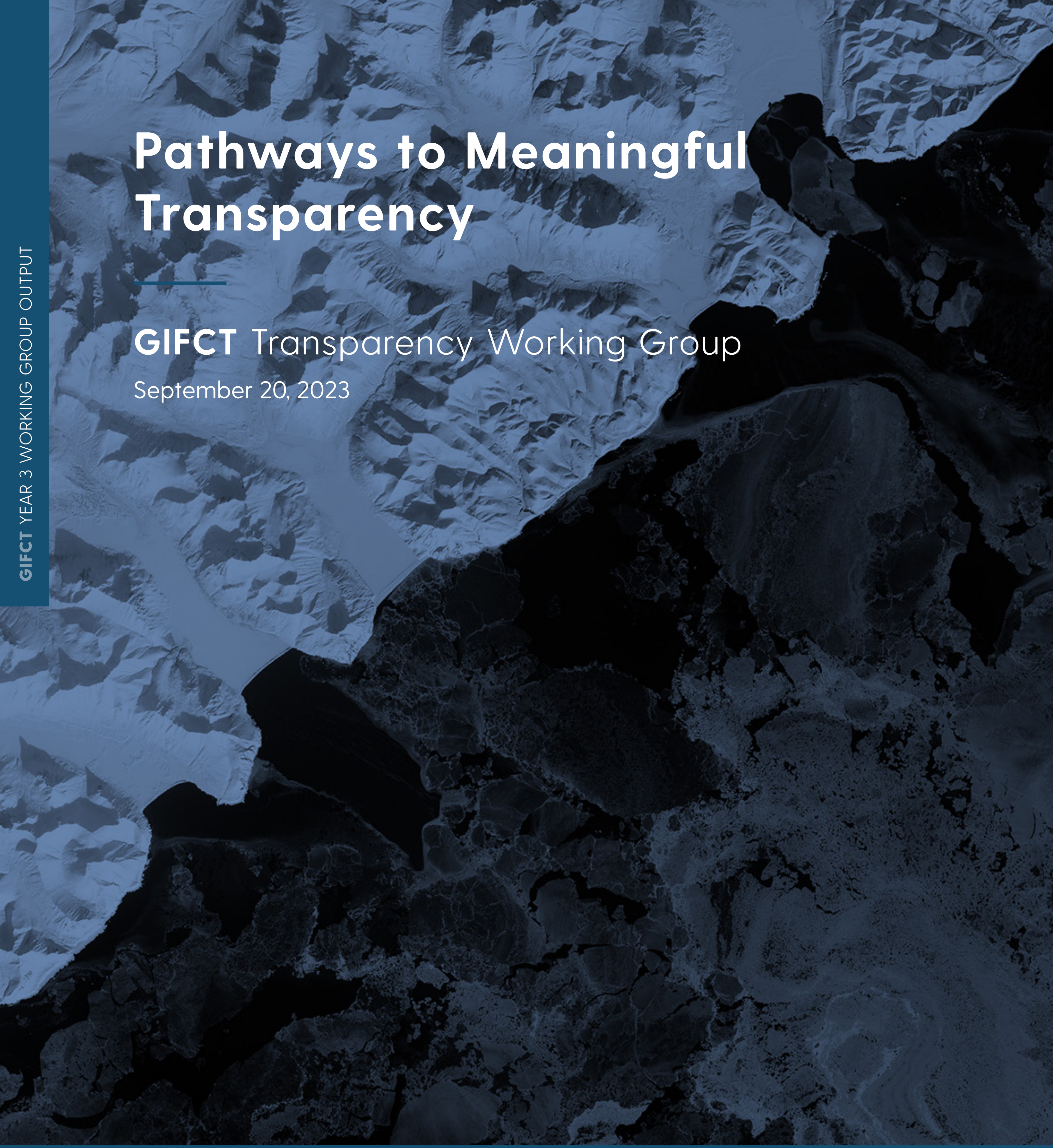
GIFCT YEAR 3 WORKING GROUP OUTPUT



Pathways to Meaningful Transparency

GIFCT Transparency Working Group

September 20, 2023

GIFCT
Global Internet Forum
to Counter Terrorism

Chris Beall, Carnegie Endowment for International Peace
Martin Cocker, Online Safety Exchange
Dr. Abbee S. Corb, Simon Wiesenthal Center
Jonathon Deedman, Richmond the American Intl. University in London

About GIFCT Year 3 Working Group Outputs

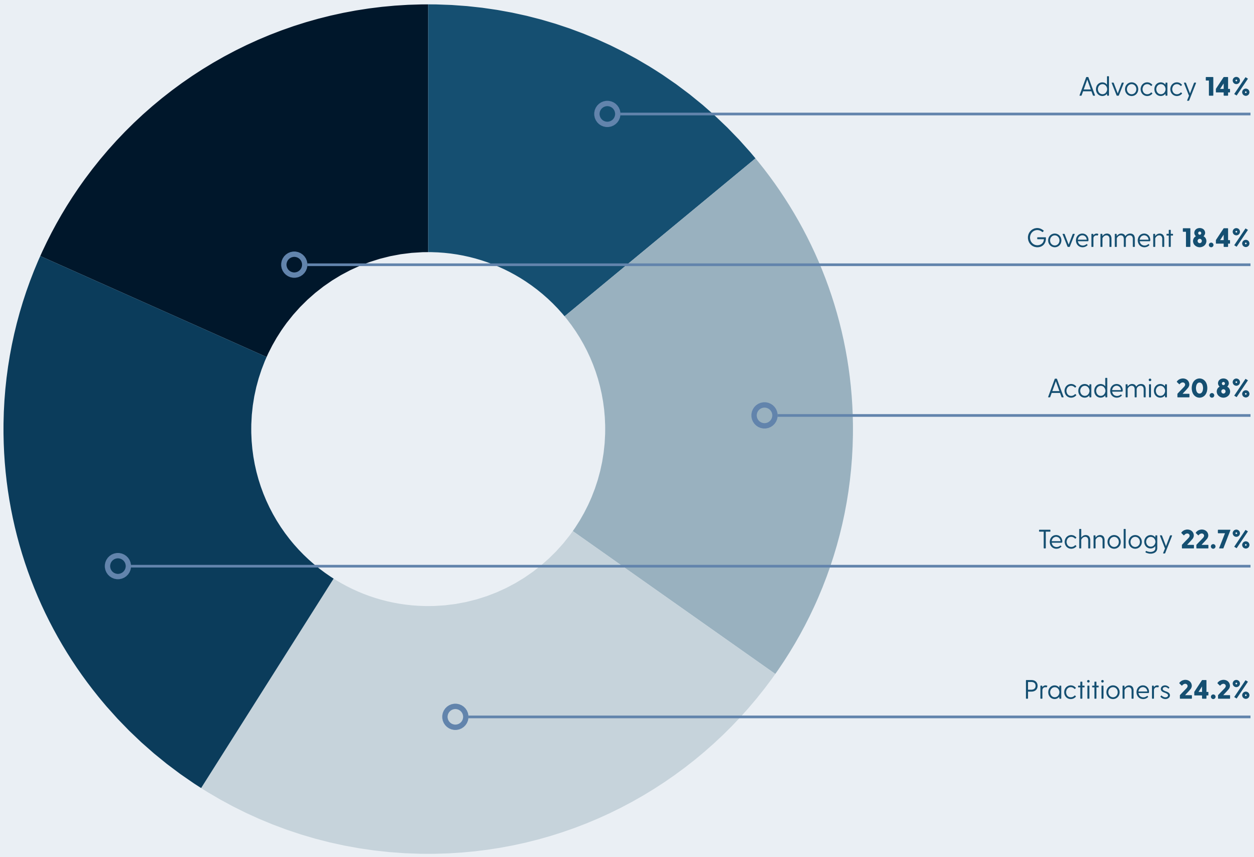
By Dr. Nagham El Karhili, Programming and Partnerships Lead, GIFCT

In November 2022, GIFCT launched its Year 3 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.

Overall, this year's five thematic Working Groups convened 207 participants from 43 countries across six continents with 59% drawn from civil society (14% advocacy organizations, 20.8% academia, and 24.2% practitioners), 18.4% representing governments, and 22.7% in tech.

WG Participants

Sectoral Breakdown



Beginning in November 2022, GIFCT Year 3 Working Groups focused on the following themes and outputs:

1. **Refining Incident Response: Building Nuance and Evaluation Frameworks:** This Working Group explored incident response processes and protocols of tech companies and the GIFCT resulting in a handbook. The handbook provides guidance on how to better measure and evaluate incident response around questions of transparency, communication, evaluation metrics, and human rights considerations.
2. **Blue Teaming: Alternative Platforms for Positive Intervention:** After recognizing a gap in the online intervention space, this GIFCT Working Group focused on highlighting alternative platforms through a tailored playbook of approaches to further PVE/CVE efforts on a wider diversity of platforms. This included reviewing intervention tactics for approaching alternative social media platforms, gaming spaces, online marketplaces, and adversarial platforms.
3. **Red Teaming: Assessing Threat and Safety by Design:** Looking at how the tech landscape is evolving in the next two to five years, this GIFCT Working Group worked to identify, and scrutinizes risk mitigation aspects of newer parts of the tech stack through a number of short blog posts, highlighting where safety-by-design efforts should evolve.
4. **Legal Frameworks: Animated Explainers on Definitions of Terrorism and Violent Extremism:** This Working Group tackled questions around definitions of terrorism along with the impact that they have on minority communities through the production of two complementary animated videos. The videos are aimed to support the global counterterrorism and counter violent extremism community in understanding, developing, and considering how they may apply definitions of terrorism and violent extremism.
5. **Frameworks for Meaningful Transparency:** In an effort to further the tech industry's continued commitment to transparency, this Working Group composed a report outlining the current state of play, various perspectives on barriers and risks around transparency reporting. While acknowledging the challenges, the Working Group provided cross sectoral views on what an ideal end state of meaningful transparency would be, along with guidance on ways to reach it.

We at GIFCT are grateful for all of the participants' hard work, time, and energy given to this year's Working Groups and look forward to what our next iteration will bring.

To see how Working Groups have evolved you can access Year One themes and outputs **HERE** and Year Two **HERE**.

Pathways to Meaningful Transparency

Introduction

Background

Transparency reporting can play a central role in preventing terrorists and violent extremists from exploiting digital platforms. It can help foster understanding and trust among stakeholders and create a basis for collective action against that exploitation. Effective transparency can underpin the design, evaluation, and continuous improvement of interventions, improve accountability, and ensure the protection of human rights and fundamental freedoms – both online and offline.

GIFCT convened the Year 3 Transparency Working Group with over 50 experts from a diverse range of organizations, stakeholder groups, geographies and disciplines to build upon the 2021 and 2022 work on transparency best practices and implementation.[1] The 2023 Group set out to understand and articulate what meaningful transparency reporting is in relation to terrorist and violent extremist content (TVEC). In particular, the group sought to articulate the barriers to its implementation and make recommendations for the development of meaningful TVEC transparency reporting across the industry.

The 2023 Transparency Working Group members believe that transparency reporting could be utilized to make a more substantial contribution to the collective action against the terrorist and violent extremist exploitation of technology than it currently does. Accordingly, we have sought to make a meaningful and practical contribution to its development. The findings in this report reflect the majority position of active participants from the GIFCT Transparency Working Group, the balance of existing desk research, and dominant survey findings from a collaborative Pol.is online survey of TWG members and their organizations.[2] The report outlines the current state of play, discusses sectoral barriers and risks to achieving effective transparency, explains how to reach an ideal end state for meaningful transparency, lays out what sectoral practitioners see such an ideal end state as looking like, and concludes by providing recommendations to improve progress toward that state. However, it should be noted that there are divergent views that would also benefit from further exploration.

Summary

Violent extremists and other criminals have leveraged and continue to manipulate available technology to suit their needs. GIFCT recognizes transparency as an important tool to empower collective action against that exploitation.

[1] "Introducing 2022 GIFCT Working Group Outputs," GIFCT.com, https://gifct.org/wp-content/uploads/2022/07/GIFCT-22WG-ResearchAgendaScopingPaper-1.1.pdf.

[2] Pol.is presents an opportunity to explore the different ways that a group of people thinks about a divisive or complicated topic such as transparency. To date, 39 different respondents have voted on the statements, presenting some degree of unanimity on the overall value of transparency reporting to industry, government, and civil society in countering violent extremism. See the Ideal End State section below for further details.

There has been significant growth in TVEC reporting in the last three years. The OECD found 15 of the top 50 most popular tech platforms issued TVEC-specific transparency reports in 2022, tripling the number published in 2020. Despite this growth, the lack of progress toward *meaningful* transparency continues to frustrate civil society organizations (CSO) and governments.

As governments, tech companies, and CSOs attempt to make progress on the practical implementation of TVEC transparency, they are confronted by both common and sector-specific challenges. While all sectors generally agree that transparency is always an important fixture, there are variations in expectations and definitions at almost every juncture. There are substantial differences in the way technology companies are categorized or defined, the designation of terrorist or extremist groups, expectations on the protection of user rights, and how harmful content is defined and categorized. Combined with a mostly voluntary reporting environment, the result has been inconsistency across reports. That has, in turn, led to criticism of both tech companies (for not meeting standards) and governments (for not enforcing them). The issue is further compounded when defining extremist content is aligned with the government of the day, thus creating not only a problem of definitional application, but also conflict between the business and policy arms of tech companies.

In addition to a need for sectors to jointly resolve the variations in definitions and expectations of transparency, there is a need for each sector and individual organization to resolve challenges specific to them. Governments and tech industry organizations are especially challenged by a need to find the right balance between transparency and protecting the privacy of citizens and platform users. CSOs will often have to balance opportunities for access to sensitive data and information against a need to remain independent.

A number of actors are seeking to improve the uniformity of transparency reporting, both uni- and multilaterally, in collaboration with emerging and established platforms. However, no organization or group has yet developed a widely accepted 'standard' for transparency that allows for meaningful interpretation of data across sectors. Simultaneously, the EU Digital Services Act[3] is seen as the beginning of a transition from a primarily voluntary reporting environment to a more regulated one, whereby legislators may set the standard rather than sectoral stakeholders themselves.

To achieve a "healthy information environment" where meaningful transparency can be better attained, industry, government and civil society partners must limit the spread and impact of violent extremist content and support a strong, well understood, and clearly articulated balance between rights and freedoms, security and accountability. As a primary tool to track progress toward this desired end state, transparency reporting needs to enable industry and government partners to measure the efficacy of their interventions designed to safeguard rights and protect users and provide CSOs the knowledge

[3] The European Parliament and Council of the European Union, "Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)," EUR-Lex, accessed July 16, 2023, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065.

to hold both to account for their commitments. To that end, transparency reporting should be based on trust, good faith, shared knowledge, common outcomes, and interoperability and comparability between industry and government reporting structures.

Recommendations

To evolve the current reporting landscape toward a safer, healthier, and more trustworthy information environment, transparency reporting should be provided in a manner that is more comparable across sectors and includes the information necessary to limit the spread of TVEC content while protecting rights and freedoms.[4]

We recommend:

1. GIFCT encourage its partners and members to work collaboratively toward articulating a clear set of desired outcomes and to weigh the acceptable level of risk inherent with balancing human rights, freedom of expression, and limiting the spread of TVEC content.
2. GIFCT support this approach by launching an international citizens assembly[5] to surface global considerations and expectations for balancing risks and benefits related to combating TVEC in the information environment. This would provide a broader set of views that could enhance and challenge expert positions. For example, Canada's Canadian Citizens' Assemblies on Democratic Expression model released meaningful, citizen-generated advice on digital tech oversight designed to strengthen legislative decision-making for the digital ecosystem.[6] Using globally tested methods, like that of the recent Global Assembly on Climate Change,[7] and leveraging its extensive, international civil society partners, GIFCT could spearhead a consultation on ideal goals, outcomes, and risk tolerance to help inform the community's ultimate goals.
3. GIFCT work with its partners and members to enhance existing reporting templates to create an adaptable basic format and skeleton design of a transparency report that focused on intended outcomes. This work should build upon existing guides and frameworks like those provided by Tech Against Terrorism[8] or Susan Ness and Chris Riley's "Modularity" framework (as a form of

[4] Further Information pertaining to the recommendations is provided in the full recommendations section later in the document.

[5] Citizens' Assemblies, "Discover democracy that works," accessed August 9, 2023, https://citizensassemblies.org/; Citizens' Assemblies, "UK Citizens' Assemblies," accessed August 9, 2023, https://citizensassembly.co.uk/.

[6] 3rd Canadian Citizens' Assembly on Democratic Expression, "Canadian Citizens' Assembly on Democratic Expression: Recommendations for reducing online harms and safeguarding human rights in Canada," Ottawa Public Policy Forum, 2022.

[7] Global Assembly Team, "Report of the 2021 Global Assembly on the Climate and Ecological Crisis" accessed July 14, 2023, http://globalassembly.org.

[8] Tech Against Terrorism, "Tech Company Transparency Reporting on Online Counterterrorism Efforts," accessed August 7, 2023, https://static1.squarespace.com/static/609d273957ee294d03d8dadf/t/60fe84b9f629417f8e459847/1627292859743/TAT+Guidelines+-+Tech+company+transparency+reporting+on+online+counterterrorism+efforts.pdf.

multi-stakeholder, co-regulatory governance)[9] to enable governments to align their approaches across different legal structures.

4. Recognizing the challenges for smaller industry partners and the barriers to entry that changes to reporting can create, GIFCT should continue to support new entrants by building capacity and leveraging industry and government support. eSafety's Safety-by-Design initiative demonstrates the value in running industry-leading, capacity-building work alongside strict enforcement activities that could be emulated.[10]
5. GIFCT take a leadership role by requiring its members to demonstrate how they are enhancing their reporting practices to include reporting against previously recommended outcomes and to enable comparability of findings across sectors.

Current State of Play

Terrorists and violent extremists have leveraged available technology to suit their needs. The GIFCT Transparency Working Group believes that transparency can empower collective action against that exploitation.

The effectiveness of specific transparency reporting in empowering that collective action depends on the reported content meeting the specific needs of multiple audiences. Each audience's needs depend on their communities' individual challenges and the interventions they are undertaking (or seek to undertake). To better understand the needs and expectations around transparency, the GIFCT Working Group utilized the survey tool pol.is to examine the perceptions regarding how tech transparency can help civil society, governments, and industry reach their goals.

The survey revealed several areas of inconsistency that currently exist in transparency reporting. For example, there was a substantial divergence among respondents between those prioritizing greater transparency around decision-making processes for content removal and those who emphasize a need for more data about the volume of removals. Interestingly, those respondents who argued for more data on volumes of removals were also the ones that indicated that current reporting structures had already contributed to their work.

The survey also identified several shared areas for improvement when it comes to the current state of transparency reporting in tech. For example, to collate and combine transparency information across platforms, respondents suggested that transparency reporting needs to be as consistent as can be realistically attained. It is recognized that transparency reporting frameworks should be flexible enough

[9] Chris Riley and Susan Ness, "Modularity for International Internet Governance," Lawfare, July 19, 2022, https://www.lawfaremedia.org/article/modularity-international-internet-governance; Chris Riley, "A Module Playbook for Platform-to-Researcher Data Access," Tech Policy Press, November 20, 2022, https://techpolicy.press/a-module-playbook-for-platform-to-researcher-data-access/.

[10] Government of Australia, "Basic Online Safety Expectations," Australian eSafety Commissioner, July, 2022, https://www.esafety.gov.au/sites/default/files/2022-07/Basic%20Online%20Safety%20Expectations%20regulatory%20guidance.pdf.

to accommodate the different capabilities and reporting metrics each tech platform has. Efforts to create a comprehensive or uniform transparency reporting framework has yet to successfully find a balance between a consistent structure that allows for cross-platform and sector-wide comparability and a working flexibility that meets the various needs of those within the TVEC response ecosystem. As a result, transparency reports have become more commonplace throughout the tech sector, without perceived corresponding successes in the fight against TVEC online within the same period.

The challenge within the current state of play is to find meaningful transparency standards without settling for the minimum requirements – and create a reporting environment that provides incentives for better reporting. As all respondents indicated in the survey, there need to be standards across the tech industry about what is meaningful transparency. Currently, there is considerable discretion as to what 'transparency' itself refers to, with a corresponding variation in what is included in transparency reporting. Reports vary from high-level descriptions of trust and safety efforts through to summary data on intervention events (such as the aggregated outcomes of takedowns and appeals). A significant variance in the meaning of transparency also exists across countries and cultures. This is especially true in how transparency is balanced against related and complementary concepts such as censorship, privacy, and freedom of speech.

As a result of the above discrepancies, reports that include data meaningful to practitioners can often be seen as inaccessible to individuals outside of that community. Reports that are pitched at a user level can fail to provide the depth of information and analysis that more expert practitioners desire. Even if the level of reporting is appropriately designed, the report may still satisfy one state's conception of transparency without satisfying that of another. These issues, when amalgamated, have prevented more uniform transparency reporting.

TVEC transparency reporting has advanced substantially in the past five years. Of the global top 50 content-sharing services, five issued TVEC transparency reports in 2020 rising to eleven in 2021, and fifteen in 2022.[11] Those 15 reports represent just over half of the 28 platforms on which TVEC was known to have appeared.

[11] OECD, "Transparency reporting on terrorist and violent extremist content online 2022," OECD Digital Economy Papers, no. 334, 2022, https://doi.org/10.1787/a1621fc3-en.

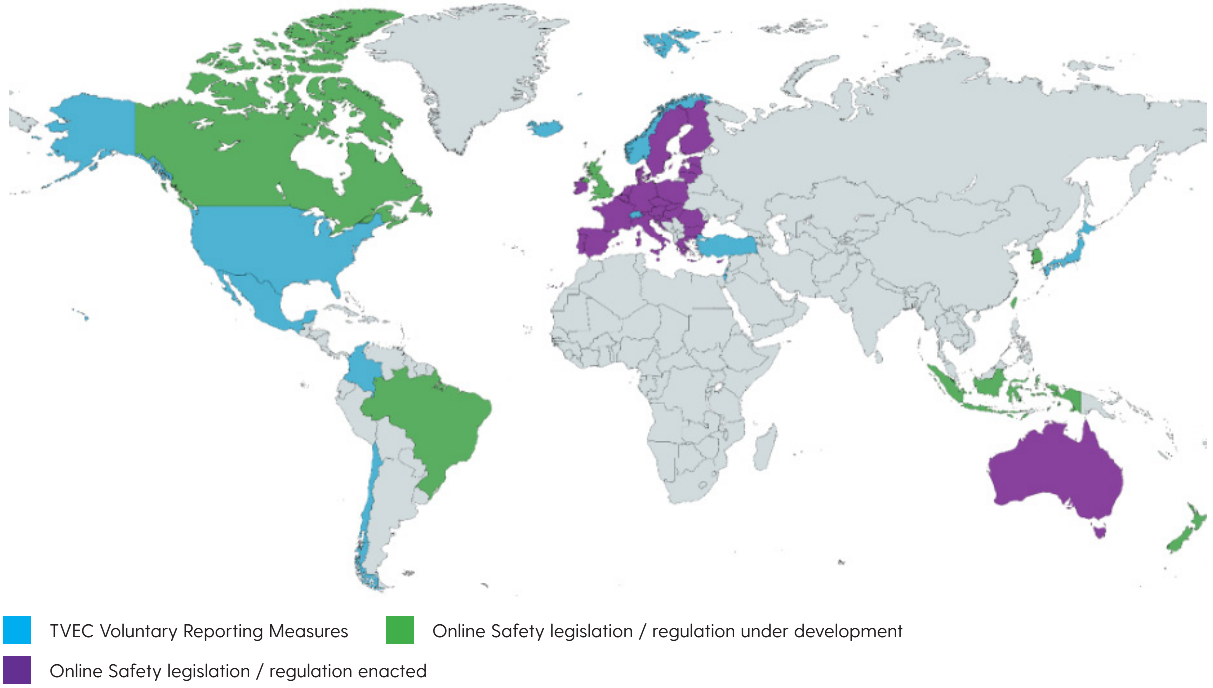

🟦 TVEC Voluntary Reporting Measures
🟩 Online Safety legislation / regulation under development
🟪 Online Safety legislation / regulation enacted

Table 1: Types of Transparency Reporting Measures

The growth of individual transparency reporting efforts has also been driven by the exploitation of individual tech platforms by violent extremists. Terrorism is theater, with hopes of reaching the greatest number of people in the shortest amount of time, and larger technology platforms represent an effective conduit. For example, the live-streaming platform Twitch was exploited by the Buffalo shooter in May 2022 when he broadcast his attack via a helmet camera. The platform – impacted by numerous hate raids, incidents of terrorism, pornography, extremism, and gore – responded by implementing new moderation and reporting tools to aid with transparency.[12]

The quality of TVEC transparency reporting has benefitted (and will continue to benefit) from the lessons and experiences across the wider field of transparency reporting. For example, the EU and Australian Codes of Practice on Disinformation have coordinated previously fractured reporting environments and driven improvements in data presentation, explanations about year-on-year variations, and ensuring reports are accessible.[13]

Improvements in transparency reporting techniques – including the contextualization of data, differentiating proactive detection from reactive removal, linking between reports and associated

[12] Twitch, "H1 2022 Transparency Report," accessed August 8, 2023, https://safety.twitch.tv/s/article/H1-2022-Transparency-Report?language=en_US.

[13] The European Commission, "The 2022 Code of Practice on Disinformation," June 16, 2023, https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation; Digital Industry Group Inc. "Australian Code of Practice on Disinformation and Misinformation," DI, December 22, 2023, https://digi.org.au/wp-content/uploads/2022/12/Australian-Code-of-Practice-on-Disinformation-and-Misinformation-FINAL-_-December-22-2022.docx.pdf.

policies, providing more detail on government takedown requests, and including data on reinstatements – are helping to better meet the needs of government and CSO stakeholders. The increased use of data visualizations and jargon-free language have also improved accessibility for a wider audience. An example of this new era of transparency reporting can be seen in Microsoft's Xbox, which released its first transparency report in the second half of 2022. The report extensively details the actions it undertakes (and has undertaken) to protect players from inappropriate behavior, which includes cheating, profanity, sexual content, harassment, hate speech, and misconduct on its platform.[14]

Efforts to centralize and index transparency reports, such as the GIFCT Member Resource Guide,[15] have assisted stakeholders working across the industry. The GIFCT membership requirements and partnership with Tech Against Terrorism[16] ensure the reports listed in that resource guide meet minimum standards.

Despite this substantial growth in the volume and quality of transparency reporting, governments and civil society continue to criticize tech companies for failing to ensure clarity regarding specific audiences and objectives for their transparency reporting. And it is true that as a substitute for investing in more substantive efforts against TVEC, transparency reports from tech platforms can be reduced to a public relations exercise or used solely to demonstrate compliance with government requirements.

Civil society organizations often drive new regulatory approaches across sectors, but do not necessarily remain at the forefront of transparency reporting and regulation when it is engaged in by tech or by governments.[17] This creates a sectoral disconnect, whereby those who conceptualize and prescribe updated regulatory approaches are not able to ensure that transparency reporting from the tech sector aligns with these updated approaches. This leaves tech companies almost unilaterally exposed to criticisms regarding a lack of clarity. Greater ongoing involvement of CSO transparency practitioners could ensure clarity regarding the specific audiences and objectives within tech's transparency reporting.

To date, the majority of TVEC transparency reporting has been undertaken voluntarily, with individual technology companies making key choices about what information to present and how. The EU Digital Services Act[18] (which takes full effect in February 2024) and the U.K. Online Safety Bill (which will likely pass into law in late 2023) mark a likely turning point. They will force large platforms to be more transparent about how their algorithmic systems work and seek to hold them to account for the wider societal harms linked to the exploitation of their services. Efforts to enforce transparency reporting have focused on big tech, and increasingly on their use of algorithms for the purposes of content surveillance,

[14] Microsoft, "Xbox Transparency Report," December, 2022, https://www.xbox.com/en-GB/legal/xbox-transparency-report.

[15] GIFCT, "Resource Guide," accessed August 7, 2023, https://gifct.org/resource-guide/.

[16] GIFCT, "Membership," accessed August 7, 2023, https://gifct.org/membership/.

[17] Bridget M. Hutter and Joan O'Mahony, "The Role of Civil Society Organisations in Regulating Business," ESRC Centre for Analysis of Risk and Regulation Discussion Papers, no. 26, 2004, https://www.lse.ac.uk/accounting/Assets/CARR/documents/D-P/Disspaper26.pdf.

[18] EU, "DSA," 2022.

moderation, monitoring, and feed curation. Through this administrative capacity, algorithms clearly play a central role in TVEC,[19] and there is an increasing demand for greater algorithmic transparency within comprehensive tech transparency reporting.[20] To date, this has not occurred, and substantive concerns remain about the possible harms that could arise from any algorithmic transparency being exploited, as even positive transparency mechanisms can be misused by bad actors.

As a result of the greater sectoral focus on the algorithmic debate and transparency efforts of larger tech platforms, smaller platforms that also host and invigorate TVEC activity are often able to avoid intense scrutiny, while the public understanding of TVEC activity becomes skewed. Indeed, within the current state of play, the inequities in prescribed transparency reporting between large and small tech companies (driven in particular by a distinct use of algorithmic content monitoring almost exclusively by larger tech platforms) continues to allow TVEC to filter through distinct platforms, and so still find a medium upon which to be digested by the radical milieu it aligns with. The voluntary nature of transparency reporting, even by the larger tech companies that are faced with more intense scrutiny, has prevented greater examination of this process.

To add to the evolving challenge, the current focus on industry does not hold governments accountable for their lack of transparency. Yet, as the majority of survey respondents reported, transparency related to counterterrorism cannot only be about industry. Governments need to be open about their actions online, too. Openness and transparency, especially about government data and information requests to industry, help protect everyone's freedom and privacy. The Christchurch Call to Action is currently working on guidance to aid government transparency reporting.[21]

Within the current state of play, many actors are seeking to improve the uniformity of transparency reporting, both uni- and multilaterally. For example, the 2022 GIFCT annual review included a new component of GIFCT membership whereby attempts have been made to create a more consistent transparency framework for its member organizations to adhere to.[22] In addition, recent GIFCT definitional frameworks – such as those put forward within the Definitions & Principles Framework Project – have attempted to increase taxonomic uniformity and help member organizations increase the external comparability of their transparency reports within the wider sector.[23] The OECD, Christchurch Call, Action Coalition on Meaningful Transparency, Tech Against Terrorism, and GIFCT all have ongoing programs designed to improve TVEC transparency reporting. There are also a multitude

[19] Joe Whittaker, "Recommendation Algorithms and Extremist Content: A Review of Empirical Evidence," Global Internet Forum to Counter Terrorism, July, 2022, https://gifct.org/wp-content/uploads/2022/07/GIFCT-22WG-TR-Empirical-1.1.pdf.

[20] The Christchurch Call to Action, "Algorithms and Positive Interventions," accessed July 16, 2023, https://www.christchurchcall.com/our-work/algorithms-and-positive-interventions/.

[21] The Christchurch Call to Action, "Transparency and Reporting Workplan 2021," accessed July 3, 2023, https://www.christchurchcall.com/assets/Documents/Christchurch-Call-Transparency-Work-Plan.pdf.

[22] "2022 GIFCT Annual Report," GIFCT.com, accessed August 7, 2023, https://gifct.org/wp-content/uploads/2022/12/GIFCT-Annual-Report-2022.pdf.

[23] "The Definitions and Principles Framework Project," GIFCT.com, accessed August 2, 2023, https://def-frameworks.gifct.org/.

of ongoing efforts to improve transparency reporting across parallel trust and safety issues such as mis-/disinformation and child sexual abuse material (CSAM).

To date, no organization or framework has emerged as the clear leader, and so – in spite of the efforts of these and various other organizations – meaningful transparency is still more an abstract goal to be attained than an operational framework to be implemented.

Sectoral Perspectives: Barriers and Risks to Achieving Meaningful Transparency

Transparency has emerged as a critical issue in the context of tech sector companies and social platforms, given their substantial impact on society. This section explores sectoral barriers and risks associated with achieving adequate transparency. It examines critical stakeholders, including tech sector providers, governments, and civil society, analyzing their challenges in promoting transparency. The findings contribute to the ongoing discourse on transparency and inform strategies for achieving a more accountable and responsible tech platform ecosystem.

Tech Sector

Challenges faced by the tech sector in promoting transparency: Tech platforms each have their own perspectives on transparency. They all face challenges in balancing transparency with content removal, protecting user privacy, supporting the security of their platforms, and preservation of proprietary algorithms. Overall, the tech sector believes transparency measures should be implemented while not compromising user privacy or revealing sensitive business information.[24]

Tech companies face challenges in promoting transparency despite using tools such as intrinsic content filtering. Challenges include the need to account for their actions in public reports and the potential threat of increased regulation in response to socially harmful content, which could incentivize them to improve content moderation. Additionally, legislation has already been proposed to increase the transparency of tech sector algorithms. With respect to algorithms, global multi-stakeholder initiatives like the Christchurch Call have collaborated with tech firms to establish work plans on algorithms and positive interventions to scale up researcher access without compromising privacy, security, or proprietary information.[25] These challenges highlight the central role of transparency within effective tech sector regulation[26] in whatever form such regulation takes.

Opaque tech platforms are seen as immensely impactful upon the societies in which they operate, and

[24] Shoshana Zuboff, The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power (London: Profile Books, 2019).

[25] The Christchurch Call, "Algorithms and Positive Interventions."

[26] Dimitri Ognibene et al., "Challenging social media threats using collective well-being-aware recommendation algorithms and an educational virtual companion," Frontiers, January 9, 2023, https://doi.org/10.3389/frai.2022.654930.

tech platforms increasingly recognize the role they play offline as much as online. The algorithms behind tech platforms may contribute to the rise of extremism and online radicalization.[27] Online influences are portrayed as the main drivers of the spread and adoption of extremist ideologies, which often contain an element of collective grievance and subsequent acts of violence.[28] Violent extremists and the morosely curious have exploited tech platforms to ensure haunting brutality stays online. Platforms are often faced with adversarial instances of people trying to circumvent their rules. Following the 2019 Christchurch Mosque attacks, tech companies came together through GIFCT (founded as a tech-run multilateral initiative in 2017 but transformed into an independent non-profit following the Christchurch attacks). GIFCT was devised to counter potential technology abuses, radicalizing events, and attacks. Adapting a technology used for years to block videos of CSAM, the GIFCT Hash-Sharing Database[29] was created to assist member companies in rapidly detecting and sharing signals related to defined terrorist and violent extremist material to take action on their respective platforms.

Tech sector perspectives on transparency: Tech companies recognize that transparency is a central consideration for both consumers and regulators for providing insight into the nature of actions occurring within any platform.[30] As industry transparency is a requirement across most regulated industries, there is (to an extent) a consumer expectation of transparency that has not yet been fully realized. Conversely, there is seemingly a lack of binding legislation to address the responsibility of the tech sector and tech platforms to protect democracy at the private, enterprise, and societal levels. Platforms can minimize and mitigate risks for societies through responsible action in the fields of human rights, education, and transparency of their algorithmic decisions.

Government

Regulatory measures for promoting transparency on online platforms: Possible regulatory measures for promoting transparency in the tech sector are being examined by governments and tech companies. For example, the EU Digital Services Act includes transparency requirements for tech platforms.[31] Proposed U.S. legislation, such as the Platform Accountability and Transparency Act (PATA) and the Filter Bubble Transparency Act, aim to increase transparency by increasing access to tech firms' internal data.[32] There is a growing likelihood of government-implemented regulation in response

[27] John Herrman, "How TikTok Is Rewriting the World," The New York Times, March 10, 2019, https://www.nytimes.com/2019/03/10/style/what-is-tik-tok.html.

[28] Jens F. Binder and Jonathan Kenyon, "Terrorism and the internet: How dangerous is online radicalization?," Front Psycholgy, October 13, 2022, DOI: 10.3389/fpsyg.2022.997390.

[29] "GIFCT's Hash-Sharing Database," GIFCT.com, accessed August 7, 2023, https://gifct.org/hsdb/.

[30] Sprout Social, "#BrandsGetReal: Social media & the evolution of transparency," accessed July 16, 2023, https://sproutsocial.com/insights/data/social-media-transparency/.

[31] EU, "DSA."

[32] United States Senate, "Filter Bubble Transparency Act," accessed July 16, 2023, https://www.congress.gov/bill/117th-congress/senate-bill/2024/text; United States Senate, "Platform Accountability and Transparency Act," accessed August 7, 2023, https://www.congress.gov/bill/117th-congress/senate-bill/5339.

to socially harmful content on tech platforms. Some have suggested that a comprehensive system of regulation overseen by a government agency is needed to ensure transparency in tech platforms. Australia's eSafety Requirements, for example, refer to measures and regulations aimed at promoting online safety, particularly for children and vulnerable individuals.[33] The specific requirements may vary depending on the country or jurisdiction.[34] For example, in the context of the U.K.'s Online Safety Bill, requirements focus on ensuring a safe online environment, combating illegal content, and protecting users (predominantly children) from online abuse and harm.[35]

Ofcom's Planned Measures under the Online Safety Bill: Ofcom, the independent regulator for the communications industry in the U.K., will be appointed as the online safety regulator under the proposed Online Safety Bill.[36] Ofcom's role would involve implementing and enforcing the provisions of the bill to ensure online safety through the following measures:[37]

1. **Duty of Care:** The Online Safety Bill would likely place a duty of care on online platforms to ensure a safe online environment and protect their users from illegal or (in the case of child users) harmful content. Ofcom would oversee the implementation of this duty, including setting out specific expectations and standards for online platforms.
2. **Codes of Practice:** Ofcom is likely to develop and enforce codes of practice for online platforms. These codes would provide guidance on various aspects of online safety, such as tackling illegal content, minimizing the spread of harmful content, and protecting users from abuse and harassment.
3. **Reporting Mechanisms:** Ofcom would establish mechanisms for users to report harmful content and online abuse. Online platforms would be required to have effective and transparent reporting systems in place, allowing users to flag and report inappropriate or harmful content.
4. **Enforcement and Penalties:** Ofcom would have the authority to take enforcement actions against online platforms that fail to comply with the requirements set out in the Online Safety Bill. This could include fines, sanctions, and potentially even blocking access to non-compliant platforms in the U.K.

Regulatory challenges in promoting transparency: Regulators do, however, face challenges in promoting transparency in the tech sector. Governments and their constituent regulatory agencies are

[33] Government of Australia, "Basic Online Safety Expectations."

[34] Allistar Knott and Dino Pedreschi, "Transparency Mechanisms for Social Media Recommender Algorithms: From Proposals to Action," Global Partnership on Artificial Intelligence, November, 2022, https://gpai.ai/projects/responsible-ai/social-media-governance/transparency-mechanisms-for-social-media-recommender-algorithms.pdf.

[35] Ofcom, "Online safety: Ofcom's roadmap to regulation," Information for Industry, July 6, 2022, https://www.ofcom.org.uk/online-safety/information-for-industry/roadmap-to-regulation; Ofcom, "Update: How Ofcom is preparing to regulate online safety," Information for Industry, June 15, 2023, https://www.ofcom.org.uk/online-safety/information-for-industry/roadmap-to-regulation/0623-update.

[36] U.K. Government, "Online Safety Bill," Parliamentary Bills, July 321, 2023, https://bills.parliament.uk/bills/3137.

[37] Ofcom, "Online Safety"; Ofcom, "Update."

imposing transparency requirements to address contemporary challenges online, such as misinformation, TVEC, and hate speech.[38] While transparency is more beneficial in some instances, such as companies revealing information to regulators, it can also sometimes be seen as a violation of privacy. In addition, tech companies are contemporaneously facing the possibility of increased regulation in response to other problematic and harmful content on their respective tech platforms. For example, policymakers in the European Union and the United States are focusing on promoting transparency around algorithmic curation systems.[39] Overall, ongoing discussions and proposed legislation are aimed at increasing transparency in tech company algorithms. The other notable challenge is that some governments are inherently averse to attempts to regulate and control the wider tech industry. For example, U.S. trade negotiators have included Section 230 of the U.S. Communications and Decency Act into recent trade agreements to limit the extent to which international governments like the United Kingdom and Canada could maintain platform accountability.[40] However, Section 230 was intended to protect free speech, not platforms themselves. Overall, there are ongoing discussions regarding proposed and upcoming legislation aimed at increasing transparency across tech platforms.[41]

Regulatory bodies, such as the Federal Trade Commission (FTC) in the United States and the Information Commissioner's Office (ICO) in the United Kingdom, play a crucial role in ensuring transparency and protecting user rights.[42] They face the challenge of enforcing regulations that strike a balance between promoting transparency and fostering innovation.[43] Regulators aim to develop frameworks that hold providers accountable for their actions, protect user privacy, and prevent the spread of harmful rhetoric and offensive content.[44] Regulatory challenges in supporting transparency include providing adequate and concise disclosures to users, public reporting, and access to information. The New Zealand Christchurch, Charlottesville, Tree of Life, Tops Market, and other attacks, as well as the January 6th insurrection event at the U.S. Capitol, have highlighted the need for regulatory action and change to

[38] Mark MacCarthy, "Transparency is essential for effective social media regulation," Brookings, November 1, 2022, https://www.brookings.edu/articles/transparency-is-essential-for-effective-social-media-regulation/.

[39] European Commission, "European Centre for Algorithmic Transparency," accessed July 16, 2023, https://algorithmic-transparency.ec.europa.eu/index_en.

[40] Han-Wei Liu, "Exporting Freedom of Speech through Trade: The Global 'US Constitution First Amendment Moment' for Online Platform Liability," University of Oxford Business Law Blog, May 23, 2022, https://blogs.law.ox.ac.uk/business-law-blog/blog/2022/05/exporting-freedom-speech-through-trade-global-us-constitution-first.

[41] Melissa de Witte, "Four questions: Evelyn Douek on what Section 230 is and why it is misunderstood," Stanford News, October 7, 2022, https://news.stanford.edu/2022/10/07/four-questions-evelyn-douek-section-230-misunderstood/#:~:text=The%20main%20goal%20of%20Section%20230%20is%20not,to%20avoid%20the%20possibility%20of%20facing%20a%20lawsuit.

[42] Government of the United Kingdom, "The benefits and harms of algorithms: A shared perspective from the four digital regulators," Research and Analysis, September 23, 2022, https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/the-benefits-and-harms-of-algorithms-a-shared-perspective-from-the-four-digital-regulators.

[43] Larue Tone Hosmer, "The Ethics of Algorithmic Transparency," Business Ethics Quarterly 29, no. 3 (July 2019): 301–328.

[44] Jonathan A. Obar and Anne Oeldorf-Hirsch, "The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services," Information, Communication, & Society 23, no. 1 (2020): 128–147, DOI:10.1080/1369118X.2018.1486870.

address TVEC, hate group rhetoric, and subversive and radical content.[45]

Effective Regulation: With the advent of new and emerging technologies and tech companies, there is a growing demand for regulatory action, with some having suggested that a new era of tech platform regulation may indeed be imminent.[46] There are, however, trepidations about the efficacy of government regulation in controlling online content, particularly when it comes to issues such as hate speech, terrorism, and other crimes.[47]

Achieving transparency in the tech sector is seen as key to maintaining democratic values, as a small number of companies have enormous control over what a broad swath of the diaspora sees, reads, and hears.[48] However, there is agreement among tech companies, lawmakers, and regulatory bodies that transparency is necessary for upholding these values. Nevertheless, achieving transparency in the industry is a complex issue with challenges spanning data management and authenticity. Regulators face challenges in promoting transparency due to a constantly evolving compliance and reporting landscape. Maintaining stakeholder trust and meeting the expectations of these stakeholders is crucial.[49] Established instances of ideological violence, catalyzed by the resultant actions taken by specific platforms, imply that we may be at a turning point regarding how business leaders and government bodies will approach tech platforms and social media regulation in the near future.

The EU Digital Services Act includes transparency requirements and on-site inspections like those in the banking industry.[50] Proposed regulatory strategies include requiring tech platform disclosures regarding advertisements, promoting competition, and complex platform algorithms. Some tech companies have implemented more effective self-governance and unilaterally release transparency reports outside of a robust regulatory agenda. However, the lack of a multilateral, self–governing effort from all

[45] Centre for Resilient and Inclusive Societies, "Inquiry Into Extremist Movements and Radicalism in Australia," February, 2021, https://new.parliament.vic.gov.au/49f2e1/contentassets/eb5e56d74963429da457fffd03844368/attachment-documents/017_attach4_final-avert-research-network_redacted.pdf.

[46] Yogesh K. Dwivedi et al., "Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," International Journal of Information Management 66, (October 2022), https://doi.org/10.1016/j.ijinfomgt.2022.102542.

[47] Brian Fishman, "Dual-use regulation: Managing hate and terrorism online before and after Section 230 reform," Brookings, March 14, 2023, https://www.brookings.edu/articles/dual-use-regulation-managing-hate-and-terrorism-online-before-and-after-section-230-reform/; Martha Minow and Newton Minow, "Social Media Companies Should Pursue Serious Self-Supervision – Soon: Response to Professors Douek and Kadri," Harvard Law Review 136, no. 8 (June 2023), https://harvardlawreview.org/forum/vol-136/social-media-companies-should-pursue-serious-self-supervision-soon-response-to-professors-douek-and-kadri/; Beatriz B. Arcila and Rachel Griffin, "Social media platforms and challenges for democracy, rule of law and fundamental rights," EU Policy Department for Citizens' Rights and Constitutional Affairs, April, 2023, https://www.europarl.europa.eu/RegData/etudes/STUD/2023/743400/IPOL_STU(2023)743400_EN.pdf.

[48] Paul Waters, "Social Media Transparency is Key for Our Democracy," Democracy Fund, August 11, 2020, https://democracyfund.org/idea/social-media-transparency-is-key-for-our-democracy/; Joe Goldman, "Tackling Democracy's Cybersecurity Problem Requires Collective Action," Democracy Fund, August 17, 2021, https://democracyfund.org/idea/tackling-democracys-cybersecurity-problem-requires-collective-action/.

[49] KPMG Advisory, "Transparency and Reporting: 2023 Regulatory Challenges: Insights on reporting, market structure, and protections and control," accessed July 16, 2023, https://advisory.kpmg.us/articles/2022/ten-key-regulatory-challenges-2023-transparency-reporting.html; KPMG Advisory, "Ten Key Regulatory Challenges of 2023: Strengthening 'weak links'," accessed July 16, 2023, https://advisory.kpmg.us/articles/2022/ten-key-financial-services-regulatory-challenges-2023.html.

[50] EU, "DSA."

tech companies means there remains no conclusive answer as to the best regulatory approach for promoting transparency in the tech sector.

Civil Society Organizations (CSOs)

Civil society perspectives on transparency for tech companies: CSOs are non-governmental organizations that include proactive community-based groups, human rights defenders, activists, advocacy groups, academics, and humanitarian actors organized on local, national, and international levels. CSOs and their respective researchers can work (and have worked) with industry self-regulatory frameworks to assist in defining access parameters for transparency in tech sector platforms, but ultimately only government or intergovernmental agencies can effectively enforce them. Tech platforms must cooperate with local actors to fight issues like hate, extremism, election-related disinformation, misinformation, and the like, and international, national, and regional CSOs must encourage and support these efforts. CSOs have a critical role to play in the building of a culture of trust, safety, and integrity within the tech sectors and can act as watchdogs by exposing corruption and promoting good governance in transparency awareness, policy making, and advancing societal wellbeing.[51] Overall, CSOs can foster consistency, accountability, and balance toward technology platforms and their expansion, growth, and diffusion.

CSO's role in achieving transparency: Civil society plays a crucial role in achieving transparency. It is seen as an important agent for promoting good governance, including transparency, effectiveness, openness, responsiveness, and accountability. CSOs traditionally enhance transparency and accountability in the management of public resources.[52] Additionally, civil society, along with the media, has a critical role to play in building a culture of integrity and promoting good governance and accountability. The role of civil society could enhance more uniformity across states and cultures in maintaining their boundaries of openness toward acceptable transparency norms, dialogue, and intergovernmental cooperation in content distribution across geographic diversity.

Civil society initiatives for tech sector transparency and democracy: Civil society initiatives are also vital for promoting transparency in the tech sector because of the significant impact the platforms have on democracy, as civil society can play a central role in balancing the power of the state. Collaborative partnerships between organizations and tech companies can improve the value of public debate and the integrity of democratic practices.[53] The productive engagement of civil society with private sector

[51] Fishman, "Dual-use regulation"; Intergovernmental Council of the International Programme for the Development of Communication, "Internet transparency: A guide to applying the UNESCO principles. Follow-up to the report 'Letting the Sun Shine In: Transparency and Accountability in the Digital Age'," UNESCO, November 24/25, 2022, https://unesdoc.unesco.org/ark:/48223/pf0000383308.

[52] Government of Canada, "Implementation Plan: Canada's Civil Society Partnerships Policy – High level narrative update on progress 2022," accessed July 16, 2023, https://www.international.gc.ca/world-monde/issues_development-enjeux_developpement/priorities-priorites/civil_policy-plan-politique_civile.aspx?lang=eng; PTF Europe, "Partnership for Transparency Strategy 2023-2026," Partnership for Transparency Fund Europe, accessed July 16, 2023, https://www.ptfeurope.org/.

[53] Avril Haines, "Summit for Democracy 2023," Office of the Director of National Intelligence, April 11, 2023, https://www.dni.gov/index.php/news-room/speeches-interviews/speeches-interviews-2023/item/2374-summit-for-democracy-2023.

technology companies and government partners can help shape and produce improved rights-based practices that factor in human rights and fundamental freedoms. Future initiatives might also focus on improving the digital security practices of civil society, human rights defenders, and media, as well as increasing their engagement on internet authority issues. Stronger protections are also becoming increasingly necessary to ensure safe and sustainable engagement with civil society actors who face growing risks of reprisals and retaliation when speaking out about human rights violations.

Academics and researchers also provide valuable insights into the technical, social, and ethical aspects of transparency. Their studies to understand the impact of algorithms, biases, and privacy practices can provide valuable insight, but should not be relied upon solely as their funding and access to data is limited. Their perspectives contribute to developing best practices, evaluating existing transparency measures, and proposing new approaches to enhance transparency.[54] These organizations can use monitoring, reporting, and advocating for change to ensure improvements in policy.[55]

Additional civil society institutions can also play a crucial role in assisting the promotion of social media accountability by monitoring, reporting, and stimulating good governance. Traditional media can further strengthen social accountability by assisting and promoting the work of CSOs.

Stakeholder perspectives on the ideal end state for transparency: Stakeholders expect transparency from organizations, and it is considered an essential part of regulatory feedback rings. Stakeholders are defined as those who may be affected by or influence an effort.

[54] Samantha Lai, Naomi Shiffman, and Alicia Wanless, "Operational Reporting By Online Services: A Proposed Framework," Carnegie Endowment for International Peace, May 18, 2023, https://carnegieendowment.org/2023/05/18/operational-reporting-by-online-services-proposed-framework-pub-89776; Zeynep Tufekci, "Engineering the public: Big data, surveillance and computational politics, First Monday 19, no. 7 (July 7, 2014), https://doi.org/10.5210/fm.v19i7.4901; The Christchurch Call, "Algorithms and Positive Interventions."

[55] Thomas Häusler, "Civil society, the media and the Internet: Changing roles and challenging authorities in digital political communication ecologies," Information, Communication & Society 24, no. 9 (2021): 1265–1282, https://doi.org/10.1080/1369118X.2019.1697338; Ian Davis, "The Role of Civil Society and the Media" in Building Integrity and Reducing Corruption in Defence: A Compendium of Best Practices," ed. Todor Tagarev (Geneva: Geneva Centre for the Democratic Control of Armed Forces (DCAF), 2010): 261–280; "Department of Global Communications Approves 5 Civil Society Organizations for Association," United Nations, Meetings Coverage and Press Releases, July 14, 2020, https://press.un.org/en/2020/pi2287.doc.htm.

Ideal End State - Pol.is Survey Results

The pol.is survey revealed insights into TWG and colleagues' views of a desired end state for meaningful transparency.[56] The survey asked the participants *What is the main goal / primary outcome that tech transparency will help civil society, governments, and industry achieve?* The intention was to identify areas of consensus and note diverging views. The pol.is tool also tracked how groups of respondents vote, offering a window into how sets of responses aligned. For greater detail on responses, see Annex 1.

Overall, the survey revealed strong consensus around the need for agreement upon standards and definitions, including the whole of the multi-stakeholder community. Along the same lines, survey participants agreed that content moderation reporting required metrics related to "prevalence" and not only raw numbers. The strongest level of divergence emerged with regard to the impact of transparency reporting on addressing TVEC and the extent to which organizations' work changed in response to transparency reporting. Further differences emerged related to the overall value and effectiveness of transparency reporting, with a small number of respondents agreeing that meaningful transparency is a desirable end state, even if it has only a limited effect on countering TVEC.

Four different groups of responses emerged from the pol.is results.[57] The majority of respondents voted similarly, but three other, smaller groups held similarly divergent views (see Annex 1 for a full list of statements and votes by grouping).

- Group A respondents advocated for greater and more detailed transparency, voting more often for transparency around decision-making processes for content removal than more data about the volume of removals. They also agreed governments need to be open about their actions online, and found that openness and transparency, especially about government data and information requests to industry, help protect everyone's freedom and privacy.
- Group B diverged from the rest of the respondents in a few areas. Notably, this group was the most supportive of GIFCT's model of communication, but differed from the majority of respondents by disagreeing with the idea that openness and transparency about government data and information requests to industry help to protect everyone's freedom and privacy.
- Group C's voting pattern provides insight into how some working in this space see the relationship between industry and government responsibilities. Unlike the majority of respondents, this group found that transparency related to TVEC should only focus on industry, not government, and disagreed with the statement that "openness and transparency, especially about government data and information requests to industry, help protect everyone's freedom and privacy." Further, they disagreed with the need for transparency around decision-making processes or for industry to be required to maintain in-house counter terrorism experts.
- Group D, which represented a small number of participants, differed from the majority in how they saw the role that transparency plays in supporting the information environment. They agreed that meaningful transparency is a desirable end goal, even if it only has incremental value addressing TVEC. However, they noted that transparency reduces information asymmetry that leads to a power imbalance between platforms/governments and users. Interestingly, this group both disagreed with the idea that industry should be required to have in-house experts and did not agree that GIFCT has spurred a dynamic model of communication.

[56] The Pol.is will remain active for six months, enabling GIFCT to track emerging and diverging views on the intended outcomes of transparency reporting following the release of the report. The Pol.is can be accessed at: https://pol.is/5fmrduaj8r.

[57] As an operating feature, Pol.is provides preliminary analysis of survey results, which formed the starting point for this analysis.

Achieving an Ideal End State for Meaningful Transparency

Continuing operations in the absence of clearly defined goals limits the efficacy of interventions designed to address TVEC and the measurement tools used to track progress. As most of the pol.is survey respondents agreed, an ideal end state should not be a fantasy, but a model that balances security, individual privacy, and free expression, all founded on openness and transparency.

Increasing transparency: Increasing aggregate transparency is an effective means to advance public trust and address issues like privacy, content moderation, hate, terrorism, and bias. As the pol.is survey demonstrated, 100% of respondents indicated that meaningful transparency is good business and enables industry, government, and civil society to build healthier and safer online spaces. Increasing transparency on tech platforms is seen as an improvement for public trust and addresses issues such as privacy, content moderation, and bias.

Promoting examples of successful transparency initiatives: Transparency initiatives can improve content and increase diversity, equity, and inclusion in companies. Transparency tools such as freedom of information, transparent budgeting, and asset declarations can also have an impact on controlling corruption and promoting equity, and can help to drive powerful transparency initiatives by various actors within tech, government, and CSO sectors.

Identifying strategies for accountability: Effective strategies for tech platform accountability include centering human rights and looking globally to how platforms have failed marginalized communities and protected powerful interests. However, there is no central authority governing how tech platforms engage controversial content such as political ads, hate speech, terrorist content, conspiracy theories, and incitement to violence – all elements that evolve (and have evolved) with tech industry growth. This has left the identification of accountability strategies to individual states and platforms.

Establishing corporate social responsibility in tech sector transparency for TVEC: Corporate social responsibility has expanded to tech platforms, allowing for two-way communication between companies and end users. Transparency is a key aspect of corporate social responsibility, with consumers expecting companies to be open about all their practices, particularly those relevant to user safety and community security. Companies are under pressure to move beyond traditional marketing and prioritize corporate social responsibility,[58] ideally predicated on a dialogic transparency model.

Conclusions

To achieve a "healthy information environment," industry, government, and civil society partners must recognize the need to limit the spread and impact of TVEC and maintain a strong balance between

[58] Yeyi Liu et al., "Building a competitive advantage based on transparency: When and why does transparency matter for corporate social responsibility," Business Horizons 66, no. 4 (2023): 517–527.

rights and freedoms, security and accountability. As the report notes, the current shift to meaningful transparency is a solid step toward this goal, but more must be done to articulate a clear way forward supported by consistent, meaningful, and open reporting from industry and government.

A key factor for determining an ideal end state for transparency and the information environment comes from articulating the intended outcomes and understanding the acceptable level of risk that society is willing to collectively tolerate. Transparency reporting itself is not risk free.[59] Civil society TWG members especially noted the potential risks for users about the sharing of private and/or personal information and views. While public transparency reporting is likely to remain at a high level, reporting to governments and regulators might be made available at a level that could be de-anonymized, putting already marginalized communities at further risk. Further, it has long been recognized industry focuses its efforts on what is measured and reported.[60] To determine where efforts to improve and align transparency reporting should be directed, partners need to move beyond what is often perceived as crisis management and spell out their intended medium and long-term goals.

As a primary tool to track progress toward this desired end state, transparency reporting needs to enable industry and government partners to measure the efficacy of their interventions designed to safeguard rights, protect users, and provide civil society partners the knowledge to hold both to account for their commitments. To that end, the TWG articulated through its discussions and in the pol.is survey that transparency reporting should be based around trust, good faith, shared knowledge, outcomes, and interoperability and comparability between industry and government reporting structures. While current transparency work is moving in that direction, *meaningful* transparency requires a shared recognition of goals beyond simple "ticking the box" exercises.[61]

Trust, good faith, shared knowledge, and outcomes: Civil society colleagues in the TWG stressed the importance of good faith and trust in terms of the quality of information provided through transparency reporting and the validity and openness of the reporting processes themselves. They expressed concerns with the sometimes performative elements of transparency reporting against TVEC, such as reports noting the volume of actions as opposed to their potential impact. They noted meaningful transparency is an essential element to tracking progress toward intended outcomes.

It was also recognized that trust is central for users, ensuring that their private communications or personal views are protected, and that aggregate data presented in transparency reports to government regulators cannot be de-anonymized. Outside traditional transparency reporting, members flagged concerns that continue to emerge with the risks of individual users being exposed through data provided to governments or law enforcement. From that view, ensuring that society maintains a level of trust in industry actions means that reporting needs to continue to include industry reporting on

[59] Fishman, "Dual-use regulation."

[60] Rebecca Schultz et al., "Better Indicators for Better Regulation: The OECD iREG Experience," La Mejora de La Regulación 907 (2019).

[61] Evelyn Douek, "The Rise of Content Cartels," Knight First Amendment Institute, February 11, 2020, https://knightcolumbia.org/content/the-rise-of-content-cartels.

government requests to share knowledge of actions intended to counter TVEC among all stakeholders. Industry representatives mirrored that perspective, arguing that trust is a fundamental goal, not just for individual companies but between people that use services and platforms so they can understand (through reporting) exactly how companies are trying to enforce their values. One representative noted that this trust is the basis for their accountability to users. From that view, transparency reporting should be motivated by an acknowledgement that building communities is an important responsibility. Using the good faith approach, transparency reporting can foster valuable feedback about improving policies and systems that can aid in their development.

Industry representatives mirrored that perspective, arguing that trust is a fundamental goal, not just for individual companies but between people that use services and platforms so they can understand (through reporting) exactly how companies are trying to enforce their values. One representative noted that this trust is the basis for their accountability to users. From that view, transparency reporting should be motivated by an acknowledgement that building communities is an important responsibility. Using the good faith approach, transparency reporting can foster valuable feedback about improving policies and systems that can aid in their development.

Comparability, not standardization: An ongoing challenge that was raised in the group's discussions related to the need both for a recognition of the challenges of standardizing reporting across the different companies and industries represented in the online environment, as well as the risks of different countries implementing requirements that differed from each other.

Beyond numbers, members flagged the importance of being able to compare the impact of platform responses to enable them to understand what is and is not working. Civil society partners stressed that an ideal end state needs to ensure that transparency reporting includes a sufficient breadth in the number of companies providing reports on TVEC, and in the depth of their reporting (including noting how threats break down by ideology, geography, etc.). Further, considering that TVEC content moves between distinct platforms and even between industry sectors (e.g., social media, video games, peer-to-peer messaging, etc.), ideal reporting should enable a macro-level understanding of shifts between online services, including changes over time. Not only would this provide insight into changing behavior patterns between targeted groups, but also could reveal elements of what is working for particular platforms and whether particular interventions in one space have instigated unintended or unanticipated risks and/or challenges in others.

Recommendations

To evolve the current reporting environment toward a safer, healthier, and more trustworthy information environment, transparency reporting should be provided in a manner that is more comparable across sectors and included the information necessary to limit the spread of TVEC content while protecting rights and freedoms.

1. **To prove effective at measuring progress toward a safe, healthy, and trustworthy information environment, transparency reporting needs to provide the information necessary to simultaneously:**

    a. Allow governments and industry to assess the impact of their interventions, and

    b. Enable regulators and civil society oversight to ensure that human rights, privacy, and competition rules are being met.

To that end we recommend:

1. GIFCT should encourage its partners and members to work collaboratively toward articulating a clear set of desired outcomes and to weigh the acceptable level of risk inherent with balancing human rights, freedom of expression, and limiting the spread of TVEC content.
2. Support this by launching an international citizens assembly[62] to surface international considerations and expectations for balancing risks and benefits related to combating TVEC in the information environment. This would provide a broader set of views that could enhance and challenge expert positions. For example, Canada's Canadian Citizens' Assemblies on Democratic Expression model released meaningful, citizen-generated advice on digital tech oversight designed to strengthen legislative decision-making for the digital ecosystem.[63] Using globally tested methods, like that of the recent Global Assembly on Climate Change,[64] and leveraging its extensive international civil society partners, GIFCT could spearhead a consultation on ideal goals, outcomes, and risk tolerance to help inform the community's ultimate goals.

2. **A consistent foundation for comparability in industry reporting would drive industry reporting toward interoperability which would:**

    a. Encourage the comparability of reporting data across organizations and industry sectors,

    b. Enable supportive civil society and government oversight, and

    c. Reduce the risk of official regulatory requirements setting unachievable or unrealistic barriers for reporting practices.

To that end we recommend:

1. GIFCT should work with its partners and members to enhance existing reporting templates

[62] Citizens' Assemblies, "Discover democracy that works"; Citizens' Assemblies, "UK Citizens' Assemblies."

[63] 3rd Canadian Citizens' Assembly, "Canadian Citizens' Assembly on Democratic Expression."

[64] Global Assembly Team, "Report of the 2021 Global Assembly."

to create an adaptable basic format/skeleton design of a transparency report that is focused on intended outcomes.

2. This work should build upon existing guides and frameworks like those provided by Tech Against Terrorism[65] or Susan Ness and Chris Riley's "Modularity" framework (as a form of multi-stakeholder, co-regulatory governance)[66] to enable governments to align their approaches across different legal structures.
3. Recognizing the challenges for smaller industry partners and the barriers to entry that changes to reporting can create, GIFCT should continue to support new entrants by building capacity and leveraging industry and government support. eSafety's Safety-by-Design initiative demonstrates the value in running industry-leading, capacity-building work alongside strict enforcement activities that could be emulated.[67]

3. Meaningful transparency reporting requires all stakeholders to evolve current reporting systems and standards to capture and share information in a way that enables:

a. Progress toward a safe, healthy, and trustworthy information environment, and

b. Comparability of findings across the tech sector.

To that end we recommend:

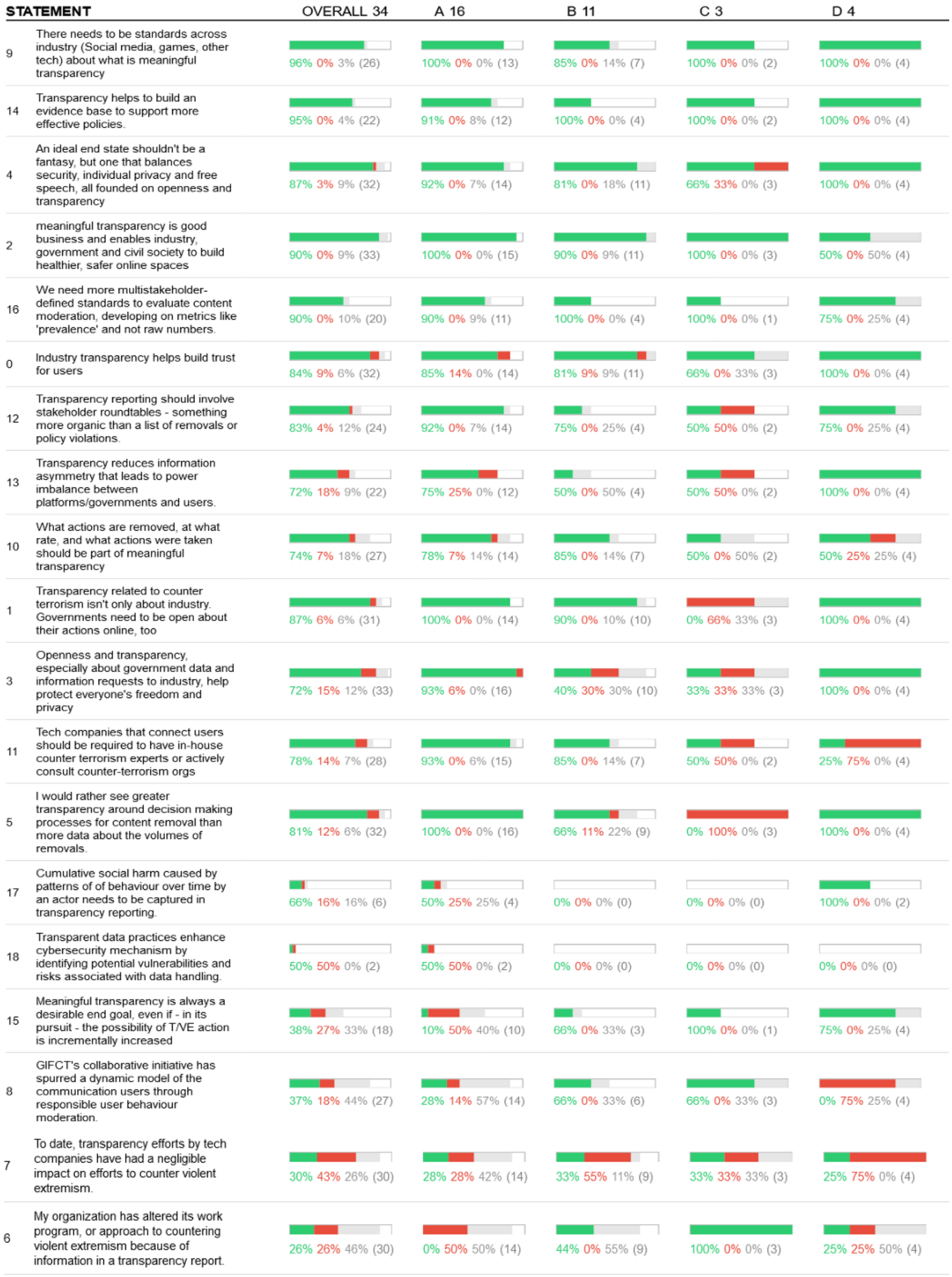
1. GIFCT should take a leadership role by requiring its members to demonstrate how they are enhancing their reporting practices to include reporting against previously recommended outcomes and to enable comparability of findings across sectors.

[65] Tech Against Terrorism, "Tech Company Transparency Reporting."

[66] Riley and Ness, "Modularity"; Riley, "A Module Playbook."

[67] Government of Australia, "Basic Online Safety Expectations."

Annex 1: Pol.is Statement (by group consensus)



STATEMENT		OVERALL 34	A 16	B 11	C 3	D 4
9	There needs to be standards across industry (Social media, games, other tech) about what is meaningful transparency	96% 0% 3% (26)	100% 0% 0% (13)	85% 0% 14% (7)	100% 0% 0% (2)	100% 0% 0% (4)
14	Transparency helps to build an evidence base to support more effective policies.	95% 0% 4% (22)	91% 0% 8% (12)	100% 0% 0% (4)	100% 0% 0% (2)	100% 0% 0% (4)
4	An ideal end state shouldn't be a fantasy, but one that balances security, individual privacy and free speech, all founded on openness and transparency	87% 3% 9% (32)	92% 0% 7% (14)	81% 0% 18% (11)	66% 33% 0% (3)	100% 0% 0% (4)
2	meaningful transparency is good business and enables industry, government and civil society to build healthier, safer online spaces	90% 0% 9% (33)	100% 0% 0% (15)	90% 0% 9% (11)	100% 0% 0% (3)	50% 0% 50% (4)
16	We need more multistakeholder-defined standards to evaluate content moderation, developing on metrics like 'prevalence' and not raw numbers.	90% 0% 10% (20)	90% 0% 9% (11)	100% 0% 0% (4)	100% 0% 0% (1)	75% 0% 25% (4)
0	Industry transparency helps build trust for users	84% 9% 6% (32)	85% 14% 0% (14)	81% 9% 9% (11)	66% 0% 33% (3)	100% 0% 0% (4)
12	Transparency reporting should involve stakeholder roundtables - something more organic than a list of removals or policy violations.	83% 4% 12% (24)	92% 0% 7% (14)	75% 0% 25% (4)	50% 50% 0% (2)	75% 0% 25% (4)
13	Transparency reduces information asymmetry that leads to power imbalance between platforms/governments and users.	72% 18% 9% (22)	75% 25% 0% (12)	50% 0% 50% (4)	50% 50% 0% (2)	100% 0% 0% (4)
10	What actions are removed, at what rate, and what actions were taken should be part of meaningful transparency	74% 7% 18% (27)	78% 7% 14% (14)	85% 0% 14% (7)	50% 0% 50% (2)	50% 25% 25% (4)
1	Transparency related to counter terrorism isn't only about industry. Governments need to be open about their actions online, too	87% 6% 6% (31)	100% 0% 0% (14)	90% 0% 10% (10)	0% 66% 33% (3)	100% 0% 0% (4)
3	Openness and transparency, especially about government data and information requests to industry, help protect everyone's freedom and privacy	72% 15% 12% (33)	93% 6% 0% (16)	40% 30% 30% (10)	33% 33% 33% (3)	100% 0% 0% (4)
11	Tech companies that connect users should be required to have in-house counter terrorism experts or actively consult counter-terrorism orgs	78% 14% 7% (28)	93% 0% 6% (15)	85% 0% 14% (7)	50% 50% 0% (2)	25% 75% 0% (4)
5	I would rather see greater transparency around decision making processes for content removal than more data about the volumes of removals.	81% 12% 6% (32)	100% 0% 0% (16)	66% 11% 22% (9)	0% 100% 0% (3)	100% 0% 0% (4)
17	Cumulative social harm caused by patterns of of behaviour over time by an actor needs to be captured in transparency reporting.	66% 16% 16% (6)	50% 25% 25% (4)	0% 0% 0% (0)	0% 0% 0% (0)	100% 0% 0% (2)
18	Transparent data practices enhance cybersecurity mechanism by identifying potential vulnerabilities and risks associated with data handling.	50% 50% 0% (2)	50% 50% 0% (2)	0% 0% 0% (0)	0% 0% 0% (0)	0% 0% 0% (0)
15	Meaningful transparency is always a desirable end goal, even if - in its pursuit - the possibility of T/VE action is incrementally increased	38% 27% 33% (18)	10% 50% 40% (10)	66% 0% 33% (3)	100% 0% 0% (1)	75% 0% 25% (4)
8	GIFCT's collaborative initiative has spurred a dynamic model of the communication users through responsible user behaviour moderation.	37% 18% 44% (27)	28% 14% 57% (14)	66% 0% 33% (6)	66% 0% 33% (3)	0% 75% 25% (4)
7	To date, transparency efforts by tech companies have had a negligible impact on efforts to counter violent extremism.	30% 43% 26% (30)	28% 28% 42% (14)	33% 55% 11% (9)	33% 33% 33% (3)	25% 75% 0% (4)
6	My organization has altered its work program, or approach to countering violent extremism because of information in a transparency report.	26% 26% 46% (30)	0% 50% 50% (14)	44% 0% 55% (9)	100% 0% 0% (3)	25% 25% 50% (4)

Copyright © Global Internet Forum to Counter Terrorism 2023

Recommended citation: Chris Beall, Martin Cocker, Dr. Abbee S. Corb, and Jonathon Deedman, Pathways to Meaningful, Transparency (Washington, D.C.: Global Internet Forum to Counter Terrorism, 2023), Year 3 Working Groups.

GIFCT is a 501(c)(3) non-profit organization and tech-led initiative with over 20 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**