline terrorist attack or violent extremist event. Given the extraordinary nature of the Christchurch attacks, GIFCT took the unprecedented step of creating a "crisis bank" of hashes pertaining to the video and manifesto to mitigate their cross-platform spread.[4] To date, the CIP has been activated a total of nine times, including the Christchurch attacks.[5]

Beyond the CIP, in 2021, GIFCT undertook a multi-stakeholder effort to expand its HSDB taxonomy with the goals of keeping pace with the changing nature of terrorist and violent extremist activities, as well as addressing underrepresented violent ideologies in the HSDB. Due to its reliance on the United Nations Security Council's Consolidated Sanctions List, the original HSDB lacked content produced by non-designated terrorist and violent extremist actors, including those who mobilize through loose networks and leaderless movements.[6] Additionally, the reliance on the UN Sanctions List meant that large volumes of the content in the HSDB pertained to designated violent Islamist extremist groups like ISIS and Al-Qaeda, while other TVEC related to white supremacy, Hindutva, and militant accelerationism were left out of the database.

After commissioning extensive feedback from experts and GIFCT members, GIFCT settled on a taxonomic expansion of the HSDB to include behavioral parameters, as well as the hashing of attacker manifestos and URLs from Tech Against Terrorism's (TAT) Terrorist Content Analytics Platform (TCAP).[7] In the current taxonomy, any hashes included in the HSDB based on behavioral parameters must meet all of the following criteria:

1. Be a non-governmental entity;
2. Have a violent extremist identifier (e.g., a distinguishable visual or textual symbol that clearly signals organizational or ideological affiliations);
3. Have a core hate-based ideology; and
4. Advocate for, or call to, violence.[8]

Additionally, hashed content based on behavioral parameters must be either an attacker manifesto, a

[4] GIFCT Staff, "Background."

[5] That is a total of 9 CIP activations and 2 CI activations as of the time of this report in late 2024. For an updated list, please see GIFCT, "Content Incident Protocol." CIP activations include: Christchurch, New Zealand in 2019; Halle, Germany in 2019; Glendale, AZ in 2020; Buffalo, NY in 2022; Memphis, TN in 2022; Louisville, KY in 2023; Perry, IA in 2024; Levittown, PA in 2024; and Eskişehir, Türkiye in 2024. CI activations include Udaipur, India in 2022; and Volgograd, Russia in 2024.

[6] Daniel Byman and Chris Meserole, "Expanding the Hash-Sharing Database," in Broadening the GIFCT Hash-Sharing Database Taxonomy: An Assessment and Recommended Next Steps (Global Internet Forum to Counter Terrorism, 2021), 25–41.

[7] GIFCT Staff, "Conclusion & Initial Next Steps," in Broadening the GIFCT Hash-Sharing Database Taxonomy: An Assessment and Recommended Next Steps (Global Internet Forum to Counter Terrorism, 2021), 150–53.

[8] GIFCT, "HSDB Taxonomy v1.0" (Global Internet Forum to Counter Terrorism, 2022).

branded terrorist publication, or a TCAP URL.

With the targeted expansion of behavioral parameters, there are now three ways content can be added to the HSDB:

1. It is associated with a named entity on the United Nations Security Council's Consolidated Sanctions List;
2. It meets the behavioral inclusion criteria; or
3. It leads to activating the Content Incident Protocol (CIP) or Content Incident (CI) levels of GIFCT's Incident Response Framework (IRF).

## Ongoing Challenges

Despite efforts to expand the HSDB taxonomy and provide guidance on inclusion criteria to member companies, GIFCT members and partners are experiencing challenges implementing the full taxonomy expansion into their workflows. This has resulted in offending content eligible for hashing sometimes not making its way into the HSDB.

An internal review by GIFCT identified three potential obstacles. First, while the 2021 taxonomy expansion provided much-needed improvements to the labels that users can apply to hashes, the expansion has not entirely improved the representativeness of data appearing in the HSDB. In particular, the HSDB remains dominated by content produced by designated organizations, particularly those tied to Islamist extremism, while underrepresenting content associated with violent extremist and terrorist actors not on the UN Sanctions List. Importantly, this does not signal a failure of the HSDB, as Islamist extremist groups like Al-Qaeda and ISIS historically have, and continue to produce large volumes of TVEC, and companies often have legal obligations to remove this material from their platforms. Nevertheless, it is likely that a sizable amount of in-scope TVEC is not currently being hashed and shared by members.

Second, GIFCT suggested that the expanded taxonomy's inclusion criteria might lack sufficient clarity, making it difficult for users to ground their decision-making in content features rather than contextual judgment calls that often require subject matter expertise. GIFCT found that it was especially difficult to determine if content advocates for, or is making a call to, violence. To avoid detection, those disseminating TVEC often include cultural or coded language, dog whistles, or co-opted signs and symbols that make implicit threats of violence. This content is carefully crafted to stay within the boundaries of what is deemed to be legally permissible speech in most countries, and it can be hard to identify when it meets the HSDB's inclusion criteria without detailed guidance. GIFCT also found that to operationalize the expanded taxonomy's inclusion criteria successfully, more clarity was needed on what constitutes a hate-based ideology or hate-based group.

Finally, GIFCT hypothesized that the lack of a systematic process for reviewing and updating the HSDB taxonomy could make it difficult to keep pace with the changing threat landscape. As the landscape continues to evolve from hierarchical groups with clear messaging to amorphous communities with mixed (and at times inconsistent) ideologies and ambitions, it is becoming increasingly rare for branding and clear strategic goals to be explicit features of TVEC. Thus, to identify content for the HSDB, moderators must have substantial expertise in the terrorist and violent extremist threat landscape, especially in the ideological fluidity of contemporary TVEC. In addition to reducing the amount of relevant content in the HSDB, this limitation could hinder the use and accurate application of group, ideology, and behavioral labels in the database. GIFCT found that content in the HSDB frequently lacks labels or is sometimes labeled inconsistently or incorrectly.

## Goals for the Hash Sharing Working Group

The 2024 HSWG was created to evaluate these challenges, assess if they are the primary obstacles preventing TVEC in the HSDB from being more representative of the threat landscape, and provide GIFCT members and partners with more guidance on identifying and labeling content for the database. The Working Group had three primary goals:

### Goal 1: Understand How Member Companies Decide Which Content to Hash in the HSDB

First, this Working Group aimed to understand how members are operationalizing the inclusion criteria of the HSDB and deciding which content to send to the database. This effort was designed to help identify where members feel there is needed clarity, as well as investigate the feasibility of developing an updated set of guiding principles or best practices that all members can reference.

### Goal 2: Understand the Challenges to Making the Hashes in the HSDB More Representative of the Threat Landscape

Secondly, the Working Group aimed to better understand the universe of in-scope content that is not being hashed. This applies to the type of content included in the taxonomic expansion of the HSDB (e.g., ensuring that attacker manifestos are being added to the database) to expand the ideological representativeness of the HSDB, thus increasing the real-world relevance of the database.

### Goal 3: Understand How Members are Keeping Pace With the Evolving Dynamics of Terrorism and Violent Extremism

Finally, the Working Group aimed to understand how members are keeping pace with the changing nature of terrorist and violent extremist threats and activities online in order to improve inter-coder reliability among members' content moderation teams. A review by GIFCT found that there are gaps in label usage, as well as inadequate guidance for how to accurately apply them. Data quality has suffered as a result, and labeling errors have accumulated. While the flexibility of the labeling system is important for keeping up with TVEC developments, the lack of clarity may undermine its practical utility.

### Additional Considerations

While the goals above describe the core focus of the HSWG, the START team considered additional factors, including:

- **Changes to how GIFCT handles the labeling process for hashed content shared by members.** Currently, GIFCT is only allowed to add additional alternative opinions to records and lacks the ability to modify the labels added to hashed content by members.[9] Given the subject matter expertise needed to accurately label hashed content in the HSDB, is it reasonable to allow GIFCT subject matter experts to modify records to improve label consistency and quality?
- **Changes regarding other types of content for inclusion in the HSDB.** Currently, logos of non-designated entities do not qualify on their own for inclusion in the HSDB because they do not explicitly advocate for violence. Should this exclusion rule be reconsidered, especially when the logo represents a group or movement with a hate-based ideology?
- **Changes to improve how content is coded in the database.** URLs are currently listed in the HSDB as a separate data category that only includes data from TCAP. In order to improve the comprehensiveness of the hashing of URLs that host TVEC, should URLs be treated as a content type rather than a category, permitting the hashing of any URL that meets the taxonomy's inclusion criteria?

## Section 2: A Review of the Hash-Sharing Database

Identifying, categorizing, and hashing TVEC is not a straightforward process. Disagreements about definitions of terrorism, differences among platforms' terms of service and content moderation policies, and legal obligations all combine to pose significant challenges to member companies' efforts to collectively combat terrorism online. Added to these obstacles is the challenge posed by the scalability of taxonomic solutions. Analytically precise taxonomies that offer extensive and nuanced parameters can be developed, but if they cannot be implemented in practice, their concrete value to members is diminished. Importantly, efforts to effectively, fairly, and transparently combat TVEC must balance counter terrorism strategies with the human and free speech rights of a diverse, global, and online public.[10]

[9] GIFCT's Hash-Sharing Database is accessed through ThreatExchange.

[10] Brittan Heller, "Combating Terrorist-Related Content Through AI and Information Sharing," Transatlantic Working Group Paper, Annenberg Public Policy Center, April 26, 2019, https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/05/Combating_Terrorist_Content_TWG_Heller_April_2019.pdf.

Despite these baseline challenges, the HSDB and its recent taxonomic expansion remain a counter terrorism success. With millions of discrete hashes, the HSDB is a unique tool that supports members' ability to moderate TVEC on their platforms at scale. While any change to the HSDB is always aimed at making the tool more comprehensive, representative, and useful for members, broadening the criteria for inclusion should not be expected to produce 1:1 ratios of various subtypes of TVEC. Similarly, while the behavioral inclusion parameters have created the opportunity to capture more comprehensive ideological content, due to historical tendencies in TVEC, as well as the related challenges of collecting more niche extremist content that is associated with emerging extremist actors (e.g., neo-fascist accelerationists), this does not mean that the hashes in the HSDB will be distributed proportionally among ideologies and groups.

With these considerations in mind, the following section provides an overview of the hashes in the HSDB and identifies areas for future growth. Feedback and insights from members that emerged during two focus groups held by researchers from START in collaboration with GIFCT is incorporated. This section is organized around the key research questions that guided these meetings:

- What data is currently in the HSDB? What data that meets the taxonomy's inclusion criteria is potentially underrepresented in the HSDB?
- What expansions to, or clarifications of, the HSDB inclusion criteria might be needed to incorporate more varied data and to help member companies understand what TVEC looks like? Do the current inclusion criteria accomplish the goal of providing member companies with an effective tool for deciding what content should be provided to GIFCT for inclusion in the HSDB?
- How do members and GIFCT assess and evaluate on an ongoing basis whether the HSDB and its inclusion criteria are keeping pace with the evolving landscape of TVEC online while still considering speech rights?

In addition to these questions, member companies expressed a desire for greater clarity on how two emerging content types, generative artificial intelligence (AI) and non-perpetrator produced content, might impact their hashing decisions. To assess the implications of these content types for the HSDB, in consultation with the HSWG, the START team examined two brief case studies which appear in Section 3.

## Current Data Representation

As of July 2024, the HSDB consists of approximately 2.2 million unique hashes, much of which come from visual media. As Table 1 shows, about 71 percent of hashes are from videos, around 28 percent are from photos or images, and around 1 percent each are from PDF/text hashes and URL hashes (the comparatively lower numbers of PDF/text and URL hashes are to be expected, as those content types have only been eligible to be included in the HSDB since 2021). The HSDB also contains labels for behavioral and severity indicators. Prior to being assigned a severity label, hashes must first be bucketed in a core category (e.g., UN Sanctions List, CIP, manifesto, etc.). After being bucketed, hashes can then receive behavioral indicators. Of hashes that have been assigned behavioral labels, 76 percent are labeled for Glorification of Terrorist Acts, and 17 percent are labeled for Graphic Violence Against Defenseless People. Recruitment and Instruction is labeled as occurring in 5 percent of behaviorally labeled hashes, and Imminent Credible Threat in 2 percent.

**Table 1. Distribution of media types in the HSDB**

| Media Type | Percent |
|---|---|
| Video | ~71% |
| Photo | ~28% |
| PDF/Text | <1% |
| URL | <1% |

While the 2021 taxonomy expansion aimed to increase the comprehensiveness of the TVEC represented in the HSDB, it was not designed to require member companies to perform type-for-type quantitative assessments of the content they are hashing. For instance, while attacker manifestos became eligible for inclusion in the database with the 2021 expansion, that change was not made with the expectation that companies will ensure that manifestos have the same representation in the database as other content types, like videos displaying graphic images. Put simply, GIFCT does not expect the number of manifestos to exceed the number of graphic and violent images or videos. Rather, the taxonomy expansion was designed to give member companies more flexibility when it comes to content inclusion in the hopes that the HSDB would represent more of the violative content that appears on their platforms.

As it currently stands, there are limitations around assessing how well the HSDB is doing in capturing all the TVEC that is eligible to be included in the database. For instance, quantitatively assessing the ideological diversity of the HSDB is not possible for two reasons. First, ideological labels were developed as a part of the 2021 taxonomy expansion, and thus earlier hashes are unlabeled. Second, ideology is not a topline label in the HSDB's taxonomy or a requirement for inclusion in the database, and it may not be applied to new hashes that are submitted to the HSDB.[11] While the number of hashes labeled with ideological information has increased annually,[12] ideological labels are missing for most hashes.

[11] Ideology labels are not required in the HSDB because it is often not feasible for member companies to determine the ideology of TVEC when it is being hashed. For example, during a CIP, members may lack the necessary contextual information to apply granular ideological detail to hashes.

[12] GIFCT Staff, 2023 GIFCT Annual and Transparency Report (Global Internet Forum to Counter Terrorism, 2024), 21.

While START's researchers cannot provide statistics on the representation of violent ideologies in the HSDB, given the original mission of the database, it is clear that the HSDB includes a sizable number of hashes associated with content produced by Islamist extremist groups and their supporters. Moreover, member companies stated that most of the content they are hashing continues to be associated with groups on the UN Sanctions List and that there is additional TVEC associated with non-designated actors that meets the HSDB inclusion criteria that they are not currently hashing. These actors generally fall outside the ambit of Islamist extremism, suggesting that additional content related to racial supremacy, anti-semitism, and violent misogyny could be hashed in the database.

While improving ideological comprehensiveness was one of the motivating factors that prompted GIFCT to embark on the taxonomy expansion in 2021,[13] members indicated that adding more varied TVEC to the HSDB remains a difficult task. There are a few reasons for this:

1. Member companies often have legal obligations to remove content with respect to designated entities that they do not have with non-designated groups, which in certain cases can shift their focus to TVEC produced by particular groups.
2. Making TVEC in the HSDB more ideologically representative of the threat landscape requires member companies to modify their internal technical processes and workflows, but due to resource constraints and other priorities, these changes have not been made. This problem has been further exacerbated by limited engagement between member companies and GIFCT on how best to tackle the technical challenges associated with developing scaled pipelines to hash content that meets the expanded HSDB inclusion criteria.
3. Content produced by non-designated entities can only be included in the HSDB if it meets the expanded taxonomy's behavioral criteria, which are rooted in core concepts like "violence," "hate," and "extremism" that are loosely defined in the research literature and not always straightforward to operationalize in practice. Member companies suggested that clarity on these concepts would help them adjudicate gray and edge cases for inclusion in the HSDB.

Finally, member companies noted the need for changes to the labeling system that would allow them to map their internal classifications to the HSDB taxonomy and hash more content at scale.

In addition to ideology, labels in the HSDB can also pertain to hashes arising from CIPs or CIs. As Table 2 shows, there have been nine such CIPs and two such CIs that have resulted in hashes. The two largest sets of incident labels pertain to a filmed attack against a civilian in Udaipur, India in June 2022 (40.6%) and the March 2019 Christchurch, New Zealand mosque attacks (30.9%). Hashes pertaining to the May 2022 Buffalo, New York Top's Grocery attack account for 15.7 percent of incident labels, the October 2019 Halle, Germany synagogue attack 10.9 percent, and the 2022 live-streamed Memphis, TN shootings 1.2 percent. All other CIP labels account for less than 1 percent of incident hashes.

[13] Byman and Meserole, "Expanding the Hash-Sharing Database."

| Table 2. Distribution of labels across incident labeled hashes (CIP and CI)+ | |
|---|---|
| **Incident Label** | **Percent** |
| Udaipur, India (CI) | 40.6% |
| Christchurch, New Zealand (CIP) | 30.9% |
| Buffalo, New York, USA (CIP) | 15.7% |
| Halle, Germany (CIP) | 10.9% |
| Memphis, Tennessee, USA (CIP) | 1.2% |
| Levittown, Pennsylvania, USA (CIP) | 0.2% |
| Glendale, Arizona, USA (CIP) | 0.2% |
| Perry, Iowa, USA (CIP) | 0.01% |
| Eskişehir, Türkiye (CIP) | 0.01% |
| Volgograd, Russia (CI) | 0.01% |
| Louisville, Kentucky, USA (CIP) | <0.01% |
| +**Note**: only includes hashes that have been labeled for an incident. | |

Overall, the HSDB has substantial coverage of video and image hashes. Based on member companies' feedback, hashes associated with Islamist extremism are well represented in the HSDB. The ideological representativeness of the database has improved since the 2021 taxonomy expansion, but according to the members the START team spoke with, this continues to be an area for future growth. The incident labels provide good coverage of some of the most notorious and widespread terrorist and mass violence events against civilians that have been broadcasted online.

## HSDB Inclusion Criteria

Focus groups held with member companies revealed important insights about the HSDB's inclusion criteria and its implications for members' utilization of the database. Interestingly, the issue of defining TVEC arose, with member companies highlighting the continued need for a definition of TVEC that all members agree on. However, it remains true that a singular agreed-upon definition of TVEC for member companies may not be legally or practically possible due to a lack of consensus on definitions of terrorism. It is with this reality in mind that GIFCT does not provide a singular definition of TVEC, but instead offers a definitional framework of inclusion criteria based on general inclusion parameters, the UN Consolidated Sanctions List, the Incident Response Framework, behavioral inclusion parameters, and the supported media types reviewed in Section 1.[14]

[14] GIFCT, "HSDB Taxonomy v1.0."

GIFCT is aware of definitional challenges, and has therefore expended resources developing tools like the Definitional Frameworks Principles Project to help members clarify their own definitions of terrorism and TVEC.[15] Importantly, member companies noted that the HSDB's expanded criteria are sufficiently flexible to capture the most important types of TVEC on their platforms. Practical guides to help member companies know when the criteria have been met emerged in focus groups as a potential way to reduce inconsistencies in how members hash content, especially when it comes to borderline or gray area cases. This could include more nuanced guidance on what phrases, images, symbols, and logos constitute a "call to violence," a "hate-based ideology," an "extremist identifier," or the other required elements of the behavioral inclusion criteria. However, member companies reported that the HSDB's inclusion criteria are relatively easy to apply to most forms of TVEC on their platforms.

The HSDB might also benefit from further expansion of permitted content types. The inclusion of logos, for example, is one broadly desired expansion that would equip members with an efficient way to identify insignia and other symbolic markers of terrorist and violent extremist groups and actors. While beneficial, logos do raise ethical and practical challenges arising from the context in which they emerge (e.g., journalism, research reports, actually violative contexts).[16]

Members also expressed a strong desire to standardize labels and their application. In general, there was agreement that a core set of labels should be developed that would be used universally by all members and applied to all hashes in the HSDB. Beyond this set of universal labels, members expressed support for a modular labeling system that will allow them to flexibly and optionally apply different "packages" of labels to describe the content they are hashing. Modules could relate to certain features of the content (e.g., ideology, group), as well as include information about the context in which the content was discovered and hashed. Overall, they could serve one of two goals: either (1) providing a clear labeling system that will allow members to selectively ingest hashes into their systems if the modules align with their platforms' policies, or (2) provide a method for aggregating a variety of content types that would be broadly applicable to different types of services (e.g., social media platforms, online marketplaces, search engines, cloud file storage). In either case, members would be free to use these modules as they see fit for their particular platforms and policies.

Beyond the development of a modular labeling system, another avenue for improvement could be for GIFCT to provide a mechanism for mapping members' own internal labeling systems to the inclusion criteria of the HSDB. This would provide clarity to members on specific alignments between their platforms' policies and the hashes available in the database. Some members expressed concerns about the difficulty in scaling to the HSDB's labeling system if such a mapping from the database to

[15] "The Definitions & Principles Framework Project," Global Internet Forum to Counter Terrorism, https://def-frameworks.gifct.org/

[16] Erin Saltman and Tom Thorley, "Practical and Technical Considerations in Expanding the GIFCT Hash-Sharing Database." (Global Internet Forum to Counter Terrorism, 2021).

their internal policies did not occur.

The discussion of a modular labeling regime for the HSDB prompted additional questions about how GIFCT could make tools like the HSDB more relevant to members' day-to-day operations. Some members suggested expanding the concept of modularity to encompass tooling beyond the HSDB itself. For example, databases of non-hashed logos, symbols, co-opted imagery, URLs, and other media hosted by GIFCT and developed in partnership with experts could offer an additional way to share information about TVEC with members.

GIFCT has already developed important auxiliary resources beyond the HSDB for members to utilize. Offering additional mechanisms for sharing information about TVEC that does not require members to develop new—or dramatically alter—existing pipelines for querying the HSDB could provide significant value. This is not a suggestion to limit the HSDB, but to specialize it as a counter terrorism tool that operates at scale and can easily integrate into existing pipelines and synchronize with member policies. If additional resources are made available to members that are subsequently widely used and implemented into their own systems to moderate TVEC, this may actually make the HSDB more functional as a scaled counter terrorism tool.

## GIFCT Member Integration

When asking members how they keep pace with the changing landscape of TVEC, it became apparent that issues of scale were paramount. Members with large operations tend to have automated and hybrid systems. When operating at scale, potentially violative content gets classified into general buckets (e.g., "violence/gore"), flagged as violative, and may not be further classified. Rather than delving into subsequent labeling, members operating at scale tend to view a single policy violation as both an efficient and sufficient way of weeding out violative content. Members also suggested that they currently do not have processes for determining what subset of the content they flag as violative meets the HSDB's expanded inclusion criteria. Additionally, they expressed concerns about the potential for overlapping classifications, leading to the invocation of more severe flags even when they are not necessary.

For its part, GIFCT has engaged in internal assessments and pilot data collection projects to understand the evolving TVEC landscape, its implications for the HSDB, and operational challenges with applying the inclusion criteria and labeling taxonomy.[17] Additionally, GIFCT funds and supports its academic wing, the Global Network on Extremism and Technology (GNET), to support and publish ongoing research into terrorist and extremist use of technology, and also runs annual working groups that convene experts, stakeholders, and members in an effort to stay up-to-date about relevant threats. GIFCT therefore has

[17] GIFCT highlighted some of these findings in its 2022 Transparency Report: GIFCT Staff, 2022 GIFCT Transparency Report (Global Internet Forum to Counter Terrorism, 2022).

access to a wealth of expert-informed information that could be mobilized to help members remain current with emerging trends in the TVEC landscape. These resources could be deployed to support development of the additional tools that live next to and augment the HSDB, and which are periodically updated as a part of GIFCT's ordinary operations, ongoing consultations with experts, or even through Working Groups themselves.

## Section 3: Case Studies

Beyond the review of the HSDB provided above, researchers from START were also asked by GIFCT to evaluate two emerging phenomena: (1) AI-produced TVEC and (2) non-perpetrator produced content. Members specifically requested that GIFCT and the Working Group consider the implications of AI-produced and non-perpetrator produced content for the HSDB to make recommendations about their suitability for hashing. In consultation with the HSWG, the START team considered these questions through two brief case studies.

### AI-Produced Content

Generative AI raises fresh challenges to countering terrorist content online and was highlighted as a key concern among academic, government, and civil society stakeholders. These challenges have accelerated in the past few years due to breakthroughs in natural language processing (NLP) by large language models (LLMs), image and video generation, and text-to-speech tools. AI models served through APIs and platforms (e.g., OpenAI, Claude) offer provisional protection against prohibited, violent, and offensive use-cases (albeit imperfectly).[18] However, a litany of tools exist for users to run AI models privately on their own local machines and with unscreened private data. For example, jailbroken GPTs promise easy access to "uncensored" language models that can be run on a user's personal computer.

While preliminary research has found that AI-generated content appears to constitute only a small proportion of currently known TVEC, it remains an emerging threat space that must be monitored.[19] Terrorist and extremist actors across the ideological spectrum are experimenting with generative AI in a variety of ways. Far-right actors have seized generative AI to generate memes and images depicting racist violence, spread symbols associated with white supremacy (e.g., Nazism, neo-fascist

[18] Kris McGuffie and Alex Newhouse, "The Radicalization Risks Posed by GPT-3 and Advanced Neural Language Models," Middlebury Institute of International Studies at Monterey, Center on Terrorism, Extremism, and Counterterrorism, September 9, 2020, https://www.middlebury.edu/institute/sites/default/files/2020-09/gpt3-article.pdf.

[19] Tech Against Terrorism, "Early Terrorist Experimentation with Generative Artificial Intelligence Services" (Tech Against Terrorism, November, 2023), https://techagainstterrorism.org/hubfs/Tech%20Against%20Terrorism%20Briefing%20-%20Early%20terrorist%20experimentation%20with%20generative%20artificial%20intelligence%20services.pdf.

accelerationism), and exploit LLMs to receive instructions on weapon fabrication.[20] Islamist extremists are using generative AI to produce propaganda, recruitment materials, and even in one example to manufacture pseudo-news broadcasts by deploying image generators to create artificial characters, text-to-speech models to produce voiceovers, and animation models to synchronize characters' lip movements with synthetic audio.[21] In another example, a leak of Meta's Llama model to 4chan enabled far-right actors to distribute the LLM via torrent, write tutorials on how to deploy the model, and teach others how to develop their own versions.[22]

How should the HSDB handle the rise of AI-generated TVEC? In the view of START's researchers, if the content produced by generative AI falls into the HSDB taxonomy, then it should be hashed just as conventional TVEC would be. When a source can clearly be identified as AI-generated, it would be beneficial to label the hash accordingly. Some AI-generated TVEC will bear clear markers of designated group affiliations, such as certain types of AI propaganda produced by supporters of ISIS, and can be dealt with by the HSDB efficiently. Likewise, an emerging trend among far-right actors is to include gore and violence explicitly in AI imagery, such as racist memes depicting the murder of racial minorities by white supremacist actors and meme characters.[23] In cases such as these, the taxonomy's behavioral inclusion parameters are clearly met. The content: 1) is produced by non-governmental entities; 2) has extremist identifiers; 3) has a clear core hate-based ideology; and 4) advocates directly for violence.

In other cases, however, the AI imagery may be less explicit. For example, does the "fashwave" aesthetic popular with neo-fascist accelerationists count as a violent extremist identifier? Is the inclusion of a skull-masked soldier in AI imagery sufficient to count as advocating or calling for violence? These are challenging conceptual questions, but notably they are not unique to AI but TVEC in general. The START team therefore does not see any reason for excluding AI-generated content from the HSDB so long as it is consistent with the taxonomy.

[20] Federico Borgonovo, Silvano Rizieri Lucini, and Giulia Porrino, "Weapons of Mass Hate Dissemination: The Use of Artificial Intelligence by Right-Wing Extremists," GNET, February 23, 2024, https://gnet-research.org/2024/02/23/weapons-of-mass-hate-dissemination-the-use-of-artificial-intelligence-by-right-wing-extremists/.

[21] Federico Borgonovo, Alessandro Bolpagni, and Silvano Rizieri Lucini, "AI-Powered Jihadist News Broadcasts: A New Trend In Pro-IS Propaganda Production?," GNET, May 9, 2024, https://gnet-research.org/2024/05/09/ai-powered-jihadist-news-broadcasts-a-new-trend-in-pro-is-propaganda-production/; Meili Criezis, "AI Caliphate: The Creation of Pro-Islamic State Propaganda Using Generative AI," GNET, February 5, 2024, https://gnet-research.org/2024/02/05/ai-caliphate-pro-islamic-state-propaganda-and-generative-ai/; Mariam Shah, "The Digital Weaponry of Radicalisation: AI and the Recruitment Nexus," GNET, July 4, 2024, https://gnet-research.org/2024/07/04/the-digital-weaponry-of-radicalisation-ai-and-the-recruitment-nexus/; Daniel Siegel, "AI Jihad: Deciphering Hamas, Al-Qaeda and Islamic State's Generative AI Digital Arsenal," GNET, February 19, 2024, https://gnet-research.org/2024/02/19/ai-jihad-deciphering-hamas-al-qaeda-and-islamic-states-generative-ai-digital-arsenal/.

[22] Daniel Siegel, "'RedPilled AI': A New Weapon for Online Radicalisation on 4chan," GNET, June 7, 2023, https://gnet-research.org/2023/06/07/redpilled-ai-a-new-weapon-for-online-radicalisation-on-4chan/.

[23] See Borgonovo, Lucini, and Porrino, "Weapons of Mass Hate Dissemination"; ADL, "Moon Man," Anti-Defamation League, n.d., https://www.adl.org/resources/hate-symbol/moon-man.

## Non-Perpetrator Produced Content

Another area for the HSDB to consider is the proliferation of non-perpetrator produced material that may depict what appears to be conventionally violative content.[24] It is increasingly common for bystanders and survivors of terrorist attacks and offline extremist violence events to document attacks as they are unfolding, as well as their aftermath.[25] To take one recent example, victims of the March 2024 Crocus City Hall ISIS-K terrorist attack in Moscow captured harrowing video of the attack. This footage, depicting the murder of dozens of civilians by ISIS-K terrorists, circulated widely on social media and has been referenced and used in news reporting.[26] At the same time, ISIS itself released a propaganda video depicting their attack, which included recorded footage from body cameras worn by the perpetrators.[27] Both types of content—victim footage and perpetrator footage—depict the violence and carnage of the attack. However, they do so from radically different points of view.

How should the HSDB handle content which depicts the consequences of terrorist violence but is not itself produced and shared by the perpetrators? In the START team's view, such non-perpetrator content, particularly when it involves victims and survivors of attacks, is out of scope for the HSDB. In the example above, these are media produced by non-governmental entities, but there is no core hate-based ideology or extremist identifier associated with the producer of the content, nor does the documentary evidence on its own call for or advocate for violence.[28] The footage also became an integral component of journalistic reporting about the attack, and exists (albeit often in an edited form) as video embeddings across a sprawling nexus of news articles.

This does not mean that non-perpetrator footage must exist un-moderated. Member companies are free to consult and enforce their policies as they see appropriate. It is perfectly reasonable for platforms to

[24] See GIFCT Incident Response Working Group, Handbook on Measuring and Evaluating Incident Response Online (Global Internet Forum to Counter Terrorism, November 2023).

[25] Stuart Allan, "Witnessing in Crisis: Photo-Reportage of Terror Attacks in Boston and London," Media, War & Conflict 7, no. 2 (August 2014): 133–51, https://doi.org/10.1177/1750635214531110.

[26] BBC, "Moscow Crocus City Hall Attack Videos Examined," BBC, March 22, 2024, https://www.bbc.com/news/av/world-europe-68643601; Valerie Hopkins and Alina Lobzina, "Screams and Blank Stares of Shock: Horror at a Russian Concert," The New York Times, March 24, 2024, https://www.nytimes.com/2024/03/24/world/europe/moscow-shooting-scene-survivors.html; Lazaro Gamio et al., "Maps and Diagrams of the Moscow Concert Hall Attack," The New York Times, March 23, 2024, https://www.nytimes.com/interactive/2024/03/23/world/europe/moscow-concert-hall-attack-map.html.

[27] Graeme Baker and Robert Greenhall, "Russian Court Charges Four Men with Act of Terrorism in Moscow Attack," BBC, March 24, 2024. https://www.bbc.com/news/world-europe-68652380.

[28] One tactic GIFCT should monitor is when an attack is broadcast not as part of the act of an intentional bystander or survivor recording, but as a part of a live-stream that unwittingly serves as a medium for documenting an attack. For example, a recent attack on a church in Australia was captured on the church's video feed (see Tiffanie Turnbull and Simon Atkinson, "Sydney church stabbing was 'terrorist' attack, police say," BBC, April 16, 2024, https://www.bbc.com/news/world-australia-68823240). However, the stream is neither conventional bystander footage (i.e., deliberately meant to capture an unfolding attack or its aftermath) nor perpetrator-directed footage. Would this qualify for inclusion? Such an ambiguous case warrants further consideration from GIFCT and monitoring of the TVEC landscape to see if this emerges as a common tactic by terrorist and extremist actors.

provide content warnings, age verifications, and topical context panels (as YouTube does for contentious issues to combat misinformation, such as with COVID-19 vaccines) to contextualize, describe, and warn users about the content. At the same time, the START team is aware of the potential for bystander footage to inadvertently facilitate the spread of terrorist messaging. The murder of Lee Rigby in the UK epitomizes this risk, as the perpetrators turned directly to bystanders who had been recording the attack to spread their message and rationale for the killing.[29] Nevertheless, efforts to combat terrorism online must always balance the free speech and human rights of victims, survivors, witnesses, and the general public. Given that such content does not align with the motivational requirements of the behavioral inclusion parameters, the START team does not feel it is suitable for hashing, but is certainly a candidate for content moderation and contextualization by technology companies.

## Conclusion and Recommendations

GIFCT's HSDB is a vital and unique counter terrorism tool. The database was originally created as part of a collaborative effort to address the challenge of terrorist imagery and videos associated with Islamist extremism that were flooding members' platforms during the height of ISIS's activities and reflects a commitment to share resources in order to combat the common threat of TVEC. For its part, GIFCT has shown versatility and a willingness to evolve the HSDB based on emerging threats in the TVEC landscape. In light of the HSDB's evolution and the review of the database conducted above, four recommendations for GIFCT to consider are presented below based on the START team's findings. Notably, START's researchers do not recommend an overhaul or expansion of the HSDB taxonomy. Overall, the START team finds the taxonomy to be flexible enough to cover the TVEC that the HSDB was designed to capture and feel that the labeling system is robust and covers the core elements of TVEC.

The START team sees two fundamental challenges facing the HSDB at the current juncture. The first relates to the internal technical and workflow changes that member companies need to make to add more variety to the content they are sending to the HSDB, which is discussed in Recommendation #1. Recommendation #2 pertains to how GIFCT can work with members to incentivize more active participation in the HSDB. Beyond these two core recommendations are two additional opportunities. Recommendation #3 considers at length the challenge of labeling and the possibilities for improvements. Finally, Recommendation #4 addresses members' stated desire for non-hashed resources that complement—but do not and should not replace—the HSDB, including symbology and plaintext URL databases.

### Recommendation #1. Work with members to understand their updated processes for classifying TVEC and resource constraints, focusing on developing scaled pipelines for hashing.

[29] Raffaello Pantucci, "Flight or Film? How Cameraphones Made the Bystander Key to Spreading Terror Message," The Guardian, December 6, 2015, https://www.theguardian.com/uk-news/2015/dec/06/flight-or-film-cameraphones-bystander-spreading-terror-message.

A primary goal of this Working Group has been to identify what TVEC is currently in the HSDB, what content that meets the 2021 expanded inclusion criteria is missing from the database, and determine methods to improve the comprehensiveness of its data. Due to the labeling issues detailed above, the START team was not fully able to quantify how much TVEC that meets the HSDB's inclusion criteria is currently missing from the database. However, member companies reported that there is likely a sizable amount of content that meets the behavioral inclusion parameters for non-designated entities that they are not currently sending to the database.

The underrepresentation of certain types of TVEC in the HSDB is primarily due to technical and resource challenges that member companies face (as well as their lack of engagement with each other and GIFCT). Member companies reported that they actively moderate content that satisfies the HSDB's behavioral inclusion parameters, but in most cases they reported a shortage of available resources and a lack of scaled pipelines for hashing, making it challenging for them to share content to the HSDB. However, in focus group sessions, START's researchers learned that the primary reason that member companies have not developed scaled pipelines to meet the HSDB's expanded typology is that improving the representativeness of the TVEC in the database is not currently a priority for them.

Indeed, it appears that member companies have had limited engagement with each other and with GIFCT on the technical, resource, and policy challenges that are hindering their ability to fully implement the 2021 taxonomy expansion and report TVEC that meets the criteria. For instance, in the context of their commitment as GIFCT members, companies could share details about their internal pipelines and moderation processes for the purposes of understanding the scope of the challenges they face in adopting the HSDB's expanded inclusion criteria. This information could then be used by GIFCT to help member companies identify solutions to their unique challenges. Future progress on making hashes in the HSDB more representative of the threat landscape will require more engagement and information sharing from member companies. The START team believes it would be beneficial for member companies to bring more voices to the table in their conversations with GIFCT, including the engineers who would be tasked with developing and maintaining the requisite pipelines for the HSDB.

One area where there is a critical need for greater collaboration between member companies and GIFCT is in the use of labeling systems. Some members reported that they are using different internal labeling systems that do not directly translate to the regime developed within GIFCT's taxonomy. This gap between the labeling system developed for the HSDB and those used internally by members is a foundational challenge that will have to be addressed cooperatively by both GIFCT and members. Over the long term, GIFCT should explore the feasibility of developing mappings between the HSDB taxonomy and members' labeling systems, and consider implementing this as part of the onboarding process. This would require investments of time and resources from GIFCT. But it would generate returns that would improve both the tech sector's collective response to TVEC and the strength of the HSDB.

Outside of resolving these labeling discrepancies, it will be critical for GIFCT to work with members to identify what specific technical or resource limitations are preventing them from sending more comprehensive content to the HSDB. Depending on the nature of these technical and resource limitations, GIFCT should consider how feasible it is for the organization to provide technical assistance to members when designing the necessary systems for using the HSDB. This would likely benefit smaller member companies the most, but large member companies also report resource constraints (though competing priorities have been the primary factor that has slowed their progress in embracing the 2021 taxonomy expansion).

### Recommendation #2. Work with members to understand and address barriers to HSDB participation, as the database's value depends on broad engagement.

A second recommendation is to have GIFCT work closely with members to identify how to encourage more participation in the hashing and sharing of in-scope TVEC to the HSDB. Currently, a small number of member companies are contributing most of the hashes to the database. Other members reported that they are not currently contributing to (or even using) the HSDB, but they expressed an interest in doing so. The types of TVEC encountered vary across member platforms. Thus, improving the variety of TVEC in the HSDB will require participation from all member companies. Moreover, because the HSDB is a collective endeavor that depends upon members sharing the burden of building the database, participation by all members is necessary for the HSDB to continue to be a useful counter terrorism tool.

The START team does not expect that members will all contribute to the HSDB in equal quantities. This reflects the realities that (a) the largest platforms are targeted by extremist and terrorist actors the most simply due to their size and large global user bases; and (b) resource limitations, which necessarily circumscribes the volume of content that smaller member companies can contribute relative to larger member companies. Nevertheless, based on what was said in the focus groups, more could be done to encourage participation in the HSDB from members. Some of this may pertain to labeling and technical challenges identified in the previous recommendation. If so, then assisting members with the labeling process and, if feasible, providing technical support for onboarding into the HSDB may itself foster more participation. Focus groups also revealed that there is an opportunity for actively participating members to work on improving the representativeness and ideological coverage of the types of TVEC they are hashing and sharing from their platforms.

In short, the START team suggests that member companies and GIFCT view improving the comprehensiveness of the HSDB as an opportunity to renew the collaborations that spawned the database and made it a valuable tool for countering Islamist extremist content on members' platforms. Member companies need to work together to find technical solutions to hashing more comprehensive content at scale, and they need to encourage each other to be active participants in addressing their shared challenges. GIFCT can facilitate future collaboration and provide technical support when possible, but ultimately the continued success of the HSDB requires engagement from all members.

### Recommendation #3. Work with members on standardizing and improving the use of labels within the HSDB.

In focus groups with members, labels emerged as a core point of discussion, and there was widespread agreement that standardizing the use of labels is necessary for improving the HSDB. In the START team's assessment, GIFCT has developed a relevant set of labels that describes crucial features of hashes. One limitation of the labeling system, however, is that outside of a general label used for internal tracking within the HSDB, the only other required label pertains to the type of behavior captured in a hash.

While these behavioral descriptors are necessary for understanding why a hash was included in the HSDB, the lack of additional labeling requirements has left a significant proportion of hashes unlabeled across core contextual dimensions. This has made it challenging to understand how many hashes in the HSDB are associated with specific groups, movements, or ideologies (and, as START's researchers discovered, non-designated entities). As the TVEC landscape continues to evolve, it is imperative that hashes be populated with descriptive labels capturing ideological and group or movement affiliation so that both GIFCT and members have a sense of how well the HSDB is capturing in-scope content.

In discussions with member companies and other stakeholders, there was a general sense that a significant amount of the TVEC that qualifies for inclusion in the HSDB is relatively straightforward to judge in comparison to the behavioral inclusion criteria. However, participants recognized that there are borderline and edge cases that stress the conceptual boundaries. Thus, addressing conceptual ambiguity around the taxonomy, specifically in the context of non-designated entities that would enter the database via the behavioral inclusion parameters, could have an impact on improving the representativeness of the hashes in the HSDB. One fruitful way this could be accomplished is through illustrative case studies of what GIFCT sees as "obvious" and "edge" candidates for inclusion into the HSDB. In whichever format chosen, GIFCT could illustrate how each feature of the behavioral inclusion parameters was applied in particular cases.

Ambiguities around these core conceptual issues can lead to labeling errors in the HSDB. However, as it currently stands, GIFCT itself cannot directly remediate labeling mistakes for hashes submitted to the HSDB. In agreement with internal assessments from GIFCT based on their own pilot data collection projects and tests of the taxonomy, the START team recommends that GIFCT should be endowed with the proper authority to directly audit, quality control, validate, and fix labeling errors in the HSDB. Because the HSDB is composed of hashes, however, this would likely require additional technical development to make back-labeling possible. This effort would almost certainly require members to provide GIFCT access to pre-hashed content, when available, so GIFCT can perform audits and revisions. It would also require the infrastructure housing the HSDB, the ThreatExchange API, to support these routine quality control mechanisms. Nevertheless, as keeper of the HSDB and its taxonomy, it would be reasonable to allocate such authority to GIFCT.

Members also expressed a strong desire for a "modular" labeling system. A modular approach to labeling was conceived as being based around "packages" (i.e., "modules") of labels that pertain to—and describe—different features of the hashes. From their perspective, members could then flexibly ingest specific packages of labels that they are interested in based on their policies and needs. This modular system would be rooted around a core set of labels that are universally applicable to every hash and used by every member to ensure that fundamental features associated with hashes are adequately captured, while then allowing flexibility around additional features that members would be free to use—or not—as they deem appropriate.

Members did not indicate what specific modules should be developed to support such a system, nor what labels would constitute the universal and required set. While it is beyond the scope of this current Working Group to devise such a system, due to the strong desire members expressed for modularity, GIFCT should work closely with members on formulating the specifics of such an approach and the implications it would have for the current HSDB taxonomy and members' existing pipelines.

**Recommendation #4. Work with members to develop non-hashed resources, including a symbology database of logos, imagery, co-opted symbols, and phraseology with contextual explanations, as well as a plaintext URL database that would better serve member needs than hashed URLs.**

The HSDB has performed well countering some of the most violent and widespread depictions of extremist and terrorist violence through hashing. At the same time, members expressed a desire to complement the HSDB with additional non-hashed resources that could serve as a knowledge base about TVEC without the technical requirement of setting up a hashing pipeline. Two specific potential initiatives emerged: 1) a symbology database; and 2) a plaintext URL database.

A non-hashed symbology database would consist of logos and other symbolic markers, perhaps even slogans and phrases, used by violent extremist and terrorist actors that members could query directly. This database would provide contextual information about the movements or ideologies associated with a symbol, and allow companies to use that information flexibly in their own systems (e.g., perhaps they can select and export a set of symbols in a suitable format and include them in existing content moderation pipelines on their platforms).

A symbology database would support member companies in hashing logos in the HSDB. In the existing HSDB taxonomy, when a logo is attached to an inclusive piece of TVEC, it serves as a violent extremist or terrorist signifier. If members have access to a resource like a symbology database that they can reference in order to understand the symbolic practices of varied violent extremist actors, then this could help with the identification of logos, co-opted imagery, and coded language in potentially violative content. This should help members more confidently apply the existing taxonomy to branded materials.

For its part, a non-hashed URL database would exist as a repository of URLs pointing to known

TVEC online that members could access directly without hashes. Members expressed that URLs are particularly difficult to hash due to their variability and permutations (e.g., the use of link shorteners and use (or non-use) of “https” or “www”). Consequently, members communicated a preference to handle and share URLs in a plaintext format.

2024 GIFCT Working Group Participant Affiliations[30]

| Academia | Advocacy | Practitioner & Researcher | Government & Intergovernmental | Tech |
|---|---|---|---|---|
| ACUNS, ISACA | ADL | Brookings Institution | Aqaba Process, Jordan Government | Amazon Web Services (AWS) |
| American University | All Tech is Human (ATIH) | Centinel | Australia, Department of Home Affairs | Discord |
| Center for Cyber Strategy & Policy, School of Public and International Affairs, The University of Cincinnati | ARTICLE 19 | Digital Security Group | Christchurch Call | Dropbox |
| Central University of Gujarat | ASEAN Coalition to Stop Digital Dictatorship | Fem AI | Department of Internal Affairs NZ (Digital safety and illegal harms) | ExTrac AI |
| Collaboratory Against Hate, Carnegie Mellon University and University of Pittsburgh | Association of british muslims | Global Disinformation Index | eSafety Commissioner Australia | Giphy |
| Columbia University School of International and Public Affairs (SIPA) | Internet Society | Hedayah | European Commission | GoDaddy.com |
| Extremism and Gaming Research Network (EGRN) | KizBasina (Just-a-Girl) NGO | Jihadoscope | Federal Bureau of Investigation (FBI) | Insikt AI and Dataietica.org |
| Georgetown University | Koan Advisory | Moonshot | Federal Ministry of the Interior and Community, Germany | ISACA Kenya |
| Hesse State University of Public Management and Security | Moroccan Observatory on Extremism and Violence | Online Safety Exchange | Netherlands Ministry of Justice and Security | Meta |
| Macquarie University | Policy Center for the New South | Peace Research Institute Frankfurt (PRIF) | New Zealand Classification Office | Microsoft |
| Royal Holloway, University of London | Search for Common Ground | Swansea University | Ofcom | Mozilla Corporation |

[30] This table highlights participants across all Year 4 Working Groups.

| | | | | |
|---|---|---|---|---|
| RUSI | Southern Poverty Law Center | Tech Against Terrorism | OSCE Secretariat, Action against Terrorism Unit | Nexi Group |
| Sapienza University of Rome (Italy) | Take This | The Millennium Project (South Asia Foresight Network) [SAFN] | Public Safety Canada | Niantic Labs |
| Swansea University | | Tremau | U.S. Department of Homeland Security | Resolver, a Kroll business |
| Trinity College Dublin | | | U.S. Department of State | SoundCloud |
| University of Cambridge | | | UK Home Office | SpaceYaTech and Africa ICT Alliance |
| University of Essex, Department of Government | | | UNICRI - United Nations Interregional Crime and Justice Research Institute | Twitch |
| University of Paris Cité (France) | | | United Nations | X |
| University of South Wales | | | United Nations Office of Counter Terrorism (UNOCT) / United Nations Counter-Terrorism Centre (UNCCT) | Xbox |
| University of Sussex | | | Virginia State Police, USA | YouTube |
| University of Waterloo | | | | Yubo |
| Vox-Pol Institute | | | | |

Recommended citation: Global Internet Forum to Counter Terrorism and National Consortium for the Study of Terrorism and Responses to Terrorism, Hash Sharing Database Review: Challenges and Opportunities (Washington, D.C.: Global Internet Forum to Counter Terrorism, 2024), Year 4 Working Groups.

GIFCT is a 501(c)(3) non-profit organization and tech-led initiative with over 30 member tech companies offering unique settings for diverse stakeholders to identify and solve the most complex global challenges at the intersection of terrorism and technology. GIFCT's mission is to prevent terrorists and violent extremists from exploiting digital platforms through our vision of a world in which the technology sector marshals its collective creativity and capacity to render terrorists and violent extremists ineffective online. In every aspect of our work, we aim to be transparent, inclusive, and respectful of the fundamental and universal human rights that terrorists and violent extremists seek to undermine.

**www.gifct.org** **outreach@gifct.org**