# 10

# Drawing Lessons from Efforts at Moderating Extremism

*Farah Pandith and Simone Lipkind*[*]

## 10.1 INTRODUCTION

During the summer of 2023, as interest in AI technology skyrocketed, so did the public's fear of its implications. Pew's August 2023 poll found that 51 percent of those surveyed were more concerned about AI than excited about it, marking not only a majority of respondents but also a 51 percent jump from 2021, when Pew had first polled Americans on this AI question.[1] A 2023 Reuters poll found that 61 percent of Americans believed AI could threaten civilization.[2]

Much of the fear of AI revolves around a future doomsday-scenario in which the technology becomes as smart as humans and then turns against us on its own accord. But what about bad actors and subsets of society that may already be working to wield AI against their fellow humans? Tristan Harris, an ethicist, has sounded the alarm on Silicon Valley's overly rosy perception of AI's ability to democratize access and specifically warned against the dangers of unqualified democratization.[3] Unqualified democratization, introduced to a world teeming with hate and extremism, makes for an ominous mix. The technology is new, but the threat is old, and the past offers insights into how extremists leverage technological advancements for their own nefarious purposes. This chapter will delve into the intricacies of violent

[*] Mitakshi Lakhani '24, candidate for Masters in Law and Diplomacy at the Fletcher School of Law and Diplomacy at Tufts University, contributed to the development of this chapter.

[1] A. Tyson & E. Kikuchi, *Growing Public Concern about the Role of Artificial Intelligence in Daily Life*, Pew Research Center (Aug. 28, 2023), https://www.pewresearch.org/short-reads/2023/08/28/growing-public-concern-about-the-role-of-artificial-intelligence-in-daily-life/ (last visited Feb. 18, 2024).

[2] A. Tong, *AI Threatens Humanity's Future, 61% of Americans Say: Reuters/Ipsos Poll*, Reuters (May 17, 2023), https://www.reuters.com/technology/ai-threatens-humanitys-future-61-americans-say-reutersipsos-2023-05-17/ (last visited Feb. 18, 2024).

[3] Tristan Harris, *AI Myths and Misconceptions, Your Undivided Attention*, Center for Humane Technology (May 11, 2023), https://www.humanetech.com/podcast/ai-myths-and-misconceptions (last visited Feb. 18, 2024).



Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010

extremism, its historical evolution and visible amplification through online disinformation, and the critical need for a whole-of-society approach to combat these threats.

## 10.2 DEFINING VIOLENT EXTREMISM

One must define violent extremism before attempting to speak about its union with disinformation. "Violent extremism" essentially involves threats or behaviors indicating a movement toward violence. As J. M. Berger, a writer and researcher focused on extremism, argues, there are extremist activities that can lead to violence without being violent themself, such as "segregation, laws against interracial marriage, and employment bans."[4] Beyond the actual violence, violent extremism includes preparation, organization, and motivation, involving law enforcement. Ideologically, it stems from religious, political, or ethnic grounds, which have been growing since the tragedy of 9/11. In today's context, we are discussing non-state actors using an "us versus them" ideology that assumes, "the success of 'us' is inseparable from hostile acts against 'them,'"[5] but various ideologies can lead to destructive actions, posing a worrisome threat.

Countering violent extremism (CVE) is a new strategy that emerged after 9/11. It aims to dismantle ideology's appeal and hinder non-state actors' recruitment by strengthening communities against extremist ideas, promoting alternative identities, and establishing social, mental, and cultural support systems. Crucially, CVE is non-kinetic – it does not rely on military action against radicalized individuals.

## 10.3 A HISTORICAL TIME LINE OF VIOLENT EXTREMISM IN THE DIGITAL AGE

Starting in the 1980s and '90s, websites and blog posts emerged as platforms for spreading hateful ideologies, but there was minimal effort from the private or public sector to regulate or take down such content. The novelty and absoluteness of publishing their ideology on a website gave violent extremists like StormFront and Al Qaeda the legitimacy they craved. Extremists discerned early on that the power to persuade goes to those who can reach the most people, and communicating your point of view on a scale not previously possible gave violent extremists what they most desired: possible ideological soldiers. Through the 9/11 era, and even for some time afterwards, Al Qaeda leveraged the internet, as well as various mediums like videotapes and CDs, to disseminate sermons and spread their ideology. It was a form of communication, and it was powerful. It created ripple effects and acted as a

[4] Julia Simms, *Are We Talking About Extremism All Wrong?*, New America (Sept. 27, 2018), http://newamerica.org/weekly/are-we-talking-about-extremism-all-wrong/ (last visited Feb. 16, 2024).

[5] J. M. Berger, Extremism (2018).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010

precursor to the "shares" we see now. This method of communication helped create a brand and a powerful emotional response. Indeed, it facilitated radicalization and significant new opportunities for organizations like Al Qaeda. But there were limitations. It was a static one-way communication tool.

The shift from stagnant websites to interactive social media platforms marked another pivotal milestone, not just in the platforms themselves but also how extremist groups utilized them. The transition to two-way communication on platforms like Facebook and YouTube transformed the landscape, allowing for the engagement necessary to create communities that provided intellectual and emotional support. This shift, post-9/11, was exemplified by video posts of Bin Laden on platforms like YouTube, which gave groups like al Qaeda the opportunity to foster emotional connections in a way they could not before on static websites. What is powerful here is the lift off from an individual who has some ideas, to a larger group of people who can *act* on these ideas after. Thus, there was now a new dimension to ideological affinity.

Furthermore, Facebook, YouTube, and Google, through their fast-changing algorithms, started shaping online behavior – creating a rapid, addictive cycle. With single searches triggering millions of results, these platforms' frictionless pace prevented (and continues to prevent) thoughtful absorption, pushing users in directions influenced by emotional responses to content, regardless of their rational instincts. These manipulative code nuances, nudging users oblivious to their intent down unforeseen rabbit holes, were unprecedented and society was not trained to recognize their subtle pull. Users could find themselves exposed to extremist content after a single innocuous search.

The 2005–6 Danish Cartoon Crisis acted as another flashpoint altering the landscape of digital extremism. Satirical cartoons of the Prophet Muhammad drawn in Denmark sparked an unforeseen global phenomenon.[6] The cartoons' virality showcased digital spaces' unprecedented force multiplier effect. What happened in Copenhagen resonated in Kabul, creating an interconnectedness that was previously unimaginable and highly consequential. Three years after the cartoons' initial publication, Al Qaeda bombers attacked Denmark's embassy in Pakistan killing six and injuring dozens.[7] Five years later, a Somali axeman tried to kill one of the Danish cartoon artists in his own home.[8]

The advent of the so-called Islamic State, also known as ISIS, marked another paradigm shift in extremists' exploitation of digital tools for international influence.

[6] C. S. Smith & I. Fisher, *Firestorm over Cartoon Gains Momentum*, New York Times (Feb. 2, 2006), https://www.nytimes.com/2006/02/02/international/europe/firestorm-over-cartoon-gains-momentum.html (last visited Feb. 18, 2024).

[7] *Dead after Pakistan Embassy Blast*, CBS News (June 3, 2008), https://www.cbsnews.com/news/6-dead-after-pakistan-embassy-blast/ (last visited Feb. 18, 2024).

[8] *Somali Charged in Attack on Cartoonist*, NBC News (Jan. 2, 2010), https://www.nbcnews.com/id/wbna34662389 (last visited Feb. 19, 2024).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010

ISIS leveraged the digital domain to reshape extremist narratives and emerged as online superstars. Their mastery of this "new" human landscape was pivotal, creating a brand through content like their *Dabiq* magazine. The magazine, with its slick design, served as a powerful recruitment tool, helping the group attract members from far beyond Iraq and Syria. Al Qaeda's *Inspire* magazine, which also attracted a wide online readership, took a different approach, acting more as a manual for adherents, instructing them on various aspects of violent extremism, including bomb-making. The Boston Marathon bombers, Tamerlane and Dzhokhar Tsarnaev (twenty-six and nineteen years old, respectively, in 2013) claimed they used this digital content as a guide in their attack, marking a critical juncture in understanding extremist instruction manuals' impact in the digital era.[9] These online resources facilitated both far-reaching recruitment efforts on behalf of a distant caliphate and lone wolf homeland attacks.

The transition to encryption ushered in yet another new challenge – smaller platforms, not just major social media players – emerged as potent threats, serving as sanctuaries for those banned from larger platforms, allowing them to continue their activities.[10] In 2016, the rise of QAnon showcased the power of misinformation and conspiracy theories, even on relatively unknown sites. Originating from platforms like 4Chan and Telegram, QAnon, the theory that "a group of Satan-worshipping elites who run a child sex ring are trying to control our politics and media,"[11] initially seemed too absurd to be taken seriously. This analysis was further compounded by gender biases that underestimated a group with so many female adherents.[12] Critically, these smaller platforms were also encrypted allowing users to raise money, spread their ideology, organize, and plan in a way that initially went undetected by authorities. The QAnon fringe movement gained momentum during the Trump era by exploiting these smaller, ungoverned spaces to both eventually infiltrate mainstream networks and maintain a way of communicating with their audiences in the event they were banned from those larger, more mainstream platforms. Their example would prove to be essential learning for how the online space was

[9] On April 15, 2013, Tamerlan and Dzhokhar Tsarnaev detonated two homemade pressure cooker bombs near the finish line of the Boston Marathon killing three and injuring hundreds. *See Unclassified Summary of Information Handling and Sharing Prior to the April 15 2013 Boston Marathon Bombings*, United States Department of Justice (Apr. 10, 2014), https://oig.justice.gov/reports/2014/s1404.pdf (last visited Feb. 14, 2024).

[10] Jacob Ware, *The Third Generation of Online Radicalization*, George Washington University Project on Extremism (June 16, 2023), https://extremism.gwu.edu/third-generation-online-radicalization (last visited Feb. 19, 2024).

[11] J. Rose, *Even If It's "Bonkers," Poll Finds Many Believe QAnon and Other Conspiracy Theories*, NPR.org (Dec. 30, 2020), https://www.npr.org/2020/12/30/951095644/even-if-its-bonkers-poll-finds-many-believe-qanon-and-other-conspiracy-theories (last visited Feb. 18, 2024).

[12] M. Bloom, F. Pandith & J. Ware, *Female Extremists in QAnon and ISIS Are on the Rise. We Need a New Strategy to Combat Them*, NBC News (Dec. 11, 2020), https://www.nbcnews.com/think/opinion/female-extremists-qanon-isis-are-rise-we-need-new-strategy-ncna1250619 (last visited Feb. 18, 2024).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010

changing. This evolution in technique involving linking profiles across platforms marked a paradigm shift in the dissemination of extremist ideologies. This dynamic shift underscores the imperative to comprehend the evolving stages of technology and their profound impact on human interaction.

In 2017, the Burmese military triggered the Rohingya Genocide through a years-long campaign of misinformation, lies, and conspiracy propaganda on Facebook.[13] This tragic event served as a stark revelation, illustrating the profound impact achievable not only by terrorist organizations or extremist movements but also by state actors themselves employing digital platforms.

Two years later, the 2019 Christchurch attack, live-streamed and accompanied by an advanced coordinated online presence of memes, hashtags, and crowdsourcing, showcased extremists' nuanced understanding of online communication.[14] The strategic adaptation of words and images aimed to attract more followers, utilizing platforms like Facebook for engagement, and the shared manifesto, with disturbing parallels to Oslo killer Anders Brevik's, prompted the Christchurch Call, a commitment from governments and online service providers to eliminate terrorist and violent extremist content online in order to prevent further such attacks.[15] Five years later, such attacks are still too common.

In summary, the internet initially offered simplistic one-way interactions. Although primitive compared to current technology, this advancement was a milestone nonetheless allowing regular individuals to share content in unprecedented way. The next stage involved two-way interaction. The subsequent wave introduced encryption, empowering smaller platforms to take ownership and provide refuge for malicious actors. Understanding these stages is imperative, as they unveil differences in human interactions with technology. Some involve static absorption of information, others incorporate emotional engagement, while the third stage is

[13] In 2017, Myanmar's security forces embarked on a campaign in the country's western region under the pretext of restoring security. The campaign killed 6,700 Rohingya in the first month alone and would go on to displace hundreds of thousands in what UN Secretary-General Antonio Guterres has labeled as an ethnic cleansing attempt. *See* Eleanor Albert & Lindsay Maizland, *The Rohingya Crisis*, Council on Foreign Relations (Jan. 23, 2020), https://www.cfr.org/backgrounder/rohingya-crisis. (last visited Februart 18, 2024). Leading up to the campaign, Myanmar's military had as many as 700 people tasked with spreading anti-Rohingya propaganda on Facebook. *See also* Paul Mozur, *A Genocide Incited on Facebook, with Posts from Myanmar's Military*, New York Times (Oct. 15, 2018), https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html.

[14] In March of 2019, Brenton Tarrant killed 51 people at a mosque and Islamic center in Christchurch, New Zealand. New Zealand authorities eventually uncovered that Tarrant had visited alt-right websites, posted alt-wing content on his Facebook page, and donated money to alt-right groups outside of New Zealand. *See Christchurch Massacre: Inquiry Finds Failures ahead of Attack*, BBC News (Dec. 8, 2020), https://www.bbc.com/news/world-asia-55211468 (last visited Feb. 19, 2024).

[15] *Christchurch Call Story*, Christchurch Call, https://www.christchurchcall.com/about/the-christchurch-call-story (last visited Feb. 19, 2024).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010

characterized by clandestine interactions in ostensibly secure spaces.[16] The comfort, ease, and safety of these encrypted spaces pose a substantial challenge to national security, as activities ranging from fundraising to spreading ideologies can transpire beyond the purview of authorities.

Undoubtedly, recounting pivotal moments when digital technology became a conduit for propagating violent extremism is disheartening. They serve as a stark reminder that we are consistently lagging behind in our efforts to address these issues, playing catch-up instead of leading the way. In the internet's nascent days, policymakers underestimated its potential for harm and consequently lacked the regulatory framework needed to navigate its wild and uncharted terrain. Opportunistic actors, spanning from human traffickers to extremists, seized the chance to establish their operations online. The digital age, in a manner of speaking, democratized terrorism – making it accessible to anyone anywhere – dismantling barriers and enabling the localized dissemination of ideologies.

## 10.4 EXTREMISM WITHOUT BORDERS

Even though America chooses to define violent extremism with the artificial parameters "international" and "domestic," ideology has no borders. We understand and have witnessed how important, and far-reaching, inspiration can be. It moves people into action regardless of whether that inspiration comes from a region that is different from yours. The universal desire for identity and affiliation transcends cultural disparities and physical borders. Extremist ideas and techniques seamlessly traverse online landscapes, impacting movements across continents.

In regions as disparate as India, Kenya, and Japan, cultural norms have undergone transformation due to the pervasive influence of online platforms. Japan had QAnon marches happening in its streets, even though we thought QAnon was a distinctly American phenomenon. Yet it was making sense to some Japanese young people. The conspiracy theory was taking root in a country extremely different from the United States. But the rage of the "us vs them," or plausibility of the outrageous conspiracy, or belief in a human hierarchy can still have local meaning in unexpected places, regardless of the specifics of what skin color or religion, for example, are desirable within a specific country. The current epoch, characterized as the "Rage Era," exhibits global ramifications affecting politics, campaigns, and the dissemination of disinformation. Previously unimaginable, America's role as a source of inspiration for online branding and organizational strategies for ideological movements like the Far-Right holds significance, with these ideas disseminating to various corners of the world. There will likely soon come a time when the United States is designated a state sponsor of terror.

[16] *Supra*, note 10.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010

The convergence of lifestyle dynamics with their impact on discourse and identity expression, especially among the youth, through social media, adds another layer of complexity to this global phenomenon. Observing the interplay between the creativity exhibited by the youth in expressing themselves and the negative forces compelling them in specific directions is indeed a captivating phenomenon. This intricate dynamic, unfolding on a global scale, demands a nuanced understanding of how these elements shape narratives and influence societies across the world. The intertwining of marketing tools and tactics with ideological and personal branding has transformed into a lifestyle choice. This extends beyond online platforms to influence various aspects of individuals' lives, from their clothing choices to the news sources they prefer and even the tattoos they display.[17] The dynamic surge we observe in nations, especially in response to American trends, highlights the adaptability and amplification of these ideologies globally.

The exploration of rules, norms, and the evolving dynamics of lifestyle choices provides valuable insights into the complex tapestry of global influences shaping the narratives of our time. The examination of the "developing" world's role in propagating violent extremism through digital means, the global dissemination of extremist ideas, and the intertwining of lifestyle dynamics with ideological expressions illuminates the intricate challenges faced by societies worldwide. As we navigate this complex landscape, a nuanced understanding of the interconnected nature of these phenomena is crucial for developing effective strategies to counteract the influence of extremist ideologies on a global scale. Far too often, our responses have failed to adequately take these complexities and nuances into account.

## 10.5 PREVIOUS ATTEMPTS AT MITIGATING ONLINE HARMS

There was widespread concern after 9/11 that as large online platforms evolved, they could serve as both a haven and recruiting system for extremists; but social media companies still failed to establish a standard for user protection. It was only after 2013, amid the rise of ISIS, that companies and policymakers began to address the issue seriously. Finally, some of the biggest tech companies involved publicly recognized the significance of their platform to extremists.[18]

The government realization struck with urgency, acknowledging that extremists were exploiting these platforms for recruitment globally. Approximately 40,000 individuals from across the globe were recruited by ISIS to the so-called

[17] Farah Pandith, How We Win: How Cutting-Edge Entrepreneurs, Political Visionaries, Enlightened Business Leaders, and Social Media Mavens Can Defeat the Extremist Threat (2019).

[18] S. Higham & E. Nakashima, *Why the Islamic State Leaves Tech Companies Torn between Free Speech and Security*, Washington Post (July 16, 2015), https://www.washingtonpost.com/world/national-security/islamic-states-embrace-of-social-media-puts-tech-companies-in-a-bind/2015/07/15/0e5624c4-169c-11e5-89f3-61410da94eb1_story.html (last visited Feb. 18, 2024).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010

Caliphate,[19] and many of these individuals were radicalized online. Governments mounted pressure on tech companies, urging them to act. Although this was a welcome step, the government's initial response was to remove specific content, which ultimately proved ineffective as bad actors adapted, circumventing tracking systems and manipulating hashtags. Consider *Mein Kampf*, once confined to the realm of print, now perpetually accessible online. The dissemination of its ideology and its influence on nefarious actions exemplify the lasting impact of online content. Once something exists online, it is almost impossible to remove it completely or permanently. Even though an individual's, and the US government's, first instinct is often "take content down!" this strategy does not work for long and misses that social media platforms have emerged as colossal information repositories fundamentally reshaping the landscape. This metamorphosis has not only revolutionized the interconnection of ideas but also introduced enduring challenges with far-reaching consequences. The videos of America-born Al Qaeda Islamic "scholar" Anwar Alawlaki continued to radicalize individuals for years after his death despite repeated attempts to block them, underscoring the persistence of ideologies and the internet's pivotal role in recruitment.

As the US government shifted from a predominantly content removal focused policy, it pressured tech companies to act on algorithms enabling harmful content. Terms of service became a focal point, with diverse platforms having different standards. The US government emphasized not just ethical behavior but companies' responsibility to national security. However, aligning private-sector incentives with the nation's need to counter extremists posed challenges. Tech companies began modifying terms of service and algorithms but with misaligned incentives regarding the nation's need to curb extremism. This dualistic dynamic impacted the speed and nature of responses, entangling legal, moral, and ethical considerations in the quest for a solution. Silicon Valley's response lacked innovation and proactivity, because it was primarily driven by bottom-line considerations. Efforts like Facebook's Oversight Board[20] and the Global Internet Forum to Counter Terrorism (GIFCT)[21] were steps

[19] Lila Hassan, *Repatriating ISIS Foreign Fighters Is Key to Stemming Radicalization, Experts Say, but Many Countries Don't Want Their Citizens Back*, PBS (Apr. 6, 2021), https://www.pbs.org/wgbh/frontline/article/repatriating-isis-foreign-fighters-key-to-stemming-radicalization-experts-say-but-many-countries-dont-want-citizens-back/ (last visited Feb. 19, 2024).

[20] "The purpose of the board is to promote free expression by making principled, independent decisions regarding content on Facebook and Instagram and by issuing recommendations on the relevant Meta content policy. When fully staffed, the board will consist of 40 members from around the world that represent a diverse set of disciplines and backgrounds. These members will be empowered to select content cases for review and to uphold or reverse Facebook and Instagram content decisions. The board is not designed to be a simple extension of Meta's existing content review process. Rather, it will review a select number of highly emblematic cases and determine if decisions were made in accordance with Meta's stated values and policies." Oversight Board, https://www.oversightboard.com/ (last visited Feb. 18, 2024).

[21] "The Global Internet Forum to Counter Terrorism (GIFCT) is an NGO designed to prevent terrorists and violent extremists from exploiting digital platforms. Founded by Facebook, Microsoft, YouTube and X (formerly Twitter) in 2017, the Forum was established to foster

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010

taken to self-govern and foster collaboration. However, the pace of change remained (and continue to remain) unsatisfactory. As threats continued to spiral, civil society gained prominence, fueled by citizen activism and concerns about technology's impact on minors. The challenge extended beyond content removal, requiring a systemic overhaul.

While mechanisms like GIFCT aimed at containing bad actors' influence, the broader landscape demanded more. The lack of a clear champion among tech companies and the absence of substantial proactive measures highlighted the need for a fundamental reboot of the American approach to what had become a mainstream utility. So stuck was the discourse around online harms that it was almost impossible to think creatively or to apply common sense changes to social media's daily functions.

Addressing online harms demands an innovative, comprehensive approach. Companies, lawmakers, and citizens must collaboratively navigate the challenges posed by extremist content. Facebook's Oversight Board and GIFCT demonstrate attempts at self-regulation, but the pace of change remains a concern. US politics and legislative stalemate make it almost impossible to make any kind of change, even if one is talking about protection of vulnerable children as young as Gen Alpha.

This multifaceted challenge involves not only preventing harmful content but also reducing the appeal of its underlying extremist ideologies. It will require something from all of us.

## 10.6 PATHS FORWARD

The battle against online extremism in its many current and future iterations requires a comprehensive, scaled, multigenerational funded, whole-of-society approach. Policymakers, tech companies, nongovernment organizations (NGOs), and citizens all need to work together, addressing issues of identity and belonging and other causes of radicalization, implementing effective regulations, and fostering a digital environment that prioritizes the well-being of society over individual interests. Behavioral change requires attention to the entire societal eco-system and the patience to stick with a strategy beyond short political cycles. The challenges are vast, but with strategic collaboration and forward-thinking initiatives, it is possible to mitigate the spread and dangers of online extremism. The following elements can decrease these harms:

> technical collaboration among member companies, advance relevant research, and share knowledge with smaller platforms. Since 2017, GIFCT's membership has expanded beyond the founding companies to include over two dozen diverse platforms committed to cross-industry efforts to counter the spread of terrorist and violent extremist content online." *About*, Global Internet Forum to Counter Terrorism, https://gifct.org/about/ (last visited Feb. 18, 2024).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010

*Redefining the Government's Role*: Governments, despite their regulatory power, have limitations in combating extremism. The need for collaboration between tech companies, governments, and NGOs becomes evident here. To make a real difference, society needs a multipronged strategy that involves various sectors and targets different audiences. Policymakers, corporate pioneers, civil society leaders, and regular citizens must collaborate.

While governments can use regulation to influence the online space, designate terrorist groups, and impede funding flows, NGOs often take on the crucial task of preventing the appeal of extremist ideology. They actively engage in preventing the radicalization process by addressing the appeal of extremist ideologies. Local NGOs bring unique legitimacy, insight, and acumen to deploy effective initiatives. However, their efforts are hindered by the scale of the issue; NGOs are relatively small compared to the magnitude of the problem. Governments should do more than just acknowledge the importance of local initiatives and empower civil society by providing support in terms of grants and resources.

A government is best suited to implement incentives for stakeholders whose priorities are not always inherently aligned with the government's own. Analogous to the guardrails in physical spaces, the digital realm requires well-defined standards and laws. Governments should play a proactive role in establishing and enforcing regulations that are better, broader, and forward leaning. Learning from the European Union's Digital Service Act,[22] which imposes stricter requirements and penalties for

[22] The Digital Services Act is the most important and most ambitious regulation in the world in the field of the protection of the digital space against the spread of illegal content, and the protection of users' fundamental rights. There is no other legislative act in the world having this level of ambition to regulate social media, online marketplaces, very large online platforms (VLOPs) and very large online search engines (VLOSEs). The rules are designed asymmetrically: Larger intermediary services with significant societal impact (VLOPs and VLOSEs) are subject to stricter rules. After the Digital Services Act, platforms will not only have to be more transparent, but will also be held accountable for their role in disseminating illegal and harmful content. Amongst other things, the DSA:

1. Lays down special obligations for online marketplaces in order to combat the online sale of illegal products and services;
2. Introduces measures to counter illegal content online and obligations for platforms to react quickly, while respecting fundamental rights;
3. Protects minors online by prohibiting platforms from using targeted advertising based on the use of minors' personal data as defined in EU law;
4. Imposes certain limits on the presentation of advertising and on the use of sensitive personal data for targeted advertising, including gender, race and religion;
5. Bans misleading interfaces known as "dark patterns," and practices aimed at misleading.

Stricter rules apply for very large online platforms and search engines (VLOPs and VLOSEs), which will have to:

1. Offer users a system for recommending content that is not based on profiling;
2. Analyse the systemic risks they create: Risks related to the dissemination of illegal content, negative effects on fundamental rights, on electoral processes and on gender-based violence or mental health.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010

data privacy violations, could serve as a model for the United States to follow. As explained by Christoph Busch in Chapter 3, the Digital Services Act will not replace the EU's Terrorist Content Online Regulation (TCO Regulation), which went into effect in 2022. Under the TCO Regulation, platforms may be required to remove specified "terrorist content" within specific timeframes. As Busch explains, "national authorities have the power to issue 'removal orders' requiring hosting service providers to remove or disable access to terrorist content in all EU Member States 'as soon as possible and, in any event, within one hour of receipt of the removal order.'"

Encouraging positive behavior and rewarding responsible conduct might too be a method worth exploring more deeply. Carrots, in the form of benefits or rewards, could motivate companies to prioritize ethical practices, aligning their interests with societal well-being. Historically, the issue has been complicated by many things, including the pace of technology and the age of US lawmakers, most in the legislative branch are not digital natives,[23] which impacts the way they use and understand social media. This often results in their inability to identify exactly what they want tech companies to do outside of change the algorithms, which will cause companies to lose money and an overall weakness in the government's ability to make the case to regular citizens about online harms. The US government has approached this issue by first trying to convince then force tech companies to change. But one could look at this issue not just with a stick in hand, but also a carrot. Early implementation of incentives for technology companies, such as tax benefits offsetting the loss in revenues due to algorithm changes, could have mitigated the spread of extremist content. The delay in implementing these incentives allowed extremists to exploit the online space, underscoring the importance of timely and effective measures.

There are also a myriad of opportunities for more effective partnerships between governments and private sector tech companies in the fight about societal manipulation. One example: Can AI, in its predictive capacity, help us recognize key opportunities to avert potential harm? The interconnectedness of extremist ideologies poses a formidable challenge. AI, by deciphering patterns, could offer insights. Consider the Christchurch shooter – could AI have detected signs across various online platforms, creating a digital flag for governments? The potential for AI to act as a preemptive tool in preventing attacks is an enticing prospect worth exploring. The current policy landscape often sticks to familiar tools without embracing novelty. Unlike other fields that adapt to new devices, policy tends to rely on

> In the context of the Russian invasion of Ukraine, involving grave and widespread violations of the human rights of the Ukrainian people, and the particular impact on the manipulation of online information, the Digital Services Act introduces a crisis response mechanism. This mechanism will make it possible to analyse the impact of the activities of VLOPs and VLOSEs on the crisis and rapidly decide on proportionate and effective measures to ensure the respect of fundamental rights." *EU Digital Services Act*, https://www.eu-digital-services-act.com/ (last visited Feb. 18, 2024).

[23] G. Skelley, *Congress Today Is Older Than It's Ever Been*, FiveThirtyEight (Apr. 3, 2023), https://fivethirtyeight.com/features/aging-congress-boomers/ (last visited Feb. 19, 2024).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010

traditional methods while expecting different outcomes. The key to fostering trust lies in collaborative efforts, innovative thinking, and a willingness to experiment with different approaches rather than clinging to established ways under the assumption that existing knowledge alone can guide decision-making. This requires cooperation among key stakeholders, domestically and internationally.

*Collaborating on a Global Scale*: Hate and extremism, as borderless issues, demand a global response. However, the current global efforts have been weak, characterized by outdated solutions, small-scale programs, and a lack of creativity. Meanwhile, "us versus them" ideologies have proliferated across border and bridged the domestic-foreign divide.

Policymakers need to collaborate on a global scale, thinking differently, scaling programs, and creating innovative, bold solutions that suit the purpose of countering extremism. Multilateral organizations are a critical tool. We have come together on natural or humanitarian disasters in the past with many different nations pitching in, bringing in tools and insights. We sprint into action when an earthquake or a tsunami or a fire happens. Nations that do not like each other find ways to work together. But we still do not have those systems in place in an effective way on issues of extremist ideology.

Likewise, we have come together on hard power in the past – fighting a common enemy. We have alliances and treaties and other mechanisms to discuss and coordinate. We share hardware and knowledge and information sharing through diplomatic and military channels among governments has improved greatly since 9/11. But even though nations now understand the risk environment better because of this coordination and partnership, on the innovation side we are still behind. There is so little that is out of the box thinking. We keep doing what we have been doing and we are not getting ahead of the threat due in part to the lack of willingness to share and adopt best practices globally, particularly regarding societal aspects.

Unfortunately, nations do not extensively share information on societal aspects of challenges including extremism. There is a need for increased interest in human behavior, fostering learning from other countries, as it intricately ties to online dynamics. The EU stands out in comprehending and addressing these societal impacts more effectively. While Americans discuss the issue, there is a notable gap in translating discussions into actionable measures.

If you are talking about hard power everyone knows what that means. Ideology is far more wiggly and squishy. It makes people uncomfortable and they do not know how to handle it. Multilaterals can make a difference not just by sharing best practices, but by really scaling the innovation piece. There are examples of governments who have decided that they are going to go hard on this issue. The EU is not waiting for the tech companies to tell them what to do. They are saying society comes first! We are going to protect our citizens, and toward that end, companies must abide by our definition of the public good. They show us a roadmap of what new policy can do. There are other non-technology related examples as well. Take the Danes and their successful recycling culture – they ingrained it into society and

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010

incentivized the population effectively.[24] Exploring such successes and understanding the underlying factors can inform strategies for behavioral change in the American context.

This is not the first time that humans have been faced with a difficult problem or a common threat that extends beyond their borders. Have there been other moments in time in human history when people from other parts of the world understood a threat and all came together and said, "Our security is interconnected. What are we going to do about it now that we have not done before" Yes, there have been. And we know firsthand that kind of response has taken place when the realm was science, for example. During the pandemic (while it took what seemed like forever but was actually remarkably fast) we found a vaccine for COVID-19 by sharing data and working together in unprecedented ways across borders. We also troubleshot ways of increasing societal buy-in for things like mask wearing and social distancing. What would it mean to apply these resources and this thinking to the extremist threat we all face? What would it mean not to forsake the exploration process but to know when to solidify an approach and move fast because we have consulted a diverse group of stakeholders? Leading without lagging demands this and a better understanding of what is coming next.

*Imagining and Forecasting the Future*: In the aftermath of 9/11, there was a mad scramble to address the immediate threats without much contemplation on the evolving nature of extremist prowess. We missed the opportunity for deep early ideation and exploration, which could have better prepared our nation for subsequent versions of extremist ideologies or the possibility that old hatreds would reemerge through new variants. We also missed the types of tools that could be deployed and the cultural changes that were a consequence of the new threat environment. Today's challenges draw from old and new tools, various demographics of digital natives, and limited scaled systems to deal with the current landscape.

Disinformation is a type of weapon. We need to respond to disinformation just as if an adversary had a new kind of missile, or tank, or gun. We need to develop a robust analysis of this weapon's potential to do harm.

In the realm of extremism, adept predictions require cultural intelligence. Cultural intelligence is a tool that aids in assessing communities, recognizing cultural shifts, and understanding what is important to specific demographics. One needs to know the reasons *why* a particular demographic acts in a particular way, not just that they are producing a particular type of action. Each generation responds differently, and policymakers should be clear on the nuances within varying societies because it can help build critical knowledge about preventative

[24] *Record-Breaking Year for Danish Recycling: 92 Per Cent of Bottles and Cans Returned for Recycling*, State of Green, Apr. 1, 2020, https://stateofgreen.com/en/news/record-breaking-year-for-danish-recycling-92-per-cent-of-bottles-and-cans-returned-for-recycling/ (last visited Feb. 18, 2024).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010

posturing going forward. When that cultural intelligence is not used at all in policymaking, in how we think about this threat, we miss a huge opportunity to succeed. As technology and society entwine, a cultural shift emerges – one to which bad actors are acutely attuned. Shaping culture becomes a potent weapon in the hands of violent extremists, reshaping beliefs, identities, and actions. We all understand how important culture is in general, and we should be applying that element to the work of defeating extremists as well.

*Focusing on Digital Literacy*: Online misinformation and digital extremism are addictive, and therefore digital literacy and hygiene should be prioritized, just as we prioritize responses to other addictive substances. This is imperative, especially for the three generations of digital natives: Millennials, Gen Z, and Gen Alpha. The understanding of how these generations are manipulated needs to be highlighted. Finland sets an example by teaching digital literacy in kindergarten, recognizing the importance of individuals comprehending manipulated content.[25] This emphasis on digital education can create a more discerning online population that questions and critically evaluates the information they encounter on an almost 24/7 basis. By preparing youth in this way, we make it harder for bad actors to win them over to their ideological armies.

## 10.7 A CAUSE FOR HOPE

Optimism emerges in these younger generations in the form of Gen Alpha, Gen Z, and Millennials, digital natives who are better attuned to the intricate workings of technology through better grasp of digital literacy and cybersecurity. Their potential solutions, combined with policymakers accepting the need for change, could pave the way for a more secure online environment.

While data may not conclusively support the claim, younger generations appear more tech-savvy, discerning fact from fiction. They grasp concepts like deepfakes,[26] CGI,[27] and AI,[28] understanding their presence in daily life. This familiarity might

[25] E. Mackintosh, *Finland Is Winning the War on Fake News. What It's Learned May Be Crucial to Western democracy*, CNN (May 17, 2019), https://edition.cnn.com/interactive/2019/05/europe/finland-fake-news-intl/ (last visited Feb. 19, 2024).

[26] A deepfake is an image, video, or audio recording that has been generated or altered by machine learning, specifically deep learning, technology. Deepfakes are hyper realistic and can convincingly "depict someone appearing to say or do something that they in fact never said or did." *Science & Tech Spotlight: Deepfakes*, Government Accountability Office, last modified Feb. 20, 2020, https://www.gao.gov/products/gao-20-379sp.

[27] CGI, or computer-generated imagery, refers to "the creation of still or animated visual content using imaging software." *Definition: Computer-generated imagery (CGI)*, TechTarget, https://www.techtarget.com/whatis/definition/CGI-computer-generated-imagery (last visited Jan. 30, 2024). In the context of violent extremism, CGI poses a threat similar to deepfakes.

[28] AI, or artificial intelligence, refers to "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments." *Artificial Intelligence*, US Department of State, https://www.state.gov/artificial-intelligence/ (last visited Jan. 30, 2024).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010

empower them to question the authenticity of, for instance, a talking bear on the back porch or a deep fake of President Joe Biden admitting to various crimes. This technological gap affects all generations, but the younger demographics seem more attuned to the ongoing changes.

Every generation faces transformative communication technologies. For Gen Alpha, the internet and AI represent only the latest chapter. While unique, it follows a historical pattern. The crucial message for these digital native generations is that they have agency. Lessons from past technological shifts can empower them to shape a future when responsibility and action lie with individuals who demand more from all of us.

Similar questions arose in the past when new technologies, such as the telephone or radio broadcasts, reshaped the human experience. Each generation witnesses transformative advancements that alter their connection to the world. Understanding this continuum is critical.

Considering the multitude of examples demonstrating how technology can shape societies for better or worse over time, the question arises for digital natives: Could this be the moment for your generation to participate actively? Directly addressing today's challenges requires a collective effort that goes beyond the expectation of external interventions. It is a call to embrace agency, whether one is an entrepreneur, educator, government employee, or any other profession. Society's rules are not dictated by external forces; we collectively shape them. The current challenges arise from a lack of societal responsibility to tackle complex issues, emphasizing the need for immediate action.

The imperative is to act now rather than adopting a passive stance. Reflecting on the example of David Hogg, a Gen Z gun control activist who transformed a personal experience into a commitment to work on gun reform, underscores the impact of individual agency.[29] Waiting for older generations to enact sweeping laws may not be the most effective approach for the moment we find ourselves in. All over the world, youth are activating a new era of empowerment and standards on issues like the environment, human rights, equity, democracy and freedom. Instead of standing by, Gen Z is proactively pursuing the changes they envision for the world.

They are not the first generation to want a more fair, more peaceful world, but they have a leg up on generations before them. The data, insights, and speed of information give them a unique and powerful moment even though challenges may persist. While imperfection is inevitable, the application of lessons learned over the last two decades can mitigate harm and bring us forward into a different era where AI, cultural intelligence, and human dedication across sectors can make the threat of violent extremism less viable.

[29] David Hogg, *March for Our Lives* (Jan. 31, 2024), https://marchforourlives.com/david-hogg/ (last visited Feb. 19, 2024).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010

Waiting for change has not produced protection for any of us. Bad actors like the White Supremacist group Attomwaffen are already working on luring in Gen Alpha.[30] The moment is now to accelerate the application of lessons learned over two decades and build stronger, more resilient communities online. The question is, will we use this moment or squander it?

[30] Helen Young, *Extremists Use Video Games to Recruit Vulnerable Youth. Here's What Parents and Gamers Need to Know*, The Conversation, November 9, 2022, https://theconversation.com/extremists-use-video-games-to-recruit-vulnerable-youth-heres-what-parents-and-gamers-need-to-know-193110.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 04:47:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009438636.010