Congressional Testimony

Social Media Platforms and the Amplification of Domestic Extremism & Other Harmful Content

David L. Sifry

Vice President, Center for Technology and Society

ADL (Anti-Defamation League)

SENATE HOMELAND SECURITY
AND GOVERNMENTAL AFFAIRS COMMITTEE

Washington, DC

October 28, 2021

10:15 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. Anyone who has been paying attention to extremist activity across the country knows that online hate and extremism are amplified, reinforced, and spread by the chorus of disinformation and hatred that runs rampant across social media. New adherents to extremism are recruited, fed a radicalizing diet of conspiracy theories, and connected to others who share those views. Most recently—as underscored by the recent explosive proof provided by leaked, internal Facebook documents—we've seen horrifying evidence of how the influence of social media spurs extremist activity, terror, and anti-democratic offline violence in this country and around the globe.

For ADL, the spread of conspiracy theories and on-the-ground domestic terrorism has been shocking but not surprising. For example, what the Facebook leaks have provided is proof of what so many of us knew to be the case—that the platform and its top executives made intentional choices to allow and even spur online harm (with offline consequences) in service of growth and revenue. The difference between what these executives say and what they do is revealed in all its chilling hypocrisy. What is clear now is that companies like Facebook are not only prioritizing profit—they are doing so at the expense of our safety, security, and democracy because it is good for their bottom line. In the end, that seems to be the most important principle for the few individuals who run the largest, most powerful, and most unaccountable communications, news, entertainment, and surveillance system the world has ever known.

Social media's amplification of extremism, disinformation and conspiracy theories is 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 terrorism 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 minimize the threat of social media's algorithmic amplification of extremism and hate. We need a bipartisan "whole of government approach"—indeed, a "whole of society" approach—to interrupt domestic extremism and harmful content amplified by social media companies in their pursuit of profit.

ADL brings unique expertise to the table in the fight against online hate. Our Center on Extremism (COE) examines the ways extremists and white supremacists exploit digital affordances to spread their messages, recruit adherents and commit acts of terrorism. Our Center for Technology and Society, which has deep policy and technical product expertise, generates advocacy-focused solutions to make digital spaces safer and more equitable. Our proficiency in these spaces, in addition to more than a century of work to fight 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, why hate-filled and extreme content is favored by platforms such that the entire enterprise is engineered and operated for its expansion—because it drives profit—and the ways in which online extremism can lead to offline violence. Finally, ADL lays out several policy recommendations for lawmakers and the technology sector to fight online hate and extremism meaningfully and significantly.

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." For decades, 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 white supremacists and other far-right violent extremists, producing research to inform the public on the scope of the threat, and working with law enforcement, educators, the tech industry, and elected leaders to promote best practices that can effectively address and counter these threats. As ADL has said time and time again, where people go, hate follows—including online.

ADL has invested resources and become a leader in fighting online hate since we launched the Center for Technology and Society (CTS) in 2017. CTS is a leader in the global fight against online hate and harassment. In a world riddled with antisemitism, bigotry, extremism, and disinformation, CTS acts as a fierce advocate for making digital spaces safe, respectful, and equitable for all people. CTS also plays a unique role among civil society organizations working on fighting online hate. It brings to bear decades of lived experience rooted in a community that has been targeted, often lethally, by bigots and extremists and leverages ADL's decades of expertise in tracking and fighting extremism and antisemitism.

One of the signature differentiators of CTS is the fact that it works 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 fringes 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.

In our direct engagement with platforms, CTS has emphasized the need for them to adopt anti-hate-by-design principles. This concept was first popularized in the area of privacy (known as privacy-by-design) but can and should be applied to building less hate-filled platforms. Our recommendations include several steps that would help inculcate a culture of anti-hate-by-design to can be implemented across social media company systems, policies, and product developments.

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 and neo-Nazi "Replacement Theory," the Big Lie about the 2020 presidential election, and COVID conspiracies, all are examples of extremism and hate that has become increasingly normalized and mainstreamed—in large part because of its viral spread online. Leaked documents from the Facebook whistleblower show this trend, our own research and that

of others confirm it, criminal prosecutions demonstrate it, and government and news investigations continue to provide a firehose of evidence.

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 began this week, 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.

A. Mainstream Social Media Platforms

Fringe platforms, despite having relatively small userbases, make use of Big Tech platforms like Twitter and Facebook to increase their reach and influence. Since Twitter's 320 million and Facebook's 2.85 billion users dwarf the hundreds of thousands of users on fringe sites, extremists leverage these mainstream platforms to ensure that the hateful philosophies which often began to germinate on message boards like Gab and 8chan (now 8kun) find a new and much larger audience. Mainstream platforms serve as a gateway for extremists to recruit curious individuals. Extremists use strategies like creating private pages and events; using coded language (called dog whistles) to imply and spread a hateful ideology on mainstream platforms; and linking to hate-filled sites to avoid content moderation.

Facebook and Twitter generally allow users to link to pages on fringe hate-filled sites, which allows visitors to mainstream sites to get to highly problematic content with little to no effort. ADL's COE found in an October 2021 study that despite Twitter's ban on external links to hate speech, extremist material and conspiracy theories, this content is frequently shared on Twitter via links from the far-right "free speech network" Gab. More than 112,000 tweets were posted containing links to Gab content between June 7 and August 22, which included antisemitism, misinformation relating to COVID-19 or the vaccines, and content promoting QAnon.

Big Tech platforms are not unwitting accomplices or merely tools for extremists to link to fringe platform content. On the contrary, platforms' algorithms amplify misinformation, extremist, and white supremacist content; connect adherents; and host and recommend anti-democratic, extremist and hate-focused groups and events. For example, last fall 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.

Big tech companies know their platforms' product features are problematic. At a congressional hearing in March 2021, Twitter Chief Executive Officer 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, 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, however, documents disclosed to the SEC by Facebook whistleblower Frances Haugen make it 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 stated 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 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. It should be noted that when alerted to the content that ADL found, TikTok took down the specific content, but they are still woefully inadequate when handling reports from ordinary users. While we appreciate their removing the specific content and their stated commitment to a zero-tolerance policy on antisemitism and hate, we are concerned that it took our notification to do so and urge them to systematically address this serious issue. Earlier this year ADL's CTS released a report that showed TikTok is still 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.

Recordings of Louis Farrakhan, Rick Wiles (founder of TruNews), and Stephen Anderson—all antisemitic individuals whose bigotry has been thoroughly documented by ADL—were readily available on TikTok in 2021. One such post, shared on May 23, 2021, showed a clip of a TruNews segment in which Rick Wiles states: "And our leaders are lowlife scum that screw little girls so the Jews can screw America...we've allowed Kabbalah practicing Jews to defile the nation." TruNews—a fundamentalist Christian streaming news and opinion platform that produces antisemitic, anti-Zionist, anti-LGBTQ+, and Islamophobic content—has been banned from YouTube and Facebook for violating the platforms' content rules.

B. Gaming Platforms

Online video games share many of the attributes of social media platforms. Games spread hate and extremism and operate at a comparable scale to social media platforms. 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 model have on domestic terrorism and extremism, we must consider the way online video games have similar consequences.

IV. DOMESTIC TERRORISM AND EXTREMISM ARE GOOD FOR PLATFORMS' BUSINESS MODELS

Big Tech's fundamental business model—targeted advertising—maximizes profits by optimizing the product mechanics of the platform to increase user engagement. AI and algorithms, surveillance advertising, subscription models and 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 social media companies have created incentive structures that employ AI and algorithms, surveillance advertising, subscription models and product affordances to exploit people's predilection for clicking on incendiary content and sharing misinformation and divisive material.

Hate speech, conspiracy theories, and misinformation—amplified and recommended by algorithms—put corrosive and false content at the tops of personalized news feeds. Platforms benefit from the existence and spread of this content because it drives their engagement metrics by motivating users to spend as much time on the platform as possible, to increase the amount of data that can be extracted about users and, in turn, enable 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.

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

way these platforms collect data and target ads. As mentioned above, social media platforms are so successful because they collect and analyze troves of user data, based on user activity on the platforms, and across the internet. This user data is collected for two key purposes: first, to keep users engaged on platforms (e.g., viewing and interacting with content) for as long as possible, so users see as many advertisements as possible; and second, to deliver highly targeted advertisements to users based on what platforms know about each users' behaviors, habits, and preferences. 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 are 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 platform and across the internet. Platforms track "likes," shares, navigation paths, hover time, watch time, and other user engagement actions. Some platforms collect additional customer data from activities off the platform. This practice has been referred to as surveillance advertising: closely tracking and profiling individuals and groups in detail and then narrowly targeting ads at them based on behavioral history, relationships, and identity. 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.

One key problem with surveillance advertising is that dominant digital advertisers (namely, Facebook, which owns Instagram, and Google, which owns YouTube) curate the content each person sees on their platforms using the data collected. The goal of surveillance advertising is to keep users engaged, to serve them more ads and mine them for more data. Big Tech platforms amplify extremism, hate, and conspiracies because they know that this content generates the most engagement and, therefore, the most profit. As discussed in more detail below, platforms' algorithmic tools have significantly boosted extremist content, from white supremacist groups and Holocaust denialism to COVID-19 hoaxes and misinformation.

Surveillance advertising, which sometimes allows for microtargeting of 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 incensing them with outrageous, divisive content. For example, prior to Twitter banning political ads in October 2019, ADL Belfer Fellow Sam Woolley—an assistant professor at the University of Texas—conducted a study of computational propaganda, Jewish Americans, and the 2018 elections.

Woolley conducted interviews and analyzed Twitter data in order to understand both the scope of the issue on a national scale and the repercussions faced on the individual level. For the interviews, Woolley spoke to five Jewish Americans involved in politics as elected officials, policymakers, journalists, political consultants, and commentators. Woolley found that political

advertising on platforms was susceptible to being gamed by bots and taken advantage of by anonymous groups. Interviewees said tech companies seemed reluctant to remove bot-driven and harassing content and posited that the companies' reluctance came from not wanting to affect user growth—a metric used to determine company value. Alongside interviews, Woolley collected and analyzed 7,512,594 tweets related to U.S. politics from August 31, 2018 to September 17, 2018. The collected tweets showed the prevalence of political bots and highlighted political groups within the U.S. political spectrum most involved in antisemitic attacks.

Political advertising has also been considered a key source of misinformation, 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 has made promises to be transparent about all of the U.S. political ads on its platform—and about who paid for them. However, it routinely misidentifies political ads and also fails to disclose important information about them. 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 a combination of artificial intelligence (AI) and machine learning (ML) models—and it also heavily relies on voluntary compliance, making it easy for bad actors to slip through enforcement gaps and also over-enforcing (and removing) legitimate ads.

Alarmingly, Edelson and her team have been able to demonstrate that extreme, unreliable news sources get more engagement on Facebook, and that the archive of political ads that Facebook makes available to researchers is missing more than 100,000 ads.

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. Notably, on August 3, 2021, after Edelson and her team started studies intended to determine whether Facebook was contributing to vaccine hesitancy and sowing distrust in elections—as well as trying to determine the role the platform may have played leading up to the January 6 insurrection—they were suspended by Facebook from accessing its data.

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 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 but 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," Edelson 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.

B. AI and Algorithms

AI and 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 forwarded, commented on, or replied to and then combine those signals to almost immediately, show that content to more users. These algorithms predict if the piece of content will increase engagement, and thus increase advertising revenue. 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 privilege this incendiary content, creating 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.

In a forthcoming peer reviewed study, Laura Edelson and a team of academic researchers consider how Facebook users interact with unreliable and partisan news sources. The team found that posts from sources known for misinformation are *six times* more likely to get engagement than factual ones. Notably, most of the misinforming content was generated by far-right sources.

In fact, in Edelson's findings, far-right engagement with misinformation made up 68 percent of total engagement. A much smaller share of publishers in other partisan categories were purveyors of misinformation. On the far right, 109 misinformation publishers accounted for almost *1.2 billion interactions*, which was more than twice the total engagement that the 154 non-misinformation news sources garnered. These findings confirm that a small number of misinformation publishers have outsized influence, generating far more interactions and audience reach than factual sources.

Extremist groups are empowered by the existence of powerful algorithms that amplify the hateful voices of a few to reach millions around the world. The persistent presence and amplification of hate, bigotry, and conspiracy theories on social media platforms has created an environment for extremism to flourish. This content, in turn, inspires individuals to commit acts of violence and domestic terrorism. While an individual who naturally engages in 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.

In another example, exposure to videos from extremist or white supremacist channels on YouTube remains disturbingly common. In January 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 study found that when participants watch 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.

Algorithmic amplification of divisive and hateful content by Facebook, YouTube and other big tech platforms creates an environment prone to inspire those curious about extremism. ADL has reported on the clear connection between online antisemitic, extremist, and hateful images and tropes reverberating on social media and offline hate and violence directed at marginalized communities. In the United States, calls to violence in the name of white supremacy and "The Great Replacement" theory, which has proliferated online and been amplified through algorithms, correlate to fatal shootings in Poway, El Paso, Pittsburgh and more, and led to the injuries and deaths at the white supremacist attacks in Charlottesville in 2017 and on the United States Capitol on Jan 6, 2021.

*C. **Revenue Sharing and Monetization***

Revenue sharing models on social media platforms, such as subscription services and direct donations for livestreaming, allow extremist content creators to monetize the spread of hate. These revenue sharing models are designed for influencers and celebrities to earn income from the content they generate but are exploited by extremists and domestic terrorists as new sources of fundraising. Because mainstream platforms like YouTube sometimes attempt to remove violent extremist content, extremists also use niche platforms with permissive content policies, such as DLive, a video-sharing platform that makes financial transactions publicly visible, and BitChute, another video-sharing platform favored by extremists. Even if extremists are suspended from big tech platforms, they often promote their fringe channels on mainstream social media. For example, well-known antisemites E. Michael Jones (800,000 views) and Brother Nathanael (970,000 views) have been banned from YouTube, but actively promote their BitChute channels on Facebook.

ADL Belfer Fellow Dr. Megan Squire, professor of computer science at Elon University, researches monetization and de-platforming (that is, restricting or removing creators) among far-right extremists and domestic terrorists. In a 2021 study, Squire analyzed extremist monetization strategies on DLive. Squire found that a small number of "megadonors" disproportionately fund extremist content creators. These megadonors spend large amounts of money financing their favorite streamers and gain their own visibility and notoriety by doing so. Squire also analyzed content creators like Nicholas Fuentes, a well-known white supremacist who participated in the 2020 "Stop the Steal" campaign as well as the January 6 U.S. Capitol insurrection. Fuentes shrewdly optimizes his donations through his reliable livestreaming schedule.

These new forms of revenue sharing allow extremist content creators to monetize their propaganda, especially livestreamed audio content, which is more difficult to detect and remove quickly. On DLive, according to Dr. Squire's study, far-right actors can earn over $100,000 in donations in less than a year through a combination of megadonors and small donors. Extremist groups, such as the "Groypers" of America First and the Proud Boys, also earn money through revenue-sharing models. Platforms like DLive make it easy for creators to cash out funds, making it a reliable income stream for extremists.

*D. **Policies and Policy Enforcement***

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 public perception. For example, despite repeated urging 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.

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.

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 public perception. 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 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.

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. 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, 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 enforce their existing policies to remove such content from their platforms—were laudable. Ultimately, however, the fact that it took such intense public pressure for them to create policy and enforce improvements is unacceptable and, frankly, dangerous. When viewed through the lens of these companies, as primarily optimizing their business models and generating profit, these behaviors come into a much clearer focus. It's also why it is so clear that self-regulation will never work to solve this pernicious issue. What is needed is the establishment of a set of clear disincentives when platforms prioritize profit over people's safety.

E. ***Product Features***

Social media platform policies are only one part of the equation when it comes to mitigating online hate and extremism. Platform product features, like groups/pages, reporting and content moderation systems, often interact to create an environment ripe for extremists and domestic terrorists to exploit.

i. 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.

Facebook amplifying and recommending extremist groups like Boogaloo has led directly to offline violence. For example, in May 2020, Dave Patrick Underwood, a Federal Protective Services Officer, was killed in a drive-by shooting carried out by two Boogaloo adherents who were connected through Facebook and discussed the idea to commit the crime in a Facebook group. These assailants had never met prior to being connected on Facebook. This was one of many extremist-related shootouts ADL's COE tracked in 2020.

In June 2020, Facebook announced that it had taken down hundreds of groups and pages on its platform associated with the violent anti-government boogaloo movement, one of several major purges of extremist material by Facebook that year to address extremists' use of its platform. In recent months, however, several new boogaloo pages have emerged on Facebook, hiding among libertarian groups and pages that also share memes advocating for violence. One of the ways these groups have been able to remain on the platform is by using unconventional naming structures for their pages (such as "Char Broil Tru infrared grilling"). Though these new Facebook boogaloo groups typically are far smaller and produce less content than their predecessors did in 2019-20, the emergence of such pages highlights the need for Facebook to take a proactive stance to ensure that boogalooers do not successfully reestablish groups on the platform.

Perhaps most concerning, Facebook algorithms appear to be recommending these boogaloo pages and groups to like-minded users, despite 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.

Clearly, Facebook's recommendation algorithms, filters and other detection methods for boogaloo groups and pages need upgrading. COE has found that even Facebook pages not directly associated with extremist groups are rife with violence. The public agrees: according to 2021 ADL data, 77 percent of Americans think laws need to be made to hold social media platforms accountable for recommending that users join extremist groups. 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.

ii. 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, and making determinations about content. 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, ADL's Center for Technology and Society 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 first assessed a platform's response time, asking whether the platform investigated the report and promptly responded to the user. We did not include auto-generated messages to the user affirming receipt of a report. Instead, ADL considered whether any follow-up messages indicated that the platform investigated and made a content moderation decision. Second, ADL assessed whether the platform explained the reason for their decision, recording whether a user was notified about why a platform made a certain decision based on its stated policies. 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, we established six categories to evaluate how well platforms responded to user reports. Four of these six categories focused on platforms' responses to reports from an ordinary user: Metric #1: Does the policy explicitly mention religion, race, or ethnicity? Metric #2: Did the platform respond within 24 to 72 hours? Metric #3: Was the user notified regarding whether the content they reported violated/did not violate a specific platform policy? Metric #4: Was the content removed or otherwise actioned as a result of the report? Two categories focused on platforms' responses to reports from trusted flaggers: Metric #5: Does the platform have a trusted flagger program? Metric #6: Did the platform take action on content reported through its trusted flagger program?

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. 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 for platforms to take, at best, piecemeal approaches that do little to address the rapidity and depth of online hate and harassment. The results of the investigation can be seen in the image 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

V. EXTREMISM ON SOCIAL MEDIA IS A DOMESTIC TERRORISM THREAT

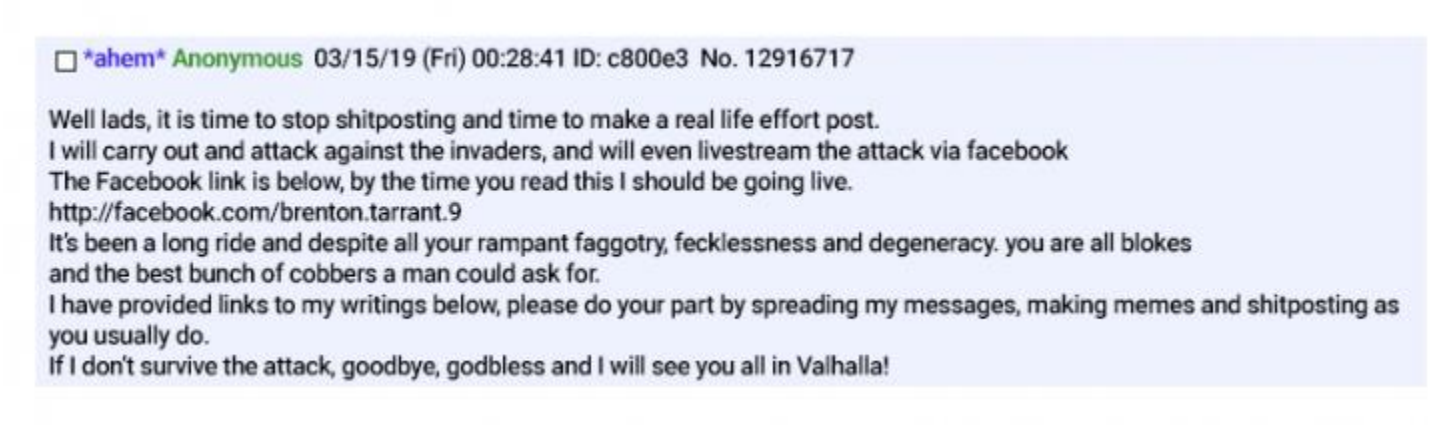
Today, extremists are enmeshed in online communities where content designed to increase their propensity for hatred and violence often circulates freely. As noted above, extremist content boomerangs from fringe websites to mainstream platforms—in part because of social media's immense power, amplification of "engaging" content, and sophisticated recommendation algorithms. However, extremism and hate that start on social media do not always stay there.

A. Examining Social Media's Role in Extremist Massacres

ADL COE research fellow Joel Finkelstein conducted an in-depth examination of genocidal language and conspiracy theories pervasive on fringe platforms. Comparing the online behavior of the perpetrators of the Pittsburgh and Christchurch massacres (Robert Bowers and Brenton Tarrant) suggested that the two killers had similar ideological motivations and were subject to similar radicalization methods. Both killers announced on fringe platforms that they were about to commit violence and seemed to identify their fellow forum participants as community members who might share their propensity to commit violence. Both killers were consumed by the conspiracy of a "white genocide."

Gab and 8chan (now 8kun)—the go-to forums for Bowers and Tarrant, respectively—are rife with white supremacist, hateful, antisemitic bigotry. Examining Bowers' and Tarrant's online actions demonstrates how online propaganda can feed acts of violent terror. Additionally, violent terror can itself create online propaganda. In both Bowers' and Tarrant's cases, the shooters strongly signaled back to their fringe web communities as though they were including them as knowing co-conspirators to their criminal acts. In both cases, the participation of these fringe web communities was key to the scope, sensationalism, and ideological thrust of the act. Moreover, both shooters claimed the same twisted notion of "white genocide"—or the imminent destruction of the white race by Jews and people of color—as the motive behind their terrorist acts, suggesting a shared ideological motivation. In fringe online communities, many members indoctrinate other users based on the conspiracy propaganda of a "white genocide."

In Tarrant's 8chan manifesto, he actively and publicly sought to use his web community as co-conspirators and identified recruitment as a goal of his violence. He also told his community that he would be livestreaming his attack on Facebook. Notably, ADL's COE also reported on antisemitic, racist, and even violent content on what appeared to be Tarrant's Facebook profile.


ahem Anonymous 03/15/19 (Fri) 00:28:41 ID: c800e3 No. 12916717

Well lads, it is time to stop shitposting and time to make a real life effort post.
I will carry out and attack against the invaders, and will even livestream the attack via facebook
The Facebook link is below, by the time you read this I should be going live.
http://facebook.com/brenton.tarrant.9
It's been a long ride and despite all your rampant faggotry, fecklessness and degeneracy. you are all blokes
and the best bunch of cobbers a man could ask for.
I have provided links to my writings below, please do your part by spreading my messages, making memes and shitposting as you usually do.
If I don't survive the attack, goodbye, godbless and I will see you all in Valhalla!

Image: Tarrant's post on 8chan about the New Zealand massacre showed his plans to use Facebook to livestream the attack and his desire to make his messages mainstream.

8chan was also the platform of choice for the extremists who carried out the murderous attacks in Poway and El Paso. White supremacist John Earnest, who allegedly opened fire inside a Chabad synagogue in Poway, California, killing one person and wounding three, also allegedly posted a manifesto to 8chan before his attack that admiringly referred to Tarrant and to Robert Bowers. White supremacist Patrick Crusius—who allegedly carried out a horrific attack that left 22 people dead and more injured at a Walmart in El Paso, Texas—is believed to have also posted a manifesto on 8chan. After the shooting, extremists discussed the attack across social media platforms like Twitter and Telegram.

Platforms like 8kun and Gab (which remains incredibly popular among extremists) force us to reassess our understanding of how violence may be inspired by hateful echo chambers. Even more broadly, as we have reported, mainstream platforms can push such individuals from an open community, such as Twitter, into fringe environments like Gab that foster acceptability of dangerous views. Relatedly, some studies have similarly demonstrated that ethnic hate expressed on social media can cause surges in on-the-ground hate crimes. The implications of this online-offline dynamic are highly concerning.

B. From Memes to Mobilization

White supremacists and extremists consistently co-opt innocuous or popular slogans by layering on exclusionary messaging and then using them as a call to action. They take advantage of cultural trends to infiltrate mainstream conversations and disguise their racist beliefs as irony or jest. Coded language, symbols, and narrative manipulation are key tactics white supremacists use to appeal to mainstream people, or "normies." For example, in March 2021 Chet Hanks (son of actor Tom Hanks) published a series of social media posts critiquing white men's attire and behavior, culminating a month later with the release of his song, "White Boy Summer" (WBS). A play on Megan Thee Stallion's 2019 hit song, "Hot Girl Summer," "White Boy Summer" took

the internet and meme culture by storm. While "White Boy Summer" was not initially intended to be hateful, white supremacists adopted the slogan and leveraged it for their own purposes. Some WBS memes included white supremacist/Neo-Nazi symbols, such as swastikas or references to the white supremacist 14-words slogan. Other WBS promoters, however, were tactically diluting their content disseminate their content widely across the internet and appeal to a broader audience. One Telegram account outlined this plan, encouraging users to strategically create content "without the use of Fascist/NS symbols so normies can get in the mood for white boy summer, and not get scared away."

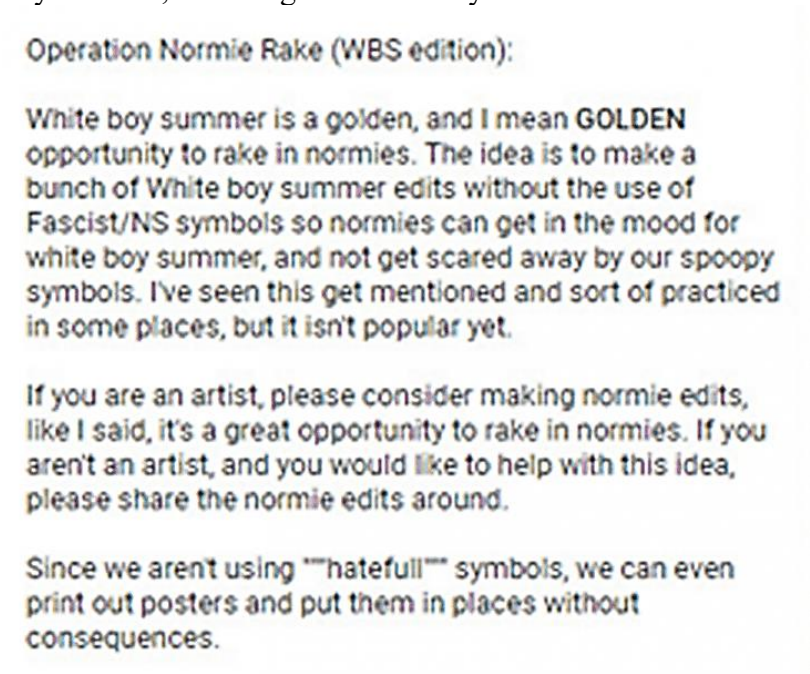


Image: Telegram post encouraging extremists to normalize posts to attract mainstream visibility.

Online users also invoked "White Boy Summer" to incite offline action. When the Derek Chauvin verdict was handed down on April 20, 2021, extremist posts encouraged people to engage in violence related to the guilty verdict, or "start celebrating #WhiteBoySummer a little early." In these cases, "White Boy Summer" was framed as a justifiable material response to "threats" against whiteness. ADL's COE tracked a variety of WBS-themed activity across the United States including merchandizing, physical propaganda distribution, extremist meetups, and proposed events. Online channels and influencers in the white supremacist movement also proposed "White Boy Summer" marches, rallies and road trips.

Social media has become a toxic ground for hosting, amplifying, and recommending corrosive content. Polarizing and bigoted language can become viral overnight, going from fringe discussions to Facebook and Twitter newsfeeds. The presence of hateful, racist and extremist content on social media thrusts bigoted ideas into the mainstream and normalizes otherwise extreme concepts and language. The leaked internal documents from Facebook have illustrated what we already knew—social media platforms cannot be trusted to make decisions that put people over profit. The result of their decisions: hate, mis- and disinformation, conspiracy theories, and extremist ideologies fester online and spread until they animate into acts of physical

violence. We cannot ignore the fact that extremism on social media is (or can be) a domestic terrorism threat.

VI. 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 violence and domestic terrorism. ADL calls for urgent action to prevent and counter domestic violent extremism. Two frameworks that ADL has created — the REPAIR plan and the PROTECT plan— promote comprehensive strategies to mitigate the threat posed by social media's impact on domestic extremism and domestic terrorism while protecting civil rights and civil liberties. Together, these strategies can have an immediate and significant impact in stopping Big Tech from amplifying and fomenting extremism that leads to domestic terrorism. Our suggestions include:

A. 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—especially for those targeted by extremists, whom are disproportionately women and members of marginalized communities. ADL's REPAIR Plan presents an integrated agenda to fight hate online and push hate, gender-based 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 and extremism. Further, officials at all levels of government can use their bully pulpits to call for better enforcement of technology companies' policies.

Regulation and Reform

Platforms play a role in fomenting hate and violence, both by providing 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 are almost completely shielded from legal liability due to Section 230 of the Communications Decency Act (CDA 230) and the lack of other legislative or regulatory requirements—even where their products, actions, or omissions may aid and abet egregious civil rights abuses and

criminal activity. Because there are no third-party/independent audits of tech companies' internal systems, 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. In an absence of transparency and oversight, online spaces have been toxic for young women and a breeding ground for extremism.

- Congress must **effectively reform, not eliminate, CDA 230** to hold social media platforms accountable for their role in fomenting gender-based violence, extremist 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.

- Section 230 reform should:
    - Stop immunizing platforms for algorithmic amplification of terrorism and discrimination. Tech companies are immune even when extremists/terrorists are recruited, radicalized, and/or are introduced to each other and plan acts of violence on their platforms. Senator Lujan's bill, the Projecting Americans from Dangerous Algorithms Act, and Reps. Eshoo and Malinowski's companion bill works to address this issue. The current broad interpretation of Section 230 means that plaintiffs alleging harm do not even get any discovery to determine what role the platform played in aiding or abetting the crime or unlawful conduct.
    - Stop immunizing tech companies from accountability for paid political and advertising content (e.g., where there is revenue-sharing or payment made from content creators to platforms). Tech companies are immune from accountability for their role in any harm caused even when they directly profit from platform-approved advertisements and other revenue-sharing agreements. Important components of Sens. Warner/Hirono's SAFE Tech Act address this issue.
    - More carefully differentiate between "conduct" and "content/speech" and eliminate immunization of the former**.** In 1996, when Section 230 was enacted, the internet was primarily text-based and noncommercial. Additionally social media platforms did not even exist. Today, however, people use the internet (and social media specifically) for far more than publishing speech. Platforms should not have wholesale immunity for everything that is produced online—especially information or conduct they create, amplify, control or profit from.
    - Ensure platforms take reasonable steps to prevent or address unlawful uses of their services. While reform cannot and should not be one-size-fits-all, every platform should do more to prevent or address unlawful content.

- Many tech policy experts have focused their efforts on reforming CDA 230 in pursuit of a non-existent one-stop solution. Importantly, reforming CDA 230 is only one essential step in a much larger process. CDA 230 reform will make platforms liable for certain unlawful third-party content; nevertheless, it is unlikely 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, increased transparency, and independent audits regarding content moderation, algorithms, and engagement features while looking for other incentive-based or regulatory action (e.g., potentially conditioning (narrower) Section 230 immunity on the steps platforms take to fight and mitigate egregious hate and disinformation).

- There is a strong need for **systematized, regulated, and easily accessible transparency** efforts from social media platforms. These platforms claim to have strong policies against hate, gender-based violence, and extremism, when in fact, most are unclear, hard to find, or have perplexing exceptions; enforcement is inequitable and inconsistent; and transparency reports are irregular and opaque.

- Additionally, Congress should encourage the Administration to **establish centers of expertise regarding online hate, gender-based 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 knowledgebase of the industry role in online harms.

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 scale issue. Moreover, the ability of tech companies to comply with global privacy regulations after first arguing that scale made such compliance impossible is instructive. Equally or more significantly, when it comes to enforcement, too often platforms miss something completely, intentionally refrain from applying the rules for certain users (like elected officials), or have biased algorithms and human moderators who do not equitably apply community guidelines. Companies also make it difficult for users to effectively lodge complaints and receive redress. Indeed, existing business models make enforcement difficult, and instead, the Administration must empower and encourage "anti-hate by design" models of online product innovation.

- **Platforms need to develop a civil rights infrastructure,** so the companies 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.

- Whole-of-government must exercise oversight by ensuring tech companies adopt and consistently enforce policies and community guidelines designed to identify and combat gender-based violence, hate, and harassment. While there is likely not a one-size-fits-all set of guidelines or enforcement given the force of law, incentives for effective standards and guidelines, transparency regarding them and their impact, and independent research evaluating these efforts can be imposed or supported by 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.

- We urge Government to consider basic consumer protection rules to product features like Facebook Groups that have amplified extremism, antisemitism, and misogyny; scaled racism and gender-based violence; and launched destructive conspiracy movements. As ADL's CEO Jonathan Greenblatt said in the Stanford Social Innovation Review, "If Mark Zuckerberg and his engineers can't improve Facebook Groups, we need to put it out to pasture permanently."

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, which 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.

- Platforms need to review and make adjustments to product features, like groups/pages, reporting, and content moderation systems, that exploit people's predilection to respond to outrage. They must consider processes that impose "friction" into product features to give users the opportunity to critically think about the content they share. Currently, split second sharing and virality are prioritized—this contributes to the amplification of highly problematic content.

- 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. IP blocking, preventing users who repeatedly engage in hate and harassment from accessing a platform even if they create a new profile, helps protect victims.

- Transparency reports must evaluate success and provide evidence that independent researchers can use; such independent researchers must be granted 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 stated actions and to assess the effectiveness of company efforts across time.

- We urge Congress and the administration 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. 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 when tech platforms host harassing content and enable perpetrators to abuse their targets. 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 Government to provide more resources and pressure on agencies to pursue investigation 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 and focusing on passing anti-cyberharassment legislation at the state level.

- 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 gain 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 hatful 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.

- Government must provide resources to civil society organizations working to counter online disinformation. **It must 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 and support concerted research to identify new ways of countering dangerous disinformation that leads to violence—especially gender-based violence. Government must not abuse this imperative to surveil vulnerable communities or to crack down on its non-violent critics and adversaries.

Research and Innovation

Government, 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 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) 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. States play a key role in this innovation, notably because our understanding of how hate impacts communities is most observable among those most familiar with their friends, neighbors, and others. Those community members are also individuals who have the most credibility in communicating with friends, family, etc. to prevent hate from taking root. States can invest in prevention, community engagement, and other tools to better understand how communities are dealing with the challenge.

B. ADL's Protect Plan

In response to the attack on the U.S. Capitol and in an effort to address the overall increase in domestic terrorism, while protecting civil liberties, ADL announced the PROTECT Plan. Domestic terrorism is a threat that impacts everyone.

P Prioritize Preventing and Countering Domestic Terrorism

R Resource According to the Threat

O Oppose Extremists in Government Service

T Take Public Health and Other Domestic Terrorism Prevention Measures

E End the Complicity of Social Media in Facilitating Extremism

C Create an Independent Clearinghouse for Online Extremist Content

T Target Foreign White Supremacist Terrorist Groups for Sanctions

Prioritize Preventing and Countering Domestic Terrorism

First, we urge Congress to adopt a whole-of-government and whole-of-society approach to preventing and countering domestic terrorism.

- In mid-June, the Biden-Harris Administration released the first-ever National Strategy to Counter Domestic Terrorism. The strategy is laudable, and a step in the right direction. However, many critical details were left unaddressed. Congress must press for further details into how the plan will be implemented, and the steps that will be taken to ensure protections for civil rights and civil liberties. Further, Departments and Agencies must create their own implementation plans for the Strategy. DHS can illuminate many of the implementation details of the Strategy by releasing its own plan. While we welcome the reinstatement of the domestic terrorism team within the Intelligence and Analysis (I&A) unit, additional initiatives and further details are needed.

- The Department of Homeland Security rightfully prioritized domestic violent extremism as a National Priority Area for the FY2021 Homeland Security Grant Program. We urge Congress to carefully oversee the effectiveness of these grants and continue the prioritization of the issue. Based on what is the most effective from this tranche of grants, the program should grow proportionate to the domestic extremist threat.

Resource According to the Threat

We must ensure that the authorities and resources the government uses to address violent threats are proportionate to the risk of lethality of those threats. In other words, allocation of resources must never be politicized, but rather, transparently based on objective security concerns.

- Congress should immediately pass the Domestic Terrorism Prevention Act (DTPA) to enhance the federal government's efforts to prevent domestic terrorism by formally authorizing offices to address domestic terrorism and requiring law enforcement agencies to regularly report on domestic terrorist threats. Congress must ensure that those offices have the resources they need and can deploy those resources in a manner proportionate to existing threats. Further, the transparency that comes with regular reporting is crucial for civil society, Congress, and the public at large to help oversee the national security process and hold leaders accountable.

- Congress must exercise careful oversight to ensure that no resources are expended on counterterrorism efforts targeting protected political speech or association. Investigations and other efforts to mitigate the threat should be data-driven and proportionate to the violent threat posed by violent extremist movements.

- The Department of Homeland Security can ensure it is resourcing proportionately by expanding data and transparency into how they see the threat and sharing with the public how the Department is aligning resources with the most lethal threats.

Oppose Extremists in Government Service

It is essential that we recognize the potential for harm when extremists gain positions of power, including in government, law enforcement, and the military.

- To the extent permitted by law and consistent with Constitutional protections, take steps to ensure that individuals engaged in violent extremist activity or associated with violent extremist movements, including violent white supremacist and unlawful militia movements, are not given security clearances or other sensitive law enforcement credentials. Appropriate steps must be taken to address any current employees, who, upon review, match these criteria. Law enforcement agencies nationwide should explore options for preventing extremists from being among their ranks.

- DHS announced that it will be vetting employees for extremist sympathies. ADL applauds this effort and welcomes any details on how the implementation of this vetting will take place, as well as any findings from the review.

Take Domestic Terrorism Prevention Measures

We must not wait until after someone has become an extremist or a terrorist attack has happened to act. Effective and promising prevention measures exist, which should be scaled.

- Congress can provide funding to civil society and academic programs that have expertise in addressing recruitment to extremist causes and radicalization, whether online or offline. By providing funding for prevention activities, including education, counseling, and off-ramping, Congress can help empower public health and civil society actors to

prevent and intervene in the radicalization process and undermine extremist narratives, particularly those that spread rapidly on the internet.

- These initiatives must be accompanied by an assurance of careful oversight and safeguards. They must also meaningfully engage communities who have been targeted by domestic terrorism and the civil society organizations embedded within them, and who have been unfairly targeted when prior anti-terrorism authorities have been misused and/or abused. They must be responsive to community concerns, publicly demonstrate careful oversight, and ensure that they do not stigmatize communities. Further, DHS should not be the only agency working on prevention; ADL urges the Department to partner with Health and Human Services and other non-security Departments whenever possible.

- While Congress has funded a small grant program for prevention measures domestically, the program is too small to have an impact at scale and, in some cases, DHS' implementation of the program has lost the confidence of communities. Now that the Administration has launched the Center for Prevention Programming and Partnerships, Congress should immediately authorize that office in statute and significantly scale its grant program; ADL has recommended a $150 million annual grant level.

End the Complicity of Social Media in Facilitating Extremism

Congress must prioritize countering online extremism and ensuring that perpetrators who engage in unlawful activity online can be held accountable. Online platforms often lack adequate policies to mitigate extremism and hate equitably and at scale. Federal and state laws and policies require significant updating to hold online platforms and individual perpetrators accountable for enabling hate, racism and extremist violence across the internet. In March 2021, ADL announced the REPAIR Plan, which offers a comprehensive framework for platforms and policymakers to take meaningful action to decrease online hate and extremism.

Create an Independent Clearinghouse for Online Extremist Content

Congress should work with the Biden-Harris Administration to create a publicly funded, independent nonprofit center to track online extremist threat information in real-time and make referrals to social media companies and law enforcement agencies when appropriate.

- This approach is needed because those empowered with law enforcement and intelligence capabilities must not be tasked with new investigative and other powers that could infringe upon civil liberties – for example, through broad internet surveillance. Scouring online sources through an independent organization will act as a buffer but will not prevent the nonprofit center from assisting law enforcement in cases where criminal behavior is suspected. This wall of separation, modeled in part on the National Center for Missing and Exploited Children (NCMEC), will help streamline national security tips and resources while preserving civil liberties. The current draft appropriations bills allocate

$500,000 toward a feasibility study for the Center; this appropriation is an excellent first step.

Target Foreign White Supremacist Terrorist Groups

Congress must recognize that white supremacist extremism is a major global threat of our era and mobilize with that mindset.

- To date, no white supremacist organization operating overseas has been designated as a Foreign Terrorist Organization. Only one has been designated as a Specially Designated Global Terrorist (SDGT). Congress should review how these designation decisions are made, whether any additional racially or ethnically motivated extremist groups outside the United States, particularly white supremacist groups, have reached the threshold for either designation, and whether such designations would help advance U.S. national interests.

- The Biden-Harris Administration must mobilize a multilateral effort to address the threat of white supremacy globally. Multilateral best practice institutions, such as the Global Counterterrorism Forum, the Global Community Engagement and Resilience Fund, and the International Institute for Justice and Rule of Law, may be helpful mechanisms through which to channel some efforts. Moreover, the Global Engagement Center should be charged with undermining the propaganda of violent extremist groups—not just designated terrorist organizations, but overseas white supremacist violent extremists as well. DHS should participate in these efforts, supporting overseas exchanges, partnerships, and best practices to engage in learning from other countries and sharing U.S. best practices, where applicable.

CONCLUSION

Thank you for the opportunity to testify before this body and for calling a hearing on this urgent topic. ADL data clearly and decisively illustrates that social media's business model directly correlates to hate rising across the United States, and fuels domestic extremism and terrorism, which continues to pose a grave threat. It is long past time to acknowledge these threats and to allocate our resources to address the threats accordingly. We must also address these threats holistically rather than piecemeal. This is precisely what ADL's **PROTECT and REPAIR** plans do, applying a whole-of-government and whole-of-society approach to push hate and extremism to the fringes of the digital world. On behalf of ADL, we look forward to working with you as you continue to devote your attention to this critical issue.