Congressional Testimony

Holding Big Tech Accountable: Legislation to Build a Safer Internet

Jonathan Greenblatt

CEO

ADL (Anti-Defamation League)

HOUSE COMMITTEE ON ENERGY AND COMMERCE
SUBCOMMITTEE ON CONSUMER PROTECTION AND COMMERCE

Washington, DC

December 9, 2021

10:30 a.m.

Working to stop the defamation of the

Jewish people and to secure

Justice and fair treatment to all since 1913

I. INTRODUCTION

During the past several years, there has been a tectonic shift in the way communities across the world integrate digital and social networks into their daily lives. Whether related to education, worship, family gatherings, social entertainment, or news, the online world is integral to our way of life. The way news and current events reverberate online matters. For ADL, the spread of conspiracy theories and hate online, often leading to on the ground violence, has been shocking but not surprising. We're seeing what used to be fringe extremist and bigoted narratives become normalized and mainstreamed. Americans have increasingly become radicalized and incited to action by the nonstop drumbeat of online hate and conspiracy theories. This is in large part because of social media's toxic business model that favors growth and engagement over public safety. No industry has ever exercised the sheer power and control over how the world communicates than social media platforms do now.

Social media's amplification of extremism, disinformation, and conspiracy theories—and the complete lack of transparency and accountability about how that amplification takes place—poses one of the greatest threats to democracy in this country, and to the safety of vulnerable individuals and communities worldwide. Hatred spread online has resulted in deadly violence in this country: from Charleston to Charlottesville to Pittsburgh, to Poway and El Paso, we have seen the fatal consequences of white supremacist extremism that often has a clear nexus to social media. We cannot afford to understate the damage from social media's algorithmic amplification of misinformation and hate, and the complete lack of accountability even when platforms aid and abet unlawful activity. We need a bipartisan, "whole of government approach"—indeed, a "whole of society approach"—to interrupt the contagion of hate spread by social media companies in their pursuit of profit. We need to finally hold Big Tech responsible for their role in fracturing democracy and inciting violence. Reform must be smart and effective. It must mitigate the unintended consequence of consolidating even more power, and ability to wreak harm, into the hands of a few impenetrable actors.

These billion-and trillion-dollar social media companies have the resources to improve systems, hire additional staff, develop better products, and provide real transparency. Yet they claim it is too burdensome. We know that's simply not true. Evidence from the tens of thousands of internal Facebook documents leaked by Facebook whistleblower Frances Haugen confirms it. Research from ADL's Center for Technology and Society confirms it. Still, without a clear understanding of what is really going on inside of these companies, we cannot begin to address the real danger posed to the public by letting platforms spread antisemitism, misinformation, and hate twenty-four hours a day, at lighting speed. And without changes to their incentive systems, social media companies will continue to operate under a business model that focuses on generating record profits at the expense of the safety of the public and the security of our republic.

ADL brings unique expertise to the table in the fight against this cycle of hate online. Our Center on Extremism examines the ways extremists across the ideological spectrum exploit the online ecosystem to spread their messages, recruit adherents, finance hate, and commit acts of terrorism. We work directly with threatened communities on the ground as well as law enforcement agencies across the country. This year alone, the Center on Extremism assisted law

enforcement over 1100 times in connection with issues related to violent extremists, and helped several community institutions prevent attacks, both online and off. Our Silicon Valley-based Center for Technology and Society, which has deep policy and technical product expertise, generates research and advocacy-focused solutions to make digital spaces safer and more equitable. CTS engages directly and regularly with major social media platforms to push for policy and product changes; and this has made measurable differences in fighting online extremism. ADL's Education team provides resources for schools and parents to teach children to counter extremist recruitment. And our International Affairs Department monitors how cyberhate in other countries and other languages impacts threatened communities around the world, including vulnerable Jewish communities.

Our expertise in these spaces, presence on the ground in communities across the country, and roots in one of the most targeted communities—combined with more than a century of work fighting against hate and for civil rights—informs ADL's analysis of the online hate and extremism ecosystem and what we can do to combat it.

This testimony will explore how platforms spread hate and extremism. It will show the link between hate-filled extremist content and user engagement, and will explore how and why this content becomes favored by platforms. While we will never eradicate hate and extremism, as this testimony addresses, lawmakers can act meaningfully and significantly to push hate and extremism back to the fringes of the digital world.

II. ADL'S FIGHT AGAINST ONLINE HATE

Since 1913, the mission of ADL (the Anti-Defamation League) has been to "stop the defamation of the Jewish people and to secure justice and fair treatment to all." One of the most important ways in which ADL has fought against bigotry and antisemitism has been by investigating extremist threats across the ideological spectrum, including from white supremacists and other far-right violent extremists, which have posed the biggest domestic terrorism threat to this country over the past decade.

Since its inception over a century ago, ADL has been the leading organization fighting hate. As we have said time and time again, where people go, hate follows—including online. That is why, in the early days of dial-up, ADL anticipated the ways in which hate speech could poison the internet and made certain we were investing our time and resources to communicate to the key players in the industry the need for clear and understandable terms of service on hate speech and encouraged them to enforce these policies aggressively. In 2017, we doubled down on our efforts and launched the Center for Technology and Society (CTS). CTS is a leader in the global fight against online hate and harassment. In a world riddled with antisemitism, bigotry, and extremism, ADL has worked with the tech industry and elected leaders to promote best practices that can effectively address and counter these threats.

ADL has become a leader in fighting online hate and disinformation. CTS acts as a fierce advocate for making digital spaces safe, respectful, and equitable for all people. We have created product interventions to slow down or stop viral hate, and have launched our own Artificial Intelligence tool to measure hate on social media platforms and evaluate policy enforcement. We

are also deeply committed to working in coalition in order to build a more equitable internet. In 2020, ADL teamed up with the NAACP, Color of Change, LULAC, Common Sense Media and other partners to launch Stop Hate for Profit, a campaign targeting Facebook because of the hate, racism, and misinformation reverberating across its platform. Joining the Stop Hate for Profit campaign, over a thousand companies worldwide pulled their advertising spends on the platform for a month, and celebrities and sports stars staged an Instagram walkout.

CTS also plays a unique role among civil society organizations working on fighting online hate in five key areas: policy, research, advocacy, incident response, and product development. It recommends policy and product interventions to elected officials and technology companies to mitigate online hate and harassment; drives advocacy efforts to hold platforms accountable and push hate back to the peripheries of society; produces data-driven applied research by analysts and a network of fellows; sheds new light on the nature and impact of hate and harassment on vulnerable and marginalized communities; brings to market technical tools and products that meet the crucial need for independent data measurement and analysis to track identity-based online hate and harassment; and empowers targets of harassment by responding to online incidents and pushing platforms to create safer online spaces for all. Our combination of technical and policy expertise—and decades of lived experience embedded in a community that has been targeted, often lethally, by bigots and extremists—informs our approach to fighting online hate, protecting targets of online harassment, and holding platforms accountable.

III. PLATFORMS SPREAD HATE AND EXTREMISM

There is no question that the prevalence and impact of online extremism is growing. The spread of QAnon and its consistent elevation of antisemitism, the mainstreaming of the foundational white supremacist "Replacement Theory," #StoptheSteal, and COVID conspiracies all are examples of extremism and hate that have become increasingly normalized and mainstreamed—in large part because of their viral spread online. Last fall, for example, a single "Stop the Steal" Facebook group gained more than 300,000 members within 24 hours. Thousands of new members joined this group by the minute and some of them openly advocated for civil war.

Discovery in civil cases, like the lawsuit against the neo-Nazi and white supremacist organizers of the 2017 Unite the Right rally in Charlottesville, which recently resulted in a $25M+ verdict for nine people injured during the rally, provide still more chilling examples. Extremists' online presence has reverberated across a range of social media platforms. This content is intertwined with hate, white supremacy, racism, antisemitism, and misogyny—all through the lens of extreme ideologies. Such content is enmeshed in conspiracy theories and explodes on platforms that are themselves tuned to spread disinformation.

We need to look no further than the deadly insurrection at our Capitol, which ADL has repeatedly called the most predictable terror incident in American history because it was planned and promoted out in the open on mainstream platforms such as Facebook, Twitter, Instagram, YouTube, and Reddit as well as fringe platforms such as Parler, Gab, 4Chan, and Telegram. As confirmed by leaked internal Facebook documents, the insurrectionists' actions were the product of weeks, months, and years of incitement, spread across the social media ecosystem that

services nearly 300 million people in the U.S. and billions around the world. This was an act of domestic terrorism that was plotted, publicized, recruited for, and financed online.

Finally, ADL considers the dramatic increase in cyberhate in recent years to be one major contributor to the domestic and international spike in antisemitic incidents in the physical world. This was especially the case during the brief war between Israel and the terrorist group Hamas this past May, which had a sharper uptick in on-the-ground antisemitic incidents targeting Jewish communities in America and around the world as compared to previous conflicts in the Middle East—such as in 2014 when Israel and Hamas fought a much more protracted war. The proliferation of antisemitism and other forms of hatred online threatens our communities around the world, including with increased physical attacks.

A. Hate and Extremism on Mainstream Social Media Platforms

Big Tech platforms are not unwitting accomplices or merely tools for bad actors to spread hateful, racist, extremist or conspiracy-related content. On the contrary, Big Tech companies know their platforms' product features are problematic and some have acknowledged it. At a congressional hearing in March 2021, former Twitter CEO Jack Dorsey admitted that his platform had "contributed to the spread of misinformation and the planning of the attack" on the U.S. Capitol on January 6, 2021. In the same hearing, however, Facebook's CEO Mark Zuckerberg disagreed with the assessment that Facebook had profited from the spread of disinformation and touted his platform's efforts to combat it.

Importantly, documents disclosed to the SEC by Facebook whistleblower Frances Haugen—who testified last week before the Communications and Technology Subcommittee for House Energy and Commerce—make clear that Facebook was aware of both the specific role its platform played in the insurrection and the broader role the platform plays in the spread of disinformation, extremism, and hate. The SEC disclosure includes statements from Facebook's internal documents. These documents acknowledged Facebook's role in augmenting "combustible election misinformation," noting "we amplify them and give them broader distribution." Internal Facebook documents also stated that the company had "evidence from a variety of sources that hate speech, divisive political speech, and misinformation on Facebook and the family of apps are affecting societies around the world . . . Our core products' mechanics, such as virality, recommendations, and optimizing for engagement, are a significant part of why these types of speech flourish."

Over the last few years, TikTok—a social media app that allows users to create and share short videos—has also hosted hate and extremism. As ADL's Center on Extremism (COE) documented in August 2020, while much of the content on TikTok is lighthearted and fun, extremists have exploited the platform to share hateful content and recruit new adherents. A recent review of the platform found that antisemitism continues to percolate across the app, including content from known antisemitic figures as well as posts perpetuating age-old antisemitic tropes and conspiracy theories. Earlier this year ADL's CTS released a report that

showed TikTok continues to be far too slow in taking down antisemitism reported by ordinary users and it still has plenty of work to do to ensure that hate is adequately remediated.

B. Gaming Platforms

Online video games share many of the attributes of social media platforms, including spreading hate and extremism. According to the Entertainment Software Association, there are approximately 227 million gamers in the United States. Gaming analytics firm NewZoo's global market report put the gaming industry's revenue at approximately $176 billion globally. With those figures in mind, the importance of addressing hate and extremism in gaming is critical.

ADL's 2021 study of hate, harassment, and positive social experiences in online games explored players' in-game exposure to topics such as extremism and disinformation. Alarmingly, 8 percent of adult gamers (18-45) and 10 percent of teen gamers (13-17) witnessed discussions about white supremacist ideology in online multiplayer games. Seventeen percent of adult gamers saw hateful messaging linking the COVID-19 pandemic to the Asian community, and 13 percent of adult gamers saw hateful anti-immigrant messages spread in online games. The survey also showed that nearly one-in-ten online multiplayer gamers (7 percent) come across Holocaust denial discussions while playing. As we continue to pay deeper attention to the impact social media's algorithms and business models have on hate and extremism, we must consider the way online video games have similar consequences.

C. Cyberhate around the World

Many of these problems are compounded when you broaden the aperture to look outside of America's borders or content in languages other than English. Facebook whistleblower Frances Haugen recently testified that an estimated 87 percent of Facebook's spending to address misinformation was on English-language content when only 9 percent of users are English speakers. For example, ADL recently identified 39 separate Arabic-language Facebook groups or pages with hundreds or thousands followers or likes that had titles specifically aimed at promoting The Protocols of the Elders of Zion, an infamously anti-Jewish hoax. In another recent ADL study, we found 25 Spanish-language antisemitic posts on Facebook that were from groups with a total of 666,728 followers and viewed 55,911 times.

Further, antisemitic content seems to be given even more of a free pass when it comes from foreign leaders or when it is couched in language that makes passing allusion to Israel or Zionists. In light of this, Twitter enables accounts that are attributed to major media outlets for U.S.-designated terrorist groups, such as Hamas, Hezbollah and Palestinian Islamic Jihad, to have a platform before thousands of followers. They use their accounts to glorify terrorism and spread absolutely horrendous antisemitic hate and conspiracies. Ayatollah Ali Khamenei, the Supreme Leader of Iran and the head of the number one state sponsor of antisemitism and terrorism in the world today, has well over a dozen current accounts on Twitter that he uses to

promote Hamas and Hezbollah terrorists and spread hateful antisemitic tropes across a broad array of world languages.

IV. HATE AND EXTREMISM ARE GOOD FOR PLATFORMS' BUSINESS MODELS

Big Tech's fundamental business model, targeted advertising, maximizes profits by keeping its users engaged on the platform for as long as possible to sell as many advertisements as possible. Platforms keep users engaged by optimizing product mechanics like how often and on which posts we click, share, like, and comment—whether in support, opposition, otherwise. Data from tracking user behavior is analyzed to build detailed advertising profiles and find as many opportunities as possible to serve users targeted ads. AI and algorithms, surveillance advertising, subscription models, and other product affordances work together to increase user engagement—positioning these companies as some of the most profitable businesses in the world. What is problematic, however, is that the goal of the platforms, and the algorithms they deploy as a result, is to exploit people's predilection for engaging with incendiary and controversial content and sharing misinformation and divisive material. This user behavior is core to the revenue model.

Hate speech, conspiracy theories, and misinformation—amplified and recommended by platform algorithms—put corrosive and false content at the tops of personalized news feeds, right next to pictures of our families and friends. As mentioned above and detailed below, platforms benefit from the existence and spread of this content because it drives their engagement metrics. It motivates users to spend as much time on the platform as possible, increasing the amount of data that can be extracted about users and, in turn, enabling platforms to serve more and more targeted advertising to users—ultimately increasing revenue. In this way, social media is the most successful extraction industry the world has ever known. When critics say that the existence and viral amplification of hate content and disinformation is a feature, not a bug, of social media platforms, this is what they mean. And until the incentives are changed, we cannot meaningfully mitigate the threat of mainstreamed hate and extremism.

A. Surveillance Advertising and Political Advertising

Like other industries, social media platforms profit from delivering advertisements to users. Tech platforms are distinct from other advertising-based businesses, however, because of the specific and unprecedented way these platforms collect data and target ads. As mentioned above, social media platforms are so successful because they collect and analyze enormous troves of user data based on user activity on the platforms and across the internet. As many experts have noted, Big Tech knows more about us than we know about ourselves.

User data is collected for two key purposes: first, to learn and constantly improve how to keep users engaged on platforms (e.g., viewing and interacting with content) for as long as possible, so that users see as many advertisements as possible; and second, to ensure that the advertisements the platforms deliver are highly targeted to users based on the huge volume of

information that platforms know about each users' behaviors, habits, and preferences (and that of their families and friends). Platforms use this data to develop highly specific advertiser-focused user segments. Then, algorithms deliver ads to specialized demographic segments through personalized content feeds.

While some user data is provided directly by users to platforms (e.g., age and location), social media companies also surveil users to gather extensive information from their profiles (e.g., friends/followers, contacts, connections, groups) as well as their online activity—both on the specific platform and across the internet. Platforms track "likes," shares, comments, navigation paths, hover time, watch time, purchases, and other user engagement actions. Some platforms also collect additional customer data from activities off the platform and combine it with their own data. This practice has been referred to as surveillance advertising, which is described as closely tracking and profiling individuals and groups in detail and then narrowly targeting ads at them based on behavioral history, relationships, and identity and the astounding predictive power that data offers. Surveillance advertising allows platforms to dominate the digital advertising market by offering both big and small businesses an extremely efficient and effective form of advertising—far more than other options such as newspaper or local TV advertising.

Surveillance advertising, which sometimes allows for microtargeting of sliced-and-diced demographic segments, can become even more problematic when used for political and "social issue" advertising. Political advertising often disseminates disinformation and fuels hate by narrowly targeting particular user segments and infuriating and activating them with outrageous, divisive content. Unlike a newspaper or TV spot, where everyone sees the same ad, these campaigns often fall under the radar and are not disclosed to users, researchers, or the public at large. Therefore, no one can meaningfully evaluate the harm of these targeted misinformation campaigns.

Misinformation is considered a key source of political advertising, according to Laura Edelson, ADL Belfer Fellow and PhD candidate in computer science at New York University. Edelson and her team have specifically focused on how misinformation spreads on Facebook. Facebook made promises to be transparent about all of the U.S. political ads on its platform—and about who paid for them. The platform, however, routinely misidentifies political ads and also fails to disclose important information about these ads. Facebook does not have humans overseeing every ad that is published on the platform—even though ads must be submitted for review. Instead, the company uses artificial intelligence (AI), including machine learning (ML) models, and also heavily relies on voluntary compliance. This makes it really easy for bad actors to slip through enforcement gaps. At the same time, biased AI systems can result in over-enforcing against (e.g., removing) legitimate ads. Alarmingly, Edelson and her colleagues have been able to demonstrate that extreme, unreliable news sources get more engagement on Facebook. This, in turn, has the doubled impact of increasing reach and, thus, becoming less expensive to bad-actor advertisers. Edelson and her colleagues also found that the archive of political ads that Facebook makes available to researchers is missing more than 100,000 ads.

As an ADL Belfer Fellow, Edelson is currently working to measure misinformation and hate speech aimed at U.S. Spanish-speaking and Asian American communities by analyzing political advertising on Facebook from the platform's Ad Library and from CrowdTangle, a research and data collection tool owned by Facebook. This research is part of a broader investigation into misinformation in political advertising on Facebook by Edelson and her team at NYU's Cybersecurity for Democracy project.

On August 3, 2021, Facebook placed an enormous obstacle in Edelson and her team's path when it suspended Edelson and her colleagues from accessing its data. The suspension occurred after Edelson and her colleagues started studying whether Facebook was contributing to vaccine hesitancy and sowing distrust in elections, and considered what role the platform may have played leading up to the January 6 insurrection. Facebook cited privacy concerns based, in part, on an FTC order. The federal agency itself disputed this concern publicly. In any case, the careful privacy protocols of the research, which only sought information about advertisers, clearly showed Facebook's ostensible justifications to be pretext. As ADL said at the time, one wonders what Facebook didn't want the public to know.

It's no surprise Facebook attempted to block Edelson's access to data seeking to uncover Facebook's role in the insurrection. According to reports, based on internal documents submitted to the SEC by the Facebook whistleblower, analysis of the January 6 insurrection illustrated that the company was fundamentally unprepared to manage the "Stop the Steal" movement, which turned violent and played a pivotal role in the insurrection. Facebook's own internal analysis found that the policies and procedures put in place were not strong enough to prevent the growth of groups related to "Stop the Steal." The report noted that Facebook treated each piece of "Stop the Steal" content individually, rather than as part of a greater whole. The result of this decision was that only some "Stop the Steal" content or groups were taken off the platform. Much of the content and many of the groups were left up and, ultimately, amplified by Facebook's own algorithms.

On September 28, 2021, Edelson testified before the House Science, Space, and Technology Committee's Investigations and Oversight Subcommittee. At the hearing, titled "The Disinformation Black Box: Researching Social Media Data," she spoke about the harms caused by misinformation on social media and the difficulties researchers face in trying to study this threat to the public. Platforms like Facebook provide independent researchers little access to advertising data, so it is difficult to understand the full impact of political and "social issue" advertising. We need more transparency about Facebook and other platforms' data collection, ad targeting, and algorithmic systems. As discussed in more detail below, proposals like the Social Media DATA Act would ensure academic researchers like Edelson have access to data related to the targeting of online digital advertisements in order to study discrimination, manipulation of youth, election interference, and other consumer harms. Additionally, proposals like the FTC Whistleblower Act of 2021 would ensure whistleblowers can safely disclose wrongdoing at social media companies, including when the companies make dangerous business decisions that harm consumers.

B. AI and Algorithms

AI and machine learning algorithms play a powerful role in the dissemination of extremism and online harm. As referenced in a report co-authored by ADL and other organizations fighting disinformation, "AI can be understood as machines that predict, automate, and optimize tasks in a manner that mimics human intelligence, while [machine learning] algorithms, a subset of AI, use statistics to identify patterns in data." Social media platforms use algorithms, largely fueled by AI and ML systems, to deliver, rank and moderate content, to determine what content should be recommended to a user, and to serve advertisements to users. Algorithms make these highly personalized decisions by collecting and synthesizing vast amounts of user data.

One primary reason algorithms amplify harmful online content on social media is that platforms optimize them for user engagement. They are tuned to keep eyeballs on the screen. Algorithms feed users tailored content based on factors including browsing activity. When a user interacts with a piece of content, algorithmic systems take note of the user's behavior to find and recommend similar content to the user. For example, if someone watches a video about an election, algorithmic systems will recognize that the user may be interested in political content and will continue to recommend related content. If someone has viewed or searches for hateful content, algorithms learn to serve the same user similar or more extreme content.

In addition to personalized recommendations, algorithmic systems focus on what pieces of content are likely to attract a wide range of users. Algorithms do this by recognizing signals—including which pieces of content are watched, shared, commented on, or replied to—and then combining those signals to show that content to more users almost immediately. These algorithms predict if the piece of content will increase engagement, and thus increase advertising revenue. Highly engaging topics such as extremism, hate, and conspiracies are known to generate engagement and thus drive profit. ADL has reported on research that controversial, hateful, and polarizing information and misinformation are often more engaging than other types of content and, therefore, receive wider circulation. Platforms' algorithmic tools significantly boosted extremist content, from white supremacist groups and Holocaust denial to COVID-19 hoaxes and misinformation. Platforms privilege and heavily promote this type of incendiary content to create a stimulus–response loop. In fact, reports of a Facebook researcher who explored how the social media platforms deepened political divides illustrated the speed with which platform algorithms get to work to recommend content rife with misinformation and extremism: less than a week.

The persistent presence and amplification of hate, bigotry, and conspiracy theories on social media platforms has created an environment for extremism to flourish. Today, extremists are enmeshed in online communities where content designed to increase their propensity for hatred and violence often circulates freely. Extremist content boomerangs from fringe websites to mainstream platforms—in part because of social media's immense power, amplification of "engaging" content, and sophisticated recommendation algorithms. However, extremism and hate that start on social media do not always stay there. This content has inspired individuals to commit acts of violence and domestic terrorism.

While an individual who naturally engages with innocuous content (e.g., cat videos, makeup tutorials, or music videos) may not be pushed toward extremist content, individuals who engage with political content, seek to understand conspiracy theories, or have existing gender/racial resentment can quickly become trapped in a negative feedback loop. For example, exposure to videos from extremist or white supremacist channels on YouTube remains disturbingly common. In February 2021, Brendan Nyhan, an ADL Belfer Fellow and professor at Dartmouth College, published a report that collected comprehensive behavioral data measuring YouTube video and recommendation exposure among a diverse group of survey participants. Using browser history and activity data, the report examined exposure to extremist and white supremacist YouTube channels as well as to "alternative" channels that can serve as gateways to more extreme forms of content. Though some high-profile channels were taken down by YouTube before the study period, approximately one in ten participants viewed at least one video from an extremist channel (9.2%) and approximately two in ten (22.1%) viewed at least one video from an alternative channel.

Moreover, the ADL/Nyhan study found that when participants watched the videos, they were more likely to see and follow recommendations to similar videos. Consumption was concentrated among a highly engaged subset of respondents. Among those who watched at least one video of a given type, the mean numbers of videos watched were 64.2 (alternative) and 11.5 (extremist). Moreover, consumption of these videos was most frequent among people with negative racial views.

Currently, platforms have no meaningful incentive to fix problematic algorithms and the public has little understanding of just how dangerous platforms' algorithms can be. As discussed in more detail below, proposals like the Algorithmic Justice and Online Platform Transparency Act, would prohibit discriminatory algorithmic processes, establish a safety standard for algorithms, and require increased transparency from platforms about the types of algorithmic processes they employ and the categories of information they collect to power their AI/ML tools.

*C. **Policymaking due to Public Relations Issues***

As of 2021, almost every major social media platform has a stated public policy prohibiting extremism, terrorism, incitement-to-violence and hate on their platform. For instance, Facebook has a policy prohibiting dangerous individuals and organizations, while Twitter has a policy prohibiting violent organizations. The path to the creation and implementation of these policies, however, was not a direct one. Platforms are too often motivated not by harm prevention but, instead, by avoiding negative public perception. For example, despite repeated demands from ADL and civil society organizations to create a policy prohibiting white nationalism, Facebook only took action to implement a policy prohibiting white nationalist content following public outcry after the 2019 massacre of 51 Muslim people by a white supremacist in Christchurch, New Zealand. This is a pattern that repeats itself over and over again: Big Tech refuses or fails to take action in the face of repeated demands—by civil society, Congress, and their own employees and researchers—then a horrific tragedy occurs, and the companies apologize and pledge to do better. This inspires some action—usually a policy change—but lack of

enforcement, lack of transparency, lack of independent verification, and exceptions to policy changes make platform actions hollow and futile…and the cycle continues. This is the playbook.

For example, in June 2020, after deep frustration with the PR-first focus of policymaking by tech platforms, a number of civil society organizations (ADL, Color of Change, Common Sense, Free Press, LULAC, Mozilla, NAACP, National Hispanic Media Coalition, Sleeping Giants) formed the Stop Hate for Profit Coalition. The coalition called on businesses who ordinarily advertise on Facebook to engage in a month-long advertising pause. Over 1,200 companies joined the July 2020 pause. Additionally, Stop Hate for Profit had a September 2020 week of action, which involved celebrities and influencers calling out hate and extremism on Facebook. Content from the September week of action had an estimated 1 billion views. In January 2021, the Stop Hate for Profit Coalition asked Facebook, Twitter, Google, and other social media platforms to #BanTrumpSaveDemocracy by permanently removing Donald Trump from their platforms.

Policy changes long demanded by civil society around militia activity, the "boogaloo" movement, and Holocaust denial were finally made by Facebook following the Stop Hate for Profit Coalition's public pressure. The campaign's success clearly demonstrates the degree to which policymaking at social media companies is too frequently driven by attempts to shift public perception—only when the companies feel they have absolutely no other choice. Other platforms, also motivated by public pressure, took similar measures in the wake of Stop Hate for Profit. Twitter banned links to hateful content on their platform, which led to the deplatforming of noted white supremacist David Duke. Reddit—which has done a better, if still an incomplete job addressing hate online than many other big platforms—released its first ever hate policy and deplatformed R/TheDonald, a forum of 800,000 users known to house hate and conspiracy theories. YouTube banned six prominent white supremacists, including Stefan Molyneux, David Duke, and Richard Spencer.

Still, social media companies' reactive practices of creating policies for public relations purposes in response to tragic events remained in full effect following the attack on the U.S. Capitol on January 6, 2021. Despite Twitter's July 2020 policy against content related to the hateful QAnon conspiracy, ADL was able to find numerous examples of QAnon on Twitter following the attack on the Capitol. It was only after increased public pressure—in light of the nexus between QAnon and the January 6 attack—that Twitter took more decisive action. After the insurrection, Twitter removed 70,000 QAnon accounts, which greatly reduced the spread of this hateful conspiracy theory on the platform. In fact, ADL found that immediately following the suspension of QAnon-related accounts on January 8, the use of QAnon-related hashtags plummeted by 73 percent.

The actions taken by tech companies—both to update their policies to better prohibit hate and extremism and to better enforce their existing policies to remove such content from their platforms—were helpful but insufficient. The fact that it took such intense public pressure for platforms to create policy and enforcement improvements is unacceptable and, frankly,

dangerous. And we're still seeing lackluster enforcement of important policy changes. For example, ADL recently found that one year after Facebook banned Holocaust denial, the majority of violative posts still accessible were posted prior to the October 2020 ban, yet never subsequently removed. These posts are located in public and private groups as well as on personal profiles, and many contain links to external, explicit Holocaust denial sources. When viewed through the lens of social media companies as primarily optimizing their business models, and generating profit, the justification for public-relations-focused decision making and subpar policy enforcement is clear. This illustrates why self-regulation will never work to solve this pernicious issue. What is needed is the establishment of a set of clear disincentives that will discourage platforms from prioritizing profit over people's safety; put differently, platforms need incentives to make changes that will significantly diminish the amount, and impact, of hate and extremism on their platforms.

*D. **Product Features***

Social media platform policies are only one part of the equation when it comes to mitigating online hate and extremism. Platform products, like groups/pages, ad targeting tools, reporting systems, and other features, often interact to create an environment ripe for bad actors to exploit.

i. **Design Features**

Manipulative design features are one significant way platforms take advantage of consumers. In November, ADL introduced its Social Pattern Library, a collection of design principles and user experience patterns intended to mitigate hateful content on social media platforms developed collaboratively with leading user experience designers. We encourage platforms to consider this living resource as it provides codified product recommendations that will help break the cycle in which hateful content is amplified through algorithms or similar features. Efforts like the Deceptive Experiences to Online Users Reduction (DETOUR) Act recognize the dangers of dark patterns and must be considered when thinking about how to repair our internet.

ii. **Groups and Pages**

"Groups" is one Facebook product feature that may have had innocent origins, but for hate and extremist groups has been foundational to offline violence and domestic terrorism. Facebook claims that it is effectively addressing hate groups on its platforms. ADL and others, however, have continued to expose egregious examples of online hate, misinformation, and extremism across the company's products.

Perhaps most concerning, Facebook algorithms have recommended pages and groups connected to the "boogaloo" conspiracy theory to like-minded users long after the company's assertion last June that it would no longer do so. That assertion was followed by broader statements (in September 2020) that the platform would not recommend groups tied to violence, and an even broader March 2021 statement that Facebook would be ending all recommendations for "civic and political groups, as well as newly created groups." A recent review found that among groups sharing violent memes and a group simply named "Let's Overthrow the Government," Facebook

was recommending groups with names like “The Hawaiian Hootenanny,” “Boogaloonia,” and “The Chaplain of the Redacted.” In addition, after one boogaloo page was “liked,” our investigation’s user received suggestions of other pages with similar content, showing how opportunities are created for users to get further steeped in the ideology.

And Facebook’s own internal reports show that their recommendation systems are powerful ways to drive engagement and that small signals—even as small as a profile showing a woman in a southern state who liked Donald J. Trump and also Fox News, got recommendations for QAnon and other conspiracy groups within 48 hours of creating the profile, even with no other interactions on the site.

iii. Content Moderation and Reporting Systems

Today, most social media companies engage in content moderation to enforce content policies. These systems enforce the policies, sometimes called Community Guidelines or Terms of Service, that determine what content, individuals, and groups are permitted on their services. Beyond having clear and comprehensive policies (which many platforms do not), platforms also communicate with their users about content management decisions. Users deserve to know that platforms will thoughtfully review their reports, especially when reporting hateful, racist, or extremist content, and deserve timely and fair decisions from those systems. Generally, companies rely on a combination of human moderators and AI and ML-based tools to carry out their content moderation efforts, which include flagging, reviewing, making determinations about content and users, and appeals. Additionally, users report violative content to platforms. Importantly, across the industry, it is hard for users to trust that their reports are being addressed.

This year, CTS developed report cards on Holocaust denial and antisemitic platform content to determine the efficacy of platforms’ reporting systems. Report cards have focused on a few different aspects of the reporting process. For ADL’s Holocaust Denial Report Card, we assessed a platform’s response time and whether the platform explained the reason for their decision. One noteworthy finding from this exercise is that platforms with explicit Holocaust denial policies did not necessarily do better enforcing those policies against our reported content, despite years of advocacy from civil society and researchers. Additionally, despite calls for greater transparency, another notable result is the opacity surrounding how platforms reported on the enforcement of their policies. The results of the investigation can be seen in the image below.

PLATFORM	EXPLICIT HOLOCAUST DENIAL POLICY?	GENERAL HATE POLICY?	EFFECTIVE PRODUCT LEVEL EFFORTS TO ADDRESS HOLOCAUST DENIAL?	RESPONSE WITHIN 24 HOURS?	NOTIFICATION OF POLICY REASON FOR ENFORCEMENT?	ACTION TAKEN AGAINST HOLOCAUST DENIAL?	GRADE
Twitch	Yes	Yes	Yes	Yes	No	Yes	B
Twitter	No	Yes	No	Yes	No	Yes	C
YouTube	Yes	Yes	Yes	No	No	No	C
TikTok	Yes	Yes	Yes	No	No	No	C
Roblox	Yes	Yes	Yes	No	No	No	C
Facebook (including Instagram)	Yes	Yes	No	Yes	No	No	D
Discord	No	Yes	No	Yes	No	No	D
Reddit	No	Yes	No	No	No	No	D
Steam	No	Yes	No	No	No	No	D

Note: In creating this framework for evaluating the efforts of digital social platforms, we weighted enforcement more heavily than policy and explicit policies more heavily than general policies. Additionally, because no platform had affirmative results in every category, we did not award any platform an "A."

Image: ADL Holocaust Denial Report Card

For ADL's Antisemitism Report Card, ADL investigators found that no platform performed above a B- in addressing antisemitic content reported to it. Also, no platform provided information or a policy rationale for why it did or did not remove flagged content. Facebook and TikTok got a "D" and "F" respectively when it came to data accessibility. The results of the investigation can be seen in the images below:

Online Antisemitism Report Card

PLATFORMS	HATE POLICY THAT EXPLICITLY MENTIONS RACE, RELIGION, OR ETHNICITY?	RESPONSE WITHIN 24-72 HOURS?	NOTIFICATION OF POLICY REASON FOR ENFORCEMENT?	ACTION TAKEN AGAINST A/S?	TRUSTED FLAGGER PROGRAM?	ACTIONED UPON TRUSTED FLAGGER REPORT?	EFFECTIVE PRODUCT-LEVEL EFFORTS TO ADDRESS ANTISEMITISM?	DATA ACCESSIBILITY GRADE	TOTAL GRADE
Twitter	Yes	Yes	No	No	Yes	Yes	Yes	B	B-
YouTube	Yes	No	No	Yes	Yes	No	Yes	C	B-
Reddit	Yes	No	No	No	No	N/A	Yes	B	C
Twitch	Yes	Yes	No	Yes	No	N/A	No	C	C
TikTok	Yes	No	No	No	Yes	Yes	Yes	F	C-
Facebook (including Instagram)	Yes	No	No	No	Yes	No	No	D	C-
Discord	Yes	Yes	No	No	No	N/A	No	C	C-
Roblox	Yes	No	No	No	No	N/A	No	D	D-

Image: ADL Antisemitism Report Card

Data Accessibility Report Card

PLATFORMS	PUBLIC API FOR PUBLIC CONTENT	RESEARCH/ NGO ACCESS TIER	REPORTING API	RETROACTIVE SEARCH FUNCTIONALITY	STREAMING SEARCH FUNCTIONALITY	RATE LIMITS	QUALITY DOCUMENTATION	GRADE
Twitter	Yes	Yes	No	Yes	Yes	High	Yes	B
Reddit	Yes	Yes	No	Yes	Yes	High	No	B
Discord	Yes	No	No	No	No	High	Yes	C
Twitch	Yes	No	Yes	No	No	High	Yes	C
YouTube	Yes	No	No	Yes	No	Low	Yes	C
Facebook/ Instagram	No	Yes	No	No	No	-	-	D
Roblox	Yes	No	No	Yes	Yes	-	No	D
TikTok	No	No	No	No	No	-	-	F

Image: ADL Antisemitism Report Card

Users deserve more transparency and greater protection from platforms than companies are inclined to provide. Such reluctance has consequences in the form of economic, emotional, mental, political, and physical abuses that affect many people's lives, as repeatedly shown in ADL's research. It is irresponsible at best, and deeply complicit and culpable at worst, for platforms to take, piecemeal approaches that do little to address the rapidity and depth of online hate and harassment.

V. POLICY RECOMMENDATIONS

We need a whole-of-government approach to address the hate and extremism on social media—especially because it can fracture democracy and lead to offline acts of hate-fueled violence. ADL calls for urgent action to prevent and counter domestic violent extremism.

ADL's Repair Plan

ADL has consistently stated that there is no single fix to the phenomenon of online hate. Whether it is in the dark corners of the internet, on the chats used by hundreds of millions of people on online multiplayer games, or a social media post that goes viral, the impact of online hate reverberates both on and offline. This is especially true for those targeted by extremists, who are disproportionately women and members of marginalized communities. The public agrees: according to 2021 ADL data, 77 percent of Americans think new laws are needed to hold social media platforms accountable for recommending that users join extremist groups. ADL's REPAIR Plan presents an integrated agenda to fight hate online and push hate, violence, and extremism back to the fringes of the digital world.

R Regulation and Reform

E Enforcement at Scale

P People Over Profit

A Access to Justice

I Interrupting Disinformation

R Research and Innovation

Congress has an important role in reducing the prevalence, impact and virality of online hate by holding social media companies accountable for their role in fomenting violence, racism, discrimination, and other harms.

Regulation and Reform

Platforms provide the means for transmitting hateful, violent, and abusive content—and, frequently, by more active enabling functions—in inciting violence, polarizing societies, spreading conspiracies, and facilitating discrimination, gender-based violence, and harassment. At the same time, tech companies have no accountability through third-party audits, no transparency requirements, and are almost completely shielded from legal liability due to Section 230 of the Communications Decency Act (CDA 230). Today, there is a complete lack of oversight and independent verification of the claims tech companies make, whether via Congressional testimony, in their transparency reports, or in related communications.

- Congress must **effectively reform, not eliminate, CDA 230** to hold social media platforms accountable for their role in fomenting violence, disinformation, and other forms of hate leading to harm—especially because of Big Tech's algorithmic amplification of dangerous content. Reform, however, must prioritize both civil rights and civil liberties concerns and not result in an overbroad suppression of free speech, nor unintentionally cement the monopolistic power of Big Tech by making it too costly for all but the largest platforms to ward off frivolous lawsuits and trolls. It is important to focus reform on targeted advertisements, egregious harms, and unlawful activity resulting in violence that has been facilitated, or actively abetted, by algorithms. This would be particularly powerful if accompanied by other laws and regulations that focus on anti-discrimination measures, increased transparency, and other means of ensuring accountability.

- Many tech policy experts have focused their efforts on reforming CDA 230 in pursuit of a non-existent one-stop solution. Importantly, this reform is only one essential step in a much larger process. CDA 230 reform will make platforms liable for certain unlawful third-party content. It is unlikely, however, to have much impact on the "lawful but awful" hate that suffuses the internet and is often protected by the First Amendment in the United States. Therefore, policymakers must also **pass laws and undertake approaches that require regular reporting of meaningful metrics, increased transparency, and independent audits regarding content moderation, algorithms, and engagement features** while looking for other incentive-based or regulatory action.

- Additionally, Congress should encourage the Administration to **establish centers of expertise regarding online hate, violence, and severe harassment across agencies**. Within every agency, there should be cross-departmental task forces to help coordinate

the work and support the necessary research, enforcement and plans of action. Agencies should work with Congress to develop research grant programs to comprehensively assess the links between Big Tech business models and online hate and build a more detailed knowledge base of the industry role in online harms.

- In an absence of transparency and oversight, online spaces have been toxic for young teenagers and a breeding ground for extremism. Proposals such as the KIDS Act would increase protections for young people (under 16) who are exposed to manipulative marketing, amplification of harmful content, and damaging design features. The Social Media DATA Act and the Algorithmic Justice and Online Platform Transparency Act would provide the public with a deeper understanding of the parameters and impact of platforms' algorithms, paid content, and other information crucial to fighting online discrimination and extremism.

Enforcement at Scale

When something goes wrong on a major social media platform, tech companies blame scale and plead impotence. The fact that millions, even billions of pieces of content can be uploaded all over the world, shared, viewed, and commented upon by millions of viewers in a matter of seconds serves as the justification for "mistakes" in content moderation—even if those mistakes result in violence and death. But scale is not the problem here; defective policies, bad products, and subpar enforcement are the root of Big Tech's lackluster enforcement.

Today, most major social media companies only publish limited information about their content policies and enforcement. Reports end up serving as a deflection away from the truth about what content proliferates on platforms. Recent revelations from leaked documents show that this isn't just a theory—organizations like Facebook had an unannounced program called XCheck that allowed over 5 million celebrities, politicians, and influencers to effectively skirt all of the published rules and policies. That meant that none of the posts by millions of public figures went through automated systems that normally flag violating content. The most influential accounts got a free pass for posting almost anything they wanted whenever they wanted. We only know about this because of a whistleblower. None of this was disclosed in a single platform report. This is why we need increased protection for whistleblowers. Proposals like the FTC Whistleblower Act of 2021 would protect whistleblowers who disclose wrongdoing at their current and former employers for issues under FTC jurisdiction.

Social media companies have little to no legal or financial incentives to give consumers comprehensive information. There is a strong need for systematized, regulated, and easily accessible transparency efforts from social media platforms. Platforms claim to have strong policies against hate, violence, and extremism, when in fact, most are unclear, hard to find, or have perplexing exceptions.

- Transparency reform would motivate platforms to be more explicit about their policies on hate, harassment, and misinformation, and apply their rules consistently. It would act as a

deterrent from making changes, exceptions, or other decisions that end up amplifying hate. It would create an environment where social media companies can compete on how well they are protecting users, not on how they can optimize the most corrosive content to keep us scrolling for as long as possible to sell as many ads as possible. Proposals like the transparency requirements offered in California's AB 587, which would require large social media companies to report on their content management policies and enforcement behaviors on a quarterly basis, can provide necessary information on how platforms are determining, codifying, and implementing their policies.

- Transparency reports must evaluate success and provide evidence that independent researchers can use; such independent researchers must be granted uninterrupted access to data, and Congress must continue an oversight role. Companies can and should increase transparency related to their products. At present, technology companies have little to no transparency in terms of how they build, improve, and fix the products embedded into their platforms to address hate and harassment. In addition to transparency reports, technology companies should allow third-party audits of their work on content moderation on their platforms. Audits would also allow the public to verify that the company followed through on its public commitments and to assess the effectiveness of company efforts across time.

- **Platforms must mitigate harm to consumers through products, designs, algorithms, and policies that further discrimination, bias, and hate.** Platforms should ensure that their design, user agreements, and policies counter the potential for bias-based discrimination and civil rights violations on the platform. To do this, platforms must regularly evaluate the way product features and policy enforcement fuel discrimination, bias, and hate and make product/policy improvements based on these evaluations. Platforms need an understanding of which populations are targeted or impacted most egregiously and why, the nature of hate content, and the path of spread; tech companies should create and maintain diverse teams to mitigate bias when designing consumer products and services, drafting policies, and making content moderation decisions. Proposals like the Deceptive Experiences to Online Users Reduction (DETOUR) Act would ensure platforms are not designing, modifying, or manipulating a user interface in a way that impairs users from making educated decisions before consenting and giving companies access to their personal data. Importantly, this proposal would also make it illegal to segment consumers of online services for the purposes of behavioral or psychological experiments without informed consent.

- A whole-of-government approach means that a wide range of legislative and regulatory bodies must exercise oversight to ensure tech companies adopt and consistently enforce policies and community guidelines designed to identify and combat violence, hate, and harassment. While there is not likely to be a one-size-fits-all set of guidelines or enforcement, incentives for effective standards and guidelines, transparency regarding them and their impact, and independent research evaluating these efforts can be imposed or supported by the government. **The FTC, State AGs, and other enforcement**

authorities also should increase consumer protection efforts, especially when tech companies engage in unfair and deceptive practices.

People Over Profit

The rapid and massive spread of extremism and hate on social media is a product feature, not a bug. Inflammatory mis- and disinformation and hate content generates growth and greater user engagement. Many tech company algorithms are wired to optimize for user engagement because the companies' business models are built around growing users and keeping people on the platform for as long as possible, to see as many ads as possible, because that is what generates revenue. As many former and current Big Tech employees have acknowledged, platforms like Facebook build and employ algorithms designed to promote engagement, thus inevitably amplifying the most corrosive content.

- **Platforms need to adjust their algorithms** and stop recommending or otherwise amplifying organizations or content from groups associated with extremism, hate, misinformation, or conspiracies to users—even if it results in less engagement from users. Platforms must invest in both AI improvements and adequately trained and resourced human content moderators—with training focused on particular cultural contexts and languages. The Algorithmic Justice and Online Platform Transparency Act and KIDS Act should be carefully considered as we look to accomplish goals of making algorithms safer for all users—especially young people.

- Platforms **also must put more resources toward protecting victims and targets of online harassment**, countering disinformation, and improving content moderation instead of prioritizing the bottom line. Platforms should provide effective, expeditious resources and redress for victims of hate and harassment. For example, users should be allowed to flag multiple pieces of content within one report instead of creating a new report for each piece of content. They should be able to block multiple perpetrators of online harassment at once instead of undergoing the laborious process of blocking them individually. Preventing users who repeatedly engage in hate and harassment from accessing a platform even if they create a new profile, known as IP blocking, helps protect victims.

- Platforms also need to pay more attention to online hate and misinformation in languages other than English. They need to invest in human content moderators who are well-trained in all of the major languages that their platforms service, and also need to devote major resources to improving AI detection of violative content in languages other than English. Leaders of state sponsors of terrorism shouldn't be given a free pass to incite Jew-hatred or glorify terrorism simply because they post in a language other than English.

- We urge Congress to focus on how consumers—and advertisers—are impacted by a business model that optimizes for engagement. Congress must focus on how both algorithmic amplification and monopolistic power can fuel hate. **They should ensure**

algorithms are ethical and fair and consider regulating surveillance advertising and increasing data privacy, so companies cannot exploit consumers' data for profit—a practice that inevitably results in greater online hate.

Access to Justice

A safer internet starts with protecting targets of harassment, not perpetrators. This means changing laws, policies, and practices that currently deny victims meaningful access to the courts and other effective avenues of redress. When tech platforms host harassing content and enable perpetrators to abuse their targets, victims of extremist violence, gender-based violence, hate, and harassment have no place to go in the face of physical threats, emotional injury, and financial and reputational harm. Victims and targets have been denied access to justice because our cyberharassment laws are outdated or don't exist at all.

According to ADL's latest data, 1 in 3 Americans who are harassed online attribute the harassment in whole or in part to their identity, referring to race, religion, gender, sexual orientation, gender identity, ethnicity, ability, and the like. More specifically, women experienced harassment disproportionately, as 35 percent of female-identified respondents felt they were targeted because of their gender. This abuse also happens in online games spaces. According to ADL's recent online gaming survey exploring the social interactions, experiences, attitudes, and behaviors of online multiplayer gamers nationwide, for the third year in a row gender was the most frequently cited reason for abuse.

Harassment intrudes into users' lives and hampers their ability to communicate, unfairly impacting marginalized communities' ability to work, socialize, learn, and express themselves online.

- We urge Congress and executive agencies to provide more resources and pressure agencies to pursue investigations and enforcement actions of bias-based cyberstalking, doxing, and swatting. Also, **Congress should update gaps and loopholes in cyber harassment laws** and the reporting of bias-based digital abuse in order to better protect victims and targets, including enacting legislation related to doxing, swatting, and non-consensual distribution of intimate imagery. One way to achieve this is by improving and passing the Online Safety Modernization Act at the federal level while also encouraging states to pass anti-cyberharassment legislation.

- According to ADL's ethnographic study of online hate and harassment, "some of the most widely reported incidents of campaign harassment (the ability of harassers to use online networks to organize campaigns of hate) and networked harassment (the weaponization of a target's online network) have been waged against women and the LGBTQ+ community." Victims and targets of cyberhate need more resources and support. Congress and the Administration should work together to create a resource center to support targets of identity-based online harassment. This center could provide tools to victims and targets seeking to communicate with social media platforms, report unlawful behavior to law enforcement, and receive extra care. Additionally, creating a hotline for victims and targets of cyberhate and harassment and requiring the platforms to

regularly report on the quantity and types of hate and harassment reported and actioned can help us tackle this issue.

Interrupting Disinformation

Hatemongers and extremists spread disinformation to harm targets and terrorize vulnerable communities; they amplify conspiracy theories to advance political aims; radicalize followers; and incite violence either intentionally as a tool to meet their goal or as a predictable outcome. Their content becomes further normalized when influential people, including high-level officeholders, spread this content further, often claiming that they are only "passing on" information they did not create for their followers to "evaluate." Hatemongers and extremists find ways to engage on mainstream social media platforms (Twitter, Facebook, YouTube), fringe platforms (Parler, Telegram, 4chan/8kun) and the Dark Web (Gab, DLive, america.win). It is a vicious cycle: this extraordinary spread is both made possible by, and helps further increase, the profound distrust of government and institutions.

The mainstreaming and normalization of hateful and extremist beliefs (including virulently misogynist, antisemitic, and racist conspiracy theories) is the foundation of much of the disinformation proliferating online. This is made evident by the fact that millions of Americans believe in QAnon conspiracies and other extremist ideologies.

Interrupting disinformation and finding/encouraging off-ramps and effective mitigation strategies to counter radicalization is no longer a marginal issue. It now requires a whole-of-government and society approach. There is a clear connection between online extremist, antisemitic, misogynist, racist, and hateful images and tropes reverberating on social media and offline hate and violence directed at marginalized communities. Further, the deadly insurrection at the United States Capitol is a key example of the violence that can erupt when extremist disinformation spreads on social media.

- The continuing spread of baseless and dangerous conspiracy theories will continue to find fertile ground. Social media algorithms recommend content to extremist-leaning users, including related groups and pages that contain harmful content. Government must join with civil society and industry to find ways to undermine, interrupt, and mitigate disinformation without undermining civil rights and liberties. **Congress should fund research on the impact of social media platforms' recommendation systems and algorithmic amplification mechanisms** on the intersection between algorithmic amplification of disinformation, misogyny, and gender-based violence.

- Congress and the executive branch must provide resources to civil society organizations working to counter online disinformation. We strongly urge you to **support widespread media literacy, digital literacy, and anti-disinformation education.** Congress should investigate the nature and impact of product designs that allow hatemongers and extremists to exploit digital social platforms and spread antidemocratic, violent, and hate-based disinformation. It should also support concerted research to identify new ways of countering dangerous disinformation that leads to violence—especially gender-based

violence. It is simultaneously vital not to abuse this imperative to surveil vulnerable communities or to crack down on its non-violent critics and adversaries.

Research and Innovation

Government actors, civil society, and the tech sector must stay ahead of the curve as emerging threats will inevitably contribute to the impact of online hate. There must be a concerted effort to focus on technology research and innovation aimed at combating online hate. Just as privacy-by-design has been promoted, with some notable success, "anti-hate by design" must be promoted and widely incorporated into social media platforms and made a fundamental consumer expectation.

Government actors and platforms must focus on research and innovation to slow the spread of online hate, including, but not limited to: (1) measurement of online hate; (2) the extent of sexism, hate and extremism in online games; (3) methods of off-ramping vulnerable individuals who may be going down a path to commit extremist and gender-based violence; (4) the connection between online hate speech and hate crimes; (5) new methods of disinformation; (6) the role of internet infrastructure providers and online funding sources in supporting and facilitating the spread of hate and extremism; (7) the role of monopolistic power in spreading online hate; and (8) audio content moderation. Congress can play a key role in this innovation to invest in improving our understanding of how hate impacts communities. Those community members are also individuals who have the most credibility in communicating with friends, family, etc. to prevent hate from taking root. Congress can invest in prevention, community engagement, and other tools to better understand how communities are dealing with the challenge.

CONCLUSION

Thank you for the opportunity to testify before this body and for calling a hearing on this urgent topic. It is long past time to acknowledge the threats social media platforms fuel and the need for increased accountability. These companies will tell you that it's too hard to address hate, extremism, and racism on the internet. They will claim the legal framework will prevent us from regulating their platforms. That is simply untrue. Time and time again, lawmakers have crafted good policies to protect consumers and industry alike—from regulation for automobiles, food, prescription drugs, and securities. There is a lot we can and must do to push hate and extremism back to the fringes of the digital world. We must address these threats holistically rather than piecemeal. This is precisely what ADL's **REPAIR** plans do, applying a whole-of-government and whole-of-society approach to fight online hate and extremism. On behalf of ADL, we look forward to working with you as you continue to devote your attention to this critical issue.