THE FUTURE OF THE CHRISTCHURCH CALL TO ACTION HOW TO BUILD MULTISTAKEHOLDER INITIATIVES TO ADDRESS CONTENT MODERATION CHALLENGES

*Rachel Wolbers**

ABSTRACT

This article explores the challenges the New Zealand Government faced after the events in Christchurch on 15 March 2019, where a violent gunman killed 51 people and live-streamed his attack on social media. The video was viewed millions of times in the days following, even as the tech companies took extraordinary efforts to reduce its virality. To find a long-term solution that ended the proliferation of this violent content while protecting human rights, the New Zealand Government decided to take a non-regulatory approach that worked alongside tech companies and civil society. The result was the creation of the Christchurch Call to Action, a multistakeholder initiative where governments and online platforms, working with civil society, committed to 25 goals to eliminate terrorist and violent extremist content while protecting a free, open, and secure internet.

This article argues that the creation of an multistakeholder initiative was not only the right option for the New Zealand Government in the aftermath of Christchurch shooting, but that multistakeholderism is the best approach for addressing all issues related to the governance of user-generated content online. The problems related to the proliferation of harmful content online cannot be solved through government regulation, and tech companies cannot, and should not, set the rules alone. Therefore, to find a solution, governments and companies must work with like-minded actors who uphold human rights principles, and meaningfully engage with civil society, technical experts, academia, and users. These solutions should be consensus-based and build in accountability mechanisms for both governments and companies. This article argues that solutions proposed addressing terrorist content could serve as a guide for other types of user-generated content where definitions remain contentious.

* Rachel Wolbers is an Adjunct Faculty member at Case Western Reserve University School of Law. This article was written as part of the Ian Axford Fellowship of Public Policy, a Fulbright program hosted in New Zealand. The author would like to thank New Zealand's Department of Cabinet and Prime Minister Christchurch Call Unit for allowing her to embed with their team for six months to write this article, including Jacinda Ardern, Paul Ash, David Reid, Elisabeth Brown, Kristina Kirk, Aline De Vincentis, and Ellen Strickland.

THE FUTURE OF THE CHRISTCHURCH CALL TO ACTION HOW TO BUILD MULTISTAKEHOLDER INITIATIVES TO ADDRESS CONTENT MODERATION CHALLENGES

TABLE OF CONTENTS

INTRODUCTION

On March 15, 2019, a gunman in the city of Christchurch, New Zealand turned on the GoPro video camera mounted on his helmet, linked the livestream to his Facebook account, and entered the Al Noor Mosque.[1] He proceeded to broadcast his brutal killing of 51 worshippers for 16 minutes and 55 seconds on Facebook.[2] This horrific attack was carefully planned to spread rapidly across the internet. And it did. In the first 24 hours, platforms such as YouTube, Twitter, Facebook, and Reddit removed millions of copies of the video.[3] The exploitation of social media compounded the tragedy of March 15, and New Zealanders sprang into action to eliminate this type of violence and horror online. As Prime Minister Jacinda Ardern wrote, "a terrorist attack like the one in Christchurch could happen again unless we change. New Zealand could reform its gun laws, and we did. We can tackle racism and discrimination, which we must. We can review our security and intelligence settings, and we are. But we can't fix the proliferation of violent content online by ourselves."[4]

In the weeks following, Arden partnered with French President Emanuel Macron to bring together governments, technology companies, and civil society to adopt a set of commitments to eliminate terrorist and violent extremist content online, known as the Christchurch Call to Action (the Call). At the core of the Call, governments and tech companies agreed to make changes to prevent the posting of terrorist content online, to ensure its efficient and fast removal, and to prevent the use of livestreaming as a tool for broadcasting terrorist attacks.[5] To succeed, the group would need to work closely with civil society to ensure freedom of expression was protected, and the voices of the victims

[1] Royal Commission of Inquiry into the Attack on Christchurch Mosques, *He Ara Waiora: Report of the Royal Commission of Inquiry into the Attack on Christchurch Mosques on 15 March 2019,* ROYAL COMMISSION OF INQUIRY IN NEW ZEALAND, 11 (Dec. 8, 2020), https://christchurchattack.royalcommission.nz/the-report/ [https://perma.cc/J92L-DBXT].

[2] Jacinda Ardern, *How to Stop the Next Christchurch Massacre*, N. Y. TIMES (May 11, 2019), https://www.nytimes.com/2019/05/11/opinion/sunday/jacinda-ardern-social-media.html [https://perma.cc/4E29-GY7C].

[3] *Id.* Facebook alone removed over 1.5 million copies of the video within the first 24 hours.

[4] *Id.*

[5] *Id.*

and survivors heard. Emerging from this coalition was a multistakeholder initiative (MSI) designed to address a complicated problem. The answer was not one that could be solved easily through government regulation, company policies and technical measures, or civil society efforts on their own. Instead, the Call engaged a whole-of-society approach whereby stakeholders worked together to tackle the problem.[6]

Four years later, the Call remains dedicated to fulfilling the initial 25 commitments governments and companies set out on May 15, 2019.[7] Over the years, the Call has added members, partnered with similar initiatives, launched new work-streams, and adapted as technology changes. As is to be expected when addressing such complicated problems, the Call has made significant progress on some commitments and is still working on others. This article seeks to discuss why New Zealand could not stop the spread of terrorist and violent extremist content online on its own – but why it may be able to meaningfully address the problem through multistakeholder solutions. This article will explain why content moderation challenges need multistakeholder solutions and how the Call can embrace this model to achieve its goals. Additionally, this article will discuss how generative artificial intelligence (GenAI) presents challenges and opportunities to eliminate terrorist and violent extremist content online and how the multistakeholder community can consider those issues.

Before discussing why New Zealand opted for a multistakeholder approach, it is important to define the problem the Call is trying to solve to prevent future attacks. As detailed by the Royal Commission of Inquiry into the Terrorist Attack on Christchurch Mosques on March 15, 2019, the Christchurch shooter[8] displayed racist views from a young age, and

[6] Jacinda Ardern, *Here's the Model for Governing AI.* WASH. POST (Jun. 11, 2023), https://www.washingtonpost.com/opinions/2023/06/09/jacinda-ardern-ai-new-zealand-planning/ [https://perma.cc/8ZHJ-6554].

[7] *The Christchurch Call to Action: To Eliminate Terrorist and Violent Extremist Content Online*, CHRISTCHURCH CALL (May 15, 2019), https://www.christchurchcall.com/assets/Documents/Christchurch-Call-full-text-English.pdf.

[8] Following the precedent set by the Royal Commission, this article will not name the individual who committed the attack and will only refer to him as the "individual" or "Christchurch shooter" to ensure his name is not glorified, *see* Royal Commission of Inquiry into the Attack on Christchurch Mosques, *supra* note 1, Volume 1 at 11.

life experiences drove his extreme and violent behavior towards people he considered a threat.[9] He legally purchased semi-automatic firearms and evaded police scrutiny throughout his planning process.[10] To address the problems brought to light by the attacks, the Royal Commission made 44 recommendations, including updating gun laws, building inclusive societies, making improvements in intelligence sharing, and rethinking hate speech frameworks.[11] This article will only address one element of the events of March 15 – the individual's use of social media platforms to broadcast his violence and the way his terrorist and violent extremist content was able to proliferate online.

Because this article will refer to several specific issues within a much broader set of problems, it is necessary to define several terms to avoid confusion. "Extremism" is defined as a belief system held together by unwavering hostility towards a specific "out-group."[12] In line with the definition provided by the New Zealand Government, a "violent extremist" is an individual who threatens to use violence or advocates for others to use violence, in support of their own agenda or to further their own set of beliefs.[13] As such, "terrorist and violent extremist content" (TVEC) refers to hateful or objectionable material that promotes harmful extreme views, such as articles, images, speeches, or videos that encourage violence.[14] People can (and do) debate at length on how to define TVEC.[15] However, in relation to the Call and this article, the two pieces of TVEC created

[9] *Id.*

[10] *Id.*

[11] *Id.* Volume 4 at 727, Part 10: Recommendations.

[12] *See* J. M. BERGER, EXTREMISM 26-27 (2018). (Berger, an expert on extremist movements and terrorism, explains that extremism arises from a perception of "us versus them," intensified by the conviction that the success of "us" is inseparable from hostile acts against "them." Extremism differs from ordinary unpleasantness—run-of-the-mill hatred and racism—by its sweeping rationalization of an insistence on violence).

[13] Department of Internal Affairs, *Countering Violent Extremism: What is terrorist and violent extremist content?*, DEPARTMENT OF INTERNAL AFFAIRS OF NZ (2022), https://www.dia.govt.nz/Countering-Violent-Extremism-What-is-terrorist-and-violent-extremist-content#_ftn1 [https://perma.cc/P4Z8-UMRC].

[14] *Id.*

[15] Issie Lapowsky, *This Big Tech group tried to redefine violent extremism. It got messy.*, PROTOCOL (Jun. 26, 2021), https://www.protocol.com/policy/gifct-erin-saltman [https://perma.cc/ZGU4-JM7E] (interview with Erin Saltman of the Global Internet Forum to Counter

by the individual – his manifesto and the video of his attack on the mosques – would fit within any reasonable definition of TVEC. The term "online service provider," which encompasses online platforms and social media companies, is defined as an online site or service that hosts, organizes, or circulates user-generated content without producing content.[16] "Content moderation" is defined as the systems and rules online platforms use to determine how they treat user-generated content on their services.[17]

For several reasons, this article is limited to discussing the challenges online service providers face when moderating TVEC and does not discuss other types of harmful content online. First, this article is meant to analyze the work of the Call, which remains limited to TVEC.[18] Second, TVEC is an area in which there is general agreement that the content itself serves little to no societal value and should therefore be extremely restricted, if not entirely prohibited, from online platforms.[19] This agreement means that TVEC can be a useful test case for broader ongoing discussions around harmful content online, which often involves types of content such as hate speech, bullying, and dis/misinformation, where there is less agreement on definitions and societal value. Third, the challenges posed by TVEC online are as old as the internet itself and have been researched and discussed

Terrorism on a months-long debate trying to define what constituted terrorist and violent extremist content online).

[16] TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 15 (2018); *see also* Robyn Caplan, *Content or Context Moderation? Artisanal, Community-Reliant, and Industrial Approaches*, DATA & SOCIETY, 8 (Nov. 14, 2018), https://datasociety.net/library/content-or-context-moderation/ [https://perma.cc/EAQ6-QAHX].

[17] Evelyn Douek, *Content Moderation as Systems Thinking*, 136 HARV. L. REV. 526, 528 (2022) (defining "content moderation" to mean platforms' systems and rules that determine how they treat user-generated content on their services. This generally accords with Professor James Grimmelmann's definition. *See* James Grimmelmann, *The Virtues of Moderation*, 17 YALE J.L. & TECH. 42, 47 (2015) (defining "moderation" as "the governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse").

[18] *The Christchurch Call to Action, supra* note 7.

[19] There has been voluminous debate around defining both terrorism and extremism that is outside of the scope of this report. For the purposes of this report, the Christchurch shooter's 74-page manifesto and video of his attack on 15 March 2019 are both considered TVEC as they have been classified as objectionable in New Zealand.

for decades.[20] As a result, many stakeholders, including governments, the tech industry, and civil society, have attempted to address the issue over the years, which allows for a thorough examination of what has worked – and what has not – when considering the next steps for the Call.

Next, it is important to define the harm that comes from the distribution of TVEC online and the broader societal problem for which the Call is trying to solve. First, harm occurs when viewers are traumatized because of their exposure to seeing violent content.[21] Second, the sharing of TVEC causes harm as a privacy invasion of both the surviving victims and the families of deceased victims. Third, both the Christchurch video and the manifesto are harmful because they may inspire others to commit similar acts of terrorism.[22] In fact, the Christchurch shooter credited a far-right extremist attack in Norway in 2011, which killed 77 people, for inspiring his own attack.[23] Unfortunately, over the past four years, several terrorists and violent extremists have been inspired by the Christchurch attacks to livestream their killing of minorities in a variety of places, including a supermarket in Buffalo, New York and a synagogue in Poway, California.[24] Therefore, the spread of TVEC online remains a complex and multifaceted problem.

---

[20] *See* Brian Fishman, *Dual-use regulation: Managing hate and terrorism online before and after Section 230 reform*, THE BROOKINGS INSTITUTION (Mar. 14, 2023), https://www.brookings.edu/articles/dual-use-regulation-managing-hate-and-terrorism-online-before-and-after-section-230-reform/ [https://perma.cc/A7YN-RUFD].

[21] *The Christchurch Call to Action, supra* note 7.

[22] *See* Office of the New York State Attorney General Letitia James, *Investigative Report on the role of online platforms in the* tragic mass shooting in Buffalo on May 14, 2022, OFFICE OF NEW YORK STATE ATTORNEY GENERAL 17-22 (Oct. 18, 2022), https://ag.ny.gov/sites/default/files/buffaloshooting-onlineplatformsreport.pdf ("But the Christchurch shooter also changed the playbook in new, deadlier ways. He was the first white supremacist to livestream his attack, and the video of the shootings went viral. He deliberately sought to create an online footprint that he hoped would be galvanizing and instructional to fellow right-wing extremists. These digital artifacts have proved to be indelible and have radicalized others, including the Buffalo shooter, who deliberately modeled his attack on the Christchurch shooter's.").

[23] *Id.*

[24] Mariana Olaizola Rosenblat & Paul M. Barrett, *Gaming the System: How Extremists Exploit Gaming Sites and What Can Be Done to Counter Them*, NYU STERN CENTER FOR BUSINESS AND HUMAN RIGHTS, 2 (May

Finally, I want to acknowledge that online platforms bring enormous societal benefits in connecting and empowering people around the world, and undue suppression of speech is a violation of human rights. As stated in the text of the Call, companies should not have to proactively scan every piece of content before it is uploaded to the internet; that would significantly restrict freedom of expression and limit the internet's ability to act as a force of good.[25] The first line of the Call is a commitment to protecting a free, open, and secure internet which is a powerful tool to promote connectivity, enhance social inclusiveness and foster economic growth.[26] Therefore, the solutions presented in this article will hopefully strike the right balance in limiting the harms caused by the spread of TVEC online while maintaining the benefits of the openness and connectivity of the internet.

This article explores the challenges the New Zealand Government faced when trying to stop the spread of TVEC online, and why it opted for a non-regulatory solution that worked alongside tech companies and civil society. Indeed, the Call is a form of multistakeholder governance – a concept built for the 21st century and the global internet age. In the first half of the 20th century, governments increasingly relied on multilateral institutions such as the United Nations (UN) and World Trade Organization to find

2023), https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/6465b2f8be2da5102bbeb2e6/1684386554096/NYU+CBHR+Gaming_ONLINE+UPDATED+May+16.pdf ("Sure enough, copycats were quick to follow. In April 2019, a little over a month after the Christchurch tragedy, a 19-year-old male shooter opened fire at a synagogue in Poway, California, while livestreaming to his followers. One of the spectators commented during the livestream, 'get the high score' – a common phrase used among gamers. In early August of the same year, a 21-year-old man shot 23 people dead in a Walmart in El Paso, Texas. In his manifesto, he echoed the Christchurch shooter's conspiracy theory of 'white replacement,' the notion that shadowy elites are plotting to destroy white populations and culture through immigration and other policies, and mentioned a desire to live out his super soldier fantasy from the video game, Call of Duty. A month later, on Yom Kippur, another far-right militant launched a livestream on Twitch, a popular site among gamers, as he prepared to murder worshippers at a synagogue in Halle, Germany. The shooter killed two bystanders and, like those before him, left a manifesto riddled with references to far-right conspiracies couched in gaming jargon.").

[25] *The Christchurch Call to Action, supra* note 7.

[26] *Id.*

consensus on policies. Nation-state actors would then implement these multilateral agreements at home. However, rapid developments in technology and trade created multinational corporations, which gradually weakened the power of states to craft policies in isolation. Furthermore, fractures between democratic and non-democratic countries eroded the ability of global institutions like the UN to address nuanced global problems. Therefore, instead of turning to multilateral institutions, in certain circumstances, like-minded governments, corporations, and civil society collaborated to address various societal problems. These collaborations are frequently called multistakeholder initiatives (MSIs).

An MSI is created when two or more types of actors (such as governments, corporations, civil society, charitable foundations, academia, technical experts, or end-users) come together in a common governance enterprise to solve a problem defined by the group. The stakeholders collectively set procedural rules for decision-making and accountability. Within an MSI, governments, especially democratically elected governments, can be understood as agents of their citizens, corporations as agents of their owners or shareholders, and civil society as agents of their members or supporters. MSIs thrive because they allow a diverse group of participants to draw on multiple perspectives to produce better-informed solutions to complex and interdependent problems. The diversity of possible challenges and outcomes means there is no single MSI model.[27] Instead, a wide variety of multistakeholder practices are adopted to solve unique problems. Some of the first MSIs addressed labor practices in "sweatshops," environmental degradation, the trade of "blood diamonds," standards for the vitivinicultural sector, and the distribution

[27] BILL GRAHAM & STEPHANIE MACLELLAN, OVERVIEW OF THE CHALLENGES POSED BY INTERNET PLATFORMS: WHO SHOULD ADDRESS THEM AND HOW? 12 (2018), https://www.cigionline.org/static/documents/documents/Stanford%20Special%20Report%20web.pdf ("There is no single definition to describe the multistakeholder approach. It would be counterproductive to stick to a single cookie-cutter approach; instead, the approach must be adapted to suit the nature of the problem being approached and the constellation of stakeholders to be involved in finding a solution.").

of development aid.[28] One area where MSIs have flourished has been relating to global internet governance challenges.

To understand why the Call chose to create an MSI in the wake of March 15, 2019, this article explores the history of how stakeholders have attempted to govern user-generated content online. Part I provides an overview of single-sided and multistakeholder governance frameworks for moderating content online. First, this part looks at single-sided frameworks created by national governments and the tech companies themselves to address the spread of TVEC online. It will examine how governments approach content online and the range of approaches taken by national regulators. This article examines a spectrum of regulation, starting with the free-speech maximalists in the US, then looking at New Zealand and the European Union (EU) as rights-respecting regimes, and finally discussing less permissive frameworks in Turkey, Russia, and China. In the absence of clear legal frameworks, tech companies have attempted to self-regulate how they moderate content to prevent TVEC online. This part also examines the history of content moderation and self-regulatory efforts. Next, because national regulation and self-regulation have not successfully addressed the problem of TVEC online, the second half of this part explores multistakeholder models. It looks at the rise of multistakeholder governance, its history in the internet governance context, and recent multilateral efforts that could undermine multistakeholderism in internet governance.

Part II distills the lessons learned from the MSIs working on internet governance issues highlighted in Part I and builds a framework for MSIs for addressing online content governance. The first section within this part proposes a taxonomy for MSIs, breaking them down as egalitarian, consultative, restrictive, and curated. It argues that a curated MSI is the best option for the work of the Call and discusses why this format works for content governance frameworks.

[28] DOROTHÉE BAUMANN-PAULY ET AL., INDUSTRY-SPECIFIC MULTI-STAKEHOLDER INITIATIVES THAT GOVERN CORPORATE HUMAN RIGHTS STANDARDS – LEGITIMACY ASSESSMENTS OF THE FAIR LABOR ASSOCIATION AND THE GLOBAL NETWORK INITIATIVE, UNSW LAW RESEARCH PAPER NO. 2015-12 10 (Mar. 10, 2015), https://ssrn.com/abstract=2576217 [https://perma.cc/K62H-EZYS].

Part III examines how the New Zealand Government should look at the history of MSIs and key best practices when charting the future of the Call. First, Part III examines New Zealand's history and culture, which provide the foundations for the multistakeholder model. Next, it covers what happened on March 15, 2019, and the progress the Call has made in the four years since. Second, this part evaluates the progress the Call has made towards building a multistakeholder community and eliminating TVEC online while protecting a free, open, and secure internet – its two overarching objectives. Third, the part discusses the evolution and adoption of GenAI technologies and their impact on the moderation of TVEC online.

I. GOVERNANCE FRAMEWORKS FOR CONTENT MODERATION

This part outlines how different actors have attempted to govern user-generated content online. The first section looks at single-sided governance initiatives created by national regulators and tech companies themselves. National governments have applied a spectrum of approaches to regulating content online, from free speech maximalism in the United States to the hyper-censorial regime in China. In the context of global online platforms, inconsistency between, and sometimes a complete lack of, national laws, often means that companies must self-regulate content moderation practices. Therefore, this section also looks at how and why tech companies have moderated user-generated content over the past 30 years. The second part of this section discusses the rise of multi-sided content governance frameworks, starting with the history of MSIs, then how multistakeholderism has evolved in the internet governance context, and concluding with recent multilateral efforts to assert government control over online internet governance.

A. Single-Sided Content Governance Frameworks

1. National Regulatory Frameworks for Content Moderation

National governments face several challenges when trying to impose legal liability on online platforms for hosting certain types of user-generated content. The first challenge arises as the technological framework that underpins the internet was designed specifically

to circumvent governmental influence. The internet's origins date back to 1969, when it was a project of the US Government's Advanced Research Projects Agency (ARPA). The internet was initially used by government and academic institutions for research and communication purposes. Given the Cold War era context, ARPA designed the internet to withstand a nuclear attack by building a system that avoids single points of failure, encourages resiliency, scales effortlessly, and restricts government control.[29] This decentralization appealed to early internet enthusiasts, who imagined a world "free of power."[30] In 1996, John Perry Barlow, a lyricist for the Grateful Dead and co-founder of the Electronic Frontier Foundation, spoke to the need for internet users to write their own rules and disparaged government control of the technology in his "Declaration of the Independence of Cyberspace."[31] Early internet protocols were heavily influenced by cyber libertarians like Barlow, who thought that the rules governing the internet should be created and enforced by online communities – not governments.[32] As a result, technologists further built the internet to interpret overt government control or censorship as damage and route around it.[33] In effect, the early internet was a multistakeholder

[29] Cade Metz, *Paul Baran, the Link between Nuclear War and the Internet*, WIRED (Apr. 9, 2012, 10:46 AM), www.wired.co.uk/article/h-bomb-and-the-internet [https://perma.cc/XV53-7JBV].

[30] Thomas Schneider, *A vision, values, principles and mechanisms for cooperation and governance fit for purpose for the digital age*, *in* 25-29 TOWARDS A GLOBAL FRAMEWORK FOR CYBER PEACE AND DIGITAL COOPERATION: AN AGENDA FOR THE 2020S (ed. Wolfgang Kleinwächter et al., Nov. 25-29, 2019), https://www.hiig.de/wp-content/uploads/2019/11/Kleinwa%CC%88chter-Kettemann-Senges-eds.-Global-Framework-for-Cyber-Peace-2019.pdf.

[31] John Perry Barlow, *A Declaration of the Independence of Cyberspace*, ELECTRONIC FRONTIER FOUNDATION (Feb. 8, 1996), https://www.eff.org/cyberspace-independence [https://perma.cc/U8LW-LHPA].

[32] Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1616 (2018) ("In the earliest days of the internet, the regulations concerning the substance and structure of cyberspace were 'built by a noncommercial sector [of] researchers and hackers, focused upon building a network'… Balkin argued that the values of cyberspace are inherently democratic — bolstered by the ideals of free speech, individual liberty, and participation."); *citing* Jack M. Balkin, D*igital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 2 (2004) and LAWRENCE LESSIG, CODE 2.0 6 (2006).

[33] NIC SUZOR, LAWLESS: THE SECRET RULES THAT GOVERN OUR DIGITAL LIVES 28 (2019) ("'the net interprets censorship as damage and routes around it,' and when Barlow said that territorial governments had no methods of enforcement that actually worked, they were in a sense correct. The internet is fantastically hard to regulate. If your goal is to permanently remove all access to a piece of information or to prevent communications between committed, but unknown, participants, you're likely out of luck.").

collaboration championed by a diverse group of actors without regard to national territorial borders or governmental controls.

By the 1990s, the internet had evolved from a communications medium owned and operated by government and academic institutions to a global platform increasingly dominated by corporations.[34] Internet adoption accelerated in the early 1990s after British computer scientist Tim Berners-Lee created the World Wide Web, which made it easier for non-technical people to access and share information online using standard protocols, thereby creating new opportunities for businesses and individuals.[35] As it grew, the internet was governed piecemeal by a variety of voluntary standard-setting bodies that empowered private companies to perform key roles as network operators and information intermediaries.[36] Throughout the 1990s, when national governments did consider regulating the internet, they largely saw the value of e-commerce and passed robust safe-harbor protections for online platforms hosting user-generated content. These legal protections led to the creation of online service providers in the early 2000s, which rapidly scaled into behemoth global companies.[37] As billions of people came online in the 2010s,

[34] *Internet Domain Names, Part 1: Hearing Before the Committee on Science, Subcommittee on Basic Research*, 105th Cong. 6–7 (1997) (Statement of Jonathan B. Postel, Director, Computer Networks Division, University of Southern California), http://commdocs.house.gov/committees/science/hsy268140.000/hsy268140_0.HTM [https://perma.cc/D44U-UNXA].

[35] *Id.* at 6.

[36] Mark Raymond & Laura DeNardis, *Multi-stakeholderism: Anatomy of an Inchoate Global Institution*, CENTRE FOR INTERNATIONAL GOVERNANCE INNOVATION AND THE ROYAL INSTITUTE OF INTERNATIONAL AFFAIRS, Research Volume 2: Global Commission on Internet Governance: Who Runs the Internet? The Global Multi-stakeholder Model of Internet Governance, 19–45 (Nov. 2016), https://www.cigionline.org/static/documents/documents/GCIG%20Volume%202%20WEB.pdf.

[37] *See* Douek, *supra* note 17, at 552; *citing* Liat Clark, *Tim Berners-Lee: We Need to Re-Decentralise the Web*, WIRED UK (Jun. 2, 2014), https://www.wired.co.uk/article/tim-berners-lee-reclaim-the-web [https://perma.cc/4AJW-PJ4U]; Adi Robertson, *Twitter's Decentralized Social Network Project Takes a Baby Step Forward*, THE VERGE (Jan. 21, 2021), https://www.theverge.com/2021/1/21/22242718/twitter-blueskydecentralized-social-media-team-project-update [https://perma.cc/99QE-6ANL]; Mike Masnick, *Protocols, Not Platforms: A Technological Approach to Free Speech*, KNIGHT FIRST AMENDMENT INSTITUTE (2019), https://knightcolumbia.org/content/protocols-not-platforms-a-technological-approach-to-free-speech [https://perma.cc/9DXP-MNEY].

many governments became wary of the free-flowing nature of the internet and started passing new regulations which threaten to undermine the decentralized internet.[38]

Starting in the early 2010s and continuing today, governments have become increasingly interested in regulating user-generated content online. However, governments have struggled to regulate online platforms for both the technical reasons described above as well as several additional reasons. First, many governments tried to fit regulation built for traditional media onto social media, which proved ineffective. The volume of content meant governments could not just hire more lawyers, police, or judges.[39] Unlike editing a newspaper, content moderation is impossible to do perfectly at scale and legal frameworks penalizing companies for every error would be impractical to enforce.[40] Second, the speed and technological complexity of online platforms limits the states' ability to commandeer or even oversee nuanced content moderation processes.[41] Indeed, some legislative proposals have become obsolete upon enactment, as companies adopted new technology and content moderation practices. Finally, regulatory frameworks typically assume a one-size-fits all approach across a particular industry. As we will explore in the next section, there is no centralized approach to the way platforms moderate content, but rather four broad approaches: artisanal or case-by-case, community-reliant, industrial or large-scale, and no moderation whatsoever. As a result, regulators have struggled to find a legal approach for

[38] Adrian Shahbaz et al., *Freedom on the Net 2022: Countering an Authoritarian Overhaul of the Internet*, FREEDOM HOUSE (2023), https://freedomhouse.org/report/freedom-net/2022/countering-authoritarian-overhaul-internet#tracking-the-global-decline [https://perma.cc/SW9P-5JZX].

[39] SUZOR, *supra* note 33, at 98.

[40] Mike Masnick, *Masnick's Impossibility Theorem: Content Moderation At Scale Is Impossible To Do Well*, TECHDIRT. (Nov. 20, 2019, 9:31 AM), https://www.techdirt.com/articles/20191111/23032743367/masnicks-impossibilitytheorem-content-moderation-scale-is-impossible-to-do-well.shtml, [https://perma.cc/887F-A4F6].

[41] SUZOR, *supra* note 33, at 98; Douek, *supra* note 17, at 532 ("Even if there were not constitutional obstacles to substantive governmental regulation of content moderation, the sheer scale, speed, and technological complexity of the task means state actors could not directly commandeer the operations of content moderation. This is a descriptive, not normative, observation: the state simply does not have the capacity to usurp platforms as the frontline of content moderation.").

a complex industry that could be reduced to a simple one-size-fits-all checklist.[42] Moreover, legislation that divides the industry based on company size, profits, or number of users has yet to be implemented successfully.[43]

These struggles to regulate online content are found in every country, however, governments have confronted these challenges in ways that reflect their views on the freedom of expression. National regulation, therefore, falls within a wide spectrum, with the United States on one end, which allows almost all speech online, and China on the other, which closely monitors almost all speech online. As this section explores, regulation in democratic governments typically aligns with international human rights principles enshrined in Article 19 of the International Covenant on Civil and Political Rights (ICCPR).[44] Under the ICCPR, content-based restrictions on the freedom of expression are only permissible when they are clearly defined by law and are necessary and proportional to justify silencing speech – a high bar for any national law to achieve.[45] In practice, in places like the US, New Zealand and the EU where human rights are respected, laws mirroring Article 19 protections give online platforms the certainty that they can host the vast majority of user-generated content without facing legal penalties.

On the other end of the spectrum, an increasing number of governments do not adhere to Article 19 of the ICCPR when regulating content online.[46] The internet's early architecture makes it difficult to block online content based on national borders, but that

[42] Douek, *supra* note 17, at 80 ("Content moderation, like data security, 'changes too quickly and is far too dependent upon context to be reduced to a one-size-fits-all checklist.'"); *citing* Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 604 (2014); Woodrow Hartzog & Daniel J Solove, *The Scope and Potential of FTC Data Protection*, 83 GEO. WASH. L. REV. 2230 (2015).

[43] The European Union's Digital Services Act has size-based requirements, but at the time of this writing, these measures have not gone into effect. *Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC*, at 8, COM (2020) 825 final (Dec. 15, 2020).

[44] The International Covenant on Civil and Political Rights, *opened for signature* Dec. 16, 1966, art. 19, S. Exec. Doc. E, 95-2, at 29 (1978), 999 U.N.T.S. 171, 178 (entered into force Mar. 23, 1976) [hereinafter ICCPR].

[45] *Id.*

[46] Shahbaz, *supra* note 38.

has not stopped authoritarian governments from trying to force online platforms to violate the human rights principles by restricting content within their country.[47] In recent years, technological advances have provided governments with solutions to break their citizens away from the global internet and control online spaces.[48] This fragmentation is sometimes referred to as the "splinternet."[49] The splinternet disrupts the previously global internet and replaces it with a system in which the internet is experienced differently by users across national jurisdictions.[50] The splinternet encompasses government restrictions on the flow of news and information, centralized state control over internet infrastructure, and barriers to cross-border transfers of user data.[51] Unfortunately, new technologies and increasing authoritarianism have led to a steady decline of global internet freedoms for the past 12 years.[52]

This next section highlights a few countries across the spectrum of national regulatory efforts: the United States, being the most-speech protective, then New Zealand, then the European Union, which has a rights-respecting framework but has passed copious amounts of legislation regulating content governance online. The section then provides examples from national regimes that subvert the protections of Article 19 of the ICCPR. There are dozens of countries that fit into this category, but this section will discuss three: Turkey, Russia, and China. Highlighting these regimes is important because China and Russia have long sought to displace the multistakeholder model of internet governance with one that promotes greater control by multilateral institutions.[53] Both countries have attempted to leverage the United Nations to endorse the right of each state to control its own "national segment of the internet."[54] As we will explore in Part II, MSIs

[47] SUZOR, *supra* note 33, at 38.
[48] Shahbaz, *supra* note 38.
[49] Dan York, *What Is a Splinternet? And Why You Should Be Paying Attention*, INTERNET SOCIETY (Mar. 23, 2022), https://www.internetsociety.org/blog/2022/03/what-is-the-splinternet-and-why-you-should-be-paying-attention/ [https://perma.cc/E6QW-HQ4F].
[50] Suzor, *supra* note 33 at 87.
[51] Shahbaz, Funk & Vesteinsson *supra* note 38.
[52] *Id.*
[53] *See Id.*
[54] *Id.*

are frequently created to fill "governance gaps," and these examples will illustrate where gaps may occur within national regulatory frameworks. In many cases, like-minded national governments will work together in multilateral or multistakeholder settings to address technological challenges. However, due to the dramatic variance of legal frameworks outlined in the next subsections, many democratic governments are unable to partner with authoritarian regimes without compromising fundamental human-rights values.

(a) The United States

The United States is undoubtedly a global outlier in its approach to free speech protections. Understanding the US legal framework is critical when discussing internet regulations, because most large global online platforms hosting user-generated content are headquartered in the US. Overwhelmingly, global online platforms were founded by US employees who built US speech values into their content moderation systems.[55] These global systems became further entrenched into the US system by US lawyers who used US legal principles to craft the global terms of service policies that dictate what a user can or cannot post on the online platform. Therefore, understanding the US system is critical for all other content governance analysis.

In the US, there are two foundational laws regarding the regulation of speech online: the First Amendment of the US Constitution and Section 230 of the Communications Decency Act. The First Amendment states that Congress shall pass no law abridging the freedom of speech and broadly protects citizens against government censorship.[56] A small

[55] Klonick, *supra* note 32 at 1621 ("A common theme exists in all three of these platforms' histories: American lawyers trained and acculturated in American free speech norms and First Amendment law oversaw the development of company content-moderation policy. Though they might not have "directly imported First Amendment doctrine," the normative background in free speech had a direct impact on how they structured their policies … Simultaneously, there were complicated implications in trying to implement those American democratic cultural norms within a global company.").

[56] U.S. CONST. amend. I ("Congress shall pass no law . . . abridging the freedom of speech.").

number of exceptions allow the government to restrict speech, including in the cases of child sexual abuse material, fraud, obscenity, incitement to violence, speech integral to illegal conduct, speech violating intellectual property law, true threats, commercial speech, and defamation.[57] Americans are fiercely protective of their "free speech culture" and courts have strongly protected this individual right.[58] As a result, many types of speech that are restricted internationally are constitutionally protected in the US. For example, content that is published by or about terrorists or extremists would be prohibited in many jurisdictions but is protected by the First Amendment as long as the content does not imminently incite violence.[59] The First Amendment applies only to the Government's restrictions on speech and does not obligate a company to allow all constitutionally protected speech on its platform. Indeed, the First Amendment protects private actors from government efforts to control speech, and the government is not allowed to compel an online platform to restrict, remove, or promote speech.[60]

In addition to the protections under the First Amendment, online platforms also benefit from the legal framework Congress created in Section 230 of the Communications Decency Act of 1996. Congress passed the Communications Decency Act to regulate pornographic material on the internet.[61] One year after passage, the Supreme Court overturned the law

[57] *See First Amendment Overview*, CORNELL L. SCHOOL LEGAL INFO. INST., https://www.law.cornell.edu/constitution/first_amendment#:~:text=The%20First%20Amendment%20guarantees%20freedoms,restricting%20an%20individual's%20religious%20practices [https://perma.cc/JXY4-RCV8]; Genevieve Lakier, *The Non-First Amendment Law of Freedom of Speech*, 134 HARV. L. REV. 2299, 2301 (2021).

[58] *See* Genevieve Lakier, *The Non-First Amendment Law of Freedom of Speech*, 134 HARV. L. REV. 2299, 2301 (2021) ("[t]he Speech Clause of the First Amendment has for decades now served as one of the most powerful mechanisms of individual rights protection in the entire federal Constitution."); *See also* Douek, *supra* note 42, at 34 ("in content moderation, the idea of prioritizing the overall functioning of the system over individual rights is dissonant with the story American society tells itself about its free speech culture.").

[59] Eric Goldman, *The United States' Approach to 'Platform' Regulation*, SANTA CLARA UNIV. LEGAL STUDIES (2023).

[60] Fishman, *supra* note 20.

[61] 47 U.S.C. § 230 (2018).

for violating the First Amendment, as it was overly broad in restricting speech.[62] However, the Court upheld the safe harbor provisions for online service providers covered in Section 230. Sometimes referred to as the "26 words that created the Internet,"[63] Section 230(c)(1) enables online platforms to host user-generated content without being held legally responsible for speech posted on their platforms by users.[64] Section 230(c)(2) empowers platforms to find and remove material they deem objectionable content without fear of legal action from users. As such, it is sometimes referred to as the "Good Samaritan" provision of the law.[65] There are several carve-outs to Section 230 protections for internet service providers, including where the platform materially contributes to criminal behavior, intellectual property claims, and promotions of sex trafficking and commercial sex.[66] Section 230 provides broad immunity for social media companies to host user-generated content and moderate that content as they see fit, as long as they do not significantly develop the content themselves.[67]

The protections under the First Amendment and the immunities granted by Section 230 work together to allow US online platforms to experiment with the type of content moderation that works best for their audience. In practice, if a social media company is sued for its content moderation decisions, it could assert a First Amendment defense, but Section 230 acts as a "procedural fast lane" to resolve litigation more quickly and cheaply.[68] The Section 230 "fast lane" made it possible for anyone to start a company and hosts user-generated content without being liable for what their users say or share.[69] This drove investment in the industry, particularly in Silicon Valley. Eric Goldman, a world-leading internet scholar, has called Section 230 a "globally unique solution" which has

[62] *See* Reno v. ACLU, 521 U.S. 844, 865 (1997).

[63] JEFF KOSSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET (2019).

[64] 47 U.S.C. § 230(c)(1) (2018).

[65] *See* Zeran v. Am. Online, Inc., 129 F.3d 327, 330 (4th Cir. 1997) (noting that the purposes of intermediary immunity in § 230 were not only to incentivize platforms to remove indecent content but also to protect the free speech of platform users).

[66] 47 U.S.C. § 230(e) (2018).

[67] Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1177 (9th Cir. 2008).

[68] Eric Goldman, *Why Section 230 Is Better Than the First Amendment*, 95 NOTRE DAME L. REV. REFLECTION 34, 39 n. 50 (2019).

[69] *Id.* at 33.

given the United States a competitive advantage when it comes to the internet.[70] As a result, the US is home to a wide diversity of online platforms that moderate user-generated content to serve different audiences including Reddit, Airbnb, Wikipedia, Yelp, and Etsy.

For over 20 years, Section 230 remained unchanged. Then, in 2018, Congress added a new carve-out to the law with the passage of the Allow States and Victims to Fight Online Sex Trafficking Act and the Stop Enabling Sex Trafficking Act, known as FOSTA-SESTA or just FOSTA.[71] Leading up to the passage of FOSTA, Backpage.com was sued by victims of sex trafficking who claimed the website had helped facilitate the criminal activity they experienced.[72] These lawsuits were dismissed by the courts, which convinced the trafficking victims to lobby Congress for an amendment to Section 230 related to promotion of sex trafficking and commercial sex.[73] FOSTA's passage was a turning point for Section 230, as it catapulted the relatively unknown and uncontroversial limited liability provisions for websites into the center of a national debate around the power of "big tech" companies. Five years later, this debate continues, without any political consensus on how to resolve it. While Democrats are pushing companies to restrict speech that is hateful or bullying, Republican states, including Florida and Texas, have passed laws requiring social media companies to leave up all constitutionally protected speech.[74] These laws are embroiled in litigation and likely to end up before the Supreme Court.[75]

[70] Caplan, *supra* note 16 at 27 (quoting comments made by Eric Goldman at the Content Moderation at Scale Conference in Washington, D.C., on May 7, 2018), https://datasociety.net/library/content-or-context-moderation/ [https://perma.cc/2ECK-JZTU].

[71] 47 U.S.C. § 230(e) (2018); Allow States and Victims to Fight Online Sex Trafficking Act of 2017, Pub. L. No. 1115-164, 132 Stat. 1253 (codified as amended in scattered sections of 18 and 47 U.S.C.) (2018).

[72] Eric Goldman, *The Complicated Story of FOSTA and Section 230*, 17 FIRST AMEND. L. REV. 279 (2019).

[73] *Id.*; *E.g.*, Backpage.com, LLC v. Dart, 807 F.3d 229 (7th Cir. 2015); Jane Doe No. 1 v. Backpage.com, LLC, 817 F.3d 12 (1st Cir. 2016); Backpage.com, LLC v. Cooper, 939 F. Supp. 2d 805 (M.D. Tenn. 2013); Backpage.com, LLC v. Hoffman, 2013 WL 4502097 (D.N.J. Aug. 20, 2013); Backpage.com, LLC v. McKenna, 881 F. Supp. 2d 1262 (W.D. Wash. 2012); M.A. ex rel. P.K. v. Vill. Voice Media Holdings, LLC, 809 F. Supp. 2d 1041 (E.D. Mo. 2011).

[74] Tex. H.B. 20 (Tex. 2021); Fla S.B. 7072, 2021 Leg. (Fla. 2021).

[75] *See* NetChoice, LLC, v. Paxton, 2023 U.S. LEXIS 4138 (2023).

Until then, divisive partisanship has entrenched a legislative stalemate and no federal laws related to Section 230 have passed since 2018.

Potentially due to this legislative stalemate, in April of 2021, Supreme Court Justice Clarence Thomas wrote a concurrence on the dismissal of a case relating to internet policies attacking Section 230 and the powers of the First Amendment.[76] As part of Thomas's concurrence he invited lawyers to bring cases challenging Section 230 to court.[77] A year later, the Supreme Court granted certiorari to two cases relating to the culpability of social media companies for a deadly Islamic State attack, which the perpetrators discussed on their platforms.[78] The family members of victims who died in an ISIS attack in Europe presented their case to the Court in February of 2023, arguing that Twitter, Facebook, and YouTube should be held liable because of ISIS's general presence on their platforms.[79] In May 2023, the Court dismissed the cases, stating that the social media companies did not provide knowing or substantial assistance to ISIS necessary to find them culpable under the Anti-Terrorism Act."[80] However, the Court expressly declined to rule on the Section 230 issues, including on whether the law applies to algorithmic promotion of content, leaving in place the broad scope of Section 230.[81]

(b) New Zealand

The next legal framework on our spectrum is that of New Zealand, which has enshrined legal provisions aligning with Article 19 of the ICCPR in the New Zealand Bill of Rights Act 1990 and the Human Rights Act 1993. These laws guarantee the right to freedom of thought, conscience, and religion, including the freedom to hold opinions without

[76] Mark MacCarthy, *Justice Thomas Sends a Message on Social Media Regulation*, THE BROOKINGS INST. (Apr. 9, 2021), https://www.brookings.edu/blog/techtank/2021/04/09/justice-thomas-sends-a-message-on-social-media-regulation/ [https://perma.cc/8XSA-59XF].

[77] Bobby Allyn, *Justice Clarence Thomas Takes Aim At Tech And Its Power 'To Cut Off Speech,"* NATIONAL PUBLIC RADIO (Apr. 5, 2021), https://www.npr.org/2021/04/05/984440891/justice-clarence-thomas-takes-aims-at-tech-and-its-power-to-cut-off-speech [https://perma.cc/6LWD-Z3WJ].

[78] Twitter, Inc. v. Taamneh, 143 S. Ct. 1206, 1210 (2023).

[79] *Id.*

[80] *Id.* at 1231.

[81] *Id.*

interference and to seek, receive, and impart information and ideas of all kinds. However, in New Zealand, freedom of expression is not absolute. There are certain limitations and restrictions, including on speech that incites violence, hatred, or discrimination; defamation; harassment; and copyright infringement.[82] This right is also limited under the Summary Offences Act 1981, which prohibits threatening or violent speech.[83] Unlike the United States and many democratically governed countries, New Zealand does not have a legal regime that specifically provides safe harbor protections for online intermediaries hosting user-generated content. Instead, it has a patchwork of laws governing content moderation, hate speech, and the distribution of TVEC online. Four primary statutes impose liability on social media companies hosting objectionable speech: the Summary Offences Act 1981, the Harmful Digital Communications Act 2015, the Broadcasting Act 1989, and the Films, Videos, and Publications Classification Act 1993.[84]

In the context of assessing user-generated content posted online, New Zealand has two statutes regulating content. First, the Broadcasting Act 1989 sets standards for traditional media 'broadcasters', but some standards apply online. [85] Second, the Films, Videos, and Publications Classifications Act creates a consumer advisory system for age suitability and warnings for content in "films". It also specifies what "publications" are illegal (or "objectionable") for distribution across mediums in New Zealand.[86] It was under this Act that the Christchurch shooter's video and manifesto were deemed objectionable in the days immediately following the attack. New Zealand's Chief Censor "called in" the livestream video and manifesto for classification, and the office decided to ban the materials on March 20 and 23, respectively.[87] This designation made it illegal to hold or distribute the video or manifesto. To comply with this legal restriction, many online

[82] Human Rights Act 1993 (N.Z.).

[83] Royal Commission of Inquiry into the Attack on Christchurch Mosques, *supra* note 1, at Part 9, chapter 4.

[84] *Id.*

[85] Broadcasting Act 1989 (N.Z.).

[86] Films, Videos, and Publications Classification Act 1993 (N.Z.).

[87] David Shanks, *Classification Office Response to the March 2019 Christchurch Terrorist Attack,* CLASSIFICATIONS OFF. (Dec. 9, 2020), https://www.classificationoffice.govt.nz/news/news-items/response-to-the-march-2019-christchurch-terrorist-attack/ [https://perma.cc/4QSY-SX9N].

platforms now work with third-party hash-sharing systems to automatically detect and remove this content, as discussed in the next section.[88]

Next, New Zealand's Harmful Digital Communications Act of 2015 regulates issues such as cyberbullying, harassment, and other forms of harmful online behavior.[89] It defines harmful digital communications as those that are threatening, intimidating, or otherwise harmful to an individual, and that are made using a digital communication device, such as a computer, smartphone, or social media platform.[90] To help enforce these rules and settle disputes with companies, the Act has an "approved agency" receive and investigate complaints about harmful digital communications.[91] The current approved agency is Netsafe, a non-profit entity that receives funding from the Ministries of Justice and Education and assists victims exposed to harmful digital content.[92] Netsafe works closely with technology companies to resolve these complaints, and with the Police and the Department of Internal Affairs, which have set up separate processes.

In May 2021, New Zealand's Government initiated the Content Regulation Review to align some of the statutory obligations of internet media companies with those of their traditional media counterparts.[93] This review is unlikely to be finalized before the October 2023 election. However, while it is under way, large online service providers in New Zealand, including Meta and Google, have worked with the New Zealand Tech Alliance to create the Aotearoa New Zealand Code of Practice for Online Safety and Harms, which provides guidance to companies on how to enhance safety and mitigate harm online.[94]

---

[88] *See Id.*

[89] Harmful Digital Communications Act of 2015 (N.Z.).

[90] *Id.*

[91] *Id.*

[92] *About Netsafe*, NETSAFE (2023), https://netsafe.org.nz/aboutnetsafe/partners/ [https://perma.cc/Z9HV-P9TL]

[93] *Media, and Online Content Regulation,* N.Z. DEPT. OF INTERNAL AFF., (Jun. 1, 2023), https://www.dia.govt.nz/media-and-online-content-regulation [https://perma.cc/FH66-WA4T].

[94] Curtis Barnes, Tom Barraclough, & Allyn Robins, *Platforms Are Testing Self-Regulation in New Zealand. It Needs a Lot of Work*, LAWFARE (Sep. 2, 2022), https://www.lawfaremedia.org/article/platforms-are-testing-self-regulation-new-zealand-it-needs-lot-work [https://perma.cc/X42Z-W5BQ].

Launched in July 2022, this self-regulatory Code of Practice requires companies to make "best efforts" towards a set of commitments that will reduce harmful content, increase transparency, and empower users.[95] "When a company signs onto the code's framework, it identifies which of its products the code will apply to, and can further choose to opt out of any measures" it feels are not relevant to the company's products.[96] Critics have argued that this is an attempt by the companies to "pre-empt regulation" and that the effort lacks legitimacy and community accountability.[97] However, due to the lack of explicit legal provisions regulating social media in New Zealand, many companies have experimented with this type of self-regulatory mechanism.

In June 2023, the Department of Internal Affairs put forward a discussion document on their proposal to regulate online platforms.[98] The document acknowledges that it can be difficult for citizens to navigate the five industry complaint bodies they can approach if they feel content is unsafe or breaches the company's terms of service.[99] The proposed regulation would create "codes of practice" which set out specific safety obligations for larger or riskier platforms and would be enforceable by an independent regulator.[100] This new independent industry regulator would provide a "clear home for consumer safety on online platforms," and industry groups would develop new codes with "input from and approval by the regulator."[101] The Department is accepting feedback on its policy proposals until July 31, 2023.[102]

---

[95] *Id.*

[96] *Id.*

[97] *Id.*; *see also* Tom Pullar-Strecker, *Social Media Firms Advance NZ's Controversial 'World First' Code of Conduct*, STUFF (Apr. 1, 2023), https://www.stuff.co.nz/business/131613136/social-media-firms-advance-nzs-controversial-world-first-code-of-conduct [https://perma.cc/3SVF-UGG9].

[98] New Zealand Department of Internal Affairs, *Discussion Document, Safer Online Services and Media Platforms*, DEPARTMENT OF INTERNAL AFFAIRS NZ (Jun. 2023), https://www.dia.govt.nz/diawebsite.nsf/Files/online-content-regulation/$file/Safer-Online-Services-and-Media-Platforms-Discussion-Document-June-2023.pdf [https://perma.cc/MEF2-EL2Q].

[99] *Id.*

[100] *Id.*

[101] *Id.*

[102] *Id.* (This report will be published on August 1, 2023, and will therefore not detail the outcome of the proposal.).

(c) The European Union

Next on the spectrum of national regulatory frameworks governing online platform liability of user-generated content is the European Union. The EU has several rights-based restrictions on speech and legal liability frameworks set out in national laws and EU-level regulations and directives. Freedom of expression is codified in Article 10 of the European Convention on Human Rights, which has been incorporated into EU law through the Charter of Fundamental Rights of the European Union.[103] This framework mirrors the ICCPR, mentioned above. However, until 2000, EU member states took different approaches to regulating content online. In most cases, online speech was subject to the same legal framework that applied to traditional media such as newspapers, television, and radio – with a great deal of variation between member states.[104] This patchwork approach created legal uncertainty for online platforms and threatened the growing e-commerce industry. As such, in 2000, the EU passed the e-Commerce Directive, which created a safe harbor for online intermediaries like the one found in Section 230, adding in a caveat that illegal content be removed "expeditiously."[105]

While the EU and US frameworks mirror each other in form and function, the definitions of "illegal" speech vary greatly. In the US, "illegal" speech exists only under the limited carve-outs of the First Amendment, and Section 230 immunity ensures that if illegal content is posted online, the user who posted the content is liable and not the platform itself. This is not the case in the EU, where member states have passed several regulations increasing liability for online intermediaries. Under the EU framework, national regulators are able to define broad categories of speech as "illegal" because there is less of a

[103] European Convention for the Protection of Human Rights and Fundamental Freedoms, *opened for signature* Nov. 4, 1950, art. 10, Euro. T. S. No. 5, 213 U.N.T.S. 221 (entered into force Sept. 3, 1953).

[104] Alexandre De Streel & Martin Husovec, *The E-commerce Directive as the Cornerstone of the Internal Market, Assessment and Options for Reform,* EUR. PARL. (May 2020), https://www.europarl.europa.eu/RegData/etudes/STUD/2020/648797/IPOL_STU(2020)648797_EN.pdf [https://perma.cc/V2E2-EBFH].

[105] Pablo Baistrocchi, *Liability of Intermediary Service Providers in the EU Directive on Electronic Commerce*, 19 SANTA CLARA HIGH TECH. L.J. 111, 111–21 (2002).

presumption against speech restrictions.[106] As a result, over the past 20 years, the EU has enacted a wide range of rules making types of speech illegal, ranging from the right to be forgotten found in the General Data Protection Regulation to hate speech laws in Germany under the Netzwerkdurchsetzungsgesetz (commonly known as NetzDG) and the restrictions on harmful speech passed recently in the Digital Services Act (DSA). Scholarly analysis of these laws will fill hundreds of textbooks; this section will only detail the regulations surrounding TVEC online. The TVEC regulations not only provide a helpful insight into the rulemaking process for content moderation more broadly in the EU; they are also relevant to the work of the Christchurch Call and this article.

As related to TVEC, the safe-harbor provisions for online platforms found in the 2000 e-Commerce Directive started to erode in 2008 after laws implementing the EU's counter-terrorism agenda were updated to criminalize the incitement to terrorism online.[107] These updates included requirements for internet platforms to cooperate with law enforcement to receive safe harbor protections.[108] The EU first explored multistakeholder options to assist online platforms with this work, including the creation of the Radicalization Awareness Network, in 2011, which provides guidance to policymakers from civil society organizations working to prevent and counter radicalization.[109] After a spate of deadly terror attacks and hate crimes in 2015, European regulators began to place more

[106] Danielle Keats Citron, *Extremist Speech, Compelled Conformity, and Censorship Creep*, 93 NOTRE DAME L. REV. 1035, 1038 (2018).

[107] Santina Musolino, *EU Policies for Preventing Violent Extremism: A New Paradigm for Action?*, REVISTA CIDOB D'AFERS INTERNACIONALS, no. 128, Sept. 2021, at 39 ("The amendment of the Framework Decision 2002/475/JHA53 in 2008 added several more activities to the list of those already criminalized and shifted the focus on criminalizing preparatory acts and incitement to terrorism. Moreover, it stressed the importance of reconsidering the potentialities of a preventive action. The adoption of the EU Internal Security Strategy in Action in 2010 and the creation, in 2011, of the EU Radicalization Action Network outlined the importance of creating a network connecting first-line experts from various EU member states.").

[108] *Id.* at 44.

[109] *Id.* at 141-42; *See also About RAN Practitioners*, EUR. COMM'N (2023), https://home-affairs.ec.europa.eu/networks/radicalisation-awareness-network-ran/about-ran_en [https://perma.cc/WAS8-KAWS].

responsibility on social media companies for the violence.[110] In December 2015, the EU created the EU Internet Forum to bring together tech platforms, law enforcement authorities, and civil society to reduce the availability of terrorist material online through programs like the EU Internet Referral Unit.[111] As part of this work, in 2016, EU regulators worked with tech companies to create a Voluntary Codes of Conduct to remove illegal hate speech, including terrorist content.[112] Under the Code of Conduct, companies agreed to voluntarily comply with any requests from the EU Internet Referral Unit and remove content within 24 hours.[113] This new framework faced significant backlash from civil society, which had been excluded from the conversation and viewed the arrangement as both overreaching and censorial because it required companies to remove speech without questioning the validity of the government's request.[114]

The EU issued its first assessment of the Voluntary Code of Conduct in December 2016, which criticized the online platform's "success rate" at actioning removal requests.[115] Moreover, EU lawmakers deemed self-regulation attempts by the online platforms to be

[110] Citron, *supra* note 106, at 1040; *See also* Lizzie Plaugic, *France Wants to Make Google and Facebook Accountable for Hate Speech*, THE VERGE (Jan. 27, 2015), https://www.theverge.com/2015/1/27/7921463/google-facebookaccountable-for-hate-speech-france [https://perma.cc/WG7L-K55J].

[111] *Migration and Home Affairs, Terrorist Content Online,* EUR. COMM'N (2023), https://home-affairs.ec.europa.eu/policies/internal-security/counter-terrorism-and-radicalisation/prevention-radicalisation/terrorist-content-online_en [https://perma.cc/V46W-ZM9Q].

[112] *EU Internet Forum: Bringing Together Governments*, *Europol and Technology Companies to Counter Terrorist Content and Hate Speech Online,* EUR. COMM'N (Dec. 3, 2015), http://europa.eu/rapid/press-release_IP-15- 6243_en.htm [https://perma.cc/ZBD7-RVKC].

[113] Citron, *supra* note 106, at 1038.

[114] Citron, *supra* note 110 at 1041 ("Although civil society organizations participated in early meetings held by the European Internet Forum, they were excluded from the negotiations that resulted in the Code. *EDRi and Access Now Withdraw from the EU Commission IT Forum Discussions*, EDRi (May 31, 2016), https://edri.org/edri-access-now-withdraw-eu-commissionforum-discussions. As the civil society group European Digital Rights (EDRi) explained, the European Commission refused to give the groups access to the negotiations and drafts of the agreement. Maryant Fernandez Perez, *New Documents Reveal the Truth Behind the Hate Speech Code*, EDRi (Sep. 7, 2016), https://edri.org/new-documents-reveal-truth-behindhate-speech-code; Jennifer Baker, *Europol's Online Censorship Unit Is Haphazard and Unaccountable Says NGO,* ARS TECHNICA (Jul. 4, 2016), https://arstechnica.com/tech-policy/2016/07/europol -iru-extremist-content-censorship-policing/ [https://perma.cc/U3MX-2EGW].

[115] *Id.* at 1042

insufficient and decided to impose legal measures to combat terrorist radicalization online.[116] In 2018, as part of an update to the Audio Visual Media Services Directive, the EU compelled member states to pass laws that would prevent the upload and dissemination of harmful material, including terrorist content.[117] Despite these changes, the EU again updated its laws again in the Regulation on Preventing the Dissemination of Terrorist Content Online (TCO) in 2021.[118] The TCO requires online platforms to remove terrorist content within one hour of receiving a removal order from a competent authority in an EU member state or face a fine of up to 4 per cent of their total revenue.[119] The TCO received significant pushback from civil society organizations for three reasons. First, civil society worried that over classification by law enforcement, combined with the tight timeline, would stifle freedom of expression.[120] Second, civil society actors noted that the TCO seems to be in conflict with the EU's ePrivacy Directive, which limits the ability of platforms to scan more private surfaces for terrorist material.[121] Finally, civil society argued that the regulation grants national governments too much power to order the

[116] *Id.* at 1042.

[117] *The Online Regulation Series – European Union (update 2021)* TECH AGAINST TERRORISM (Dec. 2021), https://www.techagainstterrorism.org/2021/12/10/the-online-regulation-series-european-union-update/ [https://perma.cc/C8L2-ENKD].

[118] Jan Penfrat, *Digital Services Act, The EDRi guide to 2297 Amendment Proposals*, EUR. DIGITAL RIGHTS (Oct. 2021), https://edri.org/wp-content/uploads/2021/10/EDRi-policy-paper-Digital-Services-Act-Nov-2021.pdf [https://perma.cc/TT2G-GUYB].

[119] Clothilde Goujard, *Online Platforms Now Have an Hour to Remove Terrorist Content in the EU*, POLITICO (Jun. 7, 2022), https://www.politico.eu/article/online-platforms-to-take-down-terrorist-content-under-an-hour-in-the-eu/ [https://perma.cc/FPW6-UMQ7]; *see also Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online*, EUR. PARL. (Apr. 29, 2021), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32021R0784 [https://perma.cc/X9YW-U7GD].

[120] EDRi, *Terrorist Content Regulation: Document Pool*, EDRI (Jan. 21, 2019), https://edri.org/our-work/terrorist-content-regulation-document-pool/ [https://perma.cc/8J4U-ECQ6] ("A major concern for the functioning and freedom of the internet is the extension of the upload filter regime the EU is currently about to introduce for copyright to terrorist content. Requiring internet companies to monitor everything we say on the web does not only have grave implications for the freedom of speech, but it also follows a dangerous path of outsourcing and privatizing law enforcement.").

[121] Fishman, *supra* note 20 ("The Terrorism Content Online regulation focuses on removing public material supporting terrorism, while the ePrivacy Directive limits the ability of platforms to scan more private surfaces for terrorist material.").

removal of speech with only minimal judicial oversight.[122] The TCO went into effect in July 2022, and there has yet to be much reporting from national authorities as to how they are implementing the regulatory tools.[123]

The other notable piece of EU legislation regarding content moderation more broadly, is the newly enacted DSA. The DSA is a sweeping legislative effort to "create safer digital space in which the fundamental rights of all users of digital services are protected."[124] While the DSA does not replace the TCO, the new rules in the DSA cover detection, flagging, and removal of "illegal content" as defined by either the member states or the EU itself.[125] For online platforms, compliance with the DSA will be extraordinarily challenging as new measures include: updating user safeguards, creating transparency and oversight processes, bans on advertising, and additional liability regimes.[126] Given the complexity of the DSA, its broader impact on content moderation and the future of MSIs will be hard to assess for years to come as pieces of the DSA are implemented both at the member state and the EU level.[127] With the DSA, the EU has increased the liability of online platforms in ways that might make them less likely to try new voluntary initiatives.

---

[122] *Id.* ("Although companies do have the ability to appeal such orders, only a few companies are likely to have the legal capacity to file such appeals at scale and they will take months, if not years, to adjudicate. The regulation effectively grants national governments extraordinary latitude to order the removal of speech with minimal judicial oversight.").

[123] *Regulation (EU) 2021/784 of the European Parliament and of the Council of April 29, 2021, on addressing the dissemination of terrorist content online, supra* note 119.

[124] *The Digital Services Act package, Shaping Europe's Digital Future*, EUR. COMM'N (Sept. 25, 2023), https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package [https://perma.cc/QAQ9-YTTE].

[125] *Questions and Answers: Digital Services Act*, EUR. COMM'N (Apr. 25, 2023), https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_2348 [https://perma.cc/B4NB-QW37] ("What constitutes illegal content is defined in other laws either at EU level or at national level – for example terrorist content or child sexual abuse material or illegal hate speech is defined at EU level. Where a content is illegal only in a given Member State, as a general rule it should only be removed in the territory where it is illegal.").

[126] Daphne Keller, *The EU's new Digital Services Act and the Rest of the World*, VERFASSUNGSBOLG (Nov. 7, 2022), https://verfassungsblog.de/dsa-rest-of-world/ [https://perma.cc/E3T2-2EPC].

[127] *Id.*

(d) Turkey

As Turkey is becoming increasingly less free, it is next on the spectrum of national regulations governing user-generated content. Turkey is a democratic regime that has become more authoritarian in recent years by passing restrictive speech laws, heavily monitoring speech online, and increasingly threatening online platforms like Wikimedia and Twitter.[128] Starting in 2016, the country has implemented several laws that allow for the censorship of online content, including the 2016 Law on the Regulation of Publications on the Internet and Suppression of Crimes Committed by Means of Such Publication.[129] This law grants authorities the power to block websites and social media accounts that are deemed to be harmful to national security or public order. Online criticism of the government or the president can result in prosecution, and many journalists and social media users have been arrested for their online activities. Additionally, the government has required social media companies to establish local offices in Turkey and to comply with government requests to remove content.[130] In 2022, lawmakers went a step further, ahead of upcoming elections, enacting new amendments which gave the government power to implement severe penalties against tech companies for failure to comply with take-down requests, ensuring companies will be complicit in censorship.[131] In the days before the 2023 election, Twitter restricted access to content in

[128] *Freedom in the World 2023: Turkey*, FREEDOM HOUSE (2023), https://freedomhouse.org/country/turkey/freedom-world/2023 [https://perma.cc/ZRH5-S9SB].

[129] *The Online Regulation Series – Turkey*, TECH AGAINST TERRORISM (Oct. 23, 2020), https://www.techagainstterrorism.org/2020/10/23/the-online-regulation-series-turkey [https://perma.cc/2V5U-J3JW] ("The Regulation of Publications on the Internet and Suppression of Crimes Committed by means of Such Publication, 2007, widely known as the "Internet Law 5651" or "Law No. 5651." This regulates prohibited content, such as child abuse images and obscenity, on the Internet and enables the blocking of websites.").

[130] *See Freedom in the World 2023: Turkey*, *supra* note 128.

[131] *Turkey: Dangerous, Dystopian New Legal Amendments*, HUMAN RIGHTS WATCH (Oct. 14, 2022), https://www.hrw.org/news/2022/10/14/turkey-dangerous-dystopian-new-legal-amendments [https://perma.cc/X828-ZZQT].

Turkey at the request of the government.[132] Regarding the restriction of TVEC online, the Turkish state has adopted a very broad definition of terrorism that increasingly covers peaceful acts of dissidence.[133] Unfortunately, Turkey's legal framework no longer complies with many of the provisions in Article 19 of the ICCPR.

(e) Russia

The second most restrictive national regulatory system on our spectrum is Russia. However, most commentators agree that Russia would be just as restrictive as China, if it had the technological capabilities to enact those restrictions.[134] In Russia, the government has the power to block websites if the state deems the content as extremist or harmful to the country's security or sovereignty.[135] Many global online service providers proactively left the Russian market in 2019, when Russia introduced a law that required all online communications to be stored for six months and made accessible to the government upon request.[136] In 2022, the government used this law to issue massive fines on platforms that refused to remove content and localize user data.[137] After Russia's invasion of Ukraine, authorities passed more restrictive legislation that granted more powers to state bodies tasked with regulation of the internet, expanded the grounds for what content could be deemed illegal, and required media outlets to refer to the war as a "special military

[132] Ashley Belanger, *Musk Defends Enabling Turkish Censorship on Twitter, Calling It His "Choice",* ARS TECHNICA (May 15, 2023), https://arstechnica.com/tech-policy/2023/05/musk-defends-enabling-turkish-censorship-on-twitter-calling-it-his-choice/ [https://perma.cc/57HH-4SDF].

[133] Nazli Ozekici, *Turkey's Broad Definition of Terrorism Does Nothing to Halt Radicalisation*, OPENDEMOCRACY (Jan. 20, 2022), https://www.opendemocracy.net/en/north-africa-west-asia/turkeys-broad-definition-of-terrorism-does-nothing-to-halt-radicalisation/ [https://perma.cc/M5P7-6JV6].

[134] *Russia Is Trying to Build Its Own Great Firewall*, THE ECONOMIST (Feb. 19, 2022), https://www.economist.com/business/russia-is-trying-to-build-its-own-great-firewall/21807706 [https://perma.cc/4679-NJMJ].

[135] *Freedom in the World 2023: Russia*, FREEDOM HOUSE (Mar. 9, 2023), https://freedomhouse.org/country/russia/freedom-world/2023 [https://perma.cc/7BEH-LXCU].

[136] *Russia: Growing Internet Isolation, Control, Censorship,* HUMAN RIGHTS WATCH (June 18, 2020), https://www.hrw.org/news/2020/06/18/russia-growing-internet-isolation-control-censorship [https://perma.cc/4L4M-MX8H].

[137] *Freedom in the World 2023: Russia, supra* note 135.

operation."[138] While Russia is still connected to the broader global internet, the Russian government has hastened its progress toward infrastructural isolation. Regarding the moderation of TVEC online, in 2022, the Russian government blocked prominent social media platforms, including Facebook, Instagram, and Twitter, and labelled the companies as "extremist organizations."[139] Time and again, Russia has shown little regard for human rights principles when it comes to protecting the freedom of expression.

(f) China

On the furthest end of the regulatory spectrum is China which has demonstrated little interest in protecting human rights. China is home to one of the world's most restrictive media environments and its most sophisticated system of censorship started in the late 1990s with the banning of pornography and media sites.[140] The country has a comprehensive censorship system known as the "great firewall", which blocks access to foreign websites and restricts content that is deemed politically sensitive or harmful to the country's social stability. As a result, almost no foreign global platforms are allowed to operate in the country, and domestic international platforms are tightly regulated. The government actively monitors online activities and requires online service providers to store user data within the country's borders, making it easier to monitor and censor content.[141] Additionally, the government has introduced laws that hold internet companies accountable for the content shared on their platforms, resulting in self-censorship by these

[138] David Ignatius, *Russia Hasn't Stopped Maneuvering for a Role in Internet Oversight*, WASH. POST (Jul. 6, 2023), https://www.washingtonpost.com/opinions/2023/07/06/russia-internet-governance-united-nations/ [https://perma.cc/MLQ7-LDY9]; *see also* Shahbaz, Funk & Vesteinsson, *supra* note 38 ("Internet freedom in Russia reached an all-time low following the government's brutal invasion of Ukraine.").

[139] *Freedom in the World 2023: Russia, supra* note 135.

[140] Rogier Creemers, *Internet Information Service Management Measures*, DigiChina, STANFORD UNIVERSITY (Sep. 25, 2000), https://digichina.stanford.edu/work/internet-information-service-management-rules/ [https://perma.cc/RQY7-8XES] ("In September 2000, State Council Order No. 292 created the first set of content restrictions for Internet content providers. China-based websites cannot link to overseas news websites or distribute news from overseas media without separate approval.").

[141] *Freedom in the World 2023: China*, FREEDOM HOUSE (2023), https://freedomhouse.org/country/russia/freedom-world/2023 [https://perma.cc/5Q2G-3FMS].

companies to avoid legal repercussions.[142] China has been rated as the world's worst environment for internet freedom for eight straight years.[143] Regarding the regulation of TVEC online, Chinese officials and state media label a wide range of activity as terrorism or violent extremism, including protests in Hong Kong, uprisings in Xinjiang and Tibet, and even a tennis star's accusation of a high-ranking Chinese Communist Party official of sexual assault.[144]

2. Self-Regulation by Social Media Companies

This section provides a brief history of company self-regulation of content moderation practices which took place in phases: early efforts before 2009, the rise of industrial content moderation in 2009–2017, and improved technology alongside increasing legal requirements beginning in 2017, through to today. As the analysis of national laws did, this section will also specifically look at how platforms moderate TVEC. Many platforms look to the UN Human Rights Guiding Principles on Business and Human Rights and Article 19 of the ICCPR to guide their governance practices.[145] However, all online platforms moderate user-generated content slightly differently, and self-regulation efforts have varied between companies over the years. Generally, online platforms have

[142] *China to Tighten Grip on Social Media Comments, Requiring Sites to Employ Sufficient Content Moderators*, SOUTH CHINA MORNING POST (Jun. 18, 2022), https://finance.yahoo.com/news/china-tighten-grip-social-media-093000585.html [https://perma.cc/7R3D-8LUY].

[143] Shahbaz, Funk & Vesteinsson, *supra* note 38 ("In China, the government has been fairly successful in pairing systematic censorship of foreign services with robust investment in domestic platforms that are beholden to the ruling party.").

[144] *See* Murray Scot Tanner & James Bellacqua, *China's Response to Terrorism*, CAN, June 2016, https://www.uscc.gov/sites/default/files/Research/Chinas%20Response%20to%20Terrorism_CNA061616.pdf.

[145] T.G. Thorley & E. Saltman, *GIFCT Tech Trials: Combining Behavioural Signals to Surface Terrorist and Violent Extremist Content Online, Studies in Conflict & Terrorism*, TAYLOR & FRANCIS ONLINE, https://www.tandfonline.com/doi/full/10.1080/1057610X.2023.2222901 ("Focused on companies' applications of policies, the UNHR's Guiding Principles on Business and Human Rights is a bedrock for tech companies. Guiding principles for tech companies in moderation practices and data collection dictate that policies should dictate actions deemed as necessary, lawful, legitimate, and proportionate, and that the right to restriction should be tied to a defined and defendable threat.").

moderated content through four broad approaches: case-by-case, community-reliant, industrial, or large-scale, and no moderation whatsoever.[146]

Self-regulation by online platforms hosting user-generated content started in the earliest days of internet bulletin board services, when companies like CompuServe and Prodigy set rules for their subscribers to follow when posting content.[147] Indeed, it was specifically to protect the content moderation practices of these early internet companies that Congress passed Section 230 in 1996.[148] When modern-day online platforms launched in the early 2000s, content moderation was largely ad hoc, and most companies presented themselves as neutral intermediaries to avoid being held responsible for what their users said and did.[149] However, even in the early days, all commercially viable platforms moderated some content, to ensure their services were not overrun with spam, nudity, or other toxic content.[150] As Charlotte Willner, one of Meta's first content moderators, noted, the ethos of the pre-2008 moderation guidelines was, "if it makes you feel bad in your gut, then go ahead and take it down."[151] This ethos, still found in the artisanal approach to content moderation, shifted to become more industrial as online platforms expanded internationally and companies sought to make their products attractive to global users.[152]

---

[146] Caplan, *supra* note 16, at 16 ("We identify three major categories of platform companies according to their size, organization, and content moderation practices: (1) The artisanal approach, where case-by-case governance is normally performed by between 5 and 200 workers; (2) Community-reliant approaches, which typically combine formal policy made at the company level with volunteer moderators; and (3) The industrial approach, where tens of thousands of workers are employed to enforce rules by a separate policy team."). (Caplan's analysis misses an emergent set of companies that claim to do not content moderation whatsoever, including platforms like 4chan, 8kun, and Gab).

[147] *Section 230: Legislative History*, ELECTRONIC FRONTIER FOUNDATION, https://www.eff.org/issues/cda230/legislative-history [https://perma.cc/KVD6-TQGT].

[148] Kosseff, *supra* note 63 at 75-76.

[149] Suzor, *supra* note 37 at 15; *see also* Tarleton Gillespie, *The Politics of Platforms*, 3 NEW MEDIA & SOCIETY 12, 12, 347-364 (May 1, 2010), https://doi.org/10.1177/1461444809342738.

[150] James Grimmelmann, *The Virtues of Moderation*, 17 YALE J. L. & TECH. 42, 45 (2015).

[151] Klonick, *supra* note 32, at 1631; *citing* Telephone Interview with Dave Willner, Former Head of Content Policy, Facebook & Charlotte Willner, Former Safety Manager, User Operations, Facebook (Mar. 23, 2016).

[152] Gillespie, *supra* note 20, at 4; *see also* Caplan, *supra* note 16, at 16.

Starting in 2009, and continuing through to 2016, companies began to craft global platform rules. This led to a more industrial process of content moderation where companies enforced rules globally on millions of pieces of content.[153] The rules are sometimes referred to as the "terms of service" or "community standards" which a user agrees to follow when signing up for a platform.[154] As described by Kate Klonick, an academic focusing on internet policies, in her seminal article on content moderation, "The New Governors", social media companies developed global standardized content rules to manage: "(1) the increase in both users and volume of content; (2) the globalization and diversity of the online community; and (3) the increased reliance on teams of human moderators with diverse backgrounds."[155] Klonick argues that online platforms self-regulated because they were economically motivated to create a hospitable environment to incentivize engagement.[156] She goes on to say that companies try to keep up as much speech as possible while upholding their ideals of corporate responsibility.[157] As processes developed, many self-regulatory models adopted a "common-law" approach, to maintain consistency in the decision-making process.[158] Even with "common-law" precedent,

[153] Klonick, *supra* note 32; *see also* Douek, *supra* note 17, at 537 ("Once those rules are written, it's simply a matter of applying them over and over … and over again—the standard picture conceives of content moderation as simply the aggregation of millions of daily paradigm cases. The scale is hard to comprehend: in Q3 2021, Facebook took down 933,426,800 pieces of content, YouTube took down 4,806,042 channels and 6,229,882 videos, and in Q2 2021 TikTok removed 81,518,334 videos. These figures do not include every time these platforms decided to leave up content flagged for review (which would greatly exceed decisions to remove content) or appeals.").

[154] Citron, *supra* note 106, at 1037 ("From the start, tech companies' commitment to free expression admitted some exceptions. Terms of service and community guidelines banned child pornography, spam, phishing, fraud, impersonation, and copyright violations. Threats, cyber stalking, nonconsensual pornography, and hate speech were prohibited after extended discussions with advocacy groups. The goal was to strike an appropriate balance between free expression and abuse prevention while preserving platforms' market share.").

[155] Klonick, *supra* note 32, at 1635.

[156] *Id.* at 1618 ("[companies] are private, self-regulating entities that are economically and normatively motivated to reflect the democratic culture and free speech expectations of their users.").

[157] *See Id.*

[158] Caplan, *supra* note 16, at 18 ("One legal counsel compared the model they took to a "common-law system" based on precedent, while others described a process similar to a grounded theory approach, a methodology used in the social sciences to inductively build up categories, through the aggregation of individual cases or data points.") (*citing* Interview with Alex Feerst, head of Legal at Medium).

platforms were constantly updating their policies to adapt to global norms in response to: "(1) government request, (2) media coverage, (3) third-party civil society groups, and (4) individual users' use of the moderation process."[159] Throughout the 2000s and 2010s, companies were largely free to write their own rules, because Section 230 and other safe harbor regimes did not draw clear lines around acceptable or unacceptable content.[160]

As the large social media companies came to dominate the global landscape and to draw increased scrutiny, a societal shift was taking place. Starting in 2017, events such as Russian meddling in the 2016 election, the genocide in Myanmar, the Cambridge Analytica scandal and, subsequently, the Christchurch attack, raised the public awareness of the potential harms of "big tech". As a result, social media companies adopted a defensive posture, and many platforms looked to self-regulatory solutions as a low-cost way to repair reputational damage and stave off government regulation.[161] Self-regulatory efforts were frequently championed by tech company employees who wanted to create change from the inside, and sky-high profits meant the companies had cash to spend on these experiments.[162] In this vein, tech companies used their money and soft power to work with civil society, journalists, and academics to institutionalize self-regulation practices through organizations like the Global Internet Forum to Counter Terrorism (GIFCT), Meta's independent Oversight Board,[163] and Alphabet's Jigsaw project, which researched how to curb extremism and misinformation across products.[164]

[159] Klonick, *supra* note 32 at 1649.

[160] Caplan, *supra* note 16, at 27 ("Within the United States, Section 230 of the Communications Decency Act provides platforms like those discussed above with the freedom to organize their content moderation teams as they see fit, as long as they are taking care to remove copyright protected and illegal content. As platforms deploy the other right given to them by Section 230 and the "Good Samaritan" provision, platforms told us they are finding it difficult to draw lines in ways that make sense both ethically and organizationally.").

[161] Kate Klonick, *The End of the Golden Age of Tech Accountability*, THE KLONICKLES (Mar. 4, 2023), https://klonick.substack.com/p/the-end-of-the-golden-age-of-tech [https://perma.cc/UZH8-NAHV].

[162] *Id.*

[163] Kate Klonick, *Inside the Making of Facebook's Supreme Court*, THE NEW YORKER (Feb. 12, 2021), https://www.newyorker.com/tech/annals-of-technology/inside-the-making-of-facebooks-supreme-court.

[164] Jane Wakefield, *TED 2018: Alphabet firm's tools to combat extremism*, BBC (Apr. 13, 2018), https://www.bbc.com/news/technology-43760213 [https://perma.cc/J2ZJ-AE8J].

Another shift began in 2017 when new technologies transformed the content moderation industry. Online platforms started to deploy automated tools to detect and filter harmful content alongside predictive models that relied on AI to learn and recognize patterns.[165] Many companies incorporated automation into their content moderation systems, with the encouragement of policymakers, who were increasingly calling on them to restrict content they deemed harmful.[166] In addition to automation, moderation itself became more nuanced, as companies thought beyond the binary decision of keeping up or taking down content. As Evelyn Douek, an internet law academic, notes, platforms adopted a variety of tools, including "sticking labels on posts; partnerships with fact-checkers; greater platform and government collaboration; adding friction to how users share content; giving users affordances to control their own online experience; looking beyond the content of posts to how users behave online to determine what should be removed; and tinkering with the underlying dynamics of the very platforms themselves."[167] Indeed, over the past few years, the evolving work of ensuring the safety and security of online platforms has become so sophisticated that it has created an entire industry of "trust and safety" professionals.[168]

While these new tools affected a wide range of content, preventing the spread of TVEC online was one area of content moderation in which companies invested significantly in

[165] CAREY SHENKMAN, DHANARAJ THAKUR & EMMA LLANSÓ, DO YOU SEE WHAT I SEE? CAPABILITIES AND LIMITS OF AUTOMATED MULTIMEDIA CONTENT ANALYSIS 19 (Center for Democracy & Technology, 2021), https://cdt.org/wp-content/uploads/2021/05/2021-05-18-Do-You-See-What-I-See-Capabilities-Limits-of-Automated-Multimedia-Content-Analysis-Full-Report-2033-FINAL.pdf.

[166] *Id.* at 10. ("Policymakers worldwide are increasingly calling on social media companies to identify and restrict text, photos, and videos that involve illegal, harmful, or false information. Many services are voluntarily incorporating automation into their content moderation systems, and government agencies are also exploring the use of automated content analysis."). *See also* Robert Gorwa, Reuben Binns & Christian Katzenbach, *Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance*, 7 BIG DATA & SOC. 1 (2020).

[167] Douek, *supra* note 17 at 5.

[168] Fishman, *supra* note 20; s*ee also* Trust & Safety Professional Association, *Trust & Safety Curriculum*, TRUST & SAFETY PROFESSIONALS ASSOCIATION (2023), https://www.tspa.org/curriculum/ts-curriculum/ [https://perma.cc/6ZTV-L6YL].

self-regulation after years of government pressure.[169] Terrorist and violent extremist use of the internet is not a new phenomenon. Indeed, al-Qaeda was operating online by the mid-1990s,[170] and the prevalence of white supremacism online was so great by 1996 that the Anti-Defamation League started tracking it.[171] Like the early cyber libertarians, early terrorists and violent extremists saw the internet as a great place to find like-minded individuals and discuss ideas free from government censorship.[172] By the early 2000s, terrorists and violent extremists were drawn to social media for the same reasons as everyone else: social media platforms are a simple and reliable way to share ideas and connect with a vast network of people.[173] But terrorist use of social media did not go unnoticed. In 2008, during a Senate hearing, US Senator Lieberman demanded that YouTube remove Al-Qaeda training videos.[174] In a response many would now find shocking, the company's representative defended the terrorist organization's right to express unpopular viewpoints on their platform.[175] Indeed, the companies did not seriously try to self-regulate until 2016, when the so-called Islamic State began using social media to recruit and inspire violence in Europe, leading lawmakers to threaten regulation.[176]

In the years following 2015, social media companies attempted to self-regulate TVEC on their platforms by establishing robust internal processes and industry collaboration. First,

[169] Fishman, *supra* note 20.

[170] Gabriel Weimann, *Terror on the Internet: The New Arena, the New Challenges*, U.S. INSTITUTE OF PEACE (2010), https://www.usip.org/publications/2010/05/terror-internet [https://perma.cc/Z9KV-NA3E].

[171] Fishman, *supra* note 20 *citing* David H. Strassler, et al., *The Web of Hate: Extremists Exploit the Internet*, ANTI-DEFAMATION LEAGUE (1996), https://www.adl.org/sites/default/files/documents/assets/pdf/combating-hate/ADL-Report-1996-Web-of-Hate-Extremists-exploit-the-Internet.pdf.

[172] Fishman, *supra* note 20.

[173] *Id.*

[174] Timothy B. Lee, *YouTube Rebuffs Senator's Demands to Remove Islamist Videos*, ARS TECHNICA (May 20, 2008), https://arstechnica.com/tech-policy/2008/05/youtube-rebuffssenatorss-demands-for-removal-of-islamist-videos/ [https://perma.cc/3UR9-JPNT].

[175] *Id.*

[176] Liat Clark, *Facebook and Twitter Must Tackle Hate Speech or Face New Laws*, WIRED U.K. (Dec. 5, 2016), http://www.wired.co.uk/article/us-tech-giants-must-tackle-hate-speech-orface-legal-action [https://perma.cc/T9JT-ZFMZ].

social media companies cleaned up their platforms individually in several ways, including: writing rules defining what constitutes a terrorist organization and TVEC; identifying and removing policy violations; providing data on TVEC in transparency reports; and limiting access to product features to decrease the virality of TVEC.[177] Additionally, platforms started to proactively work with governments and law enforcement officials to remove content from entities who were designated as terrorist organizations.[178] As companies implemented these measures, they were quick to share results with lawmakers in an attempt to stave off regulation. For example, Twitter, a company who branded itself the "free-speech wing of the free-speech party" since its founding, reported that it had suspended over 125,000 ISIS-related accounts in 2016, and Meta announced it had hired 3,000 more people to stop the spread of terrorist propaganda.[179]

Second, tech companies started to work together as an industry to self-regulate through several projects. The most notable TVEC-related self-regulatory initiative was the Global Internet Forum to Counter Terrorism (GIFCT). In 2016, the idea was floated that companies should create a shared database of banned TVEC, which would operate like PhotoDNA, a tool developed to remove child sexual abuse material.[180] At first, online platforms and civil society organizations were wary of the idea of a TVEC database, as there was no agreed-upon definition for what constituted "terrorist content."[181] However, the tech companies reversed course in December 2016, the day before the European

[177] Fishman, *supra* note 20.

[178] Klonick, *supra* note 32, at 1638; *See also* Natalie Andrews & Deepa Seetharaman, *Facebook Steps Up Efforts Against Terrorism*, WALL STREET J. (Feb. 11, 2016, 7:39 PM), http://on.wsj.com/1T; Joseph Menn & Dustin Volz, *Google, Facebook Quietly Move Toward Automatic Blocking of Extremist Videos*, REUTERS (Jun. 24, 2016), https://www.reuters.com/article/us-internet-extremism-video-exclusive/exclusive-google-facebookquietly-move-toward-automatic-blocking-of-extremist-videos-idUSKCN0ZB00M [https://perma.cc/DN9Y-9JHB].

[179] *Id.* at 1638.

[180] Kaveh Waddell, *A Tool to Delete Beheading Videos Before They Even Appear Online*, THE ATLANTIC (Jun. 22, 2016), https://www.theatlantic.com/technology/archive/2016/06/a-tool-to-delete-beheading-videos-before-they-even-appear-online/488105/ [https://perma.cc/N8C6-D6DZ].

[181] Citron, *supra* note 106, at 1044; *noting*, lawmakers were uninterested in hearing reasons why TVEC was a fundamentally different problem to child sexual abuse material which was universally considered to be illegal and abhorrent.

Commission released a damning report condemning their efforts to remove TVEC.[182] In 2017, Facebook, Microsoft, Twitter, and YouTube launched the GIFCT as an industry initiative to apply technology, share knowledge, and support research on terrorists' abuse of the platforms.[183] This new project included the creation of a database to which companies could upload terrorist content found on their platforms and "hash" the images and videos. These "hashes", frequently called "digital fingerprints", were entered into the database, and the technology prevented upload of hashed images on any of the cooperating platforms.[184] By 2019, this database included over 200,000 pieces of content.[185] Despite claims of success by tech companies, the hash-sharing database was frequently criticized by civil society for not being more transparent in regard to the content in the database and by governments that wanted to ensure the images they perceived as TVEC were included.[186] As Part III will explore, after the Christchurch attack in 2019, reforming the GIFCT from an industry-led project into an MSI became a top priority.

B. Multi-Sided Content Governance Frameworks

Single-sided efforts by national lawmakers and online platforms were successful to some extent in reducing the proliferation of TVEC online. However, many argued that a new framework was necessary because democratic countries were limited in their ability to

[182] *Partnering to Help Curb Spread of Online Terrorist Content*, META NEWSROOM (Dec. 5, 2016), https://about.fb.com/news/2016/12/partnering-to-help-curb-spread-of-online-terrorist-content/.

[183] Christchurch Call News & Updates, *Significant progress made on eliminating terrorist content online*, CHRISTCHURCH CALL TO ACTION (Sep. 24, 2019), https://www.christchurchcall.com/media-and-resources/news-and-updates/new-news-article-page-8/ [https://perma.cc/G2K4-WC29].

[184] Global Internet Forum to Counter Terrorism, *Who we are: Story, 2017 Year in Review*, GLOBAL INTERNET FORUM TO COUNTER TERRORISM (2023), https://gifct.org/about/story/#2017-year-in-review [https://perma.cc/2EKV-YAD2].

[185] Global Internet Forum to Counter Terrorism, *Who we are: Story, May 2019, the Christchurch Call to Action*, GLOBAL INTERNET FORUM TO COUNTER TERRORISM (2023), https://gifct.org/about/story/#may-2019---christchurch-call-to-action [https://perma.cc/YU8X-R5HW].

[186] Courtney Radsch, *GIFCT: Possibly the Most Important Acronym You've Heard Of*, JUST SECURITY (Sep. 30, 2020), https://www.justsecurity.org/72603/gifct-possibly-the-most-important-acronym-youve-never-heard-of/ [https://perma.cc/4PVC-27KX].

regulate content, and self-regulation was falling short.[187] One potential solution was the creation of MSIs, which would bring together governments, companies, civil society, and outside experts to identify a solution and implement it across sectors. Through an MSI, stakeholders can harness the capabilities of different actors and co-design solutions through participatory processes. This section will first explore the rise of multistakeholder frameworks and their suitability to addressing global challenges. It will then explore the successes of multistakeholderism within the internet governance space as a template for the content moderation governance problem.[188] Finally, it looks at emerging multilateral initiatives and their effort to frame themselves as multistakeholder without truly being MSIs.

1. The Transition from Multilateral to Multistakeholder

Unlike our concept of multistakeholderism, modern-day concepts of multilateralism can be traced back to 1648 and the signing of the Peace of Westphalia, which recognized the sovereignty of individual states and promoted the idea of non-interference in the affairs of other states.[189] That treaty created a world order based on interaction, negotiation, and cooperation among sovereign states.[190] John Gerald Ruggie, in his seminal article on

[187] Douek, *supra* note 17, at 603 ("An underlying theme and motivation of this Article has been that the limits of direct governmental regulation of online speech are significant, making it necessary to find an approach that leverages and legitimates platform self-regulation. Governmental oversight of platforms should aim to maximize the private sector's resources, expertise, and dynamism in finding innovative and effective methods for tackling content moderation challenges while requiring platforms to explain, justify and verify those methods. By allowing platforms to experiment, government oversight would avoid locking in the status quo at the major platforms.").

[188] *See* Raymond & Denardis, *supra* note 36, at 19–45. (This report uses the definition from Raymond & DeNardis for "internet governance" to broadly describe six technical functions for the internet including: critical internet resources such as domain names and IP addresses, internet standards for interoperability, interconnection between networks, cyber-security, information intermediation, and intellectual property rights enforcement. Content moderation happens primarily at the information intermediation layer of the internet stack, so while many of these MSIs include some coordination on content, they typically address a wider range of internet functions.).

[189] *See* Leo Gross, *The Peace of Westphalia, 1648-1948*, 42 THE AM. J. OF INT'L L. 20, 24 (Jan. 1948) https://www.jstor.org/stable/2193560.

[190] *Id.*

multilateralism, defines multilateralism as "the practice of coordinating national policies in groups of three or more states, through ad hoc arrangements or by means of institutions."[191] In this system, government representatives act on behalf of their citizens and implement the terms of any agreement within their borders to resolve international issues.[192] Following two devastating World Wars, nations strengthened multilateral institutions and developed new ones, notably the United Nations, to prevent further violent conflict and set human rights standards.[193] However, by the 1980s, multilateral frameworks were failing to address many global issues as governments lacked the internal capacity to implement policies due to a gradual erosion of trust and (in many cases) extensive corruption.[194]

Rapid globalization in the 1980s and 1990s compounded geopolitical tensions and exposed many of the underlying problems with multilateral frameworks. During this timeframe, national governments found their monopoly on public policy making increasingly contested, with the emergence of three powerful groups: transnational corporations, civil society, and an independent media.[195] First, transnational corporations grew so large that their economic power and cultural authority sometimes exceeded that of many states. Second, growing international links between civil society organizations connected disparate movements, which provided a larger platform for human rights advocacy. As such, civil society and non-government organizations came to be viewed as legitimate

[191] *Id.*

[192] Harris Gleckman, *Multistakeholderism: a corporate push for a new form of global governance*, TRANSNATIONAL INSTITUTE (Jan. 19, 2016), https://www.tni.org/en/publication/multi-stakeholderism-a-corporate-push-for-a-new-form-of-global-governance [https://perma.cc/ZF8Z-XTDG].

[193] *Id.*

[194] Christopher Ansell & Jacob Torfing, HANDBOOK ON THEORIES OF GOVERNANCE, 7 (Jun. 24, 2016); explaining the rise of multistakeholder governance ("in the fields of public administration, public law and public policy, this question arose out of the attempt to address challenges posed by administrative complexity, poor policy implementation and fiscal austerity. In the field of development studies, it developed in response to the frustration of achieving development goals in partnership with weak or corrupt developing states. In the field of international relations, economics and environmental studies, the question grew out of the need to address collective action problems and the management of common pool resources.").

[195] *Id.*

actors in the formulation, implementation, and evaluation of public policy.[196] Third, as literacy rates and access to information increased, citizens became more skeptical of state-run media organizations, lending credibility to independent journalists. The newly empowered media was quick to expose governmental inability to hold corporations responsible for their wrongdoing, which increased public pressure on corporations to respect human rights.[197]

Increased pressure on transnational corporations and the governments that failed to hold them accountable forced parties to consider collaborative approaches with new groups of stakeholders.[198] For corporations, stakeholder collaboration was valuable in jurisdictions where governments could not or would not uphold basic human rights, leaving governance gaps for unregulated business practices.[199] Additionally, these discussions could provide corporations with local knowledge and new insights into diverse problems, which sometimes yielded better return on investment.[200] Next, democratic governmental actors were quick to participate in stakeholder discussions, as results from this type of governance could demonstrate an impact on constituents without the need to pass legislation.[201] Finally, civil society found this collaboration beneficial when it provided new

[196] *Id.*

[197] Baumann-Pauly et al., *supra* note 32, at 10 ("media interest focused on headline grabbing issues, such as the use of sweatshops by well-known brands like Nike, Disney and Levi Strauss.").

[198] Ansell & Torfing, *supra* note 198, at 7.

[199] Gleckman, *supra* note 196.

[200] Ariel Babcok, Nathan Barrymore, Christopher Bruno, Allen He, et al., *Walking the Talk: Valuing a Multi-Stakeholder Strategy*, FCLT GLOBAL AND WHARTON UNIVERSITY OF PENNSYLVANIA, 5–6, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4023510 ("while there is often a natural gravitational pull to prioritize one set of stakeholders over another (shareholders in many cases), prioritizing one group continuously is not a winning long-term strategy … Future-fit, long-term companies need more durable performance to succeed – and that requires attention to a broader group of stakeholders.").

[201] Nick Buxton, *Multistakeholderism: a critical look*, TRANSNATIONAL INSTITUTE (Jan. 19, 2016), https://www.tni.org/en/publication/multistakeholderism-a-critical-look [https://perma.cc/RW3P-Q6KP]; *see also* Lawrence E. Strickling & Jonah Force Hill, *Multi-stakeholder Governance Innovations to Protect Free Expression, Diversity and Civility*, CENTRE FOR INT'L GOVERNANCE INNOVATION & STANFORD GLOB. DIGIT. POL'Y INCUBATOR, Special Report: Governance Innovation for a Connected World Protecting Free Expression, Diversity and Civic Engagement in the Global Digital Ecosystem, 45 (2018)

opportunities to demonstrate their soft power.[202] Over time, this consultation with stakeholders provided a helpful form of checking and balancing.[203] As a result, collaboration between stakeholders led to the co-production of public solutions which increased their legitimacy.[204] While often not legally binding, if executed properly, this type of collaboration establishes and reinforces standards that one party could not achieve by acting on its own. These successes contributed to the formalization of multistakeholderism and declining reliance on multilateralism.[205]

Multistakeholderism is defined as two or more classes of actors engaged in a common governance enterprise to solve a wider problem, where decision-making authority is distributed between actors based on procedural rules.[206] An MSI is created when two or more types of actors come together in a structured organization to solve a problem

("Multistakeholder processes can be resource intensive, but they are still generally less financially burdensome than traditional regulatory proceedings or litigation. Reaching multi-stakeholder consensus can be difficult and time-consuming but compare the time it takes to achieve consensus to the time it takes the US Congress to enact legislation. New entrants may have a strategic disadvantage in multi-stakeholder settings, but they at least have a seat at the table and a say in the outcome. Traditional government and multilateral rulemaking settings afford them no such right.").

[202] Raymond & DeNardis, *supra* note 40.

[203] Admin. Conference of the U.S., Recommendation 2018-7, Public Engagement in Rulemaking, 84 Fed. Reg. 2139, 2146 (Feb. 6, 2019) ("Robust public participation is vital to the rulemaking process. By providing opportunities for public input and dialogue, agencies can obtain more comprehensive information, enhance the legitimacy and accountability of their decisions, and increase public support for their rules.").

[204] Ansell & Torfing, *supra* note 194, at 7.

[205] Gleckman, *Multistakeholderism: a new way for corporations and their new partners to try to govern the* world, CIVICUS (Oct. 2018), https://www.civicus.org/index.php/re-imagining-democracy/overviews/3377-multistakeholderism-a-new-way-for-corporations-and-their-new-partners-to-try-to-govern-the-world [https://perma.cc/E8C8-VGS3] ("Even for the proponents of multistakeholderism, the transition from the nation-state as the actor in international affairs to 'stakeholders' as global governors has been an uneven process. One major element of the transition for these new claimants as global leaders is learning to work with a heterogeneous group of organizations, some of which were, or still are, institutional opponents. The differences in types of power external to an MSG group create a fundamental asymmetry of power within the group.").

[206] Raymond & DeNardis, *supra* note 36, at 20 ("Multi-stakeholderism is defined here as two or more classes of actors engaged in a common governance enterprise concerning issues they regard as public in nature, and characterized by polyarchic authority relations constituted by procedural rules.").

defined by the group. Different types of actors with a potential stake in an MSI include businesses, civil society, governments, universities, academics, technical experts, investors, and consumers.[207] In recent decades, multistakeholderism has emerged to enhance multilateral processes as well as becoming an alternative to, and sometimes a direct competitor with, traditional multilateral approaches, for several key reasons. First, MSIs are frequently created when an industry or government finds itself facing a significant amount of public pressure to fix a problem that it cannot solve on its own.[208] Sometimes this occurs shortly after a tragic event. For example, after rampant human rights violations in the diamond trade were made public, the Kimberley Process created an MSI that urged governments to pass regulation, companies to certify the source of the diamonds, and civil society to oversee the process.[209] Second, MSIs are created to help fill governance gaps in regulatory frameworks. In this situation, MSIs establish guidelines or best practices for stakeholder behavior where local or national regulators cannot or do not uphold human rights principles. For example, many MSIs were created in the 1990s to address the use of "sweatshops" in countries where governments did not enforce fair labor practices. [210] Finally, MSIs are frequently created to address technological advances where

[207] Baumann-Pauly et al., *supra* note 32.

[208] John Ruggie, Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises, Business and Human Rights: Mapping International Standards of Responsibility and Accountability for Corporate Acts, U.N. Doc. A/HRC/4/35, at 17 (Feb. 19, 2007) ("[d]riven by social pressure, [and]… seek to close regulatory gaps that contribute to human rights abuses. But they do so in specific operational contexts, not in any overarching manner. Moreover, recognizing that some business and human rights challenges require multi-stakeholder responses, they allocate shared responsibilities and establish mutual accountability mechanisms within complex collaborative networks. These can include any combination of host and home States, corporations, civil society actors, industry associations, international institutions, and investors groups.").

[209] *What is the Kimberley Process?,* KIMBERLEY PROCESS (2023), https://www.kimberleyprocess.com/en/what-kp [https://perma.cc/YDC5-HJ2P].

[210] Baumann-Pauly et al., supra note 28, at 2 ("The absence of state regulation presents major business challenges for corporations. Clothing retailers like Walmart and H&M face unsafe factory conditions in Bangladesh in the wake of the Rana Plaza tragedy. Internet service providers like Facebook and Google wrestle with their users' expectations to guarantee freedom of expression in China and other non-democratic regimes. Oil and mining companies like Shell and Newmont operating in conflict zones from the Congo to Iraq struggle to provide security for their people and facilities in these inherently dangerous places. In these contexts, multistakeholder initiatives (MSIs) have become the default response for addressing so-called "governance gaps.").

development of technology requires broader societal input to protect human rights. The need to address technological advancement is one of the primary drivers of the proliferation of MSIs in the internet governance space.

2. Multistakeholderism in Internet Governance

The 1990s were a pivotal decade for both the development of MSIs and the growth of the internet. It is therefore not surprising that the two rose to prominence together. As noted above, early internet adopters identified with many of the core principles of multistakeholderism, including the need to be collaborative, interconnected, and global.[211] The internet of today is a byproduct of multistakeholder collaboration between engineers, individuals, government agencies, and businesses. Internet standards were created by an international group of stakeholders who shared a common goal to create a decentralized network.[212] Over the years, internet governance MSIs have encompassed a wide range of approaches, procedures, formats, and outcomes.[213]

One early example of an MSI for internet governance was the Internet Engineering Task Force (IETF). Initially started in 1987 as a quarterly meeting at which a dozen US researchers could exchange ideas, by 1992 over 750 stakeholders from government, civil

[211] Internet Society, *Internet Governance: Why the Multistakeholder Approach Works*, INTERNET SOCIETY (Apr. 26, 2016), https://www.internetsociety.org/wp-content/uploads/2016/04/IG-MultiStakeholderApproach.pdf.

[212] Konstantinos Komaitis, *Global Digital Compact – Additional submission,* UNITED NATIONS TECH ENVOY FILES (Apr. 6, 2023), https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/GDC-submission_Konstantinos-Komaitis.pdf ("The Internet and multistakeholder governance are tightly interwoven. The Internet is a byproduct of a pure collaborative process between engineers, individuals, government agencies and businesses. It emerged because this different set of people shared a common goal despite their often diverse and distinct viewpoints; that goal was to create a network that would be decentralized and could respond to any type of failure.").

[213] Strickling & Hill, *supra* note 201, at 45 ("There is no one single concept of what is appropriately viewed to be a multi-stakeholder approach. There are, instead, numerous models currently in use today, each with its own unique contours. Few, if any, of the models currently in use are static; rather, they are constantly evolving to meet new and yet uncharted governance challenges.").

society, and industry were attending to set internet standards.[214] Around the same time, in 1991, Vint Cerf, Bob Kahn, and other internet entrepreneurs created the multistakeholder Internet Society to "promote the open development, evolution and use of the internet for the benefit of all people throughout the world."[215] The two organizations merged in 1992, with the Internet Society providing a legal umbrella for the IETF to help manage its growth and maintain independence from the US government.[216] Both organizations are still relevant for our discussion of successful MSIs in internet governance, as they demonstrate an early focus on multistakeholderism in the community. At the IETF, stakeholders set standards through a bottom-up process whereby decisions are based on what has been called "rough consensus and running code."[217]

Another important, albeit controversial, internet governing body created during this period was the Internet Corporation for Assigned Names and Numbers (ICANN). The roots of ICANN date back to 1969, when researchers began sending electronic messages to each other through the Arpanet.[218] To make it easier to track and send messages via the network, Jon Postel, a researcher in southern California, created a registry to manage the coordination of messages, which ultimately became the Domain Name System.[219] In practice, the Domain Name System became the "phonebook" of the internet – allowing people to easily look up other people. In the beginning, it was possible for Postel to maintain this function on his own, but the burden of providing for the technical management of the Domain Name System increased rapidly. As Postel testified to Congress, in 1993 there were 30,000 domain names; by 1997 there were 1.6 million globally.[220]

---

[214] Scott Bradner, *The Internet Engineering Task Force*, OPEN SOURCES: VOICES FROM THE OPEN SOURCE REVOLUTION, 1ST EDITION (Jan. 1999), https://www.oreilly.com/openbook/opensources/book/ietf.html [https://perma.cc/BH3H-MP5M].

[215] Henry Bennie, *25 years of the Internet Society*, CERN (Sept. 26, 2017), https://home.cern/news/news/computing/25-years-internet-society [https://perma.cc/KVS3-T2MV].

[216] Bradner, *supra* note 218.

[217] Raymond & DeNardis, *supra* note 36, at 32.

[218] *ICANN History Project*, INTERNET CORPORATION OF ASSIGNED NAMES AND NUMBERS (Oct. 2016), https://www.icann.org/history [https://perma.cc/E2D7-JZTH].

[219] *Id.*

[220] Internet Domain Names, Part 1: Hearing Before the Committee on Science, Subcommittee on Basic Research, *supra* note 38.

As the scale and complexity of the Domain Name System grew, the US Government sought to relinquish its historic control over technical internet functions by creating ICANN, a non-profit entity dedicated to the task. After lengthy stakeholder engagement, in 1998, the Department of Commerce signed a memorandum of understanding with ICANN which outlined how ICANN would manage key functions, including by allocating IP number blocks, overseeing the root server system, and coordinating technical parameters.[221] Most critically, under this agreement, ICANN managed the Internet Assigned Numbers Authority (IANA), which administers functions of the Domain Name System. It is important to note that while the IANA function serves the global internet, at that time its funding came from the US Government, and it was considered a government asset.[222] As a result, ICANN was beholden to the US Government's reporting requirements, with the understanding that the organization would eventually become fully independent.[223] While the US Government's role was largely procedural, there was mounting resentment from other nations over perceived "American control of the internet."[224] This issue threatened to divide the global internet space.

Tensions surrounding ICANN's structure escalated in the early 2000s. Many governments wanted to see the UN manage ICANN's responsibilities through the multilateral International Telecommunication Union (ITU). The ITU is a body within the UN that regulates radio spectrum, satellite orbits and certain worldwide technical standards.[225]

---

[221] Joe Sims & J. Beckwith Burr, *Memorandum of Understanding Between the U.S. Department of Commerce and Internet Corporation for Assigned Names and Numbers*, INTERNET CORPORATION OF ASSIGNED NAMES AND NUMBERS RESOURCES (Dec. 31, 1999), https://www.icann.org/resources/unthemed-pages/icann-mou-1998-11-25-en [https://perma.cc/8383-7ACS].

[222] Steve Crocker, *On Creating Internet Governance Organizations: A Comment on the ICANN Experience*, INTERNET GOVERNANCE FORUM BERLIN, "Towards a Global Framework for Cyber Peace and Digital Cooperation: An Agenda for the 2020s," 148 (Nov. 25-29, 2019), https://www.hiig.de/wp-content/uploads/2019/11/Kleinwa%CC%88chter-Kettemann-Senges-eds.-Global-Framework-for-Cyber-Peace-2019.pdf.

[223] Sims & Burr, *supra* note 225.

[224] Raymond & DeNardis, *supra* note 36 at 27.

[225] International Telecommunication Union (ITU), *About the ITU,* International Telecommunication Union (2023), https://www.itu.int/en/about/Pages/default.aspx [https://perma.cc/Y37Q-QKVQ].

Standard-setting at the ITU is top-down and bureaucratic: a method preferred by governments that were seeking to control their citizens' access to information and tax the burgeoning internet economy. This structure was anathema to the bottom-up, decentralized, and interoperable internet governance system that had developed since 1969. As a result, many stakeholders from civil society, industry, and democratic governments saw the possibility of ITU control over the IANA function as undermining both the functionality and freedoms of the global internet. These tensions came to a head in 2003, when the battle over the future of ICANN and the Domain Name System root zone management was brought up at the ITU's World Summit on the Information Society (WSIS). The entrenchment of both factions meant that no agreement was reached. However, two years later, at the second phase of WSIS in Tunis, UN members agreed on a compromise that would forestall giving ICANN oversight to the ITU by creating the Internet Governance Forum (IGF). The IGF is an MSI still under the oversight of the UN that identifies and defines the public policy issues that are relevant to internet governance.[226]

Following the directive set for the IGF in Tunis in 2005, the UN held two rounds of consultations to establish the objectives and format of the IGF.[227] The first meeting of the IGF was in 2006 in Athens. Over 1,200 participants attended from government, the private sector, civil society, academia, and technical communities.[228] In the years following, the IGF created processes to be more inclusive, including the creation of a dedicated Multistakeholder Advisory Group to help with planning and participation, starting an open consultation process to allow the public to submit suggestions regarding the program for the IGF, and instituting a host country selection process whereby countries could bid to host the event.[229] Over the years, the IGF expanded its stakeholder

[226] United Nations Secretariat of the Internet Governance Forum (IGF), *About the IGF*, INTERNET GOVERNANCE FORUM (2023), https://www.intgovforum.org/en/about [https://perma.cc/548B-E3QY].
[227] *Id.*
[228] United Nations Secretariat, *Internet Governance Forum to Hold Inaugural Session in Athens from 30 October to 2 November, Press ReleasePI/1747*, UNITED NATIONS: MEETINGS COVERAGE AND PRESS RELEASES (Oct. 25, 2006), https://press.un.org/en/2006/pi1747.doc.htm [https://perma.cc/MMH7-Z6RS].
[229] United Nations Secretariat of the Internet Governance Forum (IGF), *supra* note 229.

engagement and helped develop a broader sense of multistakeholderism throughout the internet governance community.

In 2013, events threatened to undo the successes of multistakeholderism. Edward Snowden's leak of thousands of documents revealed an extensive spying program the US National Security Agency conducted over internet infrastructure. Many world leaders turned to the UN in hopes of finding a multilateral solution to government surveillance.[230] One voice calling for multilateral intervention was then-President of Brazil Dilma Rousseff, who had her personal cell phone targeted for the content of calls, emails, and messages by the National Security Agency.[231] In the days following the leaks, she urged the UN and government actors to get involved to enforce rules governing the internet. However, shortly after her speech to the UN, Brazil instead decided to organize the Global Multistakeholder Meeting on the Future of Internet Governance, which came to be known as NETmundial.[232] The NETmundial conference took place in April 2014, bringing together over 1,400 people from all over the world.[233] Stakeholders collaborated in small working groups over several days to create an outcome document which outlined principles for internet governance and a roadmap for the future of the internet governance ecosystem.[234] By all measures, this was a significant achievement for the multistakeholder model, which had not traditionally produced consensus-driven outcomes.

NETmundial's successes fostered goodwill among stakeholders in the internet governance ecosystem. Seeking to maintain momentum, just a few months later, the conference organizers teamed up with ICANN and the World Economic Forum to start the

---

[230] Deborah Brown & Anriette Esterhuysen, *Extracting lessons from NETmundial: Achieving bottom-up and multistakeholder outcomes from global internet governance policy discussions*, ASSOCIATION FOR PROGRESSIVE COMMUNICATIONS (2016), https://www.apc.org/sites/default/files/ExtractingLessonsFromNETmundial.pdf.

[231] *Id.* at 6.

[232] *Id.*

[233] cgi.br (Brazilian Internet Steering Committee) & /1net, *NETmundial Multistakeholder Statement*, NETMUNDIAL GLOBAL MEETING ON THE FUTURE OF INTERNET GOVERNANCE (Apr. 24, 2014), https://netmundial.br/about/ [https://perma.cc/Y68L-ANB6].

[234] *Id.*

NETmundial Initiative (NMI).[235] The NMI was meant to "carry forward the cooperative spirit of São Paulo and work together to apply the NETmundial Principles."[236] However, it ran into trouble almost immediately when it was revealed that the three lead organizers had awarded themselves "permanent seats" on its 25-member council, isolating key stakeholders and directly undermining the NMI's claims to be a bottom-up MSI.[237] The NMI was further undermined by a lack of transparency, accountability, and inclusivity – all values called for in the outcome documents from NETmundial.[238] Finally, it was hard to justify the need for a separate initiative when the reforms outlined in the outcomes document had been enacted by the IGF and ICANN.[239] As a result, NMI's "mandate" to ICANN and the World Economic Forum expired in 2016, and the initiative was shut down.

One issue that received a lot of attention at NETmundial was the ongoing debate related to the US Government's oversight of ICANN, and the IANA functions. Stakeholders contended that internet governance could never be multistakeholder as long as the US

[235] World Economic Forum and Internet Corporation for Assigned Names and Numbers, *NETmundial Initiative for Internet Governance Cooperation & Development*, WORLD ECONOMIC FORUM (Aug. 28, 2014), https://www3.weforum.org/docs/WEF_1NetmundialInitiativeBrief.pdf.

[236] *Id.*

[237] Internet Society, *Internet Society Statement on the NETmundial Initiative, Press Release,* INTERNET SOCIETY (Nov. 17, 2014), https://www.internetsociety.org/news/press-releases/2014/internet-society-statement-on-the-netmundial-initiative/ [https://perma.cc/577N-2W3W] ("Based on the information that we have to date, the Internet Society cannot agree to participate in or endorse the Coordination Council for the NETmundial Initiative. We are concerned that the way in which the NETmundial Initiative is being formed does not appear to be consistent with the Internet Society's longstanding principles, including: Bottom-up orientation, Decentralized, Open, Transparent, Accountable, Multi-stakeholder.").

[238] Larry Strickling, *Remarks of Assistant Secretary Strickling on the Self-Governing Internet at Georgia Institute of Technology*, NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION (Oct. 26, 2016), https://ntia.gov/speechtestimony/remarks-assistant-secretary-strickling-self-governing-internet-georgia-institute [https://perma.cc/95JU-KT7C] ("Yet despite support from the United States government and others, the NetMundial Initiative never got off the ground. Why? Because it lacked the support and participation of all the relevant stakeholders, most notably the business community and the Internet Society. It was developed in a top-down way, without bottom-up support and input from the community. In the eyes of many key stakeholders, the initiative lacked the legitimacy it needed to succeed.").

[239] *Id.; see also* cgi.br & /1net, *supra* note 233 (citing outcomes document text).

Government still maintained oversight.[240] However, six weeks before NETmundial, the US Government announced its intent to transition its stewardship role of the IANA function to the global multistakeholder community.[241] In June 2014, ICANN started a multistakeholder process to transition the IANA function away from US Government oversight.[242] Over the next two years, participants held more than 600 meetings to finalize the details and on October 1, 2016, the process was completed.[243] As part of these negotiations, ICANN added another layer of governance, called the Empowered Community, which promoted multistakeholderism in its processes supporting its internet governance activities.[244]

Another issue debated at NETmundial was the future of the IGF, which was set to be reviewed in 2015 by the UN General Assembly in a process called "WSIS+10." In December 2015, much of the advice provided in the NETmundial outcomes document was incorporated into the 10-year renewal of the IGF.[245] Part of this renewed mandate included a commitment to the UN's Sustainable Development Goals, an ambitious blueprint for global peace and prosperity established in 2015.[246] Goal 17 recognizes multistakeholder partnerships as important vehicles for mobilizing and sharing knowledge,

[240] cgi.br & /1net, *supra* note 233.

[241] ICANN's Major Agreements and Related Reports *Transition of NTIA's Stewardship of the IANA Functions*, ICANN, https://www.icann.org/resources/pages/process-next-steps-2014-06-06-en#:~:text=On%2014%20March%202014%20the,to%20the%20global%20multistakeholder%20community [https://perma.cc/DV95-TLKB].

[242] Fiona Alexander, *Global Digital Cooperation: Conditions for Success*, *in* TOWARDS A GLOBAL FRAMEWORK FOR CYBER PEACE AND DIGITAL COOPERATION: AN AGENDA FOR THE 2020s, 70–73 (INTERNET GOVERNANCE FORUM BERLIN, Nov. 25-29, 2019).

[243] Larry Strickling, *Remarks of Assistant Secretary Strickling at The Internet Governance Forum USA*, NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION (Jul. 14, 2016), https://ntia.gov/speechtestimony/remarks-assistant-secretary-strickling-internet-governance-forum-usa-1 [https://perma.cc/CL3H-9RBW].

[244] Crocker, *supra* note 222, at 152.

[245] Internet Society, *Understanding the WSIS+10 Review Process*, INTERNET SOCIETY (May 2015), https://www.internetsociety.org/wp-content/uploads/2017/08/WSISplus10-Overview.pdf [https://perma.cc/7DXN-5U6Z].

[246] United Nations Department of Economic and Social Affairs, *The 17 Goals*, UNITED NATIONS (2023), https://sdgs.un.org/goals [https://perma.cc/ZCF4-7NCQ].

expertise, technologies, and financial resources to support the Sustainable Development Goals in all countries.[247] With this mandate, IGF continued its role as a global convener for multistakeholderism in internet governance with no additional powers to bind stakeholders to standards or rules. One stakeholder, Nnenna Nwakanma, spoke to the consensus: "IGF is not what we want it to be. But we do not have a better option. We all wish to be happy, but since we cannot all be happy in our own ways, we settle for collective dissatisfaction."[248] The IGF remained a worthwhile initiative for many stakeholders.

Despite the criticism, the IGF helped entrench an ethos of multistakeholderism in the internet governance space for three key reasons. First, global multistakeholder attendance at the conference brought together people from around the world and across sectors who shared common goals and beliefs. These connections were invaluable to the internet governance ecosystem, which requires a high degree of trust between disparate groups. Second, the language of multistakeholderism was so pervasive that many organizations sought to adopt similar messaging to increase the credibility of their policy solutions. Third, the IGF was purposefully created to avoid regulatory approaches. Instead, it encouraged bottom-up, collaborative solutions. This allowed new organizations to fill the policy vacuum and start new MSIs that could create policy between smaller groups of stakeholders. As a result, in the past 18 years, hundreds of internet governance MSIs have been created to address global issues. There are too many to name and analyze here, but it is worth mentioning a few of the pivotal MSIs that formed following the creation of the IGF in 2006 that are still relevant today.

One topic frequently discussed at the IGF is government censorship and privacy violations. In 2008, there was a series of high-profile instances where technology

[247] United Nations Department of Economic and Social Affairs *Sustainable Development, Multi-stakeholder partnerships*, UNITED NATIONS (2023), https://sdgs.un.org/topics/multi-stakeholder-partnerships [https://perma.cc/A79X-TCMR].

[248] Nnenna Nwakanma, *Because I am involved!, in* Towards a Global Framework for Cyber Peace and Digital Cooperation: An Agenda for the 2020s, 198-200 (Internet Governance Forum Berlin, Nov. 25–29, 2019).

companies legally complied with the Chinese Government's requests for access to data. This data was then used to jail journalists and activists. Finding themselves in a no-win situation – either they violated a legally issued government order or they undermined human rights – certain global companies teamed up with civil society, investors, and academia to create the Global Network Initiative (GNI).[249] GNI established the Principles of Free Expression and Privacy to create a baseline of human rights commitments that participating stakeholders agreed to uphold globally.[250] As an MSI, GNI collaborates to find solutions to the challenges of protecting digital rights globally by drawing on the perspectives, leverage, credibility, and expertise of many different stakeholders.[251] One unique aspect of GNI is its independent assessment process, through which participating companies undergo a third-party review of their efforts to implement the GNI Principles and their more detailed Implementation Guidelines. These assessments focus on internal company systems and emblematic case studies, providing insights to non-company GNI members on sensitive, non-public information and scenarios. GNI's Board is then charged with determining whether each company has implemented the Principles and Implementation Guidelines "in good faith, with improvement over time." Over the past 15 years, GNI has continued to be a leading MSI on internet governance issues, fostering multiistakeholder collaboration to push back on government censorship, enhance shared learning, and provide tools to support responsible decision making by tech companies.

Another topic frequently discussed at the IGF was how global technology companies should operate if national regulations conflict. In 2011, Internet & Jurisdiction Policy Network (I&J) was formed to address the idea that governments, internet companies, civil society, and academics should come together to advance legal interoperability online.[252] I&J focused on specific issues-based problems, believing that cooperation in the internet

[249] Baumann-Pauly et al., *supra* note 28.

[250] Global Network Initiative, *The GNI Principles*, GLOBAL NETWORK INITIATIVE (June 2011), https://globalnetworkinitiative.org/gni-principles/ [https://perma.cc/PFP9-URZ2]; *see also* Baumann-Pauly et al., *supra* note 28.

[251] Global Network Initiative, *About GNI*, GLOBAL NETWORK INITIATIVE https://globalnetworkinitiative.org/about-gni/ [https://perma.cc/LWZ8-ZRFF].

[252] Internet & Jurisdiction Policy Network, *History*, INTERNET & JURISDICTION POLICY NETWORK (2020), https://www.internetjurisdiction.net/about/history [https://perma.cc/6LNC-W7G4].

governance space needed to be addressed with joint agenda setting and policy development by all relevant stakeholders to foster the mutual trust needed for implementation.[253] After four years of meetings and stakeholder consultations, I&J has launched three workstreams: Data & Jurisdiction, Content & Jurisdiction, and Domains & Jurisdiction. These workstreams eventually led to policy option papers and toolkits for governments, tech companies, and civil society, which continue to be relevant and useful for stakeholders across the internet governance sector.

Overall, each of the MSIs discussed above (IETF, Internet Society, ICANN, IGF, NETmundial, GNI, and I&J) succeeded in bringing stakeholders together to address challenging internet governance problems that could not be solved through national laws and tech industry self-regulation on their own. The past 30 years of MSIs have produced a rich, normative framework of stakeholder collaboration to ensure internet governance is a highly interdependent process.[254] However, resurgent top-down multilateral efforts in the internet governance space threaten to undermine this progress.

3. Recent Multilateral Efforts in Internet Governance

As the internet is increasingly intertwined with other global issues, the UN has tried to move internet governance away from multistakeholderism and back into a multilateral framework. To do this it has launched multi-year initiatives that will culminate in the multilateral negotiations of the Global Digital Compact in 2024.[255] The timeline and development of this work is troubling to the broader multistakeholder internet governance community. First, in July 2018, the UN Secretary-General convened a High-Level Panel

[253] Bertrand de la Chapelle, *Towards a Governance Protocol for the Social Hypergraph*, INTERNET GOVERNANCE FORUM BERLIN, "Towards a Global Framework for Cyber Peace and Digital Cooperation: An Agenda for the 2020s," 106-109 (Nov. 25-29, 2019), https://www.hiig.de/wp-content/uploads/2019/11/Kleinwa%CC%88chter-Kettemann-Senges-eds.-Global-Framework-for-Cyber-Peace-2019.pdf.

[254] Komaitis, *supra* note 212.

[255] United Nations Office of the Secretary General's Envoy on Technology, *Report of the Secretary-General Roadmap for Digital Cooperation*, THE UNITED NATIONS TECH ENVOY (May 2020), www.un.org/techenvoy/sites/www.un.org.techenvoy/files/general/Roadmap_for_Digital_Cooperation_9June.pdf [https://perma.cc/EY7X-6WVC].

on Digital Cooperation to advance proposals to strengthen cooperation in the digital space.[256] This kicked off two years of debate around the UN's role in internet governance and culminated in June 2020 with the "Roadmap for Digital Cooperation", which the Secretary-General's Office of the Envoy on Technology was set to implement.[257] Following the publication of Roadmap for Digital Cooperation in 2021, the UN Secretary-General's Envoy on Technology put forward "Our Common Agenda", which proposed a "Global Digital Compact – an Open, Free and Secure Digital Future for All."[258] In 2023, the UN sought input from all stakeholders on the Global Digital Compact, which it plans to integrate into a policy brief to help aid future negotiations on internet governance policies.[259] This negotiation will culminate in 2024, when the UN will host the "Summit of the Future," at which the member states will agree on multilateral solutions to "strengthen" global internet governance.[260] Part of this work will include rules for internet governance for "ensuring the protection of human rights in the digital era."[261] Therefore, while the Global Digital Compact will have multistakeholder input, the final agreement will be multilateral in nature. This is troubling because it creates top-down rules that give countries like Russia and China the ability to weaken the strong human rights protections put in place by a multistakeholder framework.

Unfortunately, the Global Digital Compact is only one slice of the work the UN has launched related to internet governance in the past few years. Additionally, the United Nations Educational, Scientific and Cultural Organization (UNESCO), which promotes freedom of expression, access to information, and digital transformation, has also taken an interest in internet governance issues.[262] UNESCO's "Internet for Trust" is developing

[256] *Id.*

[257] *Id.*

[258] United Nations, *Our Common Agenda Policy Brief 5: A Global Digital Compact – an Open, Free and Secure Digital Future for All*, 30 (May 2023), https://www.un.org/techenvoy/global-digital-compact [https://perma.cc/4QVL-DSDK].

[259] *Id.*

[260] *Id.*

[261] *Id.*

[262] UNESCO, SAFEGUARDING FREEDOM OF EXPRESSION AND ACCESS TO INFORMATION: GUIDELINES FOR A MULTISTAKEHOLDER APPROACH IN THE CONTEXT OF REGULATING DIGITAL PLATFORMS 2 (Apr. 27, 2023), https://unesdoc.unesco.org/ark:/48223/pf0000384031.locale=en [https://perma.cc/4QHM-4QCC].

"guidelines for regulating digital platforms: a multistakeholder approach to safeguarding freedom of expression and access to information".[263] UNESCO is trying to build upon the work it has done in the domain of broadcast regulation, which established principles for internet universality known as the ROAM principles: rights, openness, accessibility to all, and multistakeholder participation.[264] This MSI has been criticized as unnecessary as it is unclear how these guidelines will work with other UN initiatives, including the work of the Envoy on Technology and the IGF.[265] Additionally, the proposed guidelines for regulation are consistent with Article 19 of the ICCPR, which leaves stakeholders to wonder why UNESCO is trying to rewrite settled principles.[266] One theory is that

[263] *Id.*

[264] *Id.; see also Internet Universality Indicators: Background*, UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION (2019), https://www.unesco.org/en/internet-universality-indicators/background [https://perma.cc/MD8K-ZJ4C].

[265] Global Network Initiative, *Global Network Initiative Comments on UNESCO's "Guidance for Regulating Digital Platforms: A Multistakeholder Approach",* GLOBAL NETWORK INITIATIVE (Apr. 2023), https://globalnetworkinitiative.org/wp-content/uploads/2023/01/GNI-Comments-on-UNESCO-draft-Guidance-FINAL.docx-1-1.pdf [https://perma.cc/LWZ8-ZRFF] ("The shortcomings of the consultation process are underscored by the lack of any clear and compelling justification for why this process is being expedited, and the resulting lack of risk-benefit analysis or impact assessment. Perhaps due to this rushed process, the Guidance makes only passing mention of the "Our Common Agenda" report, the "Global Digital Compact," the "UN Summit of the Future," and the WSIS+20 process, and does not offer any clear articulation of how this initiative has been or will be coordinated with other relevant UN offices and initiatives, including the Tech Envoy's office, UN Human Rights (OHCHR), and the Internet Governance Forum.").

[266] David Kaye, *UNESCO Guidelines for Regulating Digital Platforms: A Rough Critique*, UCI LAW INTERNATIONAL JUSTICE CLINIC (Feb. 21, 2023), https://ijclinic.law.uci.edu/2023/02/21/unesco-guidelines-for-regulating-digital-platforms-a-rough-critique/ [https://perma.cc/9RBN-4BN9] ("Second, the draft provides limited if any guidance as to the definition of the problem it is meant to address. From a legality perspective ("provided by law"), this is deeply concerning. Early on, the draft emphasizes "content that is illegal under international human rights law and content that risks significant harm to democracy and the enjoyment of human rights." …I blanched when I saw that [definition], since generally speaking (with two exceptions) international law does not make content illegal; it provides a framework of guaranteed individual rights (Article 19: seek, receive and impart information and ideas of all kinds, regardless of frontiers) along with a set of narrow limitations as to when the state may restrict those rights. It is true that Article 20 of the International Covenant on Civil and Political Rights (ICCPR) obligates states to prohibit "propaganda for war" and "advocacy of national, racial or religious hatred

government actors who do not agree or abide by Article 19 principles and are seeking to weaken existing protections.

The underlying problem with both initiatives is that they are top-down, bureaucratic, and inherently multilateral.[267] As Konstantinos Komaitis, an internet scholar and practitioner, describes it, this model "is based purely on state actors making all decisions at the exclusion of other stakeholders who can make valuable and informed contributions … It will not advance the Internet; on the contrary, it will break it into small pieces. It will fragment it."[268] Multilateral negotiations on internet governance are particularly troubling when non-democratic nations like Russia and China are given a seat at the table. Democratic governments know that the autocrats will not uphold human rights commitments relating to freedom of expression and privacy, so any negotiation will likely weaken the legal commitments already in place. The UN understands the mistrust of civil society and democratic governments and, therefore, cloaks these initiatives as a "multistakeholder effort" by involving the private sector, civil society, and other stakeholders in consultations.[269] In this framework, the UN envisions itself as a convener for multistakeholder policy dialogues, but calls upon member states to develop and implement regulatory frameworks.[270] Therefore, both initiatives embrace the term "multistakeholder" without actually being multistakeholder. As a result, the rights-respecting internet governance community is increasingly uncomfortable with the efforts by the UN to set standards.

---

that constitutes incitement to discrimination, hostility or violence". But if that's what the draft means to address, why not say that directly? This may seem like an editing issue, but the lack of clarity opens the door to state arguments that categories of content many want to limit (e.g., defamation of religion, lèse-majesté, false information, extremism, and so on) are not merely subject to restriction but illegal under international law. This could amount to a major win for governments not, shall we say, entirely enamored of Article 19 of the ICCPR.").

[267] Komaitis, *supra* note 212.

[268] *Id.*

[269] *See* United Nations Office of the Secretary General's Envoy on Technology, *Report of the Secretary-General Roadmap for Digital Cooperation*, *supra* note 255, at 22.

[270] *Id.* at 24.

II. Creating a Typology of Multistakeholder Initiatives for Content Governance

The New Zealand government set up the Call as an MSI to address online user-generated content governance frameworks, understanding that a patchwork of national laws and self-regulation has not been sufficient to eliminate TVEC online while protecting a free, open, and secure internet. The Call drew upon the history of the internet, which is a by-product of multistakeholder collaboration between engineers, governments, tech companies, and civil society.[271] However, despite the deep history of multistakeholderism in internet governance over the past 30 years, many MSIs have only recently started to consider multistakeholder solutions for governance of user-generated content online.[272] While the early MSIs provide a good guide, current models can be slightly adjusted to better address content governance issues and new technologies.[273] This part outlines the different types of MSIs found in the internet governance space.

Overall, this part argues that stakeholders should embrace MSIs to effectively address problems with content governance for three important reasons. First, the content online crosses borders and cannot be effectively legislated by national governments, leaving governance gaps. This is particularly important because not all governments are willing to govern content in a way that respects human rights. To solve this problem, MSIs can exclude bad actors without compromising protections. Second, online platforms will continue to struggle to create their own standards without more input from governments, civil society, and technologists. These inputs can help balance national security interests with freedom of expression and provide local context and accountability. Finally, multistakeholderism is already built into the internet's foundation, and it can therefore be easily imported into new initiatives. Today, the effectiveness of multistakeholderism appears to be taken at face value; almost all internet policymaking initiatives have

[271] Komaitis, *supra* note 212.

[272] *See* de la Chapelle, *supra* note 253, at 106 ("A distributed institutional ecosystem was progressively developed for governance OF the internet58. It efficiently enabled this unique creation of mankind to now serve more than half the world's population. However, equivalent efforts were not devoted to developing the necessary policy-making tools for governance ON the internet, i.e. to organize its uses and mitigate in respect of human rights abuses it can allow. As a result, we witness a legal arms race.").

[273] Strickling & Hill, *supra* note 201.

adopted the model.[274] As the Internet Governance Project, a non-profit organization affiliated with the Georgia Institute of Technology, has argued, the embrace of multistakeholderism is generally a positive development as it ensures that civil society, governments, and tech companies all have a seat at the table.[275] But not all MSIs are created equal, and this framing can mean that the term "multistakeholder" is sometimes applied unequally, and therefore critical analysis is required.[276]

The variation of challenges, actors, and structures can make it difficult to have one definition of "MSI". As the former administrator of NTIA Larry Strickling notes, multistakeholder models have their own unique contours, but, "few, if any, of the models currently in use are static; rather, they are constantly evolving to meet new and yet uncharted governance challenges."[277] Part I defines an MSI as two or more classes of actors engaged in a common governance enterprise to solve a wider problem, where decision-making authority is distributed between actors based on procedural rules.[278] Therefore, two core elements that define an initiative as multistakeholder are: the inclusion of multiple types of actors and the distribution of decision-making authority based on procedural rules.[279]

---

[274] JYOTI PANDAY, MILTON MUELLER & FARZANEH BADIEI, *MULTISTAKEHOLDERISM & PLATFORM CONTENT GOVERNANCE: AN ASSESSMENT FRAMEWORK WITH APPLICATIONS* 2 (Jan. 20, 2022), https://www.internetgovernance.org/wp-content/uploads/MS-Content.docx-1.pdf [https://perma.cc/ZX8M-5WND] ("The term "multistakeholder" (MS) is now claimed as a legitimizing feature of various international, Internet-related policy development entities. Civil society in particular tends to demand multistakeholder governance in order to gain entry into decision-making processes otherwise controlled by business or government. While in many ways the advance of MS governance is a good thing, it also means that the term can be applied loosely or even deceptively. We need to ask what multistakeholderism really means in a particular policy environment, and we need to assess critically how these organizations are being set up.").

[275] *Id* at 5.

[276] *Id.* at 1.

[277] Strickling & Hill, *supra* note 201, at 45.

[278] Raymond & DeNardis, *supra* note 36.

[279] *See* Panday, Mueller & Badiei *supra* note 274 (Panday, Mueller & Badiei, also includes "funding" as a distinctive category. This report addresses funding considerations as part of the terms of reference rather than its own category)

These two core elements frequently take two forms within any given MSI, creating four overarching types of MSIs. In terms of the first element, the inclusion of stakeholders, there are two principal systems: either anyone who is interested can participate, or the MSI only allows stakeholders who meet certain criteria to join. The second element, regarding how the MSI distributes decision-making authority based on procedural rules, is slightly more complex. One option is for decisions to be made by the consensus of all stakeholders; in this case, the governance itself takes a multistakeholder form.[280] Consensus-based institutions are considered more multistakeholder in nature and can increase the possibility that the solution presented by the MSI is adopted in the long-run.[281] The second option is for stakeholders to serve a purely consultative purpose; in this case, decision-making happens unilaterally by the designated authority.[282] This can also be considered "ancillary" multistakeholder governance, because it involves the multistakeholder body acting as an appendage to a decision-making body.[283] While neither of these decisions are straightforward or strictly binary, it is helpful to make distinctions to examine what types of MSIs are best suited for each unique situation.

There are thus four types of MSIs:

[280] Jan Aart Scholte, *Multistakeholderism: Filling the Global Governance Gap?* 4 (Apr. 6, 2020), https://globalchallenges.org/multistakeholderism-filling-the-global-governance-gap/ [https://perma.cc/RPV3-ZJDB].

[281] Strickling & Hill, *supra* note 205 at 49 ("Also, to maximize the possibility of success, participants must be the ones who make the final decision on a particular issue, not the convening body. This feature is one of the fundamental differences between a multi-stakeholder process and consultation. If participants are not empowered to make a final decision, then a process is merely consultative. By contrast, multi-stakeholder processes that place responsibility for final decision making on the participants themselves are generally viewed as more legitimate. They also tend to be more successful because the prospect of fashioning policy, and not just offering commentary, frequently induces the participants to put in the extra effort needed to reach a consensus. Further, entrusting the participants with the power to make decisions also reduces the possibility of non-participants mounting a collateral challenge of the outcome by appealing to others who did not choose to participate.").

[282] *See* Panday, Mueller & Badiei *supra* note 278, at 4-5; the authors have a third category where MSIs have a consultative body that is advisory in status, but its formal advice triggers some kind of procedure and cannot be ignored. However, for purposes of this report, this middle ground will be included in the consultative function.

[283] Scholte, *supra* note 280 at 4.

- egalitarian: any stakeholder, consensus decision-making
- consultative: any stakeholder, unilateral decision-making
- restricted: limited stakeholders, unilateral decision-making
- curated: limited stakeholders, consensus decision-making.

A. Egalitarian MSIs: Any Stakeholder, Consensus Decision-making

This type of MSI was the vision of early internet adopters – people like John Perry Barlow thought that the rules for the internet would emerge through community engagement and consensus.[284] As such, we saw examples of this type of MSI in the early days of the internet.[285] Egalitarian MSIs look like Athenian democracy, where all stakeholders must participate directly in the decision-making. While this may sound aspirational, there are many MSIs that operate in this manner. The IETF is an example. It does not have an official or defined membership; rather, it allows anyone to participate.[286] Many of the stakeholders come from industry, government, civil society, and the technical community, but everyone participates in their personal capacity. The IETF has no formal voting process but makes decisions based on what has been called "rough consensus and running code."[287] The IETF's process of standard-setting is a significant investment in time and energy by stakeholders, but in the end, the community is able to progress with the greatest amount of input and consensus possible.

The IETF has sustained this type of MSI for decades, but not all egalitarian MSIs have succeeded. For example, in 2009, Facebook (now Meta) experimented with its own

[284] Barlow, *supra* note 35.

[285] Komaitis, *supra* note 216.

[286] P. Resnick, *On Consensus and Humming in the IETF*, INTERNET ENGINEERING TASK FORCE (Jun. 2014), https://datatracker.ietf.org/doc/html/rfc7282 [https://perma.cc/SM9R-R3SM]; *see also* Niels ten Oever, *Plus Hum Now: Decision Making at the IETF*, HACK_CURIO (Mar. 2018), https://hackcur.io/please-hum-now/ [https://perma.cc/4Q3Y-7YHS].

[287] Raymond & DeNardis, *Multistakeholderism: Anatomy of an Inchoate Global Institution*, 7 INT'L THEORY 572, 597 (2015).

egalitarian MSI after changes to its privacy policy were significantly criticized.[288] Facebook announced that it would develop the site's terms of service through consensus building, by asking users to weigh in on company policies. In 2012, Facebook tested this approach by putting forward two different privacy policies; it asked users to vote, committing that if more than 30 per cent of all active registered users participated, their decision would be binding.[289] When it came time to vote, only 665,654 people voted – about 0.3 per cent of Facebook's 200 million users at the time.[290] Facebook followed the majority opinion of the lackluster showing, but since most people voted for the proposed changes, the decision was criticized for being a cover for the company to do a thing it already wanted to do.[291] In the end, Meta scrapped the initiative, which the *Los Angeles Times* called "a homework assignment no one did."[292] Overall, an egalitarian MSI approach works best where relevant stakeholders are deeply invested in the outcome and highly motivated to find consensus.

B. Consultative MSIs: Any Stakeholder, Unilateral Decision-making

A consultative MSI is created when one stakeholder has unilateral decision-making authority but seeks input from all interested stakeholders. The stakeholder input is considered but not dispositive to the final decision. This type of multistakeholder governance is found throughout many democratic institutions, in places like the US Administrative Procedures Act, which requires a "notice and comment period" before a

[288] NICOLAS P. SUZOR, LAWLESS: THE SECRET RULE THAT GOVERN OUR DIGITAL LIVES 10 (2019); Mark, Update on Terms, Facebook (Feb. 17, 2009) www.facebook.com/notes/facebook/update-on-terms/54746167130 [https://perma.cc/NZ3G-MVM2].

[289] Adi Robertson, *Mark Zuckerberg wants to democratize Facebook – here's what happened when he tried*, THE VERGE (Apr. 6, 2018), https://www.theverge.com/2018/4/5/17176834/mark-zuckerberg-facebook-democracy-governance-vote-failure [https://perma.cc/PEN7-HUDP].

[290] *Id.*

[291] *Id.*

[292] David Sarno, *Facebook governance vote is a homework assignment no one did*, LOS ANGELES TIMES (Apr. 23, 2009), https://www.latimes.com/archives/blogs/technology-blog/story/2009-04-23/facebook-governance-vote-is-a-homework-assignment-no-one-did [https://perma.cc/R4NA-AM86].

regulatory agency can issue a final rule.[293] In these MSIs, the multistakeholder community acts as a sounding board or advisor to the decision maker.[294] One example of this type of consultative MSI is Meta's Oversight Board, which hears appeals from users regarding content moderation decisions taken by Meta on Facebook and Instagram and issues binding decisions to the company.[295] As part of the Board's work when deciding cases, it opens up a "public comment process", which allows any stakeholder to submit their thoughts on how the company should have moderated a piece of content or crafted its policies.[296] The Board seeks advice from all stakeholders to gain local context and subject matter expertise to improve the quality of their decisions.[297]

Other consultative MSIs previously discussed include UNESCO's Internet for Trust and its development of "guidelines for regulating digital platforms: a multistakeholder approach to safeguarding freedom of expression and access to information" and the UN's

[293] Justia, *The Notice and Comment Process Legally Provided for Agency Rulemaking*, JUSTIA (May 2023), https://www.justia.com/administrative-law/rulemaking-writing-agency-regulations/notice-and-comment/ [https://perma.cc/EQ83-YF2U]; *see also* Strickling & Hill, *supra* note 201 at 49 ("Notwithstanding the desire of government officials to allow a group of stakeholders to reach a consensus decision, the laws of the government, such as the Administrative Procedures Act in the United States, may prohibit giving the decision-making power to a group of stakeholders and require the agency to conduct subsequent notice and comment on the rule-making processes, thus diminishing the incentive of stakeholders to work together to reach consensus in the multi-stakeholder discussions.").

[294] JYOTI PANDAY ET AL., MULTISTAKEHOLDERISM & PLATFORM CONTENT GOVERNANCE: AN ASSESSMENT FRAMEWORK WITH APPLICATIONS 2 (2022) ("[T]he additional stakeholders serve a purely advisory or consultative function; the act as a sounding board or advisor to the decision maker.").

[295] The Oversight Bd., RULEBOOK FOR CASE REVIEW AND POLICY GUIDANCE 3 (Nov. 2020) https://oversightboard.com/sr/rulebook-for-case-review-and-policy-guidance (noting the Board opens public comment processes for both case decisions and Policy Advisory Opinions).

[296] *Id.* at 9.

[297] *Id.*; *see also* THE OVERSIGHT BD., ANNUAL REPORT 13 (2023), https://oversightboard.com/news/560960906211177-2022-annual-report-oversight-board-reviews-meta-s-changes-to-bring-fairness-and-transparency-to-its-platforms/ ("As a Board, our achievements so far have been made possible by listening to and collaborating with researchers, civil society groups and others who have worked for many years on the issues we are dealing with. To find practical solutions to our strategic priorities, and the enormously challenging issues they raise, the subject-matter expertise and local knowledge of these stakeholders is essential.").

Roadmap for Digital Cooperation.[298] In these cases, the UNESCO Secretariat or the UN Secretary-General is developing the new guidelines and, as part of the process, will conduct multistakeholder consultations.[299] In these examples, a multistakeholder community is consulted, but the authority for deciding ultimately sits with one stakeholder. The advantage of a consultative MSI is that it allows for a wide range of voices to participate in a process without relying on consensus to reach a final decision. This expedites the process and provides all stakeholders an opportunity to weigh in. However, this type of MSI can be deceptive if the term "multistakeholder" is used to legitimize a process without clearly explaining stakeholders' lack of decision-making authority.[300]

C. Restricted MSIs: Limited Stakeholders, Unilateral decision-making

A restricted MSIs allows only qualified stakeholders to participate in the initiative and decision-making to happen unilaterally. In some respects, the US Supreme Court's amicus curiae process is a restricted MSI, because it allows for the consideration of stakeholders' views on current cases but restricts those stakeholders to attorneys admitted to practice before the court.[301] Given the restrictions, many would not consider this to be a multistakeholder process. Another decision-making institution that embraces the restricted MSI model is the ITU, which allows non-state actors to join the multistakeholder processes through restricted participation. A non-state stakeholder is referred to as a "sector member" and must apply to join and be sponsored by a member state.[302] Sector members can participate in day-to-day standards-setting work within the

[298] UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION, *supra* note 262, at 1.

[299] *Id.* at 2.

[300] JYOTI PANDAY ET AL., MULTISTAKEHOLDERISM & PLATFORM CONTENT GOVERNANCE: AN ASSESSMENT FRAMEWORK WITH APPLICATIONS 2 (2022).

[301] SCOTT S. HARRIS, MEMORANDUM TO THOSE INTENDING TO FILE AN AMICUS CURIAE BRIEF IN THE SUPREME COURT OF THE UNITED STATES 4 (2023) https://www.supremecourt.gov/casehand/AmicusGuide2023.pdf.

[302] ITU Membership Terms & Conditions, INTERNATIONAL TELECOMMUNICATIONS UNION, https://www.itu.int/hub/membership/become-a-member/member-terms-conditions/ (last visited Oct. 27, 2023) [https://perma.cc/Y28C-F83M].

working groups of the ITU, but any recommendations made by the working groups are ultimately approved exclusively by member states.[303]

One example of a restricted MSI that is working to address content moderation challenges surrounding TVEC online is the Organization for Economic Co-operation and Development (OECD), which developed a Voluntary Transparency Reporting Framework for TVEC online.[304] Created after the Christchurch attack in 2019, the OECD created the Framework "in collaboration with member countries, business, civil society and academia, to develop a multi-stakeholder, consensus-driven framework" for voluntary transparency reporting by companies.[305] The OECD limited participation to its member states and invited guests, and the OECD had final decision-making authority on what criteria would be included in the Framework. In the past three years, the OECD has published an annual report that takes stock of the "current policies and procedures related to TVEC of the world's leading online platforms and other online content sharing services."[306] This MSI has provided insight into global efforts to reduce TVEC online. However, in the cases of both the OECD and the ITU, stakeholder participation is limited and decisions on the outcome are made by one actor.

D. Curated MSIs: Limited Stakeholders, Consensus Decision-making

Of the internet governance MSIs mentioned in Part I, curated MSIs are the most common type; this is the approach taken by the Christchurch Call. Stakeholders in these MSIs uphold the ideals of consensus in the same way the early internet adopters envisioned

---

[303] Mark Raymond & Laura DeNardis, *Multistakeholderism: Anatomy of an Inchoate Global Institution*, 7 INT'L THEORY 572, 597 (2015).

[304] ORG. FOR ECON. COOP. AND DEV., TRANSPARENCY REPORTING ON TERRORIST AND VIOLENT EXTREMIST CONTENT ONLINE 2022 at 9 (2022) https://www.oecd.org/digital/vtrf/.

[305] Maddie Cannon, *A Review of International Multi-Stakeholder Frameworks for Countering Terrorism and Violent Extremism Online*, GLOBAL NETWORK ON EXTREMISM & TECHNOLOGY (Mar. 16, 2022), https://gnet-research.org/2022/03/16/a-review-of-international-multi-stakeholder-frameworks-for-countering-terrorism-and-violent-extremism-online/ [https://perma.cc/B8XW-XNDU].

[306] ORG. FOR ECON. COOP. AND DEV., TRANSPARENCY REPORTING ON TERRORIST AND VIOLENT EXTREMIST CONTENT ONLINE 2022 at 2 (2022) https://www.oecd.org/digital/vtrf/.

through shared decision-making authority. However, stakeholder participation is limited to those who meet a defined set of criteria. One important reason for limiting participation in internet governance MSIs stems from the divide between democratic and authoritarian regimes and their vastly different approaches to upholding the human rights principles as they relate to content moderation practices. Therefore, in many instances, a demonstrated commitment to upholding human rights is a baseline for participation in the MSI to safeguard the consensus-driven outcomes.

GNI and I&J are two examples of internet governance curated MSIs. To participate in GNI, stakeholders must support the organization's established Principles of Free Expression and Privacy and undergo a due diligence check.[307] GNI develops governance structures for companies to implement through a consensus-based multistakeholder process.[308] Likewise, I&J works with relevant stakeholders committed to preserving "the cross-border nature of the internet, [protecting] human rights, [fighting] abuses, and [enabling] the global digital economy."[309] I&J working groups develop consensus-based toolkits that strive to fill an institutional gap in internet governance.[310] Additionally, many curated MSIs address content moderation challenges surrounding TVEC online, including the EU Internet Forum and the GIFCT. As a curated MSI, the EU Internet Forum provides a collaborative environment in which partners discuss and address the challenges posed by malicious and illegal content online – including TVEC.[311] The Forum is chaired by the European Commission, which invites stakeholders from the internet industry and civil society actors to participate.[312] The GIFCT has a multistakeholder Independent Advisory Committee that advises the Operating Board through regular

[307] *The GNI Principles*, Global Network Initiative, https://globalnetworkinitiative.org/gni-principles/ (last updated 2017) [https://perma.cc/RE8B-LGX6].

[308] *Id.*; *see also* Baumann-Pauly et al., *supra* note 28.

[309] INTERNET & JURISDICTION POL'Y NETWORK PROGRESS REP. 2021 at 23, INTERNET & JURISDICTION (2021) https://www.internetjurisdiction.net/uploads/pdfs/IJPN-Progress-Report-2021.pdf.

[310] *Id.* at 5.

[311] European Union Internet Forum Membership, EUR. COMM'N, https://home-affairs.ec.europa.eu/networks/european-union-internet-forum-euif_en (last visited Oct. 27, 2023) [https://perma.cc/K2J6-J8WE].

[312] *Id.*

meetings and makes decisions by consensus.[313] The Independent Advisory Committee appoints representatives from civil society organizations and academia as well as representatives from governments who are members of the Freedom Online Coalition, a group of governments dedicated to human rights.[314]

III. THE FUTURE OF THE CHRISTCHURCH CALL TO ACTION

This part will apply the lessons learned from multistakeholder governance to the Christchurch Call to Action. The first section discusses what happened on March 15, 2019, with an emphasis on how the Christchurch shooter exploited social media to amplify his terrorist attack. Next, this section catalogs the creation of the Call and provides an overview of the work it has done over the past four years. The second section provides an analysis of the progress of the Call, including its achievements and where the Call is still working to fulfill its commitments. Finally, the third section examines the work of the Call and its pivot towards new and emerging technologies – including GenAI.

A. History of the Christchurch Call to Action

1. March 15, 2019

On the Friday morning of March 15, 2019, the Christchurch shooter drove from his home in Dunedin, New Zealand to Christchurch, a city with a small but growing Muslim population.[315] At 1:18 pm, the individual emailed his 74-page manifesto with details of the attack plans to dozens of government officials and media organizations.[316] Eight minutes

[313] GLOBAL INTERNET FORUM TO COUNTER TERRORISM, GIFCT INDEPENDENT ADVISORY COMMITTEE: INTERIM TERMS OF REFERENCE 2 (2023), https://gifct.org/wp-content/uploads/2021/09/GIFCT-IAC-Terms-of-Reference.pdf.

[314] *Id.* at 3.

[315] ROYAL COMM'N OF INQUIRY INTO THE TERRORIST ATTACK ON CHRISTCHURCH MASJIDAIN ON 15 MARCH 2019, KO TŌ TĀTOU KĀINGA TĒNEI: ROYAL COMMISSION OF INQUIRY INTO THE TERRORIST ATTACK ON CHRISTCHURCH MASJIDAIN ON 15 MARCH 2019 at 40 (2020), https://christchurchattack.royalcommission.nz/assets/Report-Volumes-and-Parts/Ko-to-tatou-kainga-tenei-Volume-1-v2.pdf.

[316] *Id.* at 42.

later, at 1:26 pm, he updated his Facebook status with links to seven different file-sharing websites that contained copies of a manifesto he had written explaining his motivation.[317] He then posted to 8chan, an online message board frequently used by white supremacists, a link to his Facebook account with the message, "well lads, it's time to stop shitposting and time to make a real life effort post. I will carry out an attack on the invaders, and will even live stream the attack via facebook … I have provided links to my writings below, please do your part by spreading my message, making memes and shitposting as you usually do."[318]

At 1:33 pm, he linked the feed of the GoPro camera on his helmet to his mobile phone and started a Facebook livestream of the footage.[319] At 1:40 pm, he entered the Masjid an-Nur mosque and opened fire on the worshippers gathered for Friday prayers.[320] After completing the first attack, he went back to his car and drove to a second nearby location, the Linwood Islamic Centre, arriving there at 1:52 pm. There, he opened fire on worshippers again.[321] The individual got back in his car to attack a third location, the Al-Nur Early Childhood Education and Care Centre, but was arrested after two New Zealand police officers rammed his vehicle with their car.[322] In total, 51 people died and 40 people suffered gunshot injuries.[323] On August 27, 2020, the individual was sentenced to life imprisonment without parole for the murder of 51 individuals and designated as a terrorist entity under the Terrorism Suppression Act 2020.[324]

On Facebook, the live feed continued throughout the attack. It remained on the individual's page for another 12 minutes before Facebook was notified by police and removed the content.[325] The video of the attack was viewed 4,000 times before it was

[317] *Id.* at 41.
[318] *Id.*
[319] *Id.* at 42.
[320] *Id.* at 43.
[321] *Id.* at 44.
[322] *Id.* at 45, 46.
[323] *Id.* at 46.
[324] *Id.* at 47.
[325] Chris Sonderby, *Update on New Zealand*, FACEBOOK NEWSROOM (Mar. 18. 2019), https://about.fb.com/news/2019/03/update-on-new-zealand/ [https://perma.cc/RJ2B-VV6F].

taken down by Facebook.[326] Responding to the call to action from the individual's message on 8chan, like-minded extremists copied and shared the footage across the internet on platforms such as Twitter, YouTube, and Reddit. In the first 24 hours after the attack, Facebook removed or blocked over 1.5 million uploads of the video.[327] As quickly as social media platforms could take down the content, it was re-uploaded, "sometimes spliced into new video clips, making it impossible to detect quickly."[328] The video was widely viewed across New Zealand as it reappeared on social media – sometimes promoted by the company's algorithms, which amplified trending content. While it is impossible to know how many New Zealanders saw the video, in the first week after the attacks 8,000 people who saw it called mental health support lines.[329]

As Kevin Roose of the *New York Times* noted, the Christchurch massacre:

> … felt like a first – an internet-native mass shooting, conceived and produced entirely within the irony-soaked discourse of modern extremism. The attack was teased on Twitter, announced on the online message board 8chan and broadcast live on Facebook. The footage was then replayed endlessly on YouTube, Twitter and Reddit, as the platforms scrambled to take down the clips nearly as fast as new copies popped up to replace them.[330]

Clips made their way to the mainstream platforms after spreading across smaller social media sites like 8chan, 4chan, Discord, and Gab. The individual, who was steeped in internet subcultures, carefully planned his attack to go viral on these sites by providing followers with many "in-joke" opportunities to create memes. He had stated in his

[326] *Id.*

[327] *Id.*

[328] ROYAL COMM'N OF INQUIRY INTO THE TERRORIST ATTACK ON CHRISTCHURCH MASJIDAIN ON 15 MARCH 2019, *supra* note 315, at 46.

[329] Jacinda Ardern, *How to Stop the Next Christchurch Massacre*, N.Y. TIMES: SUN. OPINION (May 11, 2019), https://www.nytimes.com/2019/05/11/opinion/sunday/jacinda-ardern-social-media.html [https://perma.cc/8XND-6R7P].

[330] Kevin Roose, *A Mass Murder of, and for, the Internet*, N.Y. TIMES: THE SHIFT (Mar. 15, 2019), https://www.nytimes.com/2019/03/15/technology/facebook-youtube-christchurch-shooting.html [https://perma.cc/QM5A-YBTT].

manifesto, "memes have done more for the ethnonationalist movement than any manifesto."[331] Without any meaningful content moderation or cooperation, the Christchurch manifesto and video content circulated freely on the smaller unmoderated platforms whose users quickly produced memes and spread the individual's message.

As the video and manifesto migrated from these smaller sites, the large platforms faced several challenges in the first 24 hours. First, companies frequently rely on "hash" technology to remove objectively awful content such as child sexual abuse material and ISIS beheading videos.[332] "Hashing" works by creating a digital fingerprint of the unique pixels of an image.[333] In the case of the Christchurch attack, extremists sympathetic to the shooter were slightly altering the video before uploading it, or creating memes, which evaded detection by the hash technology.[334] Second, graphic content is primarily removed from platforms using AI. However, in 2019, the online platform's AI tools could not identify first-person shooting videos as graphic content because no large dataset of similar videos existed to train the algorithm.[335] Third, removal efforts were made more difficult as clips of the video were included in reporting on mainstream media outlets. The media clips spliced the footage from the shooter making it impossible to effectively filter the video that included the more graphic scenes of the massacre. Eventually, YouTube stopped trying to differentiate between media footage and the massacre video and blocked all videos using the footage.[336]

[331] Graham Macklin, *The Christchurch Attacks: Livestream Terror in the Viral Video Age*, 12:6 CTC SENT. 18, 25 (2019).

[332] Kate Klonik, *Inside the Team at Facebook that Dealt with the Christchuch Shooting*, THE NEW YORKER (Apr. 25, 2019), https://www.newyorker.com/news/news-desk/inside-the-team-at-facebook-that-dealt-with-the-christchurch-shooting [https://perma.cc/A6RY-68CY].

[333] *See* Part I, Section 2 for more background on how hash-sharing technology works.

[334] *See* Klonick, *supra* note 332.

[335] *See* Evelyn Douek, *Australia's "Abhorrent Violent Material" Law: Shouting "Nerd Harder" and Drowning Out Speech*, 94 AUSTRALIAN L. J. 41, 50 (2020).

[336] *See* Alex Hern, *Facebook and YouTube Defend Response to Christchurch Videos*, THE GUARDIAN (Mar. 19, 2019, 8:18 AM), https://www.theguardian.com/world/2019/mar/19/facebook-and-youtube-defend-response-to-christchurch-videos [https://perma.cc/G8KK-RCJG].

After addressing the initial technical problems, in the days following the attack, tech industry leaders issued their well-rehearsed *mea culpa* about the proliferation of harmful content online. The tech executives acknowledged that they needed to improve structures to stop this type of event from happening again. Meta's chief operating officer, Sheryl Sandberg, said in a letter responding to the attacks, "[m]any of you have also rightly questioned how online platforms such as Facebook were used to circulate horrific videos of the attack … We have heard feedback that we must do more – and we agree."[337] Microsoft's president, Brad Smith, stated, "it's clear that we need to learn from and take new action based on what happened in Christchurch."[338] YouTube's chief product officer, Neal Mohan, said, "this incident has shown that, especially in the case of more viral videos like this one, there's more work to be done."[339] While livestreaming a terrorist attack of this magnitude was unprecedented, the act itself was unfortunately all too common.[340]

After years of promises from tech companies to clean up their platforms, these statements fell short of convincing lawmakers that they could solve the issue alone. The UK Home Secretary, Sajid Javid, said, "[o]nline platforms have a responsibility not to do the terrorists' work for them. This terrorist filmed his shooting with the intention of spreading his ideology. Tech companies must do more to stop his messages being broadcast on their

[337] Julia Carrie Wong, *Facebook Finally Responds to New Zealand on Christchurch Attack*, THE GUARDIAN (Mar. 29, 2019, 6:57 PM), https://www.theguardian.com/us-news/2019/mar/29/facebook-new-zealand-christchurch-attack-response [https://perma.cc/R7LS-VWWE].

[338] Brad Smith, *A Tragedy That Calls for More than Words: The Need for the Tech Sector to Learn and Act After Events in New Zealand*, MICROSOFT (Mar. 24, 2019), https://blogs.microsoft.com/on-the-issues/2019/03/24/a-tragedy-that-calls-for-more-than-words-the-need-for-the-tech-sector-to-learn-and-act-after-events-in-new-zealand/ [https://perma.cc/FKV4-MCDT].

[339] Elizabeth Dwoskin & Craig Timberg, *Inside YouTube's Struggles to Shut Down Video of the New Zealand Shooting — and the Humans Who Outsmarted its Systems*, WASH. POST (Mar. 18, 2019, 6:00 AM), https://www.washingtonpost.com/technology/2019/03/18/inside-youtubes-struggles-shut-down-video-new-zealand-shooting-humans-who-outsmarted-its-systems/ [https://perma.cc/8LYE-UCSD].

[340] For instance, a shooter in Cleveland had livestreamed his violent attack several years ago. *See* Jane Morice, *Facebook Killer Chooses Victim at Random, Laughs About Killing in Videos*, CLEVELAND.COM (Apr. 17, 2017, 3:37 AM), https://www.cleveland.com/metro/2017/04/accused_facebook_live_killer_c.html [https://perma.cc/7C7T-TD2D].

platforms."[341] In the US, the Chairman of the House Homeland Security Committee, Bennie Thompson, told tech company executives during a congressional hearing weeks later, "[y]our companies must prioritize responding to these toxic and violent ideologies with resources and attention. If you are unwilling to do so, Congress must consider policies to ensure that terrorist content is not distributed on your platforms[.]"[342] In perhaps the most extreme response, just weeks after the attack, the Australian Government passed the Criminal Code Amendment (Sharing of Abhorrent Violent Material).[343] Without much debate or input from stakeholders, the Australian legislation created criminal and civil penalties for tech companies if users post abhorrent violent material, including the Christchurch video and manifesto.[344]

Like others, the New Zealand Government considered its options for how to move forward. Prime Minister Jacinda Ardern said in a speech to Parliament in the days after the attack, "[w]e cannot simply sit back and accept that these platforms just exist and that what is said on them is not the responsibility of the place where they are published […]They are the publisher, not just the postman. It cannot be a case of all profit, no responsibility."[345] In the aftermath of the event, Ardern would go on to say:

[341] Hern, *supra* note 336.

[342] Lauren Feiner, *House Homeland Security Chair Calls on Facebook, YouTube, Twitter and Microsoft to Explain the Spread of Mosque Shooting Video*, CNBC (Mar. 19, 2019, 5:00 PM), https://www.cnbc.com/2019/03/19/rep-bennie-thompson-asks-tech-to-explain-mosque-shooting-video-spread.html [https://perma.cc/XWC3-8VKP].

[343] *See* Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019, No. 38, § 474.31 (2019) (Austl.). *See also* Douek, *supra* note 335, at 42.

[344] *See* Douek, *supra* note 335, at 42 ("The Australian Parliament showed little regard for these complexities when it passed the *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019 (Cth) (AVM Act)* on 4 April 2019. The Act made its way through both houses of Parliament in less than two days and came into force two days later. The government did not consult with experts, civil society or industry. Proudly touting the legislation as a "world first", the government did not stop to ask itself why that was.").

[345] Matt Novak, *New Zealand's Prime Minister Says Social Media Can't Be 'All Profit, No Responsibility'*, GIZMODO (Mar. 19, 2019), https://www.gizmodo.com.au/2019/03/new-zealands-prime-minister-says-social-media-cant-be-all-profit-no-responsibility/ [https://perma.cc/X3Q8-PY6D].

> "I don't think anyone wants platitudes. We didn't want just a response to that individual act. If anything, we wanted to make sure that the pain and the horror of not just the act itself, but the fact that it was then broadcast, didn't occur elsewhere. … Governments will not be able to regulate their way out of this problem. Tech companies, perhaps, if they continue to work on their own may not find solutions, but through collaboration together, I do believe we can make progress[.]"[346]

2. The Creation of the Christchurch Call to Action

In the days following March 15, 2019, Prime Minister Ardern explored options to ensure this type of attack never happened again. Ardern and her team understood that the world's outrage would eventually dissipate, so they needed to use their moral authority to build an initiative that could have a lasting impact.[347] With this in mind, two weeks after the attacks, Ardern and her team met Microsoft President Brad Smith, who happened to be in New Zealand for a visit, planned long before the attack.[348] After discussing multiple options, the teams sketched out the idea of a "Christchurch Call to Action'' as an MSI. The solution was akin to the Paris Call for Trust and Security in Cyberspace – an MSI designed to protect international norms against cyber-attacks that had launched six months prior.[349] Due in part to France's successful leadership of the Paris Call, the New Zealand Government found a willing partner in President Macron to set up a similar initiative to address the problem of TVEC online. The timing was fortuitous, as France was set to host the Tech for Good Summit alongside the G7 Digital Ministers Meeting just a few weeks later. Organizers of the Call hoped to formally announce it at the Summit, and sign on other government leaders.[350]

---

[346] *Prime Minister Jacinda Ardern: Can We Work Together to End Violent Extremism Online?*, TOOLS & WEAPONS WITH BRAD SMITH (July 6, 2022), https://tools-and-weapons-with-brad-smith.simplecast.com/episodes/jacinda-ardern-can-tech-and-government-end-extremism-online/transcript [https://perma.cc/Z8MY-PTJQ].

[347] *See* Smith & Browne, *supra* note 189, at 125.

[348] *See* Smith, *supra* note 338.

[349] *See* Smith & Browne, *supra* note 189, at 127.

[350] *See NZ and France Seek to End Use of Social Media for Acts of Terrorism*, CHRISTCHURCH CALL (Apr. 24, 2019), https://www.christchurchcall.com/media-and-resources/news-and-updates/nz-and-france-seek-to-end-use-of-social-media-for-acts-of-terrorism/ [https://perma.cc/QT2M-78KB].

To pull off a launch just six weeks later, the New Zealand Government and Microsoft teams worked around the clock to secure additional partners in the major platforms, including Google (and its subsidiary YouTube), Facebook, Twitter, and Amazon, as well as two French companies, Dailymotion and Qwant.[351] The eight companies had very different platforms, business models, engineering capabilities, and experiences with TVEC online, but they were able to find commonality in wanting to prevent another Christchurch-type attack.[352] As a result, the companies worked with New Zealand and France to come up with nine steps they could take to address TVEC online. Five of these steps would be for individual companies to take: tighten their terms of service, better manage live videos, respond to user reports of abuse, improve technology controls, and publish transparency reports. Four of the steps were industry-wide: launch a crisis response protocol, develop open source-based technology, improve user education, and support additional research to prevent TVEC online.[353]

Throughout the initial creation of the Christchurch Call, civil society was skeptical that the initiative could produce any meaningful outcomes or commit to a human rights-respecting framework.[354] Behind this skepticism was a general frustration with the EU, which was in the process of passing a regulation on "Preventing the Dissemination of Terrorist Content Online." As discussed above, these regulations were incredibly controversial with many civil society organizations viewing them as threatening to free expression and human rights.[355] Additionally, the regulations came on the heels of the EU

[351] *See Supporters*, CHRISTCHURCH CALL, https://www.christchurchcall.com/our-community/countries-and-states/ [https://perma.cc/X78E-URTA] (last visited Oct. 20, 2023).

[352] *See* Smith & Browne, *supra* note 189, at 126.

[353] *See id.*

[354] *See* Courtney Radsch, *Taking Down Terrorism Online While Preserving Free Expression*, MEDIUM (May 15, 2019), https://medium.com/@old-cradsch/taking-down-terrorism-online-while-preserve-our-free-expression-678ab1100a67 [https://perma.cc/ZP9Z-WUQH]. *See also* Jillian C. York, *The Christchurch Call: The Good, the Not-So-Good, and the Ugly*, ELEC. FRONTIER FOUND. (May 16, 2019), https://www.eff.org/deeplinks/2019/05/christchurch-call-good-not-so-good-and-ugly [https://perma.cc/ZSG7-RCC9].

[355] See Part I(A)(1)(c), related to the EU's TCO and discussion with civil society organizations that happened in 2019. *See also* Citron, *supra* note 106, at 1038–39.

passing several laws regulating content moderation that were thought to be technologically unworkable, restrictive of human rights, ambiguously drafted and massively overreaching.[356] As a result, there was little trust between civil society organizations and European regulators, which meant that the inclusion of France as a co-lead for the Call raised concerns.

The hostility towards European regulators hung over the room as several civil society organizations met with the New Zealand Prime Minister the day before the Call was to be launched in Paris. At that meeting, civil society organizations presented a letter with input from dozens of signatories detailing their concerns.[357] Included in the letter were complaints civil society raised with European regulators in previous conversations, including the lack of clear definitions of "terrorism" and "violent extremism," the need to differentiate between social media companies and internet infrastructure providers, and the importance of governmental transparency around take-down requests.[358] Additionally, the letter expressed concern that civil society had been left out of the early stages of negotiations and a perceived lack of desire for meaningful input from civil society by governments.[359] The New Zealand and French teams worked closely with the group to resolve some of these issues, and won over a number of civil society representatives. In the end, all stakeholders pledged to work together to better incorporate civil society views into the text of the Call commitments themselves.

Civil society was not the only recalcitrant stakeholder in May 2019. Despite being the corporate home of most of the major tech platforms, the US Government declined to join the Call, stating it was "not in a position to join the endorsement" because of issues

[356] *See* Penfrat, *supra* note 118, at 13.

[357] *See* FARZANEH BADII ET AL., CIVIL SOCIETY POSITIONS ON CHRISTCHURCH CALL PLEDGE ¶ 2 (Electronic Frontier Foundation 2019), https://www.eff.org/files/2019/05/16/community_input_on_christchurch_call.pdf (noting that the letter includes signatures from 20 contributors to the document).

[358] *See id.* at 2–3.

[359] *See id.* at 3.

regarding the First Amendment.[360] However, the White House said it "stands with the international community in condemning terrorist and violent extremist content online" and supported the Call's goals.[361] In 2019, the Trump Administration had been focused on "political censorship" of speech by social media companies. As such, many conservatives saw the Call as a threat to free speech.[362] Behind the scenes, US Government officials stayed in touch with their New Zealand and French counterparts and did what they could to support the effort.[363] However, on the day the Call was signed, the White House announced the creation of a "tool" Americans could use to report if their speech was removed by social media companies due to "political bias."[364] Two years later, under the Biden Administration, the US Government joined the Call, noting that they would not take any action to "undermine the First Amendment."[365]

On May 15, 2019, just two months after the attacks, New Zealand and France formally announced the creation of the Christchurch Call to Action, a set of commitments by governments and online service providers to eliminate TVEC online while protecting the free, open, and secure internet. To support the organization, the Governments of New Zealand and France formed the Call Secretariat, which would be staffed by government officials. The original Call text of 25 commitments was supported by 17 countries, the EU,

---

[360] Adam Satariano, *Trump Administration Balks at Global Pact to Crack Down on Online Extremism*, N.Y. TIMES (May 15, 2019), https://www.nytimes.com/2019/05/15/technology/christchurch-call-trump.html [https://perma.cc/JBE8-F7E8].

[361] Tony Romm & Drew Harwell, *White House Declines to Back Christchurch Call to Stamp Out Online Extremism Amid Free Speech Concerns*, WASH. POST (May 15, 2019, 6:44 PM), https://www.washingtonpost.com/technology/2019/05/15/white-house-will-not-sign-christchurch-pact-stamp-out-online-extremism-amid-free-speech-concerns/ [https://perma.cc/9ZMN-493X].

[362] *See* Charlie Warzel, *The World Wants to Fight Online Hate. Why Doesn't President Trump?,* N.Y. TIMES (May 16, 2019), https://www.nytimes.com/2019/05/16/opinion/christchurch-online-extremism-trump.html [https://perma.cc/HP5H-BX67].

[363] As mentioned in the preface, I was one of these US government employees helping behind the scenes. This report does not contain any confidential information from my work with NTIA.

[364] *See* Warzel, *supra* note 362.

[365] *Statement by Press Secretary Jen Psaki on the Occasion of the United States Joining the Christchurch Call to Action to Eliminate Terrorist and Violent Extremist Content Online*, THE WHITE HOUSE (May 7, 2021), https://www.whitehouse.gov/briefing-room/statements-releases/2021/05/07/statement-by-press-secretary-jen-psaki-on-the-occasion-of-the-united-states-joining-the-christchurch-call-to-action-to-eliminate-terrorist-and-violent-extremist-content-online/ [https://perma.cc/8NQF-JDXT].

and eight tech companies.[366] Of these commitments, five apply to only the governments, seven apply only to the tech companies, and the other 13 apply to both.[367] The commitments include developing tools to prevent the upload of TVEC, countering the drivers of violent extremism through education, "increasing transparency around the removal and detection of content, and reviewing how companies' algorithms direct users to violent extremist content."[368] The commitments are careful to balance freedom of expression with the need for governments and companies to do more to counter extremism – both online and offline.

As a curated MSI, the text of the commitments is important for a few key reasons. First, the Call acknowledges that there are already several other forums addressing the issue of TVEC online, including multilateral efforts at the G7 and G20, and tech industry efforts such as the GIFCT and Tech Against Terrorism (TAT).[369] The drafters understood that this was not a new idea, but that it would be the first of its kind to bring a broader group of stakeholders together to address TVEC online – breaking down the silos of many of the other initiatives. Second, while the Call commitments include a provision to consider regulation, there is no commitment to impose new regulations on tech companies or law enforcement. This multistakeholder approach stood in contrast to the discussions happening in some multilateral forums at the time. Third, civil society is not formally committed to the Call; instead, several of the commitments within the Call require governments and online service providers to work with civil society to promote community-led efforts. As such, the supporters commit to recognizing the important role of civil society in offering advice and increasing transparency.

[366] *See Christchurch Call to Eliminate Terrorist and Violent Extremist Online Content Adopted*, CHRISTCHURCH CALL (May 16, 2019), https://www.christchurchcall.com/media-and-resources/news-and-updates/christchurch-call-adopted/ [https://perma.cc/4B9Q-W98U]. The original supporters of the Call were: France, New Zealand, Canada, Indonesia, Ireland, Jordan, Norway, Senegal, the UK, the European Commission, Amazon, Facebook, Dailymotion, Google, Microsoft, Qwant, Twitter, YouTube, Australia, Germany, India, Italy, Japan, the Netherlands, Spain, and Sweden.

[367] *See The Christchurch Call to Action*, *supra* note 7.

[368] Christchurch Call, *supra* note 366.

[369] *See The Christchurch Call to Action*, *supra* note 7, at 1.

3. Overview of the Work of the Christchurch Call to Action

After the launch in May 2019, work on the Call steadily increased over the next few months, culminating in September that year, when leaders reconvened at the United Nations General Assembly in New York. At this meeting, the leaders acknowledged the progress that had been made towards fulfilling the Call commitments and welcomed 31 new countries and two international organizations as partners.[370] Among the Call's accomplishments was the establishment of a Christchurch Call Advisory Network (CCAN), to advise on the implementation of the Call.[371] CCAN was initially a group of 40 organizations, including representatives from civil society, human rights defenders, technical experts, and free speech advocates. In 2019, this group was formally recognized to provide expertise to the Call's government and company supporters on how they can fulfill the commitments in the Call.[372]

Other accomplishments coming out of the Call's 2019 Leaders' Summit were in relation to the GIFCT. By 2019, the GIFCT database included over 200,000 pieces of content, but there was still a strong focus on ISIS propaganda and beheading videos.[373] Despite the growth of the hash-sharing database, and the inclusion of new social media companies, the GIFCT was not a standalone organization. Instead, the founding member companies (Microsoft, Facebook, YouTube, and Twitter) rotated leadership each year, meaning

[370] *See Significant Progress Made on Eliminating Terrorist Content Online*, CHRISTCHURCH CALL (Sept. 24, 2019), https://www.christchurchcall.com/media-and-resources/news-and-updates/new-news-article-page-8/ [https://perma.cc/UY7E-M5M7].

[371] *See id.*

[372] *See History*, CHRISTCHURCH CALL ADVISORY NETWORK, https://christchurchcall.network/about-us/history/ [https://perma.cc/C6SG-7UQP] (last visited Oct. 20, 2023).

[373] *See* GLOB. INTERNET F. TO COUNTER TERRORISM, TRANSPARENCY REPORT – JULY 2019 (2019), https://gifct.org/wp-content/uploads/2020/10/GIFCT-Transparency-Report-July-2019-Final.pdf ("At the end of 2018 the GIFCT gave itself the goal of reaching 200k hashes by the end of 2019. We are pleased to say that the Hash Sharing Consortium has reached over 200k unique pieces of terrorist content. Companies often have slightly different definitions on "terrorism" and "terrorist content". For the purposes of the hash sharing database, and to find an agreed upon common ground, founding companies in 2017 decided to define terrorist content based on content relating to organizations on the UN Terrorist Sanctions lists.").

processes were updated and staffed ad hoc by each company.[374] This had proved challenging on March 15, 2019, when the companies tried to quickly stop the spread of the Christchurch massacre video and manifesto. The GIFCT reported that it hashed more than 800 visually distinct versions of the video in the first 48 hours.[375] The attack highlighted the overall importance of this tool to the safety of billions of users around the world, and the Call's company supporters agreed to GIFCT reforms.

As part of their Call commitments, companies outlined five steps they would take as individual companies and four they would take as an industry.[376] The four industry commitments would largely be enacted through changes to the GIFCT.[377] At the Leaders' Summit in September 2019, the GIFCT announced the creation of a standalone organization with a dedicated structure and staff, as well as the creation of working groups focused on research, algorithms, and information sharing.[378] Another important announcement at the Summit was the creation of a multistakeholder Independent Advisory Committee (IAC), which would include representatives from governments, civil society, and academia to guide the GIFCT Operating Board on organizational priorities.[379] Finally, the GIFCT and governments worked together to establish a "Content Incident Protocol" to provide a more systematic way of addressing terrorist content in the wake of an attack.[380] These changes were remarkable achievements for a multistakeholder institution to accomplish in just four months.

[374] *See Governance*, GLOBAL INTERNET FORUM TO COUNTER TERRORISM, https://gifct.org/governance/ [https://perma.cc/L7G3-3GN7] (last visited Oct. 20, 2023).

[375] *See Story*, GLOBAL INTERNET FORUM TO COUNTER TERRORISM, https://gifct.org/about/story/#march-2019---cross-industry-collaboration [https://perma.cc/Q2H9-6K9H] (last visited Oct. 20, 2023) ("In response to the Christchurch mosque shootings in New Zealand, members of GIFCT utilized channels of communication that GIFCT had developed as well made use of the hash-sharing database to share more than 800 visually-distinct videos related to the attack.").

[376] *See* Smith & Browne, *supra* note 189, at 126.

[377] *See* GIFCT, *supra* note 373.

[378] *See* Christchurch Call, *supra* note 370.

[379] *See* GIFCT, *supra* note 374.

[380] *See* GIFCT, *supra* note 375.

Going into 2020, the Call made progress on several other commitments, but COVID-19 slowed its momentum, as governments and tech companies needed to prioritize their responses to the pandemic. Therefore, the Call Secretariat set out to conduct a stock-taking exercise in 2020 with input from the Call community and publish the results at the Leaders' Summit in 2021. Given the wide range of efforts happening worldwide to reduce TVEC online, this was also an attempt to understand the landscape and assess where the Call could add value. On April 14, 2021, the Call published its first Christchurch Call Community Consultation Report.[381] The Call Secretariat sent out a questionnaire to all the signatories of the Call, as well as civil society organizations affiliated with CCAN. In total, there were 99 parties contacted, and 39 participated in the study, including 24 countries, six companies, and nine civil society organizations.[382] The overarching goal of the survey was to establish a baseline of progress to inform the future direction of the Call.[383]

The stock-taking report found that the Call community had undertaken dozens of new initiatives in their home jurisdictions, and companies had created new policies to fulfill the commitments of the Call. When asked what the most important accomplishment the Call had achieved, 50 per cent of the respondents answered it was the creation of a multistakeholder approach to preventing the abuse of the internet by terrorist and violent extremists.[384] Another 26 per cent believed it was raising awareness of the issue of TVEC online.[385] The remaining 24 per cent believed it was reforming the GIFCT and creating Crisis Incident Protocols.[386] The response to where respondents wanted to go next were

[381] *See* CHRISTCHURCH CALL, CHRISTCHURCH CALL COMMUNITY CONSULTATION FINAL REPORT (2021), https://www.christchurchcall.com/assets/Documents/Chch-Call-Community-Consulation-Report-2021.pdf.

[382] *See id.* at 4 ("The consultation was open to submissions from 21 September to 30 October 2020. In total, members of the Call community submitted 39 responses.").

[383] *See id.*

[384] *See id.* at 66 ("50% of responses referenced, in some capacity, the unique multistakeholder approach embraced in the development and implementation of the Call. For the first time, governments, major tech companies, and civil society representatives have created an innovative, flexible coalition, working cooperatively to stop and prevent attacks like Christchurch being broadcast and spread online.").

[385] *See id.*

[386] *See id.*

mixed, but the majority supported increasing collaboration on a multistakeholder approach and recruitment efforts to increase the number of Call supporters.[387]

With the mandate to strengthen a multistakeholder model, the Call community reunited (virtually) in May 2021 for the two-year anniversary of the Call. The 2021 Leaders' Summit produced a concrete work plan for what the Call would accomplish in the next three years.[388] The 2021 priorities included developing an advisory function for CCAN, enhanced information sharing, increased tech company membership, sharing best practices, and strengthening the links between the Call and the GIFCT.[389] Additionally, the Call put together four work plans for community building, crisis response, tech and government transparency, and algorithms and positive interventions.[390] Each work plan detailed what the Call had accomplished since 2019 and what the working groups would do in the next six months, one year, and three years to fulfill the Call commitments.

First, the Community Work Plan outlined the work of the Call to foster multistakeholderism and give all stakeholders a seat at the table. To maintain this momentum, one of the most pressing tasks was to fund a Secretariat to assist CCAN

[387] *See id.* at 70.

[388] *See* CHRISTCHURCH CALL, COMMUNITY WORK PLAN 2021 (2021), https://www.christchurchcall.com/assets/Documents/Community-Work-Stream-Work-Plan.pdf.

[389] *See id.* at 1 ("As we move into the third year of the Call, it is incumbent on the Call community to ensure all members are equipped to participate fully in the work of the Call. The work plan developed by the Call Community work stream seeks to facilitate a flourishing community where every stakeholder has a seat at the table as envisaged by the text of the Call. The plan also seeks to foster trusted relationships between all stakeholders. Without this, no other work stream will reach its full potential. This plan identifies areas where more work is needed to achieve this ambition. It seeks to build trust across the multi-stakeholder community through improved information sharing and increased channels of communication, including through the use of technology. Recognizing the value of increased industry participation in the Call, it promotes involvement of the entire Call community in the on-boarding of new supporters, to ensure the continued integrity of the Call principles. The Call commitments are voluntary, therefore any mechanisms for understanding how supporters are carrying out the commitments in the Call must be grounded in trust-based dialogue between members. In order to achieve these objectives, resourcing will be required. It is our hope that the community will rise to this challenge, bringing their different capacities and capabilities to bear on the project.").

[390] *See Publications*, CHRISTCHURCH CALL, https://www.christchurchcall.com/media-and-resources/reports-and-publications/ [https://perma.cc/N25X-3MFF](last visited Nov. 1, 2023).

rather than expecting CCAN supporters to volunteer for administrative tasks.[391] This was crucial, as most of the civil society organizations involved are run on very tight budgets and were juggling dozens of similarly related initiatives on TVEC. In the short term, the New Zealand and French Governments provided the funding to hire the Secretariat. Next, the Call wanted to further develop the advisory function of CCAN by increasing its membership, creating a technological solution to enable intersessional dialogue between Call community members, and developing more accountability mechanisms.[392] Finally, the Community Work Plan envisioned a closer link between CCAN and the GIFCT, as the GIFCT is the Call's "primary partner for delivery against Call commitments through its multistakeholder working groups."[393]

Second, the Crisis Response Work Plan set out key objectives for improving processes for crisis response under the Call.[394] While coordination between governments and tech companies had improved remarkably since 2019, terrorists and violent extremists were still turning to social media to broadcast their attacks and promote radicalization efforts. Additionally, the Call was looking to civil society to help improve crisis response tools to reflect due process and human rights considerations.[395] Therefore, the Call set out to conduct a review of the Call's Crisis Response Protocol, along with a comprehensive mapping exercise of all content incident protocols to identify where there were overlaps or gaps.[396] Finally, as the Community Work Plan had done, the Crisis Response Work Plan

[391] *See* Christchurch Call, *supra* note 388, at 2.

[392] *See id.* (noting that medium-term objectives achievable within 6–12 months include "[d]evelop civil society advisory function of the Call, through addressing gaps in diversity and mapping and utilization of diverse expertise within the network.").

[393] *Id.* at 4.

[394] *See* CHRISTCHURCH CALL, CRISIS RESPONSE WORKPLAN 2021 (2021), https://www.christchurchcall.com/assets/Documents/Christchurch-Call-Crisis-Reponse-Workplan.pdf.

[395] *See id.* at 1.

[396] *See id.* ("Since its launch in May 2019, the Call has developed a dedicated Crisis Response Protocol (Christchurch Call Crisis Response Protocol). Other protocols at an international, domestic and organisational level have also been developed. Some of these protocols are geographically specific, whilst others are more global in nature and seek to coordinate a swift response.")

called for broadening the Call's membership and involving civil society and academia in the discussion.[397]

Third, the Transparency and Reporting Work Plan discussed how increased transparency could build trust among stakeholders, help prevent and reduce harm from TVEC online, and protect human rights and fundamental freedoms.[398] Many of the objectives of this work plan focused on the need to raise awareness of, and guide stakeholders to, the ongoing transparency reporting-related work happening at the GIFCT, TAT, and OECD.[399] However, one key initiative that the Call was undertaking that was not happening in other fora was related to how governments can be more transparent about when they ask companies to remove TVEC.[400] As one of the only MSIs in this space with governments, civil society and tech companies at the table, the Call was unique in asking government leaders to examine their practices and provide guidance on how they could improve processes in line with human rights principles.

Finally, the Algorithms and Positive Interventions Work Plan looked at ways to better understand user journeys and the role algorithms play in the radicalization process.[401] In 2021, there were several MSIs working on issues related to this topic. Two of these MSIs involved government stakeholders: the GIFCT's Content-Sharing Algorithms, Processes, and Positive Interventions Working Group[402] and the Global Partnership on Artificial Intelligence, created by Canada and France at the G7 Digital Ministerial Meeting in 2020

[397] *See id.* at 7.

[398] *See* CHRISTCHURCH CALL, TRANSPARENCY & REPORTING WORK PLAN 2021, at 1 (2021), https://www.christchurchcall.com/assets/Documents/Christchurch-Call-Transparency-Work-Plan.pdf.

[399] *See id.* at 2.

[400] *See id.* at 9.

[401] *See* CHRISTCHURCH CALL, ALGORITHMS & POSITIVE INTERVENTIONS WORKPLAN 2021, at 2 (2021), https://www.christchurchcall.com/assets/Documents/Algorithms-and-Positive-Interventions-WorkPlan.pdf.

[402] *See* GLOBAL INTERNET FORUM TO COUNTER TERRORISM, CONTENT-SHARING ALGORITHMS, PROCESSES, AND POSITIVE INTERVENTIONS WORKING GROUP — PART 2: POSITIVE INTERVENTIONS (2021), https://gifct.org/wp-content/uploads/2021/07/GIFCT-CAPI2-2021.pdf.

and hosted by the OECD.[403] Therefore, the Work Plan set out action items in this area to avoid duplicating efforts. As a result, this Work Plan included building understanding of recommender algorithms and user journeys, empowering community-driven online interventions, and mechanisms for TVEC removal including transparency and redress.[404] In line with the literature in 2021, the emphasis was on positive intervention measures to redirect a person away from extremist or terrorist content.

To carry out each of these work plans, community members met frequently throughout 2021 and 2022. Given that much of the work was happening at the working-group level, the Call Secretariat conducted a survey of community members to understand how stakeholders felt the work plans had progressed and published the results in August 2022. Additionally, the Secretariat hosted two community-wide meetings to discuss the work plans and evaluate resourcing.[405] The 2022 Community Update contains feedback from these meetings and survey results.[406] Overall, community members thought the greatest achievements of the Call since 2021 were: the creation of a new stakeholder on-boarding process for Call supporters, a review and update of the Crisis Response Protocol, better ties with the GIFCT, increased awareness of the Call's work, and improved communication through monthly calls with CCAN and additional stakeholders.[407]

Alongside the Community Update, the Call community supporters made several statements ahead of the 2022 Leaders' Summit detailing their progress in fulfilling the

[403] *See About GPAI*, THE GLOB. P'SHIP ON A.I., https://www.gpai.ai/about/ [https://perma.cc/FY7J-4H2Q] (last visited Nov. 2, 2023).

[404] *See* Christchurch Call, *supra* note 401, at 6.

[405] *See* CHRISTCHURCH CALL, CHRISTCHURCH CALL 2022 COMMUNITY UPDATE, at 2 (2022), https://www.christchurchcall.com/assets/Documents/Christchurch-Call-2022-Community-Update.pdf ("This report reflects an overview by the Secretariat of Call Community efforts and progress under each of the work plans. It draws on input from responses to our 2022 Community Survey, which asked Community members to share their thoughts on the allocation of effort and progress made under the work plans, and their assessment of risks, opportunities, and priority areas as this work continues. In addition to the Community Survey, the Community came together over the course of two meetings to collectively reflect on progress on the work plans, and opportunities and priorities for the future.").

[406] *See id.*

[407] *See id.* at 4–5.

Call commitments.[408] Five governments and the European Commission outlined actions they had taken to address TVEC online, including Australia's passage of the Online Safety Act 2021, the EU's regulation on "preventing the spread of extremist content online", Japan's efforts to improve the capacity of Association of Southeast Asian Nations countries to prevent TVEC online, and India's media standards framework.[409] Other organizations gave updates on their work, including the GIFCT, which stated it had responded to over 270 attacks since creating the Crisis Incident Protocol; the Global Partnership on Artificial Intelligence, which summarized its work on recommender algorithms; and Inclusive Aotearoa Collective Tāhono, which detailed its work in New Zealand to build more inclusive communities.[410] CCAN provided a response to the community statements document expressing the desire for civil society to play a more pronounced role in policy development and urging supporters to engage with them more frequently.[411]

In addition to their message in the community statements document, CCAN announced a separate initiative to evaluate the work of the Call.[412] The CCAN evaluation document, published in September 2022, expressed frustrations with the work of Call, including lack of transparency on commitments, lack of concrete evidence that human rights due diligence processes were in place, a failure on the part of government and company leaders

[408] *See* CHRISTCHURCH CALL, CHRISTCHURCH CALL COMMUNITY STATEMENTS SEPTEMBER 2022 (2022), https://www.christchurchcall.com/assets/Documents/Christchurch-Call-Community-Statements-2022.pdf.

[409] *See id.* at 1–7.

[410] *See id.* at 8–14.

[411] *See* CHRISTCHURCH CALL ADVISORY NETWORK, CHRISTCHURCH CALL ADVISORY NETWORK (CCAN) POSITION STATEMENT — CHRISTCHURCH CALL SUMMIT, 2022, at 2 (2022), https://christchurchcall.network/wp-content/uploads/Summit-Sept-22-CCAN-Statement.pdf ("Finally, we believe civil society should have a more pronounced role in policy development. Just as we advocate for online service providers to include civil society earlier in the design process, so too should governments in creating their policies. We urge the supporter companies and states to consult with CCAN to ensure that the Call values are incorporated and that the commitments enumerated in the Call to Action are undertaken in a manner that is consistent with the rule of law and international human rights law, and in a way that meets the needs of people and communities most impacted by TVEC.").

[412] *See* CHRISTCHURCH CALL ADVISORY NETWORK, EVALUATING THE IMPACT OF GOVERNMENT AND COMPANY COMMITMENTS UNDER THE CHRISTCHURCH CALL TO ACTION (2022), https://www.christchurchcall.com/assets/Documents/Christchurch-Call-2022-CCAN-Evaluation-Project.pdf.

to meaningfully engage civil society, and the creation of silos around the Crisis Response Protocols.[413] To remedy these problems, CCAN recommended regular reporting to CCAN from governments and companies of their actions, which could be done via publicly accessible repositories.[414] To start this work, CCAN decided to undertake an evaluation process of governments' and companies' efforts to fulfill the Call commitments.[415] This evaluation would survey six governments and four companies and cover overarching themes related to transparency, human rights due diligence, civil society engagement, and cross-Call collaboration.[416] On June 30, 2022, the survey was sent out to identified participants and CCAN members volunteered to do additional research to supplement responses.[417] As of the time of writing, this evaluation is still under way.

The CCAN evaluation followed a similar approach to another report released by CCAN in September 2022, which analyzed anti-dehumanization policies.[418] In March 2022, CCAN distributed a request for information to government and company supporters of the Call to map current approaches combating dehumanizing speech.[419] This research was important to the work of the Call, as dehumanization is a common feature of terrorist and violent extremist propaganda. Dehumanizing speech is separate from hate speech; it aims to lower an audience's moral reflexes towards a particular group, which can lead to offline

[413] *See id.* at 1 ("In contrast, it was much harder to find evidence that supporters had implemented their commitments under the Call beyond declarations of intent to do so. If work was undertaken in response to the Call, it was rarely identified as such, making measurement of the Call's impact difficult. This raises questions about the consistency of the Call's impact across its many government and company supporters.").

[414] *See id.* at 3.

[415] *See id.*

[416] *See id.* ("We also selected a small sample of the supporting governments and companies to include in this first evaluation. We chose six governments—New Zealand, France, Australia, Canada, United Kingdom, and India—and four companies—Microsoft, Meta, Twitter and Google. We chose these signatories based on their role as leaders of the Call (in the case of New Zealand and France), the longevity of their support for the Call, and our internal capacity to conduct this analysis, such as familiarity with language, legal systems, and access to resources.").

[417] *See id.*

[418] *See* CHRISTCHURCH CALL ADVISORY NETWORK, CCAN REPORT ON ANTI-DEHUMANIZATION POLICY — CHRISTCHURCH CALL SUMMIT, 2022 (2022), https://christchurchcall.network/wp-content/uploads/CCAN-Report-on-Anti-Dehumanization-Policy.pdf.

[419] *See id.* at 1.

violence, as seen in the Christchurch shooter's manifesto.[420] The evaluation found that, of the companies and governments surveyed, only Twitter had specific policies regarding dehumanizing speech. However, companies and governments alike had rules and laws that could cover dehumanizing speech if applied correctly.[421] Therefore, the report suggested that the Call members could work together on strategies to counter the production and dissemination of dehumanizing speech, including through frameworks related to hate speech, disinformation, harmful digital communications, and tort law.[422] The Call community welcomed this thoughtful feedback. The report is an outstanding example of multistakeholderism advancing policy changes.

Before turning to the discussions at the 2022 Leaders' Summit, it is important to acknowledge one event that brought renewed attention and urgency to the work of the Call – the mass shooting in Buffalo, New York on May 14, 2022. In Buffalo, a white 18-year-old male killed 10 people in a supermarket in a predominantly black neighborhood.[423] The shooter wore a GoPro camera and attempted to livestream his attack on Twitch, a gaming platform, but the company disabled the livestream within two minutes.[424] An investigative report into the incident by the Office of the New York State Attorney General details the Buffalo shooter's radicalization online and his use of social media platforms, including Reddit, Discord, 4chan, 8kun, and others to connect with violent extremists.[425] Notably, in the Buffalo shooter's manifesto, he stated that the Christchurch attack was a "catalyst" and inspired him towards ethno-nationalist beliefs.[426]

[420] *See id.* at 2 ("Dehumanization is a distinct concept from hate speech and Terrorist and Violent Extremist content (TVEC), although it often features in both. Dehumanizing language or speech (e.g., referring to a race of people as a disease) is a type of hate speech, broadly defined, and can create a heightened environment for violence.").

[421] *See id.* at 4 ("Except for Twitter, there were no existing laws, rules or policies distinctly on dehumanizing speech or language. However, there were laws, rules or policies that conceivably could cover dehumanizing speech or language.").

[422] *See id.* at 10.

[423] *See* Office of the New York State Attorney General Letitia James, *supra* note 22.

[424] *See id.* at 9 ("The shooter began livestreaming using the online platform Twitch at approximately 2:08 p.m., using a GoPro video camera attached to his helmet … Twitch stopped the livestream approximately two minutes after the first person was shot.").

[425] *See id.* at 6–9.

[426] *See id.* at 19.

Unfortunately, the Buffalo attack has not been the only Christchurch-inspired attack; there have been others in Poway, El Paso, Dayton, Halle, Glendale, Nakhon Ratchasima, Nice, and Vienna.[427] While the companies had improved their capabilities to stop the spread of the video and manifesto, it was clear more work needed to be done.

The Buffalo attack was top-of-mind at the 2022 Leaders' Summit on the sidelines of the UN General Assembly in New York in September. The meeting was an opportunity to welcome new industry supporters and partner organizations, including Roblox, Zoom, Mega, Clubhouse, the Global Community Engagement and Resilience Fund, and TAT.[428] Additionally, the meeting sought to provide a strategic direction for the upcoming year, prioritizing three areas: improving incident response, understanding how algorithms and social drivers can lead to radicalization, and future-proofing the Call.[429] With regard to the latter, leaders added two new workstreams – one exploring new technologies and the other exploring the drivers of violent extremism, including gender-based hate.[430] The Summit's joint statement also mentioned how the Call's multistakeholder model could help similar MSIs combatting disinformation, harassment, hatred online, and issues affecting youth, including Tech for Democracy, the Summit for Democracy, the Global Partnership for Action on Gender Based Online Harassment and Abuse, and the Global Partnership on Artificial Intelligence.[431]

---

[427] *See* CHRISTCHURCH CALL, SECOND ANNIVERSARY OF THE CHRISTCHURCH CALL SUMMIT — JOINT STATEMENT BY PRIME MINISTER RT HON JACINDA ARDERN AND HIS EXCELLENCY PRESIDENT EMMANUEL MACRON AS CO-FOUNDERS OF THE CHRISTCHURCH CALL, at 2 (2021), https://www.christchurchcall.com/assets/Documents/Christchurch-Call-2nd-Anniversary-Summit-Co-chair-Statement-2021.pdf ("Despite our achievements so far, the many attacks since Christchurch - in Colombo; El Paso; Dayton; Halle; Glendale; Nakhon Ratchasima; Conflans-Sainte-Honorine; Nice; and Vienna among others – bear witness to the challenge we still face.").

[428] *See* CHRISTCHURCH CALL, CO-CHAIR STATEMENT CHRISTCHURCH CALL LEADERS' SUMMIT — NEW YORK, 20 SEPTEMBER 2022, at 2 (2022), https://www.christchurchcall.com/assets/Documents/Christchurch-Call-Joint-Statement-2022-English-version.pdf.

[429] Christchurch Call to Action, *Our Work: Leaders' Summits*, CHRISTCHURCH CALL TO ACTION (2023), https://www.christchurchcall.com/about/leaders-summits/ [https://perma.cc/P28A-QRTT].

[430] Ardern & Macron, *supra* note 428.

[431] *Id.*

Another important announcement in September 2022 was the launch of the Christchurch Call Initiative on Algorithmic Outcomes (CCIAO). The CCIAO is a project funded by Microsoft, Twitter, and the Governments of New Zealand and the US to develop new technologies to understand the impacts of algorithms on people's online experiences.[432] Working with OpenMined, an open-source non-profit organization, the CCIAO is developing tools to provide access to researchers to study how individuals are radicalized across platforms. In the past, it has been difficult to carry out extensive research on TVEC because quality research requires access to sensitive information across platforms. The CCIAO is developing software through a privacy-enhancing technology that will enable data scientists to study algorithms across multiple online platforms. This technology provides cross-platform analysis which will give researchers a better understanding of how people are radicalized online and how to more effectively intervene to protect people, both online and offline.[433] If proven successful in the Call context, this could open up a new field of algorithmic research for a much wider application.[434] Work on the CCIAO is under way; researchers are beginning to access data from Twitter, DailyMotion, and LinkedIn through the privacy-enhancing technology to explore the ways in which AI and humans interact online.[435]

[432] Christchurch Call to Action, *Christchurch Call Initiative on Algorithmic Outcomes*, CHRISTCHURCH CALL TO ACTION (Sep. 2022), https://www.christchurchcall.com/media-and-resources/news-and-updates/christchurch-call-initiative-on-algorithmic-outcomes/ [https://perma.cc/QW3Q-K6L4].

[433] *Id.* ("That system will help us to answer questions such as: "What are the distinct features of a user journey for someone that engages with TVEC?" "What is the before/after impact of positive interventions, or changes to ranking systems or other platform features designed to reduce toxicity or risk of harm?" "What do user journeys for 'at risk' user types look like between and across platforms?" "How effective and fair are the automated systems that identify and remove TVEC?").

[434] Ardern, *supra* note 6 ("We're also taking on some of the more intransigent problems. The Christchurch Call Initiative on Algorithmic Outcomes, a partnership with companies and researchers, was intended to provide better access to the kind of data needed to design online safety measures to prevent radicalization to violence. In practice, it has much wider ramifications, enabling us to reveal more about the ways in which AI and humans interact.").

[435] *Id.*

On January 19, 2023, Prime Minister Ardern announced she was resigning from office and would not seek re-election.[436] However, Ardern was committed to staying involved with the work of the Call. On April 4, 2023, Prime Minister Chris Hipkins announced he was appointing Ardern as special envoy for the Christchurch Call.[437] During a virtual gathering on the fourth anniversary of the creation of the Call, Special Envoy Ardern stated the Secretariat's intention to host a 2023 Leaders' Summit in September.[438] Ardern and other speakers during the virtual meeting called for more attention in several areas, including understanding the impact of algorithmic systems on radicalization, confronting the reality of gender-based hatred and abuse as a factor in radicalization and violence, and considering emergent technologies including generative AI.[439]

B. Evaluation of the Christchurch Call to Action

Evaluating the Call is not a mere check-box exercise, as the initiative is a bottom-up, large-scale collaboration between various stakeholders who all have their own motivations and reasons for implementing the commitments of the Call.[440] Ardern summarized the accomplishments of the Call in a June 2023 op-ed in the *Washington Post*, saying:

> "… we have developed new policies and ways of working that holistically address the complexities of terrorist and violent extremist content. We have established new crisis protocols to respond effectively and in a coordinated manner to attacks with an online component. We worked as a community to establish the Global Internet Forum to Counter Terrorism as an independent NGO. This created the

[436] Beehive Press Release, *Prime Minister Jacinda Ardern Announces Resignation*, NEW ZEALAND GOVERNMENT (Jan. 19, 2023), https://www.beehive.govt.nz/release/prime-minister-jacinda-ardern-announces-resignation [https://perma.cc/KY5J-H7HX].

[437] Christchurch Call to Action, *New Zealand Special Envoy for the Christchurch Call Announced*, CHRISTCHURCH CALL TO ACTION (Apr. 4, 2023), https://www.christchurchcall.com/media-and-resources/news-and-updates/new-zealand-special-envoy-for-the-christchurch-call-announced/ [https://perma.cc/DRK5-2YK5].

[438] Christchurch Call to Action, *Four Years of the Christchurch Call*, CHRISTCHURCH CALL TO ACTION (May 15, 2023), https://www.christchurchcall.com/media-and-resources/news-and-updates/four-years-of-the-christchurch-call/ [https://perma.cc/ZQ7B-T449].

[439] *Id.*

[440] Ardern, *supra* note 6.

> opportunity for the GIFCT to become a more fully multistakeholder construct, develop integrated solutions, and share information and expertise, should it choose to. I know we still have work to do to fulfill this vision. We now better understand the online ecosystem and the experiences of affected communities, having led collaborative research across our community. And we have built a strong and diverse multistakeholder community."[441]

Rather than go through each of the original 25 commitments in the Call, this section examines overarching themes of the Call's work in two areas: building a multistakeholder community to address the drivers of TVEC and taking steps to eliminate TVEC online while protecting a free, open, and secure internet. This section will discuss the work of both the Call and some of the other MSIs that collaborate with the Call community.

1. Building a Multistakeholder Community

Among the goals of the Call is to counter the drivers of TVEC through a whole-of-society approach to addressing the problem, via a multistakeholder framework. Call supporters agreed to work collectively on 12 commitments which fall into three broader buckets of work: raising awareness to widen support for the Call, working with civil society to address the drivers of TVEC, and accelerating research.[501] This section evaluates how the Call has accomplished these three overarching goals.

(a) Raising Awareness

The events of March 15, 2019, were a harsh wake-up call to governments and tech companies alike. In their aftermath, the New Zealand Government received an outpouring of support and had the authority to lead an MSI to tackle the issue.[442] Since 2019, the Call has done a remarkable job of keeping the Christchurch shooting front and center in global content moderation discussions. As part of this effort, the Call community has

[441] *Id.*
[442] Smith and Browne, *supra* note 189.

partnered with dozens of other MSIs to confront the challenges of TVEC online, including the IGF, GNI, I&J, the Summit for Democracy, and the EU Internet Forum (to name a few MSIs we have already examined in this article).[443] Additionally, the New Zealand Government has partnered with governments and civil society in their efforts to eliminate TVEC online, including the Jakarta Centre for Law Enforcement Cooperation, the Pacific Working Group on Counter Terrorism and Transnational Organized Crime, the Global Community Engagement and Resilience Fund, the UN Office on Drugs and Crime, and the Aqaba Process (to name a few global non-MSI forums). The Call's supporters, often working alongside CCAN, have also attended a wide range of conferences to build awareness for the Call, including RightsCon, the Paris Peace Forum, and the Trust and Safety Professional Association's "TrustCon." This effort has brought together new stakeholders who may not have been impacted by the events of March 15, 2019, but are now coming together to share ideas on how to combat TVEC online.

One way the Call has ensured attention on its work has been through annual Leaders' Summits, where supporters meet to confirm priorities and identify areas of focus.[444] Ahead of these summits, the Call Secretariat convenes working groups to undertake multistakeholder efforts throughout the year and encourage stakeholders to act independently in their commitments.[445] These summits bring together stakeholder "leaders" – meaning heads of governments, CEOs, and top leaders from civil society or academia. One goal for having these conversations at the "leader-level" is to ensure the issue remains a top priority. This framing ensures the top officials are aware of the ongoing work, but it can present challenges to the overall inclusiveness of the event. Many heads of state and CEOs have incredibly busy schedules, which can conflict with the timing of the meeting, resulting in key supporters being left out of the discussion. In the

[443] The Christchurch Call to Action: Full English Text*, supra* note 7 ("Tech for Democracy, the Summit for Democracy, the Freedom Online Coalition, the Declaration for the Future of the Internet, the Aqaba Process, the Global Partnership for Action on Gender Based Online Harassment and Abuse, the Global Partnership on Artificial Intelligence, and the International Call to Stand up for Children's Rights Online, and where there is multistakeholder interest in new work programs separate to the Call.").

[444] Ardern, *supra* note 6.

[445] *Id.*

long run, this can negatively impact implementation efforts as these leaders do not typically do the day-to-day work of implementing commitments. Therefore, this framing risks disenfranchising supporters who may feel less bought-in on the process. Making these summits more inclusive and accessible could be one way to improve raising awareness around the work of the Call.

(b) Working with Civil Society

The Call's supporters – governments and tech companies – work with civil society primarily through CCAN. CCAN represents a diverse group of civil society actors, including victims of the Christchurch attack, human rights organizations, technical experts, and free speech advocates.[446] CCAN has worked closely alongside Call supporters over the years to provide expert advice in a manner consistent with a free, open, and secure internet and international human rights principles. In many ways, CCAN is a separate curated MSI that sits alongside the Call itself. It has its own website, terms of reference, and leadership structure.[447] Additionally, CCAN has its own recruitment and approval process, which has changed over the years and was most recently updated in a 2022 terms of reference.[448] Despite the growing number of civil society organizations working on content moderation problems, CCAN has not grown much in four years: from 40 members to 46.[449] One reason for this may be the terms of reference, which limits the amount of funding an organization can receive from governments or companies without showing adequate independence.[450] Unfortunately, this has the unintended consequence of

[446] Christchurch Call Advisory Network, *About Us – History, supra,* note 428.

[447] *Id.; see also* Christchurch Call Advisory Network, *Terms of Reference*, CHRISTCHURCH CALL ADVISORY NETWORK (Sep. 2022), https://christchurchcall.network/governance/ [https://perma.cc/G3GG-LVKH].

[448] *Id.*

[449] Christchurch Call Advisory Network, *About Us – Members*, CHRISTCHURCH CALL ADVISORY NETWORK (Sep. 2022), https://christchurchcall.network/about-us/members/ [https://perma.cc/Z2AV-GNTH].

[450] *See* Christchurch Call Advisory Network, *Terms of Reference*, *supra*, note 503 ("Members must be independent of governments and companies. To qualify for membership, they should, if applicable: 1. Establish that they have organizational and accountability structures in place, such as being registered as a non-governmental organization in their country or providing visibility of their operations through a published statement of purpose and meeting minutes; 2. Include in their application a statement that their work is not directed or strongly influenced by a government or private sector company. 3. To the

limiting many of the most relevant non-profit organizations researching and developing solutions for addressing TVEC online. Therefore, one way the Call has worked to ensure broader inclusion of new stakeholders has been to create a "partners" category, which includes organizations such as TAT, UNESCO, and the Council of Europe.[451] This solution helps ensure a wider variety of perspectives.

Tension between civil society, governments, and industry on policy direction is common within any MSI, because those groups tend to view their roles very differently. In many cases, civil society often see themselves as individual advocates instead of implementation partners. Indeed, over the years, CCAN has requested a "more pronounced role in policy development" from the government and company supporters of the Call.[452] Unfortunately, these tensions have been amplified by the fact that CCAN members are not formal supporters of the Call commitments, but serve an advisory role.[453] As the Call looks towards new projects and initiatives, finding ways to more directly incorporate CCAN into the structure of the Call could help address some of these tensions and increase interest among potential new supporters. One way the Call has addressed these tensions is by directly incorporating CCAN members into the Call's working groups. Additionally, the Call supporters continue to build trust between stakeholders through summits, on-going conversations, and internal transparency processes. All this work will hopefully contribute to the appeal of joining the Call and increase the diversity of supporters.

extent applicants receive significant funding (more than 25% of their operating budget) from governments or corporations, the application should disclose the total percentage of their operating revenue that comes from these sources, the specific governments or companies that provide funding, and what measures they take to ensure/maintain independence from those funders. 4. Organizations need not disclose how much funding they receive from any particular source, and financial information will not be shared with anyone not directly involved in determining membership eligibility for the applicant; such disclosure could pose serious legal, reputational, or security risks to the applicant or its partners.").

[451] Christchurch Call to Action, *Our Community – Partners*, CHRISTCHURCH CALL TO ACTION (Apr. 4, 2023). https://www.christchurchcall.com/our-community/partners/ [https://perma.cc/4E7F-R3KX].

[452] Christchurch Call Advisory Network, *Christchurch Call Advisory Network (CCAN) position statement, Christchurch Call Summit, 2022, supra* note 467.

[453] *Id.*

(c) Accelerating Research

Over the years, Call supporters have invested heavily in research initiatives addressing the problems of TVEC online. Companies, through their contributions to the GIFCT, support the Global Network on Extremism and Technology, which is the GIFCT's academic research arm, exploring the nexus between online behaviors and offline harms.[454] The GIFCT also commissions research about the evolving tactics, capabilities, and identities of violent extremist groups and shares them with the Call community more broadly. Additionally, the GIFCT, and its member companies, work with social scientists and extremism experts in various regions to help them develop the skills to identify and counter extremism. TAT receives both government and company funding to support third-party researchers on projects.[455] Individual companies have their own initiatives such as Google's Jigsaw project, which explores threats to online discourse, and Meta's Oversight Board, which is currently exploring how the company moderates content related to dangerous individuals and organizations.[456] Governments have also supported research; two examples include Canada's Centre for Community Engagement and Prevention of Violence, which seeks to counter radicalization to violence,[457] and New Zealand's He Whenua Taurikura, the National Centre of Research Excellence for Preventing and Countering Violent Extremism.[458]

---

[454] Global Internet Forum to Counter Terrorism, *Research*, GLOBAL INTERNET FORUM TO COUNTER TERRORISM (2023), https://gifct.org/research/ [https://perma.cc/RNB7-E54Y].

[455] Tech Against Terrorism, *Research and Publications*, TECH AGAINST TERRORISM (2023), https://www.techagainstterrorism.org/home [https://perma.cc/RUZ3-9USS].

[456] The Oversight Board, *Oversight Board announces a review of Meta's approach to the term "shaheed"*, THE OVERSIGHT BOARD (Mar. 2023), https://www.oversightboard.com/news/1299903163922108-oversight-board-announces-a-review-of-meta-s-approach-to-the-term-shaheed/ [https://perma.cc/BE8Y-PMG4].

[457] Public Safety Canada, *Canada Centre for Community Engagement and Prevention of Violence*, GOVERNMENT OF CANADA (Dec. 5, 2022), https://www.publicsafety.gc.ca/cnt/bt/cc/index-en.aspx [https://perma.cc/3XKE].

[458] Department of the Prime Minister and Cabinet, *He Whenua Taurikura*, DEPARTMENT OF THE PRIME MINISTER AND CABINET (Sep. 22, 2021), https://www.dpmc.govt.nz/our-programmes/national-security/counter-terrorism/he-whenua-taurikura [https://perma.cc/W5FY-2JHF].

In addition to the individual research-supporting efforts, the Call launched its own research initiative, the CCIAO, mentioned above. The CCIAO is funded by Microsoft, Twitter and the governments of New Zealand and US to create new technology to understand the impacts of algorithms on people's online experiences. This cross-industry and government project was necessary because studying the impact of algorithmic outcomes, and the way they impact a user's journey to radicalization, is incredibly difficult to do in a way that allows researchers access to highly sensitive datasets using privacy-respecting technologies. While many online platforms claim they have made progress in improving algorithmic recommendation systems, without independent study it is impossible to measure the impacts of these changes. The CCIAO seeks to address these challenges by providing researchers access to anonymized datasets to test how people are radicalized online. The technology, if proven successful in the Call context, could open up a new field of algorithmic research with a much wider application. Work on the CCIAO is now underway, with researchers beginning to access the platform to explore the ways in which AI and humans interact online. This research project is set to be a cornerstone of the Call's future work on AI and automation.

2. Eliminating TVEC Online

It would be impossible to calculate the percentage of content online that qualifies as TVEC, and whether that number has increased or decreased since 2019. Even without this data, we know that the world remains a long way from "eliminating" TVEC online. However, the Call has been an important catalyst for efforts to achieve this goal, by coming up with a plan and getting stakeholders to agree to it. In 2019, the mere fact that companies and governments could agree to work together to solve a broader societal challenge was novel. The Call deserves credit for bringing together stakeholders to work collectively to address the issue and publicly commit to a plan.[459] This planning itself represents progress, as it generated proactive thinking on solutions and highlighted key

[459] Douek, *supra* note 17 at 595 ("First, requiring planning forces platforms to think proactively and methodically about potential operational risks. The process of having to articulate a plan itself engenders proactivity and highlights blind spots. Platforms are known for failure to anticipate key risks, so "making [platforms] think" is meaningful, and a useful counterweight to the "Move Fast and Break Things" culture of Silicon Valley.").

blind spots in both company and governmental actions to eliminate TVEC online.[460] Another positive outcome from the Call's creation was that the commitments put public pressure on companies to invest in policy and technical solutions.[461] Finally, over the past four years, the Call has served as a rallying point for greater cross-industry reporting, which has helped improve compliance standards and create best practices.[462] Through convening stakeholders and publicly committing to a plan of action, the Call has helped companies find new ways to eliminate TVEC online – both individually and as an industry.

(a) Individual Company Solutions

As part of their commitments to the Call, companies outlined steps they would take to address TVEC on their own platforms, undertaking to "tighten their terms of service, better manage live videos, respond to user reports of abuse, improve technology controls, and public transparency reports."[463] There is no doubt many companies have implemented changes in all five of these areas. However, as Evelyn Douek, a scholar who fastidiously tracks changes to social media companies' policies, notes, it can be difficult to know exactly what changes companies implemented specifically as a commitment to the Call and what changes they made because it happened to align with other company priorities.[464] Unfortunately, this is part of a broader accountability problem for internet

[460] *Id.*

[461] *Id.* at 597 ("transparent plans facilitate broader policy learning for regulators and across industry. Comparative information would show industry best (or worst) practices"); *citing* Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1533–34 (2019).

[462] *Id.* at 595. Public planning efforts create some accountability on the companies and governments to improve their practices related to content moderation.

[463] Smith and Browne, *supra* note 193 at 154.

[464] Douek, *supra* note 17, at 75 ("Content moderation plans so far have largely been of this nature—often the announcement of a plan has been the end of a platform's external engagement with an issue, rather than the beginning. For example, the public has been left almost entirely in the dark about the effectiveness of platforms' exceptional COVID-19 misinformation rules released to great fanfare. Two years after the adoption of the "Christchurch Call to Eliminate Terrorist and Violent Extremist Content Online," there has been little public accounting of how companies have implemented their voluntary

companies who rarely publicly explain how they enforce their own rules and the systems they have in place.[465] However, several companies have stated that they changed their policies to fulfill Call commitments. For example, a representative from Twitter testified to Congress that the Christchurch Call made the company realize they needed a real-time communications strategy in a crisis.[466] Additionally, Meta testified in that same hearing that the company introduced reforms in line with their Call commitments to limit access to certain features – notably live streaming – for users that had violated its Dangerous Organizations policy. [467] CCAN is currently undertaking an evaluation process to track which companies and governments have implemented new policies in accordance with their Call commitments.

Another industry shift since the creation of the Call has been more transparency in the way platforms defines TVEC in their terms of service and disclose moderation of those

pledges. Therefore, any regulatory scheme must include an obligation for platforms to provide an annual public review of the implementation of their plans to create some measure of accountability for platforms' progress towards their goals."); *see also* Christchurch Call Advisory Network, *CCAN Report on Anti-Dehumanization Policy, supra* note 474.

[465] *Id.* at 71("Requiring platforms to publish and explain plans for how they will enforce their own rules may sound like a feeble form of accountability. But it's hard to overstate both how ineffective platforms are at enforcing their rules, and how little is known about what systems they have in place to do so. Despite being a purely procedural (not outcome-based) form of accountability, there are four main benefits of requiring platforms to have publicly available plans for rule-enforcement and that distinguish this form of systems-based transparency from the transparency theatre of aggregated information about individual cases.").

[466] Mass Violence, Extremism and Digital Responsibility: Hearing before the Senate Comm. on Commerce, Science, and Transp., 116th Cong. (Sep. 8, 2019), https://www.commerce.senate.gov/2019/9/mass-violence-extremism-and-digital-responsibility [https://perma.cc/SL4M-KPYE] (Nick Pickles of Twitter told the Committee, "We've grown that partnership, so we share URLs. So, if we see a link to a piece of content like a manifesto, we're able to share that across industry. And furthermore, I think an area that after Christchurch we recognized we need to improve, we now have real time communications in a crisis, so industry can talk to each other in real time operationally to say even, you know, not content related, but situational awareness.").

[467] *Id.* at 3 (Monika Bickert of Meta testified to the Committee, "For example, in response to the tragic events in Christchurch, we made changes to Facebook Live to restrict users if they have violated certain rules—including our Dangerous Organizations and Individuals policy. We now apply a "one-strike" policy to Live: anyone who violates our most serious policies will be restricted from using Live for set periods of time—for example, 30 days—starting on their first offense.").

rules through transparency reporting.[468] In 2019, a few online platforms only vaguely defined TVEC in their terms of services – and many did not even do that.[469] After the Christchurch shooting, not only did companies more clearly define TVEC, they also started to report on their TVEC content moderation practices in their transparency reports. In 2022, the 15 largest online platforms that released transparency reports included TVEC information, up from only five companies in 2019.[470] It is hard to argue that the Call is solely responsible for this industry effort, as calls for increased transparency around content moderation practices have recently come from every corner of government and civil society. In fact, organizations like the GIFCT, OECD and TAT have created programs to make transparency reporting easier and standardized.[471] However, these commitments remain a priority for the Call, because the quality of transparency reporting still needs improvement. Current transparency reporting efforts are only marginally helpful, as they provide a lot of data without revealing much information at all.[472] Additionally, critics argue that aggregate content moderation enforcement numbers do not always give the full picture of trends, because the raw numbers of removals could be affected by factors that do not always reveal underlying

[468] Ardern, *supra* note 6.

[469] OECD Publishing, *Current Approaches to Terrorist and Violent Extremist Content Among the Global Top 50 Online Content-Sharing Services*, OECD DIGITAL ECONOMY PAPERS No. 296, 11 (Aug. 14, 2020), https://www.oecd-ilibrary.org/docserver/68058b95-en.pdf [https://perma.cc/G7EJ-CBX6] ("The practice of reporting information on how companies moderate and remove content based on their own ToS and policies generally, and based on their anti-terrorism and anti-violence policies in particular, is hardly widespread. Of the 23 Services profiled in this Report that issue any transparency reports at all,18 only five (Facebook, YouTube, Instagram, Twitter and Automattic) issue reports specifically about TVEC.").

[470] *Id.* at 16.

[471] Tech Against Terrorism, *Transparency Reporting Guidelines*, TECH AGAINST TERRORISM (2023), https://transparency.techagainstterrorism.org/ [https://perma.cc/5SFX-GXXV].

[472] Douek, *supra* note 17, at 48 ("Platforms can drown observers in data while revealing little."); *citing* Sun-ha Hong, *Why Transparency Won't Save Us*, CIGI (Feb. 18, 2021), https://www.cigionline.org/articles/why-transparency-wont-save-us [https://perma.cc/MRX5-FR9Y]; Mike Ananny & Kate Crawford, *Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability*, 20 NEW MEDIA & SOC. 973, 979 (2018); *see also* Nicolas P. Suzor et al., *What Do We Mean When We Talk About Transparency? Toward Meaningful Transparency in Commercial Content Moderation*, 13 INT'L J. COMM. 1526, 1528–29 (2019).

content moderation practices.[473] Even taking into account improvements in reporting and increased attention from multiple MSIs, more work is needed for transparency reporting to meaningfully contribute to our understanding of the root causes of TVEC online.[474]

(b) Industry-Wide Solutions

Two areas where the companies committed to industry-wide solutions involved mitigating the dissemination of TVEC online and working together on a crisis response protocol.[475] When GIFCT was restructured 2019, and began running as a distinct entity in 2020, these were two of its top priorities. To mitigate the dissemination of TVEC online, the companies further invested in the GIFCT to manage and develop the hash-sharing database. Additionally, the GIFCT started to work more closely with smaller platforms through TAT, which hosts a mentoring program to help develop capability across the sector. Like the more extensive hash-sharing database, TAT runs the Terrorist Content Analytics Platform (TCAP) which automates the detection and removal of verified terrorist content on tech platforms.[476] The TCAP primarily focuses on small tech platforms, many of which may not have the capacity to moderate TVEC or lack access to automated processes.[477] Because they do not require a financial commitment to join, TAT enables many smaller companies to learn more about terrorist misuse of internet platforms and ways to mitigate this risk on their services.

Other industry-wide commitments within the Call relate to the development of a crisis response protocol. This is an area where there are now multiple different protocols and

[473] *Id.* ("But aggregate enforcement numbers, without more, do not explain relevant denominators or the cause of various trends. For example, when a platform reports an increase in takedowns, it might be intuitive to assume this is because that platform is doing a better job of finding violating content and removing it. But there could be many other reasons: there could be more content overall on the platform; there could be an increase in that kind of content; the platform might have lowered its confidence threshold for removing violating content; the platform might have broadened its definition of violating content; and so on.").

[474] Daphne Keller, *Who Do You Sue? State and Platform Hybrid Power Over Online Speech*, HOOVER AEGIS SERIES PAPER NO. 1902 13 (2019); *see also* Douek, *supra* note 17 at 47.

[475] Smith and Browne, *supra* note 193 at 154.

[476] Tech Against Terrorism, *Terrorist Content Analytics Platform*, TECH AGAINST TERRORISM (2023), https://www.terrorismanalytics.org/ [https://perma.cc/5SFX-GXXV].

[477] *Id.*

methodologies (often based on who 'owns' or 'manages' the protocol, and for what purpose) and work is needed to ensure coordination and compatibility between them. The Call and the GIFCT – operating in different ways – each contribute to crisis response protocols to stop the rapid dissemination of TVEC and quickly remove footage from many platforms.[478] The GIFCT has developed its Content Incident Protocol, which contributes to an Incident Response Framework.[479] At the beginning of 2023, the GIFCT's crisis response systems and incident management channels had been activated 306 times to monitor and assess incidents in 44 countries. The Content Incident Protocol, which deals with crises that meet strict criteria, had been activated four times, including incidents in Halle, Germany, Glendale, Arizona, Buffalo, New York, and Memphis, Tennessee.[480] Many Call supporters have their own national or regional protocols such as the Europol Protocol. The Call has its own Christchurch Call Crisis Response Protocol, which draws on developments in the wider crisis response landscape. However, since 2022, the Call has been working to map out overlapping systems as many stakeholders have different needs when handling TVEC online.[481] For example, how should crisis response protocols respond to bystander footage or if different protocols are needed based on regional needs.[482]

C. Future of the Call and Generative Artificial Intelligence

Alongside policymakers worldwide, the Call is turning its attention to the potential benefits and challenges posed by the development of GenAI. GenAI has recently become mainstream as millions of people around the world experiment with products like ChatGPT and Google's Bard. While the technological developments of GenAI are relatively new, the Call's focus on AI is not. One of the Call's initial commitments was to "review the operation of algorithms and other processes that may drive users towards

[478] Ardern, *supra* note 6.
[479] Global Internet Forum to Counter Terrorism, *Content Incident Protocol*, GLOBAL INTERNET FORUM TO COUNTER TERRORISM (2023), https://gifct.org/content-incident-protocol/ [https://perma.cc/4ZS7-X7JZ].
[480] *Id.*
[481] *See* Christchurch Call to Action, *Christchurch Call 2022 Community Update, supra* note 405.
[482] *Id.*

and/or amplify TVEC."[483] This includes designing a multistakeholder process for examining the use of algorithms and automation to remove TVEC.[484] Additionally, in 2019, the tech companies' nine-point plan to implement the Call included work to "accelerate machine learning and AI." From the beginning, the Call anticipated the emerging challenges and opportunities of AI and carved out space to discuss new technologies and TVEC online. In 2022, the Call accelerated this work by launching the CCAIO, which enables accredited researchers to examine algorithmic processes and their impact on radicalization. In recent months, Call Leaders have discussed their desire to further the work the Call has already started on understanding the impact of algorithmic systems on radicalization and consideration of emergent technologies, including GenAI.[485]

1. What is GenAI?

To understand GenAI, it is helpful to understand that an algorithm is a set of instructions given to a computer or online system that dictates how to transform a set of data into a useful informational output.[486] AI is a process that layers many algorithms and applies software code to teach computers how to understand, synthesize, and generate knowledge in ways similar to the ways in which people do it.[487] In recent months, several companies

[483] The Christchurch Call to Action: *Full English Text, supra* note 7 ("Review the operation of algorithms and other processes that may drive users towards and/or amplify terrorist and violent extremist content to better understand possible intervention points and to implement changes where this occurs. This may include using algorithms and other processes to redirect users from such content or the promotion of credible, positive alternatives or counter-narratives. This may include building appropriate mechanisms for reporting, designed in a multi-stakeholder process and without compromising trade secrets or the effectiveness of service providers' practices through unnecessary disclosure.").

[484] *Id.*

[485] Christchurch Call to Action, *Four years of the Christchurch Call*, *supra,* note 438.

[486] Jory Denny, *What is an Algorithm? How Computers Know What to Do with Data*, THE CONVERSATION (Oct. 17, 2020), https://theconversation.com/what-is-an-algorithm-how-computers-kno w-what-to-do-with-data-146665 [https://perma.cc/5U4Y-2N9N].

[487] Marc Andreessen, *Why AI Will Save the World*, ANDREESSEN.HOROWITZ (JUN. 6, 2023), https://a16z.com/2023/06/06/ai-will-save-the-world/ [https://perma.cc/F7NX-2F97] ("a short description of what AI *is*: The application of mathematics and software code to teach computers how to understand, synthesize, and generate knowledge in ways similar to how people do it. AI is a computer program like any other – it runs, takes input, processes, and generates output. AI's output is useful across a wide range

have released AI products that can generate new content through learning patterns from pre-existing data, including text, images, and video.[488] These GenAI products are built from large language models that are trained on an enormous amount of text to recognize patterns in language.[489] While predictive language models have been around since the 1980s, in 2017 Google researchers created a new architecture called transformers that allowed language models to train on massive data-sets.[490] These 2017 transformer-based language models created a much richer representation of language, but were limited by the lack of computing power available to researchers.[491] As a result, initial models were expensive to build, because they required so much data to function properly.[492] However, once the data is compiled and trained, generating text or other outputs becomes relatively cheap to do and can be fine-tuned for specific tasks.[493] Given the ease of their use, it is hard to accurately forecast how the new technologies will impact content moderation processes, but a few key trends are emerging.[494]

GenAI could both positively and negatively impact the prevalence of TVEC online and its moderation in several ways. First, online platforms already heavily rely on AI models for

of fields, ranging from coding to medicine to law to the creative arts. It is owned by people and controlled by people, like any other technology.").

[488] Kristen E. Busch, *Generative Artificial Intelligence and Data Privacy: A Primer*, CONGRESSIONAL RESEARCH SERVICE (May 23, 2023), https://crsreports.congress.gov/product/pdf/R/R47569 [https://perma.cc/FK4C-SKR2].

[489] *Id.*

[490] Gabriel Nicholas & Aliya Bhatia, *Lost in Translation, Large Language Models in Non-English Content Analysis*, CENTER FOR DEMOCRACY AND TECHNOLOGY (May 2023), https://cdt.org/wp-content/uploads/2023/05/non-en-content-analysis-primer-051223-1203.pdf [https://perma.cc/EZC5-EZ2K].

[491] *Id.* at 13 ("But in 2017, Google researchers released a paper on a new architecture called transformers, which allowed language models to train on lots of data at the same time, in parallel rather than in sequence. These transformer-based language models could ingest so much data simultaneously that they could learn associations between entire sequences of words, not just individual words.").

[492] Busch, *supra* note 488 at 3 ("for example, OpenAI's ChatGPT was built on a large language model that was trained on over 45 terabytes of text data scraped from the internet.").

[493] Nicholas & Bhatia, *supra* note 490.

[494] Tom Cunningham, *The Influence of AI on Content Moderation and Communication*, GITHUB (Jul. 7, 2023), https://tecunningham.github.io/posts/2023-06-06-effect-of-ai-on-communication.html [https://perma.cc/X3QW-TJU2].

their content moderation operations, including for the detection of spam, bots, child sexual abuse material, hate speech, TVEC, and other violating content.[495] As companies better integrate GenAI technologies into their content moderation processes, they should get better at finding and removing violating content as well as increase the accuracy of content moderation systems, because AI will be able to more closely replicate human judgment.[496] On the other hand, the widespread availability of GenAI tools will significantly reduce the costs and time it takes for bad actors to develop content.[497] Therefore, even as detection capabilities improve, the bad actors producing harmful content are likely to use GenAI to create content that can evade platform detection tools.[498] Additionally, the widespread availability of GenAI will significantly reduce the costs and time it takes for bad actors to run extensive influence operations online.[499] As a result, it will be much easier to manipulate and synthesize media, which will make it harder for people to discriminate between real and fake media.[500] Therefore, GenAI is likely to improve the tools available for both the detection and creation of harmful content.

---

[495] *Id.* ("AI classifiers are rapidly approaching human-level accuracy for these properties and this means that platforms (and governments) will be able to near-perfectly filter out content that violates their rules, even when content-producers have access to the same technology.").

[496] *Id.*; *see also* Alex Rosenblatt et al., *Unleashing the Potential of Generative AI in Integrity, Trust & Safety Work: Opportunities, Challenges, and Solutions*, THE INTEGRITY INSTITUTE (Jun. 8, 2023), https://integrityinstitute.org/blog/unleashing-the-potential-of-generative-ai-in-integrity-trust-amp-safety-work-opportunities-challenges-and-solutions [https://perma.cc/UX3T-N7DF].

[497] Josh A. Goldstein et al., *Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations*, GEORGETOWN CENTER FOR SECURITY AND EMERGING TECHNOLOGY, OPENAI & STANFORD INTERNET OBSERVATORY (Jan. 10, 2023), https://arxiv.org/abs/2301.04246 [https://perma.cc/PHN4-PNMW].

[498] *Id.*

[499] *Id.* at 8 ("Language models could drive down the cost of running influence operations, placing them within reach of new actors and actor types. Likewise, propagandists-for-hire that automate production of text may gain new competitive advantages.)".

[500] *Id.* at 8 ("Recent AI models can generate synthetic text that is highly scalable, and often highly persuasive. Influence operations with language models will become easier to scale, and more expensive tactics (e.g., generating personalized content) may become cheaper. Moreover, language models could enable new tactics to emerge—like real-time content generation in one-on-one chatbots.").

Another area in which GenAI could have both positive and negative impacts is in training content moderation systems to better understand local languages and contexts. The lack of non-English datasets remains one of the biggest challenges for content moderation systems because, without sophisticated classifiers, automated tools struggle to understand local contexts. While larger companies may hire teams of specialists with language expertise, smaller companies cannot hire moderators fluent in multiple languages.[501] To overcome this challenge, GenAI could assist in the creation of synthetic datasets to help train content moderation classifiers in non-English languages.[502] These generated datasets could fill in linguistic gaps and improve classifiers, which would increase the quality of content moderation and slow the proliferation of harmful content online.[503] However, these generated datasets need to be carefully trained and overseen by humans. If not, GenAI could have a negative impact because the large language models can have built-in biases which could undermine many human rights protections.[504] Therefore, it is necessary to build guardrails around this technology and establish norms. Multistakeholder forums offer promise for doing this; done well, they should enable the integration of the diverse perspectives needed to make this a safer process.

2. What is the Impact of GenAI on TVEC?

[501] Thorley & Saltman, *supra* note 145 at 7 ("Terrorist content is shared in a wide array of languages, and while larger tech platforms have the capacity to employ specialist teams with subject matter and language expertise, most companies have comparatively small moderation teams to review content and very few linguists with the appropriate mix of global dialects.").

[502] Nicholas & Bhatia, *supra* note 490, at 37 ("At once, companies are increasingly deploying multilingual language models to bridge the gap between the functionality in English and other languages across a myriad of tasks, such as harmful content detection, sentiment analysis, and content scanning. However, as we show in this paper, these multilingual systems are relatively new and perform inconsistently across languages.").

[503] *Id.*

[504] *Id.* at 6 ("Large language models' general use in content analysis raises further concerns. Computational linguists argue that large language models are limited in their capacity to analyze forms of expression not included in their training data, meaning they may struggle to perform in new contexts. They may also reproduce any biases present in their training data. Often, this text is scraped from the internet, meaning that large language models may encode and reinforce dominant views expressed online.").

According to Brian Fishman, a terrorism expert previously employed by Meta, tech companies have been using automation and AI for years to moderate TVEC in both simple and complex ways.[505] Simple automation is used in technologies like GIFCT's hash-sharing database, which matches static information to identify TVEC online.[506] Complex automation, powered by AI, is used to build sophisticated text classifiers that can assess new material and determine the likelihood of it being TVEC.[507] Complex AI processes will become far more sophisticated, and GenAI could help create variants of known pieces of violating content and block their upload.[508] For example, one reason the Christchurch shooter's video was so hard to remove is because sympathetic extremists regularly altered versions of it, often only slightly, to evade hash-based detection systems. Using GenAI, a computer could create variants and hash them for automated detection systems. However, to protect speech and human rights, these variants should be checked before they are automatically added to hash-sharing databases. Overall, GenAI is likely to improve detection of TVEC, increase the speed and effectiveness of human processes, and provide more transparency to users.[509]

GenAI can also compound the problem of moderating TVEC online by making it easier for bad actors to create content that is more appealing towards vulnerable groups which could lead to an increase in radicalization. GenAI will drive down the costs of running TVEC influence operations by automating the production of propaganda used to

[505] Fishman, *supra* note 20.

[506] *Id.* ("Simple automation matches static information to identify problematic content or patterns. This includes keyword searches, hash-matching, and various rule-based detection schemes. Sometimes these systems are extremely effective, especially when combined with intelligence collection and sharing.").

[507] *Id.* ("Complex automation, however, requires building sophisticated classifiers that not only match known bad content but also can assess novel material and determine the likelihood that it violates some predetermined rule. Using such tools to achieve policy ends is an art in itself—and in that way, social media companies are canaries in the coalmine for lawmakers and bureaucrats around the world who will increasingly need to both set policy constraining the use of AI and establish guidelines for implementing policy via AI.").

[508] Cunningham, *supra*, note 494 ("The prevalence of variations of known-violating content will decrease. E.g. content that is a match against databases of illegal sexual media (PhotoDNA), IP-protected content (ContentID), or terrorist recruitment content (GIFCT). Obfuscation will become harder as AI models get better.").

[509] Rosenblatt et al., *supra* note 496.

radicalize extremists.[510] Additionally, GenAI could help make TVEC more compelling and persuasive by generating individualistic messages which include specific linguistic and cultural context.[511] Furthermore, GenAI could decrease the cost of recruitment by deploying GenAI chat bots that target vulnerable persons through one-on-one conversations in online environments.[512] Finally, GenAI could help influence operations to avoid detection by hash-sharing databases as they would no longer need to use copy-pasted messaging.[513] These significant risks will require technologists to work with civil society, governments, and companies to deploy safeguards and establish norms to prevent further radicalization campaigns online.

3. Options for the Call to Address the Impact of GenAI on TVEC

A curated MSI brings together governments, companies, and civil society to address problems and propose solutions when new technologies are likely to have a profound impact on society. GenAI creates new 'tools and weapons' in the effort to combat TVEC online and the Call is strategically positioned to support solutions for problems GenAI may create as it relates to the proliferation of TVEC online.[514] The Call could tackle these challenges by expanding its ongoing efforts or by slightly restructuring its curated MSI approach. Indeed, as part of their 2022 Leaders' Summit, the Call recognized the importance of addressing new technology issues as they relate to the Call's 25 commitments, and that the Call model might assist with this work.[515] To fulfill this goal, the Call created a "New Tech" workstream, which brings together the Call's multistakeholder community to support the adoption of new technologies while promoting

[510] Goldstein et al., *supra* note 497 at 3 ("For malicious actors looking to spread propaganda—information designed to shape perceptions to further an actor's interest—these language models bring the promise of automating the creation of convincing and misleading text for use in influence operations, rather than having to rely on human labor.").

[511] *Id.* at 4 ("Generative models may improve messaging compared to text written by propagandists who lack linguistic or cultural knowledge of their target.").

[512] *Id.*

[513] *Id.* at 4; *noting* that propaganda will become less discoverable because ("[e]xisting campaigns are frequently discovered due to their use of copy-and-pasted text (copy-paste), but language models will allow the production of linguistically distinct messaging.").

[514] Ardern, *supra* note 6.

[515] Ardern & Macron, *supra* note 428.

safety and securing against TVEC.[516] This workstream is addressing a range of issues including the development of immersive, augmented and virtual reality environments, the impact of the decentralized web, the use of new AI tools, and how terrorist and violent extremists use gaming platforms. A second area where the Call could expand its work to address GenAI is through its Algorithms and Positive Interventions Workstream, where the Call prioritizes action to better understand the impacts that algorithms and other processes may have on TVEC.[517] Through this workstream the Call could explore ways to improve research insights into GenAI that could provide technical, political, and social assurance for governments, companies, and users.

Additionally, the Call could expand the work of the CCIAO to research how GenAI will impact the distribution of TVEC online. One way to do this would be to empower researchers to use the CCIAO to test safety features and develop guardrails for GenAI. In this way, the CCIAO could act as a tool to allow researchers to experiment with new products in a controlled setting. In the area of technology governance, this type of environment is frequently referred to as a "sandbox." In recent years, many stakeholders have deployed developmental sandboxes when experimenting with new technology, as they provide a conducive, contained space where governments, companies, civil society and other stakeholders can test technologies before launching them at scale.[518] Additionally, a sandbox would provide a controlled environment for stakeholders to work together to develop technologies in a responsible and ethical way.[519] A CCIAO

[516] *Id.* at 4 ("Launch a new stream of work to understand how we can support the adoption of new technologies while promoting safety and securing against terrorist and violent extremist content.").

[517] Christchurch Call to Action, *Christchurch Call Initiative on Algorithmic Outcomes, supra* note 432.

[518] United Nations Department of Economic and Social Affairs, *Sandboxing and Experimenting Digital Technologies for Sustainable Development*, UNITED NATIONS FUTURE OF THE WORLD POLICY BRIEF NO. 123 (Dec. 2021), https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/PB_123.pdf [https://perma.cc/5ZGB-8TZX].

[519] Wolf-Georg Ringe, *Why We Need a Regulatory Sandbox for AI*, UNIVERSITY OF OXFORD FACULTY OF LAW BLOGS (May 12, 2023), https://blogs.law.ox.ac.uk/oblb/blog-post/2023/05/why-we-need-regulatory-sandbox-ai [https://perma.cc/5SML-XQHU] ("A regulatory sandbox promises a number of advantages. First, it promotes innovation: AI is a rapidly evolving technology, and the regulatory environment has struggled to keep up. A sandbox allows for the development of new AI technologies in a controlled

development sandbox could have four key functions. First, researchers could study how users are exposed to TVEC and how a person could be redirected or otherwise disengaged from TVEC using GenAI. Second, researchers could explore the accuracy of the systems detecting and removing TVEC and concerns around bias. Third, researchers could test ways to create a healthier, safer online information environment that reduces radicalization and the risks of harms relating to TVEC. Finally, this sandbox could help foster multistakeholder solutions that support human rights and a free, open, secure internet. This project would leverage the existing work of the Call and provide a sustainable solution to addressing new and emerging technologies.

As explored above, the moderation of TVEC online is an area where there is consensus among stakeholders on what should and should not be allowed online in line with human rights principles. Additionally, stakeholders are highly motivated to find solutions to the problems created by TVEC online because it can lead to offline violence. Moreover, for stakeholders considering how to moderate GenAI content online, starting with a (relatively) uncontroversial type of content like TVEC can provide a framework for other areas of content moderation. The Call could bring its multistakeholder approach to the GenAI and rapidly scale up. Additionally, as discussed above, one of the greatest threats from the development of GenAI is the potential to radicalise individuals towards terrorist and violent extremism. Therefore, the Call should consider ways to deploy its resources and scale up its impact on policy governance relating to the development of GenAI.

CONCLUSION

As the Call considers its future, this article has several suggestions to help the organization build a self-sustaining MSI. These suggestions are based on an exploration of how single-sided and multistakeholder models have impacted the governance of user-generated content online over the years. Governmental regulatory frameworks have inherent problems balancing human rights and adapting to technical challenges and

environment reducing the risk of violating laws or regulations. This has proven to reduce the so-called 'time to market' for innovations, giving new businesses increased legal certainty and thereby leading to more innovation.").

companies are struggling to draw lines around acceptable and unacceptable speech. Therefore, this article argues that MSIs are the best, most sustainable, model to protecting the freedom of expression and reducing harmful content online. As demonstrated by the success of multistakeholderism in the internet governance space, the best solutions to content moderation challenges come about when MSIs bring together a diverse coalition of stakeholders and craft consensus-based policies.

The Call was set up as an MSI in the wake of the tragic events of March 15, 2019, and has made significant progress towards eliminating terrorist and violent extremist content while protecting a free, open, and secure internet over the past four years. The Call has accomplished this through a mulitstakeholder approach that brings together governments, tech companies and civil society. To sustain the momentum of the Call and advance its core mission, first, the Call should explore restructuring the MSI to ensure it has a strong foundation to scale and grow the organization. Second, the Call should further expand its work to address the challenges and opportunities posed by the development of GenAI and its impact on TVEC online.

The Call should further expand its work on GenAI as it has addressed the impact of AI on content moderation from the beginning and therefore, is in a prime position to become a leading MSI developing best practices. The Call should explore ways to expand the work of the CCIAO to foster a multistakeholder effort to better understand how GenAI will impact the prevalence of TVEC online. While TVEC is only one type of content that will be impacted by GenAI, it is a good place for an MSI to start because stakeholders generally agree on foundational definitions and the harms of the proliferation of TVEC are so great. Indeed, the risk of offline harms caused by the prevalence of TVEC online has shifted in recent years from as violent extremists have attacked democratic institutions in places like Washington, DC on January 6, 2021, in Wellington on March 2, 2022, and in the "Freedom Convoy" which turned violent in Canada in 2022. Understanding how GenAI will impact TVEC online and finding multistakeholder solutions to address the problems could be foundational to all other GenAI challenges going forward. Therefore, the Call should explore how it can expand its work in this area.

As Jacinda Ardern outlined in her op-ed in June 2023, "I see collaboration on AI as the only option… Together, we stand the best chance to create guardrails, governance structures and operating principles that act as the option of least regret. We don't have to create a new model for AI governance. It already exists, and it works."[520] The Call has the foundations and by implementing these recommendations it can better ensure its future as a self-sustaining MSI.

[520] Ardern, *supra* note 6.

APPENDIX: FREQUENTLY USED ACRONYMS

Artificial Intelligence	AI
Christchurch Call Advisory Network	CCAN
Christchurch Call Initiative on Algorithmic Outcomes	CCIAO
Christchurch Call to Action	The Call
Digital Services Act	DSA
European Union	EU
Generative Artificial Intelligence	GenAI
Global Network Initiative	GNI
International Covenant on Civil and Political Rights	ICCPR
International Telecommunications Union	ITU
Internet and Jurisdiction	I&J
Internet Assigned Numbers Authority	IANA
Internet Corporation for Assigned Names and Numbering	ICANN
Internet Engineering Task Force	IETF
Internet Governance Forum	IGF
Islamic State of Iraq and Syria	ISIS
Multistakeholder Initiative	MSI
National Telecommunications and Information Administration	NTIA
Organization for Economic Co-operation and Development	OECD
Regulation on Preventing the Dissemination of Terrorist Content Online	TCO
Section 230 of the Communications Decency Act	Section 230
Tech Against Terrorism	TAT
Terrorist and Violent Extremist Content	TVEC
The Global Internet Forum to Counter Terrorism	GIFCT
United Nations	UN
United Nations Educational, Scientific and Cultural Organization	UNESCO
United States	US
World Summit on the Information Society	WSIS