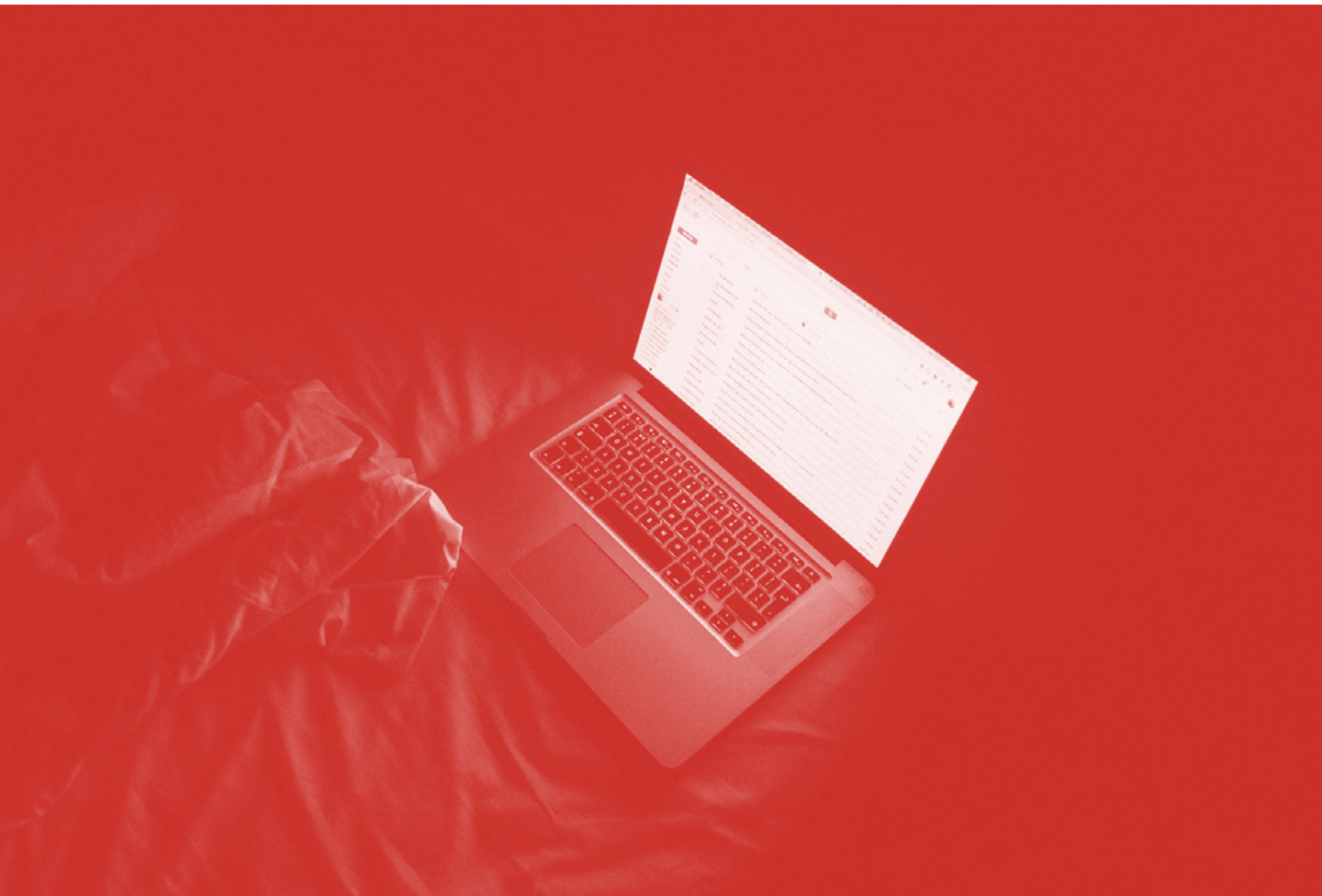
Texas National Security Review: Volume 2, Issue 2 (February 2019)
Print: ISSN 2576-1021 Online: ISSN 2576-1153

CROSSROADS: COUNTER-TERRORISM AND THE INTERNET

Brian Fishman



Brian Fishman, who leads the effort against terrorist and hate organizations at Facebook, argues that counter-terrorism researchers need to tailor their recommendations to the corporate policymakers inside tech companies who want to do far more than the bare minimum.

Public policy is traditionally thought of as the work of governments, however, private actors — including universities, health care providers, and a range of other private infrastructure operators — have long played important roles in shaping both society and national security. And while these institutions have typically operated under regulatory frameworks that set basic operating standards and compel information sharing with governments, they also make important choices on their own that affect millions of lives. Now, tech companies are counted alongside these institutions, but with a scope that is far wider — spanning the globe and crossing innumerable governmental jurisdictions — even if it is effectively virtual and doesn't involve driving specific healthcare decisions or determining security at a particular power plant. Such dynamics raise important questions both for how governments should interact with tech companies to set behavioral guidelines as well as for the companies themselves, which will inevitably determine how to manage social challenges outside of a strict regulatory framework. One of the most important of these policy areas is counter-terrorism.

Private actors have long taken part in counter-terrorism efforts: Banks, critical infrastructure operators, and airlines are important elements in societal efforts to protect against terrorism. But terrorist use of the Internet has brought an entirely new class of private actors to the forefront of the fight. Social media companies, both individually and in concert with one another, have developed robust operations to prevent terrorists from abusing their platforms. Like all counter-terrorism programs, these efforts are imperfect, but they represent a significant new component of the societal response to terrorist violence. As the head of Facebook's effort to counter abuse by terrorists and hate organizations, I have a unique vantage point on the intersection of social media and counter-terrorism, and the following essay, though it does not argue for Facebook's position on any particular issue, certainly reflects my experience at the company.

For the most part, tech companies have voluntarily initiated their counter-terrorism efforts. Nonetheless, there is an ongoing debate about what governments should require of private companies when it comes to matters of counter-terrorism. But focusing on regulation as the primary mode through which society can address terrorist activity online is misplaced. Indeed, the singular focus on government as the only actor in counter-terrorism operations online is outdated. Many tech companies actively counter terrorists online — and the effects of that work are almost certainly broader and more important to overall online counter-terrorism efforts than anything required by government regulation. The question of whether or not governments should require tech companies to conduct counter-terrorism operations is, of course, politically important. However, the voluntary efforts made by these companies are likely to have a far greater impact on addressing the problem of terrorist exploitation of the Internet. For example, in the first nine months of 2018, Facebook removed 14.3 million pieces of content related to the Islamic State, al-Qaeda, and their affiliates, only 41,000 of which were flagged by external sources, primarily regular users. The overwhelming majority of the content removed came as a result of Facebook's voluntary internal efforts. Regardless of the future regulatory environment, these efforts are likely to remain critical. This is why the most important counter-terrorism questions involving Internet technology companies are what the scope of those voluntary activities should be and the best ways to implement them.

Despite the importance of company actions, counter-terrorism experts tend to focus their recommendations narrowly to government actors rather than addressing tech companies directly. One reason for this is that very few counter-terrorism policy professionals have experience working in social media companies, whereas many of them have experience working in government. These individuals therefore have relevant domain expertise about violent groups and about how

government counter-terrorism efforts function, but they have little knowledge about corporate policymaking processes, the backend of web technology, or operating at the scale of today's social media companies. Companies, for their part, have been too slow to disclose information about their counter-terrorism efforts, which is crucial to closing the knowledge gap and enabling the counter-terrorism policy community to offer more useful guidance. The failure of tech companies to be more transparent about their ongoing efforts also reinforces the outdated belief among counter-terrorism policy experts that they are not taking any action or simply do not care about the problem.

Counter-terrorism policy experts should tailor analysis and recommendations to these decision-makers within tech companies and to the challenges they face.

The counter-terrorism policy community has been understandably slow to recognize the shift within tech companies to more aggressively address terrorist content online, in part, because these companies were late to address the threat. During that period of prevarication, counter-terrorism policy experts urged companies to do more and, in many cases, urged governments to force companies to do more. But in the wake of the Islamic State of Iraq and the Levant's (ISIL) aggressive exploitation of the Internet, many companies began to tackle the problem, and thus the important questions facing the counter-terrorism policy community have shifted. Rather than continuing to simply call for companies to do more, it's important for the counter-terrorism policy community to speak directly to the policymakers inside companies to inform and influence how they approach counter-terrorism. Leaders at all levels inside tech companies are making critical decisions about how to define, identify, and take action against terrorist actors. These decisions have tremendous reach and, in many cases, are without precedent. Counter-terrorism policy experts should tailor analysis and recommendations to these decision-makers within tech companies and to the challenges they face.

In order to do that, policy experts and policymakers need a shared lexicon for understanding the ways that terrorists use the Internet and how that manifests on different types of digital platforms. They also need a shared understanding of the scope of the policy questions faced by tech companies and the tradeoffs inherent in those policy choices. This essay endeavors to provide both, with the hope that the framework will help policymakers in technology companies, improve interactions between tech companies and the traditional national security community, and inform government policymakers considering how to structure a productive regulatory framework.

This essay does not argue that certain solutions are better across the board than others, but it does highlight key questions, illustrate tradeoffs, and encourage the counter-terrorism policy community to address some of the specific questions faced by people working on counter-terrorism inside social media companies.

The Problem: How Do Terrorists Use the Internet?

While it is tempting to think of terrorists as using the Internet as if it were a monolithic entity, such thinking is counterproductive. The reality is that terrorists use a wide range of different digital platforms for different purposes. Analysts know this, of course, and should lean into this granularity to drive a much more nuanced conversation about the threat posed by specific online behavior on particular platforms, and the techniques companies can employ to manage it. The following is both a typology for thinking about terrorist use of the Internet and a lexicon for breaking down the activities terrorists engage in on various types of technology platforms.

Terrorist Functions Online

Generally speaking, terrorists use the Internet in much the same way as other people: They send messages, coordinate with people, and share images and videos. The typology below attempts to describe terrorist behavior online in terms of the generic functions the underlying technology facilitates. So, instead of "attack planning" or "propaganda distribution," the framework below uses terms like "content hosting" and "audience development." Here's why: Technology companies never build products to facilitate "attack planning," but they do think about how to enable "secure communication." To build a terminology bridge between the counter-terrorism and tech communities, we need language that speaks to

how generic Internet functionality that is usually used for positive social purposes can be abused by bad actors.[1]

Content Hosting

Modern terrorist organizations produce a wide range of propaganda in the form of imagery, videos, and audio files. Prior to broadband Internet, this sort of material was distributed manually, either in the form of printed material or pressed into video tapes, cassettes, or DVDs. Since the advent of broadband, terrorist organizations have moved those repositories online, first via file-sharing sites where users could download media and, subsequently, via services that enable large-scale file-sharing and video-streaming. Groups like ISIL still use a variety of cloud services as media repositories and consistently use video-streaming services to distribute propaganda material. Others, like Hamas and the Atomwaffen Division, have their own websites. Not every Internet platform is well suited for hosting content. Video and audio streaming sites are used for this purpose, as are cloud-based file repositories. Some offer unique capabilities including the ability to livestream video from a phone or camera. Social media platforms that facilitate easy video and image hosting can also be used for this purpose.

Audience Development

Terrorists need an audience for all sorts of reasons: to directly engage the population they want to influence, to attract media attention in order to indirectly engage the population they want to influence, and to identify potential recruits. ISIL famously used Twitter for this purpose in 2014 and 2015 because the platform offered a vast audience for ISIL's sophisticated propaganda and easy access to journalists who, in writing about that propaganda, served as inadvertent enablers. Terrorist groups think about audience development differently depending on their goals, their ideology, and their theory of victory. Although ISIL is ideologically rigid, it imagines itself as the vanguard of a vast populist movement, whereas ISIL's ideological cousin, al-Qaeda, is less ideologically stringent but conceives of its near-term audience more narrowly. These differences influence the groups' respective rhetoric but may also drive the type of digital platform each uses for developing its audience. Organizations like ISIL aim to recruit en masse, but smaller organizations looking to establish an elite core of actors may instead concentrate on audience development within a target population. Despite the glaring lack of studies comparing how terrorists use social media versus mass media, traditional mass media is likely still a critical method for conducting audience development. Nonetheless, new digital platforms are clearly useful to these groups.

Brand Control

Terrorism has famously been called "propaganda of the deed." The desire of terrorist groups to control their political messages creates a need for well-branded information conduits that can be used to validate the initial distribution of propaganda. Thus, spokespeople, dedicated media production houses, and reliable information-distribution channels online are critical. Modern terrorist groups have used dedicated web forums (e.g., al-Hesbah), official web pages (e.g., Atomwaffen, Hamas, and Hezbollah), Twitter handles, and, most recently for ISIL, Telegram channels to help cue to their target audience that the materials being distributed there are authentic. Maintaining brand control requires consistency, which gives technology platforms a particularly important role to play in disrupting this effort among terrorist groups.

Secure Communication

Despite occasional "lone wolf" attacks, terrorist violence is usually conceived of, planned, and executed as part of a group. As such, secure communications between conspirators are paramount. The ubiquity of encrypted messaging tools has lowered the bar for communicating securely and thus prompted increased scrutiny of platforms that provide encrypted services. However, terrorists have long used a variety of techniques to ensure secure messaging on the Internet. Al-Qaeda famously employed "email dead drops," in which users would share account log-in information and leave messages for one another as drafts, thereby avoiding scanning while messages were in transit. Obscurity is often a tool for security: It can be facilitated via fake accounts, multiple accounts, and secret web forums only accessible to

1 For other frameworks, see: *The Use of the Internet for Terrorist Purposes*, (United Nations Office on Drugs and Crime: New York, 2012), https://www.unodc.org/documents/frontpage/Use_of_InternetInternet_for_Terrorist_Purposes.pdf; Maura Conway, "Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research," *Studies in Conflict & Terrorism* 40, no. 1 (Spring 2017): 77–98, https://doi.org/10.1080/1057610X.2016.1157408.

invited members. Counter-terrorism professionals might breach these techniques, if they know where to look. Steganography, or the practice of leaving a hidden message in plain sight, is often overlooked. Such messaging might come in the form of using a pre-determined but innocuous code word to send a message or obliquely referencing some shared experience to authenticate oneself online. For example, consider senior al-Qaeda commander Atiyah abd al-Rahman's instruction in 2005 to the commander of al-Qaeda in Iraq, Abu Mus'ab al-Zarqawi, on how to identify one another on Islamist chat forums: "I am ready to communicate via the Internet or any other means, so send me your men to ask for me on the chat forum of *Ana al Muslim*, or others. The password between us is that thing that you brought to me a long time ago from Herat."[2] American officials ultimately captured Atiyah's letter to Zarqawi, but they likely did not know what Zarqawi's present to Atiyah had been and thus would be unable to determine which chat thread on a crowded forum was important.

Community Maintenance

Terrorist groups often rely on "in-group" social dynamics to reinforce antipathy to "out-group" members. As such, restricted spaces where propaganda can be shared, watched in unison, and discussed, are often critical. In the real world, terrorist groups use meetings, meals, religious sermons, and rallies to build this sort of in-group cohesion. Online, closed groups in messaging applications, restricted spaces on social media platforms, and branded online forums serve to separate in-group participants from outsiders. In some cases, community maintenance can be accomplished in more open digital environments using symbols and phrases that denote in-group membership. Making such public signs is much easier, however, after the basic in-group lexicon has been established in more closeted environments. These closed spaces also offer a way to reinforce and normalize an ideological worldview that endorses violence as a means to an end. This function may be particularly important for less institutionalized radical movements, such as white supremacists, as opposed to the more structured organizations jihadists tend to create. Such environments serve not only active terrorists, but also a circle of potential supporters who may some day serve as recruits. Dedicated salafi-jihadi and white supremacist web forums are often used for this purpose, but closed groups in social media platforms or messaging applications are also used.

Financing

Sustained terrorist campaigns cost money. Digital tools offer mechanisms for both fundraising and financial transfers. The core problem with electronic money transfers is security, which has driven many terrorists to use cash or to transfer money via criminal networks, in the form of illicit goods, or through traditional money-changing networks like hawalas. But some groups do use electronic transfers, either hoping to avoid scrutiny via obscurity or, in recent years, by using crypto-currencies. In the digital space, terrorist groups may use traditional financial senders such as Western Union, electronic transfers between banks and online payment systems (e.g., Paypal or Venmo), direct fundraising for charities, or person-to-person transfers using platforms like GoFundMe or Messenger Payments. Terrorists may also facilitate financial transfers online by sharing account numbers and digital passwords for more traditional exchanges.

Information Collection and Curation

Terrorist groups also use the Internet to collect information. Militants use online mapping tools to plan attacks, monitor news, and identify potential recruits. Various platforms can be used for these purposes, including social media, traditional media, search engines, and specialized tools for identifying critical infrastructure and other sensitive targets. All of these tools are used by everyday people to find grocery stores, old friends, and the quickest ways to get across town.

What About the Platform?

It is important to recognize that some online platforms are better suited for some of the functions listed above than others, which means that terrorists often use multiple platforms for their activity online. For example, in 2014, Twitter was widely used by ISIL for audience development and brand control, but because Twitter does not allow users to upload long videos or create content repositories, ISIL propagandists used YouTube, Justpaste.it, or other platforms for content hosting. They would then post links to the content hosting site on their chosen audience-development platform. Likewise, a terrorist might use Facebook for audience development, but convince a target for recruitment to shift to Telegram to communicate

2 "Atiyah's Letter to Zarqawi," Dec. 11, 2005 (10 Dhu al-Qida 1426), *Combating Terrorism Center Harmony Program*, https://ctc.usma.edu/harmony-program/atiyahs-letter-to-zarqawi-original-language-2/.

securely if the recruit showed promise. There are broader ways to think about platform preferences as well. For example, community maintenance does not require a mainstream social platform because adherents are already interested in the group's ideology and therefore are likely willing to adopt a new tool. But audience development requires utilizing platforms with an audience or active users already in place. Telegram, for example, has become a key tool for many terrorist organizations, but it is effectively only useful for brand control, community maintenance, and secure communication. It is not ideal for audience development or content hosting.

The challenge for my platform, Facebook, is that a user can credibly perform all of these functions there. This is primarily a testament to Facebook's success at building a suite of tools that everyday users want to use. But it creates challenges because that suite of tools can be used in various nefarious ways by bad actors. Consider the following assets Facebook provides: no platform has a bigger user-set for audience development; it is easy to create specialized groups for community maintenance purposes; most (but not all) forms of media can be uploaded for content hosting; and persistent accounts can be used for brand control. Among other implications, this suite of functionality means Facebook needs a wide range of countermeasures to prevent misuse.

Just as platforms vary in their utility for various functions, terrorist groups vary in the value they place on specific functions. Al-Qaeda has always conceptualized itself as a smaller, more elite organization than ISIL, thus it was slow to abandon the use of web forums that were well suited to community maintenance and brand control, even after the rise of social media networks. ISIL, by contrast, long aimed to build a broad social movement and encourage so-called lone wolf attacks. Compared to al-Qaeda, it historically risked its brand control as a result of focusing so heavily on audience development and content hosting, relying on platforms like Twitter, Facebook, and YouTube. For example, ISIL has embraced unofficial media groups producing pro-Islamic State propaganda more so than al-Qaeda. Over time, ISIL came to understand the importance of brand control and has embraced Telegram as a core tool for achieving that goal.

Policymakers, in both government and corporate settings, and the wider counter-terrorism policy research community must understand how terrorists use specific platforms in order to effectively prescribe countermeasures.

Policymakers, in both government and corporate settings, and the wider counter-terrorism policy research community must understand how terrorists use specific platforms in order to effectively prescribe countermeasures. For example, platforms used for content hosting should prioritize mechanisms to identify terrorist propaganda — various techniques for content matching are likely to prove useful. But these techniques will not be as important for platforms used to maintain a group's community, communicate securely, and organize financing. For those platforms, identifying behavioral signals or information-sharing with partners may be more important. Platforms that support numerous functions will need to develop a variety of techniques. There is no one-size-fits-all solution to this problem and the counter-terrorism policy community must not make the mistake of suggesting otherwise. Tech company decision-makers are well aware of the differences between platforms and, in a very different way, so too are the terrorist groups that use them.

Counter-Terrorism Questions for Technology Companies

Companies developing a counter-terrorism policy need to build a strategy that is adaptable enough to keep pace with changing dynamics in the real world and evolving technical realities. They must consider the tactical implications both to terrorists and to their far more numerous benign users. They also must consider how their choices will impact more traditional counter-terrorist actors, whether in government or nonprofits. The purpose of this section is to focus on some of the key challenges counter-terrorism policymakers at technology companies face. It is crucial for counter-terrorism policy experts to understand the variety of factors that shape how a company responds to terrorism on its platform. These factors vary widely, and include the following:

- The balance between freedom of speech and the privacy and safety of users and society writ large. These principles are not always directly at odds, but the tension between them cannot be fully resolved

without tradeoffs. Some platforms have historically sought to encourage unfettered speech. Others endeavor to foster community by encouraging users to reflect themselves authentically online while attempting to enforce a stronger set of community rules.
- The particular functions, as described above, that terrorists seek to conduct on a particular platform.
- The degree to which a company understands the manner in which terrorists are misusing its tools. The reality is that some companies simply do not understand the ways in which their platforms are being misused.
- The resources available to the company. Technology companies have vastly different resources to address online ills such as terrorist activity. Policymakers often conceptualize Silicon Valley companies as behemoths with vast resources, but terrorist groups exploit a wide range of technology platforms, the smallest of which can count their employees on one hand and do not have the resources to hire counter-terrorism specialists or dedicate large engineering and operational teams to counter-terrorism.
- The willingness to address political conflicts. In general, technology companies endeavor to set policies that will apply globally, regardless of country. This urge for universality is very different from the way governments approach geopolitical questions, where modulating policy according to each country is common.

Against this backdrop, online platforms must make a series of strategic policy choices and operational decisions for addressing terrorist activity online, with far-reaching policy impacts. The purpose of describing these issues is not to argue for any one particular solution. Rather, it is to illustrate how these choices may manifest for policymakers within tech companies so that the traditional counter-terrorism community can consider and, hopefully, better advise this new crop of policymakers emerging within tech companies.

This list of questions, and potential solutions, is not intended to be comprehensive. However, it is illustrative of the key strategic and operational issues and potential solutions facing technology companies as they construct counter-terrorism strategies.

Strategic Choices

How to Determine Who Is a Terrorist?

One of the most fundamental policy decisions technology companies face is how to determine who is a terrorist. There are several options, each with its own pros and cons.

One option is to rely on international designation lists, such as those maintained by the United Nations or European Union. This approach allows companies to lean on institutions that theoretically reflect the global community's collective wisdom and allows a technology company to avoid making decisions that may be perceived as political. The problem with this approach is that international organizations, and the lists they generate, in reality reflect a politicized consensus developed after much political wrangling. Moreover, the lists are updated very slowly, and often reflect a lowest-common-denominator approach. This generally means that such lists include the most prominent global terrorists but exclude militant groups that receive less global attention or are only relevant in specific locales.

Companies that want to address a wider range of terrorists using their platforms might instead decide to rely on designation lists maintained by various governments around the world. This approach avoids the lowest-common-denominator issue and can align a company with legal authorities around the globe. The problem is that some government actors designate non-violent political groups as terrorists, so this approach may lead a company to censor groups based on a regime's political agenda. A company may try to rely only on terrorism lists from specific governments, for example, from their home country or from other democratic states. But this approach forces companies to determine which countries are suitably democratic. In addition to risking that a government will block a particular service from operating in its jurisdiction, this approach would also mean that companies, not political institutions, would be making key decisions globally about which governments are legitimate.

A final option is that companies can designate terrorist organizations themselves. This approach offers companies a mechanism to resist government pressure to crackdown on peaceful opposition groups, but it requires companies to do extensive analytical work, come up with a clear definition of terrorism, and assert a designation role traditionally reserved for governments.

How to Structure Basic Content Standards?

It may seem easy for a company to simply "prohibit terrorism" on their platform, but putting in place a robust policy is far more complex. Companies must, for example, determine whether to construct restrictions at a content, account, or user level, as well as what sort of engagement with terrorist content or groups is acceptable and what is not.

Content-level restrictions proscribe support for terrorism within individual pieces of material online. "Content" differs by platform, but on Twitter it would be a tweet; on Facebook, a post, comment, or similar piece of user-generated information; and on YouTube, a single uploaded video.

Even at the content level, companies must determine what sort of material violates their rules. One mechanism is simply to prohibit formal propaganda produced or explicitly designed to advance the message of a terrorist or terrorist group. This is a powerful approach against groups like ISIL that produce a high volume of branded formal propaganda, but it is less valuable to counter informal propaganda, which is common among a range of terrorists, including white supremacists and localized ISIL supporters in some areas of the world. However, targeting informal propaganda may create implementation challenges as this material is more difficult to identify.

Removing content produced by terrorist organizations may seem straightforward, but companies must also determine how far to take that approach. For example, should they remove praise and support for terrorist groups, even if it seems to come from people without any official ties to the group or people who support a group's political goals but not its violent tactics? Removing such support from social media will tend to produce more equity across organizations, including those with less formal support structures, but it also generates ambiguity. What exactly does "praise" mean? Does it apply even when the terrorist group is doing something seen as positive for a community — for example, providing disaster relief or negotiating a ceasefire? This approach may also implicate regular users in complex political situations who may express support for a group widely understood globally as a terrorist organization, like Hezbollah, that nonetheless maintains local political legitimacy.

Companies must also determine whether to allow some content from terrorist groups on their platforms in specific circumstances. This might come in the form of political campaigning by groups like Hezbollah or the Milli Muslim League, or Sinn Fein during an earlier time period. Platforms may also choose to allow terrorist content when it is shared for purposes of counter-speech — pushing back against the narrative of terrorist groups — or by mainstream media or academics. Content clearly condemning terrorism, raising awareness about terrorism, or advancing the study of these groups has obvious social value, but allowing even this content carries risks. Adversarial terrorist groups may use such policy carve-outs to obfuscate their true intent when posting content, and terrorist supporters may still engage dangerously with content when it is shared by a legitimate actor for legitimate purposes. Moreover, any complexity in a policy regarding terrorist propaganda will slow enforcement decisions.

Removing content produced by terrorist organizations may seem straightforward, but companies must also determine how far to take that approach.

Some companies may determine that it is inefficient or ineffective to simply prohibit terrorist content from being shared on their site. Rather, they deem it better to remove accounts that represent terrorist entities or that demonstrate support for terrorism. The most straightforward way to do this is simply to remove an account after a certain number of content violations. The benefit of this approach is simplicity. It also ensures the account is judged directly on its own online behavior. Some companies may want to assess accounts using a broader set of indicators to determine whether removal is warranted. This might include the account's IP address, its engagement with other dangerous accounts, patterns of friending behavior, or other account-level metadata, as well as technical signs gathered with anti-spam techniques that indicate an account was created disingenuously or reflects a previously removed account. Importantly, metadata-based tools may work even when content is encrypted, making them potentially very valuable for encrypted platforms.

The most aggressive approach to imposing content standards focuses on the user directly. This means that a real-world person is simply not allowed to use a platform, regardless of who they

interact with or what they post. This approach is straightforward for notorious terrorists like Osama bin Laden but is more complicated when it comes to more obscure terrorists like, for example, members of the Kurdistan Worker's Party. User level restrictions also raise important practical questions. Should a prohibition extend only to leaders of a terrorist organization or to all members? How should those categories be defined and what is the evidentiary standard for determining whether someone falls into either category? Moreover, even in the best of circumstances, a company will not be able to create, or reasonably enforce, a comprehensive list of the world's terrorists. Despite this final problem, establishing stringent restrictions at the user-level does offer a consistent standard for removing terrorist users on a given platform if the company becomes aware of them.

How to Manage Government Content Removal Orders?

Governments often report content to social media companies if they deem it illegal or unacceptable per the company's terms of service. The differences between the two types of requests are important and result in very different kinds of referrals to a technology company. The former, if legitimate, is a legal order that carries the weight of law and usually comes from a judge. The latter is simply an administrative referral that may come from a communications regulator or Internet referral unit. Companies must determine how to respond to these referrals, with each approach carrying pros and cons.

Legal Orders

The simplest approach for social media companies is to abide by government declarations that the flagged content on their platform is illegal and remove it. This approach may appeal to smaller companies in particular that do not have the resources to make in-house judgments about potential terrorist content or a legal team to validate that an order is legally binding. The downside, however, is that it risks potentially allowing governments to censor unpopular political views online. It also raises the possibility of companies erroneously taking action on orders from entities that do not actually have the legal standing to order content removal.

A company may also simply decide to ignore government legal orders. This approach limits, for example, the ability of authoritarian governments seeking to censor content, but raises the possibility of missing genuinely dangerous content on the platform. It also increases the risk that a government may sanction or block that platform, which obviously has important implications both for the business and for the ability of citizens to express themselves. This approach does not require extensive resources, however, which is a major advantage for companies with limited capacity.

A middle-ground approach is to review all government referrals against the company's own terms of service, assuming they exist. This limits the risk of both facilitating government censorship and leaving up dangerous content, but it requires time and internal resources that may only be available to larger companies. It could also lead to the company opposing a government legal order, which may involve extensive litigation or result in the platform being blocked in that country.

Tech companies may also try to apply a legitimacy standard to take into account human rights and adherence to rule-of-law in an effort to distinguish legitimate legal orders from illegitimate ones. In order to operationalize this approach, companies would likely have to evaluate orders at the country level — meaning orders from certain countries would be respected while orders from other countries would be ignored. They would also assess the legal validity of the order itself. This approach will inevitably create controversy when a company rejects certain legal orders while accepting others.

Administrative Referrals

Companies have similar options for responding to administrative referrals from governments, although the legal implications here are obviously different. As knowledge about terrorist use of the Internet has grown more prominent, both states and bodies like the European Union have developed specialized programs to identify terrorist content online and refer it to tech companies.

Some companies may treat government referrals in the same way they do legal orders and remove the content in question immediately. This approach is valuable for small companies with limited capacity, but opens the door to extensive, and potentially politicized, government censorship because such administrative orders do not require legal review. Likewise, companies could decide to ignore government referrals entirely, either by refusing to accept such referrals or deciding not to act on such information. This would limit the ability of a government to use companies as a means of exercising censorship, but creates the genuine risk

of missing dangerous content since governments — which maintain expertise on terrorism — are more likely than regular users to refer actual terrorist content to companies.

A middle-ground approach would require reviewing government referrals against the company's own terms of service. This limits the risk of government censorship and of leaving up dangerous content, but it requires time and internal resources that, again, may only be available to larger companies. Companies may also decide to split the difference and abide by legal orders to remove content but review administrative referrals against their terms of service. These more nuanced approaches typically require more sophistication from the company, including legal, policy, and operations teams working in concert at a global level. This kind of coordination may be feasible for larger companies but is very difficult for smaller platforms.

When a company receives a legal order or referral from a government, it must also determine whether removals should be applied only within the boundaries of that country or globally. Removing content globally will likely satisfy the government more fully and avoids the odd scenario of data accessibility varying by location or via a Virtual Private Network. But this approach effectively gives any country the ability to project its own legal framework onto other countries, which may result in content that is legal in many places being removed because of the dictates of more repressive systems. Only removing content locally prevents governments from imposing a global censorship regime based on local law. But this creates obvious workarounds through Virtual Private Networks or other techniques that will allow the proscribed content to still be accessed within the country demanding removal. Reviewing referrals against internal terms of service helps obviate this issue because if the content does violate a company's terms of service, it is reasonable to apply that decision globally.

When a company receives a legal order or referral from a government, it must also determine whether removals should be applied only within the boundaries of that country or globally.

Operational Choices

The policy choices discussed above are foundational, but good outcomes require more than just policy — that policy must be applied effectively. The operational counter-terrorism choices facing technology platforms vary dramatically. They depend on the nature of the product itself, how terrorists use the product, and the resources a company has to invest in countering the problem. And, as with many problems, it is not always clear that throwing more resources at combatting terrorist activities online will dramatically improve outcomes.

The broader counter-terrorism community often fails to consider the operational tradeoffs facing companies developing online counter-terrorism programs. This problem is heightened by the techno-utopianism long touted by Silicon Valley, which has created the misconception that simple technical solutions exist for most problems. Unfortunately, that does not reflect reality. In truth, decision-makers inside tech companies must balance different counter-terrorism priorities and make bets on the utility of investing in various programs with uncertain outcomes.

The operational issues facing tech companies can be broken down into four broad categories:

1. How to find potential terrorist material?
2. What to do when potential terrorist material is found?
3. Should appeals be allowed and, if so, how should they work?
4. Should counter-speech efforts be supported? If so, how?

At a high level, these questions may seem simple. In practice, they are more complex. The most important, over-arching operational challenge for tech companies is scale. Facebook took action on 14.3 million pieces of content related to ISIL or al-Qaeda in the first nine months of 2018, finding 99 percent of that content itself. Facebook's policy team writes exacting rules and rigorous implementation guidelines for identifying and removing content but does not take part in most removals. Instead, machine-learning classifiers and a team of more than 15,000 reviewers — 200 of whom are specialists on terrorist groups and other dangerous organizations — take action on

content. But achieving consistency and accuracy is challenging when these processes play out globally with all the complexities of culture, language, and political context, not to mention simple human error. Even if mistakes only occur in a small percentage of cases, the massive scale of the Internet means there will nevertheless still be a high number of errors.

Likewise, sometimes seemingly obvious solutions do not pan out. Facebook, for example, allows users to report terrorist material they encounter, but this is a very inefficient way to find terrorist content. Only 41,000 of the 14.3 million pieces of content against which action was taken were the result of reports that originated outside Facebook. Though it might seem like Facebook should prioritize user reports of terrorist content, the reality is that these reports often simply point to content that users do not like rather than to actual terrorist content. Flagging content internally is a far more accurate and efficient way to identify terrorist content online. This creates what amounts to a customer-service problem: Facebook obviously wants to be responsive to the concerns of users, but focusing on external reports — from both users and governments — means focusing on the lowest scale, least precise methods of identifying terrorist content.

How to Find Potential Terrorist Content?

The best methods for identifying terrorist content largely depend on how a platform defines terrorism and the content that violates its standards, as well as how the platform itself is built. It is useful to think about detection methods as

falling into two sub-categories: human approaches and automated approaches. Human approaches have the advantage of flexibility: People can adjust what they are looking for and quickly identify new behavioral patterns by terrorists. Automated techniques are valuable in that they scale to a global audience. However, they are not as nimble as human approaches and can potentially be circumvented by adaptive adversaries. Many of the bigger tech companies, Facebook included, utilize both human and automated techniques.

Human Approaches

As discussed above, referrals of terrorist content from governments and inter-governmental organizations can be fruitful. Relative to user reports, government referrals are generally precise — meaning they actually point to terrorist content — but they are low in volume. A company may use government reports to identify and remove terrorist content, and in doing so may mitigate external pressure from those governments, but this approach is extremely limited in scope. Moreover, government referrals almost always focus on content hosting, audience development, and brand maintenance functions. Governments may be aware of other activities conducted by terrorists online, but generally do not want to squander valuable intelligence sources or reveal the methods they use to identify such behavior.

A company may use government reports to identify and remove terrorist content, and in doing so may mitigate external pressure from those governments, but this approach is extremely limited in scope.

Tech companies may also work with external teams to identify terrorist content. For example, YouTube uses a "Trusted Flagger" program while Facebook contracts with a range of vendors to provide targeted referrals of terrorist content.[3] A company could decide to provide specialized tools or API access to facilitate the work of such partners. Like government referrals, referrals from these external teams tend to be high quality. Importantly, they can usually be produced in higher volume than government referrals. Nonetheless, these reports are still relatively small in scale and tend to focus solely on content hosting and audience development functions of terrorist groups.

For many technology platforms, user reports are a critical way of maintaining a relationship with users concerned by material they see on the platform. These reports provide a method of redress that, at best, provides both useful information to the platform and gives the user a sense of ownership and responsibility. In the real world, counter-terrorism programs remind citizens, "If you see something, say something." User reports reflect the same general instinct online. Moreover, users in the aggregate see far more content than either governments or external teams. The problem with user reports is twofold: First, users often report benign content or information they simply do not like. This means that the platform must invest significant resources to identify which reports are useful, a process that is costly, time-consuming, and may distract from higher-value efforts. Second, users must be motivated to report things. This is unlikely in closed spaces where terrorists conduct community maintenance or communicate securely because only individuals likely to support the terrorist cause will be present in such spaces.

The human approach does not always rely on external information sources. Platforms can also use internal teams of specialists to identify terrorist content. These teams may have better technical tools than outside sources, which allows them to identify a wider range of terrorist behavior than the content hosting and audience development identified by governments, users, and external teams. Nevertheless, they cannot match the scale of user reports, let alone the scale of automated techniques described below. Given the limitations on how much content these internal experts can identify, platforms have to determine whether investing in these teams makes sense or whether employee time should be reserved for other tasks.

Automated Approaches

There are many automated methods that can be used to identify potential terrorist content,

3 For more information, see: "YouTube Trusted Flagger Program," Help Center, YouTube, accessed March 10, 2019, https://support.google.com/youtube/answer/7554338?hl=en; Monika Bickert and Brian Fishman, "Hard Questions: Are We Winning the War on Terrorism Online?," Facebook Newsroom, Nov. 28, 2017, https://newsroom.fb.com/news/2017/11/hard-questions-are-we-winning-the-war-on-terrorism-online/.

Most techniques to identify terrorist content are implemented by single companies on their own platform. However, some companies have begun sharing signals of potential terrorist content with one another.

and all of them have costs and benefits. Some are only useful with certain types of content while others are unreliable unless used in conjunction with human reviewers. While such techniques are critical to a robust counter-terrorism effort online, they are not foolproof.

Content matching is one of the simplest automated detection techniques available. This approach creates a "digital fingerprint" of known bad files, whether images, video, audio, or text. These digital fingerprints, known as "hashes," manifest as unique strings of numbers, letters, and symbols that correspond to a given file. Those hashes can then be matched against hashes created when content is uploaded to a particular platform. Many hashing techniques allow a company to catch an image or video that has been altered, but these techniques do sometimes miss content that a human being would recognize as fundamentally the same. Content matching is particularly effective in countering terrorist groups that regularly release formal propaganda. The technique does have limitations, however: It requires creating hashes from content uploaded to a platform and will not work on content that has been encrypted. It also does not work for newly created content, whether live-streamed or otherwise produced in the real world and then uploaded. Content matching also requires making a range of policy choices, most notably setting thresholds for how similar a piece of content must be to another known piece of bad content to which it has been algorithmically matched in order for it to be removed or reviewed. Setting a lower threshold will capture more bad content but is more likely to result in false positives, while setting a higher threshold will result in fewer false positives but is more likely to miss some terrorist content.

Optical recognition technology allows platforms to scan for logos, weapons, and other potentially worrisome indicators in an image or video — even if the overall image or video does not match a known digital fingerprint. This technique is more sophisticated than content matching and thus harder to deploy for small companies. Like content matching, optical recognition also generates confidence scores that rate the likelihood that something identified by the algorithm is, in fact, worrisome. However, this technology can only scan content that has been uploaded to a platform, requires extensive training data, and will not work on encrypted content.

Many terrorist organizations use hashtags to identify their content or insert propaganda into mainstream conversations. ISIL, for example, often coordinates "raids" using hashtags on specific platforms. Platforms can identify these hashtags in various ways, for example, by monitoring terrorist communications where hashtags are discussed, by systematically identifying hashtags commonly associated with terrorist content and then using them to search for other content, or by identifying key themes and issues targeted by terrorist actors and searching for related hashtags. The benefit of hashtag tracking is that it allows quick tactical disruption of terrorist propaganda distribution. But hashtags are used on some platforms more than others and can easily be changed by terrorist organizations. Hashtag-based detection also requires strong coordination with a team of human experts to be sustainable.

Text classification, another automated approach to flagging terrorist content, uses machine learning techniques to identify text that is similar to content already determined to support terrorists. Text classification can be very useful for detecting potential terrorist content, but because of nuances in language it may not be precise enough to reliably delete content without some human oversight. Such approaches also require a large corpus of training data, which may be difficult to acquire for smaller companies.

All of the approaches above rely on assessments of content itself, which is only possible if it is not encrypted. But some platforms may be able to identify dangerous accounts based on account-level behavior, such as having relationships with suspect accounts, using worrisome IP addresses, using bots to auto-create accounts, or acting in conjunction with other accounts that have demonstrated similarly problematic behavior. Because platforms differ, these behavioral signs are likely to vary significantly by platform. One advantage of this behavioral approach to tracking and countering terrorist activity online is that it can be used on some encrypted platforms because it does not rely on content. But such an approach can generate high rates of both false positives and false negatives — and those rates cannot be verified in an encrypted setting.

Most techniques to identify terrorist content are implemented by single companies on their own platform. However, some companies have begun sharing signals of potential terrorist content with one another. The most notable example is the Global Internet Forum to Counter Terrorism's hash-sharing database, which allows companies to benefit from their colleagues' work

in other companies.[4] This is particularly important when terrorist groups use multiple platforms in coordination with one another. Such collaboration offers small companies a quick way to develop a relatively sophisticated counter-terrorism program, but it is not a panacea for the reasons described above.

The most sophisticated efforts to identify terrorist content rely on machine learning that looks at a variety of signals to determine whether a piece of content supports terrorism. These techniques develop a confidence score indicating the likelihood that a piece of content supports terrorism. These tools are very powerful because they can holistically assess content. However, they require extensive training data as well as difficult policy decisions to set thresholds for taking action based on the confidence scores produced by the algorithm. These tools must also be carefully maintained to sustain accuracy, which means human beings continuing to train and retrain existing algorithms. In short, even the most sophisticated machine-learning techniques require continued human maintenance to work as intended.

What to Do After Finding Potential Terrorist Content?

In the real world, deciding how to take action on content depends on various factors, including the context under which it was uploaded and the confidence with which an algorithmic classifier suggests it supports terrorism. This paper does not capture all of that variation. It does, however, describe an assortment of actions that can be taken and discusses the variety of circumstances in which those actions might be used. Inherent to this discussion is the notion that terrorist content might be shared for legitimate reasons by academics, activists decrying extremism, or journalists. The notion that there are legitimate reasons to share terrorist propaganda significantly distinguishes this kind of content from other types of harmful content found online, most notably child pornography. Legal regimes proscribe sharing or possessing such content regardless of circumstance. As a practical matter, this means that reviewing terrorist material by a company often requires a more nuanced assessment of context than when it comes to child pornography, which can slow down the review process and increase the likelihood of human mistakes.

This section is therefore broken into two parts: first, a discussion of the relative pros and cons of allowing human beings or automation to "decide" when to take action on a given piece of digital material; and, second, to assess those actions themselves.

Human Beings and Automation

Human beings assess context far better than computers, particularly when considering the linguistic breadth of the Internet and cultural specificities related to terrorism. Companies can hire people with specialized language and cultural skills, who can apply some level of judgment or cultural nuance in reviewing content. But human judgment carries costs as well. Many companies simply do not have the resources to hire large teams of human reviewers, and those that do must struggle to ensure that those teams apply policy consistently at scale. Moreover, human beings are fallible, they get tired, they have personal biases, and the work of reviewing the intense content that terrorist groups often produce can be exhausting and disturbing.

Automation avoids many of these pitfalls: Computers do not get tired or make "mistakes" in the traditional sense. Algorithms, perhaps counterintuitively, also have some advantages for small companies because, once trained, they do not require the large human teams necessary for human review. But automated systems are only as good as the training data and labeling exercises used to program and maintain them. A poorly trained algorithm may have a systemic bias around certain types of content or certain organizations and, as a result, can produce false positives and false negatives, just as humans do. This carries real risk: Counter-speech campaigns sometimes purposefully emulate the visual style and language of terrorist propaganda, which might confuse some automated detection techniques, but not a human being.

In other words, enabling an algorithm to remove content does not obviate the need to make difficult policy decisions. It just changes how those policy choices manifest. A policymaker must decide whether the computer should remove content when the confidence indicates a particular likelihood that it supports terrorism. Is a 95 percent likelihood the right threshold? How about 90 percent? Eighty percent? Fifty percent? Those decisions all lead to the inevitable result that benign content will be removed erroneously. The question is how many of those "false positives" are acceptable. Ultimately, people set these standards, not computers.

4 For more information, see the Global Internet Forum to Counter Terrorism website: http://www.gifct.org/.

What Actions Should Companies Take?

In addition to determining who or what should make the final decision about a suspected account or piece of content, companies must also determine what action to take. The simplest choice is to remove the flagged content or account. Removal is also appealing because it constitutes a consistent and visible action against terrorist material. However, when it comes to account removals, companies must determine how many instances of content violation should trigger removal. Should it be one? What about false positives? That could lead to immediate account removal. Perhaps it should be two or three? Or five or 10? Should some violations be deemed more egregious than others, or is every instance of support for terrorism equal?

If removal does not seem appropriate, platforms can instead limit the visibility of the content or account. This may be a useful tactic in situations when a human or algorithm is not completely confident that the material in question supports terrorism. Such limitations could even be employed on a temporary basis until a more definitive judgment can be made. Once again, these techniques, especially the more nuanced ones, will be much easier to implement for larger companies than smaller ones.

The Global Internet Forum to Counter Terrorism maintains a database of more than 100,000 visually distinct images and 10,000 visually distinct videos that can be used by participant companies to identify dangerous material on their own platforms. But companies must decide whether to utilize this database. It may seem like a no-brainer, but smaller companies have to make difficult decisions about where to apply limited engineering resources. Even if they decide to focus on counter-terrorism, they may determine that other techniques will be more fruitful and so decide not to spend the resources to contribute to this hash-sharing database.

Finally, companies must determine whether to refer a potentially dangerous account or piece of content to law enforcement. Not every violation of a company's terms of service deserves law enforcement attention and companies have obligations to protect user information, except in extenuating circumstances. So, companies must determine a standard for when to refer an account to law enforcement. Should they limit such referrals to accounts associated with specific groups? Should they have clear evidence of an imminent attack? What does "imminent" really mean? Should they refer individuals coordinating propaganda? Should they provide information about individuals when the only reasonable real-world action would have to come from the military rather than law enforcement? How certain should a company be that the account-holder in question poses an actual threat?

Should Appeals Be Allowed?

Even the best policies are still fallible because there will always be errors that result from both false positives and false negatives. A company must, therefore, decide whether and how to allow for redress by users. Appeals systems create policy questions of their own, however. How long should a user have to appeal? How difficult should it be to appeal? Should the user be able to introduce new evidence to an appeals process in order to justify that their intent in posting violating content was actually benign? Should the same review teams that made the potential error assess the appeal, or should companies establish an independent review body? These tricky questions are made harder because appeals of decisions involving terrorist content put a company in the uncomfortable position of potentially communicating directly with a terrorist group or their agents during the course of the appeal. Indeed, at the scale of the Internet, not only are erroneous removals inevitable, so too is the erroneous reinstatement of terrorist content and accounts after having been removed correctly. A company must decide whether the risk of inadvertently reinstating terrorist behavior

How certain should a company be that the account-holder in question poses an actual threat?

is worth the value of giving the larger digital community the ability to seek redress.

Should Companies Support Counter-speech Efforts? If So, How?

Counter-speech programs have a long and complex history in counter-terrorism. Critics question their effectiveness and suggest that efforts to "counter violent extremism" are used as cover to monitor minority communities.[5] And yet, the promise of counter-speech efforts that proactively turn people away from radicalization is compelling. Many Internet companies were founded to empower and promote speech, thus counter-speech work has an obvious appeal compared to censorship. The challenge for technology platforms is twofold: Tech companies cannot communicate credibly directly against violent extremist organizations, and companies often have legal and political incentives not to favor one political ideology over another. Nevertheless, there are a range of options for supporting counter-speech efforts short of simply producing and distributing messages directly.

The tech community was regrettably slow in taking counter-terrorism efforts seriously. But basing policy recommendations on that historical tardiness rather than on the contemporary challenge of how best to respond is worse than unhelpful — it is counterproductive.

The simplest approach is that companies can support civil society groups — students, non-governmental organizations, and activists — to develop their own campaigns against extremism. This might mean providing financial support but could also include providing training and resources as well. Offering advertising credits is a simple way to empower non-profits to expand their reach. Tech companies may also decide to introduce counter-speech to users when they engage with particularly worrisome content or concepts online. The Jigsaw Redirect program, for example, introduces counter-speech messages when users search for terms that suggest they are interested in extremist groups.[6] Finally, tech companies might also develop ad-targeting tactics for non-profits engaged in counter-speech efforts just as they would with a small business trying to reach new customers. The challenge in this case is determining which users are potentially at risk of radicalization, which could easily lead to bias.

Conclusion

This essay has developed a new typology for thinking about how terrorists use the Internet and has illustrated *some* of the strategic- and operational-level decisions that policymakers at technology companies face as they develop counter-terrorism programs. In doing so, it hopefully has established parameters that will help produce fruitful conversations between the traditional counter-terrorism policy community and a new crop of policymakers within technology companies.

It is worth briefly pointing out some areas this essay has not addressed: This discussion has not, for example, wrestled with how companies should communicate with their users about counter-terrorism work, transparency more generally, the value of various metrics for measuring success, structures and dynamics for sharing information with academics and other researchers, or the utility (or lack thereof) of broader concepts like deterrence in digital counter-terrorism. This essay has only partially raised critical issues like encryption and the persistent tension between privacy and security. It does not wrestle with the specific challenges raised by a host of emerging technologies, including crypto-currencies, live-streaming, online video gaming, and virtual reality. It also sidesteps an issue that can be central for tech companies developing counter-terrorism programs: the perception that an aggressive program against non-state militant actors effectively benefits state actors. This is a key issue, but one that is outside the scope of this

5 For a useful review of the various critiques of countering violent extremism programs, see: Robin Simcox "Can America's Countering Violent Extremism Efforts Be Salvaged?," *War on the Rocks*, Dec. 17, 2018, https://warontherocks.com/2018/12/can-americas-countering-violent-extremism-efforts-be-salvaged/.

6 For more information on the Jigsaw Redirect program, see, https://redirectmethod.org/.

paper and has already been widely discussed in other venues. Regardless, the counter-terrorism policy community can and should productively weigh in on all of these issues.

Indeed, the counter-terrorism research community should not accept the categorizations in this paper as fixed. They should be interrogated and improved on. That said, any critique must account for the tradeoffs inherent in choosing specific counter-terrorism approaches and the differences between technology platforms and the companies that run them. Failure to acknowledge, for example, that scanning content contains privacy tradeoffs, that growing review teams leads to management challenges and inconsistent enforcement, that small companies have vastly different capabilities than large ones, and that specific technical solutions are better suited for some platforms facing specific types of counter-terrorism challenges than others may produce satisfying rhetoric, but little else. Counter-terrorism policymaking online, like most policymaking, is about balancing tradeoffs. The counter-terrorism community must acknowledge those tradeoffs to productively influence real-world decision-making.

The importance of having constructive discourse about digital threats cannot be overstated. The tech community was regrettably slow in taking counter-terrorism efforts seriously. But basing policy recommendations on that historical tardiness rather than on the contemporary challenge of how best to respond is worse than unhelpful — it is counterproductive. The largest technology platforms have made great strides countering terrorist content and, although they can still do better, have genuinely committed to addressing the problem. In order to improve, they need *specific* guidance from the counter-terrorism policy community on *how* to improve. Researchers simply demanding that tech companies "do more" is no longer helpful. It suggests limited practical knowledge about the issues and should be seen for what it is — a surface-level political argument rather than useful policy guidance.

Of course, outdated policy analysis can also spill into counterproductive regulatory efforts. Regulatory policy that explicitly constrains or implicitly disincentivizes voluntary counter-terrorism efforts by tech companies — even if it compels some forms of productive engagement between companies and government — risks making the terrorism environment online worse, not better.

Perhaps even more importantly, smaller technology platforms are carefully watching the engagement between larger platforms and the policy community. As large platforms push terrorists deeper into the shadows of the web, smaller platforms will be a more important part of the counter-terrorism effort. It should be everyone's goal to bring these platforms into the conversation, not scare them away. We must convince them that such engagement is productive rather than just an exercise in exposing a digital platform to criticism or penalty.

Digital counter-terrorism efforts are daunting and therefore humbling. The scale of the challenge is massive, and every success is mitigated by adaptive adversaries working to circumvent new rules and enforcement efforts. Tech companies should have humility in the face of such a monumental challenge and reach out to the traditional policy and counter-terrorism community for advice and guidance.

Policymakers and academics must have a sense of humility as well. Studies of terrorism online are hampered by incomplete data and usually only measure content-hosting and audience-development functions, which can mislead the public and policymakers about where companies should focus their efforts. To put it bluntly, researchers cannot reliably measure how much content terrorists post online because of the confounding effect of platform countermeasures. Researchers do not see what terrorists post. Rather, they see what is left after platform countermeasures are employed. For the major platforms, this is usually a small subset of what was posted originally, and it means that there is a fundamental bias in nearly all studies of terrorist content online. This bias was not nearly as severe in the years before platforms began to respond to ISIL's broad exploitation of their platforms, but that situation has now changed. If counter-terrorism analysts fail to mention this dynamic in their research, they mislead themselves and their readership about what terrorists are doing online and what platforms are doing to counter terrorist activity.

At the same time, companies need to be more transparent about their policies and their enforcement of those policies. Whenever possible, they should provide access to data that researchers cannot otherwise get a hold of. But researchers must recognize that such sharing creates privacy risks of its own and, in some cases, is directly restricted by existing privacy constraints on tech companies. Tech companies and researchers should also endeavor to utilize shared terminology and conceptual frameworks, such as the typology presented above.

Digital counter-terrorism efforts will not and should not be driven primarily by governments, even in a more aggressive regulatory environment. Regulation may eventually set some baselines for these efforts but treating regulation as a panacea is a mistake. Indeed, regulatory efforts that compel companies to focus on narrow aspects of the problem may actually create more problems than they resolve. Regardless, companies will continue to be primary actors in the counter-terrorism effort. The operative question is not *whether* they should work to improve, it is *how*, precisely, they should go about doing so. Counter-terrorism researchers should recognize this, and tailor their recommendations not just to regulators working to set baselines, but to the corporate policymakers who want to do far more than the bare minimum.

Brian Fishman *leads efforts against terrorist and hate organizations at Facebook. He is the author of* The Master Plan: ISIS, al-Qaeda, and the Jihadi Strategy for Final Victory (*Yale University Press, 2016*)*, and is the former director of research at the Combating Terrorism Center at West Point. The views expressed in this paper are Mr. Fishman's alone.*