GIFCT Transparency Working Group:
One-Year Review of Discussions

July 2021

Introduction

This review presents the primary questions as well as the larger themes and points raised by the GIFCT Transparency Working Group's multi-stakeholder members over the last year. The conversations and conclusions from this Group reflect how greater transparency can build trust among tech companies, the public, and other stakeholders.

Moving forward, this Group will continue to share information, explore different themes and map out knowledge gaps to help better understand transparency metrics. It will also identify additional forums and processes to help enable effective communication among stakeholders, whereby tech companies can explain the limitations and capabilities of technical tooling and help contextualize the concerns of civil society organizations (CSOs) and governments.

Primary Questions

This section outlines some of the "big questions" (arranged by sector) that have been raised by the Group's members:

From All Sectors:

- What is the ultimate objective of transparency reporting? Understanding this can allow the Working Group and tech companies to work backward to define what metrics are needed.
- What does "meaningful transparency" mean and what does it look like? This is a question that needs to be teased out in a multi-stakeholder setting.
- Which elements of transparency may be required by law and regulation in a given situation, and which are voluntary in nature? How can a patchwork of requirements be avoided?
- Which transparency reporting metrics can be compared or standardized and which should be considered incomparable?
- How can a wider range of tech companies be supported in developing their transparency reporting?

From Tech Companies:

- What are the policy and other questions that governments and civil society are seeking to answer using tech company transparency reporting? These can differ depending on socio-political and geographical context, but understanding the goals of other stakeholders can support an informed conversation on what information is relevant, what can/should be shared, and potential means of standardizing data across organizations.

From Governments:

- Understanding that tech companies all operate differently, both in terms of policy and services provided to users, what is a reasonable standard baseline that allows comparisons to be drawn across companies? Is this feasible and, if so, how can it be established?
- In order to understand existing partnerships, programs, and proactive work, where can governments go to find information on what tech companies are doing in the preventing and countering violent extremism (PVE/CVE) space?
- Which transparency reporting metrics can help governments understand the level of threat stemming from terrorist and violent extremist content?

From CSOs / NGOs / Academics:

- What are the legal, privacy, or human rights tradeoffs or benefits from increased transparency? What are the potential human rights implications of expanded and/or proactive counterterrorism mechanisms? How can transparency reporting help alleviate human rights concerns?
- How can transparency reports speak and give insight to geographic nuance/diversity around actions taken in relation to terrorist and violent extremist content? It is hard for researchers to know any geographic breakdowns in most transparency reports.
- What other transparency processes can be held up as examples for GIFCT or member companies as examples of good practice?

Discussion Points and Perspectives

Aspects of Transparency

The Purpose of Transparency:

The purpose of having a transparency report is to communicate to the wider public the policies and procedures a company has in place, how and to what extent those policies are enforced, and thereby more clearly speak to accountability and progress. Transparency also allows other stakeholders to understand the scope of the relevant problem, the actions being taken by a company to address it, and to evaluate their actions in an informed way.

The ultimate aim of the Working Group is to develop good practices and provide access to resources that facilitate greater transparency from all relevant stakeholders while respecting privacy and human rights. Bearing in mind the diversity of digital platforms and their approaches to transparency reporting, the Working Group should aim to sponsor research and curate resources that can be absorbed organically into a range of platforms, allowing for variance in the nature, size, and goals of different companies.

Audiences for Transparency:

Different audiences may have different transparency needs. For example, while transparency reports (i.e., macro-level data) may be useful for policymakers and researchers, for individual users it is transparency (as opposed to transparency reporting) that is more important. Transparency around policies and reasons for decisions in individual cases can help users make informed decisions about how to use the platform and ensure that appeal processes are meaningful. The appropriate level of detail may depend on a range of factors, including privacy, legal, and other considerations.

Meaningful Transparency:

Because technology evolves quickly, it is hard to know what "meaningful transparency" is – both for governments and for tech companies. Without knowing what data is meaningful, it is hard to speak to the process. As a result, there is a push to have at least some baseline metrics that can be compared across companies. There is also often a need to pair data with qualitative explanations in order to give the metrics meaning.

Transparency at What Cost?

There is a need to ensure that transparency is not at the cost of human rights considerations, such as privacy, security, or freedom of expression. For instance, some concerns were raised over the Christchurch Call to Action's goal to "eliminate all terrorist and violent extremist content online." Is that ever achievable and, if so, at what cost? Will it result in over-censorship? In order to understand some of the potential tradeoffs between transparency and privacy, there needs to be better synergy with other GIFCT working groups, particularly with the Content-Sharing Algorithms, Processes, and Positive Interventions Working Group (CAPPI) on algorithmic transparency questions.

NGO and government participants recognized that companies are private entities that have to make their own decisions about their platform transparency based on a range of factors unique to each platform, such as size, resources, and safeguarding user privacy. Emerging legal requirements may also influence companies' decisions.

Means of Transparency:

Participants recognized that there are various mechanisms for transparency and knowledge-sharing. It is important to understand the difference between the focus and metrics included in a transparency report compared to other means of transparency (such as tweets, blog/media posts, and discussions at wider forums).

Sector-Specific Considerations

Diversity in Tech Companies and Platforms:

The online space is not just social media, and care must be taken to ensure that when talking about transparency for tech companies the conversation is not only about companies whose primary content is user-generated. Different companies have different purposes, capacities, services, and tools.

In the context of transparency, there is a recognition that if platforms could be compared to one another in terms of their metrics, this could facilitate academic research, regulation, and other stakeholder interests that concern multiple platforms. Most transparency reports do not always include the same or comparable metrics. However, tech companies point out that parity does not recognize the diversity in how tech platforms are set up, how the platform is utilized, what their policies dictate and how they are subsequently enforced. The Internet and platforms are not homogenous in function, form, or purpose. Transparency in definitions of terrorism and violent extremism for any entity that has a framework for identifying and removing content is key. What elements of transparency reporting, policies, and metrics can be compared or standardized and what should be considered incomparable?

Difficulty in Developing Transparency Reports:

It is important for external stakeholders to understand the difficulties in assembling an initial transparency report, especially for small or new companies. This is often a huge internal effort for a company, particularly with the knowledge that this effort will be expected at least annually thereafter. Any data given in a transparency report needs to be replicable and accurate. The teams putting together the transparency report will include engineering and data support, as well as policy and legal teams for reviewing.

Some smaller companies would welcome a certain level of agreed-upon minimum framing for what should/could be included in a transparency report to lean into as guidance. This would likely not have a granular focus on terrorist and/or violent extremist content in a first instance, as this may not be a high-frequency issue on their platform, or the platform may need to prioritize transparency about legal and government requests, or other categories of content that violates their rules such as child sexual abuse material. For instance, LinkedIn described that their first transparency reporting had a legal focus for compliance, specifically targeting government requests for data and removals.

Small and mid-sized companies with less internal subject matter expertise and limited staffing can learn from the experience of other companies. Size and other constraints also affect the frequency at which a company can produce transparency reports and the granularity of that reporting. Some larger companies, like YouTube and Facebook, have published quarterly reports, but most companies have annual reports.

Larger companies expressed that their aim with transparency reporting is to give top metrics to ensure there is a basic understanding of how policies were applied and ensure there are some levels of interaction available with metrics. Country and regional breakdowns of removals were difficult for most companies to produce (only YouTube reported having such nuanced metrics), though government removal and information requests were often by country.

It is also difficult to calculate the "amount of content removed" when what defines a piece of content can be contested or ambiguous. Different types of content removal can occur: a comment, a solitary piece of content, or an entire account along with all of its content. This may also evolve as a platform builds capability. Companies note nuance is lost in reporting metrics.

Government Transparency

While governments in the Working Group mentioned that their countries worked to provide some transparency, there is also a general recognition that governments are not (by design) always necessarily transparent about their own efforts in countering terrorism and violent extremism. There are some levels of mandated transparency (including through freedom of information legislation), but there is no singular equivalent to an annual transparency report. That being said, as some governments appear to be shifting to become more proactive in regulating and reporting illegal content, there is also scope for better information on similar steps taken with legal but problematic content. Some governments are also working to develop and improve legal frameworks for their own approach to transparency.

The Group noted that it would be useful to have examples of good government transparency practices to point to above and beyond terrorist designation lists. Several participants have also emphasized the need for governments to seek to provide information about the requests or referrals they make for content/account moderation or removal. In relation to this, tech companies also mentioned they are not always certain about how transparent they can be regarding government requests when governments themselves are not. Government participants mentioned that coordinating and formalizing communication between government personnel and tech companies are important next steps in ensuring better government transparency. Civil society participants also pointed to disparities across the world in government transparency and that some governments do, in fact, forcibly stop companies from divulging certain information.

Supporting Frameworks

To Regulate or Not?

Some governments have already developed (or are developing) regulations around terrorist and violent extremist content removal, including proposed transparency reporting requirements. The sanctions contemplated for non-compliance may include fines, "shaming" by listing non-compliant entities, and (in some instances) penalties under criminal law. Working Group participants noted that it would be helpful to avoid the development of conflicting requirements across different jurisdictions.

Algorithmic Transparency:

Working Group participants expressed interest in exploring this challenge further. Algorithmic transparency refers here to processes by which tech companies might share access to data and information surrounding algorithmic outcomes thought to be exploited by violent extremists or linked to radicalization processes. This is a unique challenge in the transparency space as these algorithms are often closely guarded, proprietary elements of a tech company's business model. Overcoming barriers to accessing information in this space is important not only to transparency reporting, but also to academic research aiming to better understand the relationship between algorithms, violent extremism, and radicalization. As stated above, questions on algorithmic transparency will demand increased synergy with at least the CAPPI Working Group, if not others.

Where Support is Most Needed:

The Group talked widely about the need to support mentorship for smaller companies – like the GIFCT membership process and the Tech Against Terrorism mentorship process. Recognizing that the same expectations for all companies are unreasonable, there is an opportunity to support the development and collation of guidelines, good practices, and case studies. The Group will undertake a project in the second half of 2021 to bring some of these materials together.

Definitions and Concerns of Scope Creep:

Participants pointed out the need to carefully parse out what scope creep is. GIFCT work is not about "objectionable" or wider "harmful" content. It is about definable terrorist and violent extremist content. There are cultural sensitivities and no agreed definitions for "extremism" or "radicalized" content.

Recommendations for GIFCT

- In addition to supporting tech companies in developing their own transparency reports via the GIFCT membership process, whereby companies are supported through Tech Against Terrorism's Mentorship Program, GIFCT is also responsible for producing its own annual transparency report.
- The Transparency Working Group has shared with GIFCT recommendations to be considered for its upcoming transparency report (expected July 2021) and beyond.
- The Group has also discussed the need for enhancing the reach and widening the channels for dissemination of GIFCT's transparency reporting to help stimulate engagement with diverse demographics.

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