VERBAL WARNING:

Addressing Hate Speech and Dangerous Speech in the U.S.
**TABLE OF CONTENTS**

Executive Summary  
Introduction  
Part I: Addressing Hate Speech / Dangerous Speech: An Ecosystem  
Part II: Key Approaches for addressing HS/DS  
Part III: Foundation of an Effective Field  
Part IV: Key Institutions and Industries  
Part V: Key Target Audiences  
Part IV: Recommendations
EXECUTIVE SUMMARY

The United States has recently experienced an alarming rise in hate speech and hate-based incidents targeting ethnic, racial and religious minorities, as well as refugees, immigrants, the LGBTQ community, those with disabilities, and women. In September 2017, the Center for the Study of Hate and Extremism reported a 20% increase nationally in hate crimes, compared to the previous year. The FBI reported that hate crimes targeting Muslims doubled in 2016, while the Southern Poverty Law Center has documented a striking increase in hate groups across the country over the past two years - something that had been in decline until 2015. Yet, even these striking statistics fail to communicate the extent of the problem. Most hate crimes do not get reported, often because the victims are ashamed or afraid to do so. Further, it is difficult, if not impossible, to fully capture incidents of bigotry and harassment that do not rise to the legal level of a hate crime, or the amount of hateful rhetoric online targeting people for their identity - not to mention the individual and cumulative effects these actions have.

Across the country, organizations, activists, and other actors are working tirelessly to address and counter hatred against targeted groups. In 2017, The Nexus Fund undertook a scoping of more than 400 such organizations in order to gain an understanding of existing efforts and actors addressing hatred, reportedly effective interventions and approaches, key challenges, and the most pressing needs and gaps identified by affected communities and their allies.

Given the connected rise of hate speech online, we also conducted outreach to the tech sector to learn how they are addressing these issues, the challenges in doing so, and what they can do moving forward. Targeted interviewees include tech companies, advocacy groups, coalitions, academics, and relevant nonprofit organizations. This research is documented in an extensive appendix to this report.

The primary objective of this scoping report is to provide interested donors, including The Nexus Fund, with information in order to inform planned and potential investments in this area. While this report is not comprehensive, our research team identified a critical mass or strongly representative sample of formal and informal organizations working on these and related issues. Our aim is to highlight the trends of how organizations and other actors are working to address hatred and division, and to identify the needs and gaps that were identified by those working closest to these issues on regular basis.

Throughout the report, we highlight organizations that were named in the interviews of their peers as particularly effective. These are not meant to be read as the definitive “best-in-class”, rather these organizations can offer diverse models that may be replicated or scaled up, and their efforts show how approaches are put into real-world action.
Naming the problem

The definition and usage of the term ‘hate speech’ is hotly debated. In fact, many organizations and activists working on what some would call hate speech, do not themselves use that term. Globally, the atrocity prevention community has documented the tremendous power of what has been labeled Dangerous Speech: any form of expression (speech, text, or images) that can increase the risk that its audience will condone or participate in violence against members of another group.¹ A few organizations in the U.S. have begun to adopt this term into their work. However, most organizations — particularly at the grassroots level — are not yet familiar with this concept.

Finding one term to use throughout this report that appropriately categorizes all manner of identity-based bias, bigotry, hateful rhetoric and hateful speech mentioned by interviewees proved difficult. Among the 400+ groups and experts consulted for this scoping, there is no consensus on the terminology and parameters. While definitions and correct terminology are important and should be addressed beyond this scoping, for the purposes of this report, we wanted to be more inclusive than exclusive, and to focus more on the information and recommendations than definitions. Further, given that debate over exact definitions could distract from the very real and pressing issue at hand — at least at this moment — and given that the term ‘hate speech’ has both colloquial and legal definitions, we made the decision to use the umbrella term “Hate Speech/Dangerous Speech” (abbreviated as “HS/DS”).

Supporting a healthy ecosystem

Just as no single event or person shapes our individual values and opinions, there is no silver bullet approach that can reverse or prevent hatred. Rather, there must be a healthy ecosystem of approaches, tools, organizations, efforts, institutions, and actors. Interviewees represented virtually all areas of this ecosystem. Understanding this ecosystem, the needs and challenges of those within it, and the gaps that need to be filled, is the first step to providing needed support to effectively counter hate and promote tolerance, acceptance, and inclusion.

Key approaches

We have organized the various approaches to addressing HS/DS into two broad categories: ‘shaping attitudes’ and ‘response.’ Research shows that, when done effectively, promoting positive narratives and positive social norms can reduce bigotry and lead to the rejection of identity-based hatred. The first category, shaping attitudes, encompasses efforts to do just that.

Given that the values we form as children and young adults become the values we are likely to maintain throughout our lives, education and youth engagement are critical to shaping

¹ The Dangerous Speech Project, https://dangerousspeech.org/faq/?faq=200
positive attitudes towards others. This includes curricula and programs for K-12 schools, as well as on college and university campuses, where hate groups and intolerant messaging have begun making inroads.

Efforts to positively change culture, build cross-cultural understanding, and normalize diversity in communities are also critical, beyond engaging students. Organizations across the country cite success with interfaith dialogue, facilitating interaction between groups, and working within communities to expand their own attitudes and views to be more inclusive and tolerant of others. Building public awareness and public education efforts were also mentioned as ways to shape public opinion about targeted groups.

The second category, response to HS/DS, is equally important and encompasses a wide array of efforts. Activism, including events that bring community members together in condemning hatred while showing support for the target of hate are found to be helpful. Such events give individuals a healthy way to express outrage and support; this, in turn, can help promote healing, reinforce a unified identity and send a message that hatred is unacceptable by demonstrating positive social norms. Public response actions are particularly useful to state and reset positive social norms. Efforts can include, but are not limited to: protests, demonstrations and other public events, and public statements, including public letters and op/eds.

When people are engaged by someone who holds credibility and influence over them, such as a friend, religious leader, or other influential figure, they are much more likely to begin the process of attitude and behavior change. Cultivating effective allies is a key component of efforts to counter HS/DS, as they can serve as the most effective messengers to those in other groups - whether they are bystanders or actively engaged in propagating HS/DS. Many organizations aim to cultivate and deploy such allies to speak to their own sphere of influence, in-group and/or hateful speakers on behalf of targeted individuals and communities. Cultivating allies is also done through efforts to assist in real-time, such as bystander intervention training, a method of empowering individuals to interrupt behaviors associated with HS/DS and related violence. These trainings are designed to promote positive social norms, encourage the practice of situational awareness, and recognize community-specific vulnerabilities.

Foundation of an effective field
In addition to key approaches, those working in this ecosystem need the best possible resources at their disposal. In particular, the areas of research, network building, tools and messaging are critical to building and maintaining the foundation for successful efforts.

A lack of relevant, actionable and comprehensive research and data was stated as a pressing challenge by the majority of interviewees across all communities. While they are not a panacea for effectively addressing HS/DS — people are rarely moved by information alone — research and data play a critical role, and the importance of comprehensive research and data was stated as a pressing challenge by the majority of interviewees across all communities.
Many interviewees expressed that the strategies and tools used in this nascent field are outdated and potentially ineffective. In fact, several organizations reported knowing their efforts were based on potentially ineffective strategies, and that their tools had not been appropriately examined or improved. Many simply feel ill-equipped to research and incorporate the most effective strategies and tools, while others expressed fear that funders would judge any change as an admission of failure. Several smaller groups we interviewed expressed frustration that while there are likely existing tools and programs to model and utilize, they do not know where to find them or how to implement them effectively, and often feel that they are reinventing the wheel through guesswork.

Similarly, there is a need for improved messaging by organizations and activists across the field. Messaging is an often misunderstood and overlooked component of efforts to address HS/DS. Many organizations rely on conventional wisdom and instinct when crafting their messaging, which can do more harm than good. Storytelling in particular came up as an area of messaging that organizations wanted more training on how to utilize more effectively.

The existence of strong, broad-based networks was cited by national and local organizations as a significant determinant of their effectiveness. Such networks are critical to building and sustaining momentum for issues affecting vulnerable communities at all levels from the federal policy stage to the frontlines of community-based work. By creating effective distributed networks, an organization can sustain on-the-ground engagement with targeted communities, multiply programmatic impact to address HS/DS and share effective strategies that can be replicated across different communities. It’s worth noting that networks are more often successful when they have dedicated coordinating staff, and they only work when the organizations involved are highly motivated to participate.

Key institutions and industries
Several institutions and industries are especially key to shaping public opinion about targeted groups. Engaging them - as well as providing support to them and those that engage them - should be part of a larger strategy for countering hatred.

The entertainment industry, dedicated to storytelling, has proven to be a force for influencing public opinion. On the positive side, movements for equality and tolerance have been significantly bolstered through this industry. Will & Grace and Ellen were key to the LGBTQ community’s progress, while current programs such as The Carmichael Show, Master of None, and Fresh off the Boat are praised for effectively taking on stereotypes with humor and poignant commentary. However, at times the use of stereotypical and negative portrayals of women and minorities have both reflected and contributed to bigotry. Examples include the portrayal of Muslims as terrorists on shows such as Homeland, while roughly half of all Latinx on television
are portrayed as criminals.\textsuperscript{2} Engaging with the entertainment industry to advocate for positive portrayals of targeted groups will amplify messages that organizations are promoting on the ground, reaching a larger audience than any organization can on its own.

The \textbf{press} is another key institution for this field to engage, and for similar reasons. Journalists and news outlets help to shape Americans' views of targeted groups by what they choose to cover, and how they describe these communities and their experience. Unfortunately, members of the press often use words, phrases and framing that inadvertently - and at times purposefully - send negative messages about targeted communities. Engaging them to inform and shape coverage is crucial to reaching millions of Americans with a more positive view of targeted groups.

The \textbf{technology sector} occupies an important space in both the spread and countering of hatred online, particularly on social media platforms such as Facebook and Twitter. Among other negative consequences, online HS/DS can incite real-world violence offline. For example, Dylann Roof, who murdered nine African-Americans in Charleston in 2015, cited propaganda he had seen online, as the primary driver of his beliefs and violence. While much attention has been paid to the rise of hate speech and hate groups online, advocates and researchers argue that the limited actions tech companies have taken do not correspond with the scale of the problem. (See Appendix for more on the tech sector.)

\textbf{Key target audience(s)}

While many audiences should be engaged, we chose to focus on one for this report, based on feedback from interviewees: \textbf{those in homogenous or non-urban areas} and what is dubbed the ‘moveable middle.’

Americans are self-segregating — and have been for some time — into geographic areas to live near others they perceive to be like them. This has many negative consequences: for one, studies show that people in communities with little to no immigrants or refugees report the most xenophobia, indicating that when people don’t interact with others that aren’t like themselves, they are susceptible to negative narratives about those groups. Second, people tend to become more entrenched in their views as those views are constantly reinforced by others around them who agree; people start to feel they are living in a world where it is irrational not to agree with them. Third, it is becoming harder for organizations that promote tolerance and diversity to reach the very people who need to be reached, and easier to feel that ‘preaching to the choir’ is paramount to effective action.

For these reasons and more, \textbf{those living in homogenous and/or non-urban areas} are a key target audience. Further, support for organizations based in these areas is more important than

\footnotesize
ever, as they have physical access to these crucial target audiences, who may not otherwise be engaged by those with positive messages about diversity and targeted groups.

An audience segment frequently dubbed the ‘moveable middle’ are neutral or malleable individuals who can be most easily influenced with positive messaging, and even persuaded to engage both their own group and across groups, given the right context, message and messenger. They may be bystanders, on the fence, or unaware of these issues. They may also represent people whose values are neither strictly progressive nor conservative. These individuals tend to be open to dialogue across the political aisle and can be a smart target for engagement on sensitive social and political issues. Naturally, there is overlap between the ‘moveable middle’ and those living in homogenous and non-urban areas.

**Recommendations**

Based on our findings, the following are our top recommendations for funders seeking to support the efforts of organizations and activists working to address HS/DS across the country.

In order to improve efforts, the field needs more and **improved research and data** to inform and measure their efforts, in areas such as comprehensive, data-driven analysis of strategies and tools currently being used to counter HS/DS so that we may understand if and how they are effective.

Efforts to address HS/DS are naturally siloed, as most organizations working on these issues tend to be set up to serve the needs of individual identity groups, such as immigrants, women, and religious groups. This speaks to the nascent nature of this field in the U.S., and requires concerted efforts to connect actors in order to increase their ability to be effective. Further, many “gaps” cited by interviewees are resources that do exist, but which they are not familiar with, which presents an opportunity to simply connect activists and organizations with existing resources and make them more publicly available in one shared place. Unfortunately, there are few efforts to make best practices and new learnings in this field actionable for other actors in the field. **Creating opportunities for shared learning** as part of a community of practice will be a critical area for the field moving forward if we are to expect any improvement.

Effectively countering HS/DS cannot be done through discrete projects alone; it requires a **long-term support** for the resources and staff dedicated to the often difficult, slow work of shifting social norms. Across the board, interviewees expressed a need for long-term work on these issues that often involves community engagement over a period of many years, with staff and other needed resources dedicated to that work. This is only possible if donors provide long-term support and funding, and are realistic about the slow nature of social change. Other key areas of funding include **unrestricted funds** in multi-year grants to allow organizations to focus on the work at hand rather than catering and reporting to donors. Interviewees report that understaffing is one of the primary barriers for organizations to address HS/DS. Whether
through multi-year grants or any other type of support, organizations working on this issue are in desperate need of funding for increased staff.

Often — and particularly in times of crisis — individual and institutional donors provide funds to the organizations that are the largest and most well known. These organizations also generally have by far the most funding to begin with, and the most resources dedicated to fundraising. During these times, it is important to provide support for smaller and lesser-known organizations doing good work on these issues. Further, there is an enormous gap in low-dollar funding for small, community-based projects that can affect attitude and behavior change, and whose work at the hyper-local level, in the aggregate, could have an enormous impact in creating and maintaining positive social norms across the country while addressing HS/DS. Funders that are unable to manage small grants may want to consider providing funding to re-granting organizations that are poised to do so.

Hearts and minds can only be changed if you can reach them. Geographic divisions create physical barriers to reaching key target audiences, particularly in rural areas and red states. The good news is that efforts exist to do just this, but they tend to be sorely under-resourced and are often run by untrained volunteers. Support should be provided for organizations and projects in red states and rural areas, and key actors and organizations in these states should be engaged directly for a deeper understanding of their needs.

Targeting young people, teachers and schools is critical to shaping the hearts and minds of the next generation. College campuses are currently ground zero for reaching young people due to increased efforts by the so-called alt-right to foment hatred and division. We recommend extensive, focused support for student-led organizations addressing HS/DS (or who are poised to do so) as well as organizations working with students, and campus staff. Similarly, grade-school students represent the most malleable of any target audience for pro-tolerance and pro-diversity messages. Initiatives working with students, teachers and administrators on bullying and HS/DS are key to creating the next generation of open-minded adults that celebrate — rather than fear — diversity.

Funders must also invest in the development of more effective training, strategies and tools. Organizations working on HS/DS – particularly those with minimal resources - are often using outdated and ineffective strategies and tools to address HS/DS, but lack the time to and familiarity with finding and implementing them into their efforts. Further, resources for training on new strategies and tools are scarce. Best practices and cutting-edge resources exist and are constantly being improved. Organizations must have access to the best possible, most up-to-date training, strategies and tools in order to achieve the best possible outcomes.

Exploring opportunities for cross-sectoral collaboration can help improve efforts by tapping into relevant resources and expertise. The private sector has an enormous amount of resources and expertise that could be utilized to inform and improve efforts to address HS/DS. Specifically, the marketing sector has knowledge of how to influence attitude and behavior that could be
immediately and directly applied, while the tech sector can be tapped to provide tools, trainings and other resources to organizations and activists working on these issues. From both our research and organizational experience, we have also learned that many relevant academics are eager to share their time and talent with those working to address HS/DS. Key academic areas include psychology, behavioral science and neuroscience.

Individuals addressing HS/DS, especially public-facing leadership, often face harassment and threats of violence, and increased security must be provided for those working to address HS/DS to ensure their safety. As it stands, there is limited capacity for these organizations to provide security infrastructure and the lack of unrestricted funding further hinders the ability to address urgent security needs as they arise without compromising volunteer recruitment, or leading to burnout and staff turnover, or other negative impacts on sustainability.

For engaging the tech sector (covered in the appendix), the primary recommendations are: to foster sustained engagement between platforms, civil society, and affected communities; to fund research on trends and mitigation strategies to inform programmatic decisions and allow for more accurate determination of the most effective interventions; and to support key civil society groups engaging their communities on this issue and the technology industry.
INTRODUCTION

In spring/summer 2017, The Nexus Fund undertook an in-depth scoping of more than 400 organizations across the U.S. working to address various aspects of hateful speech (terms discussed below) in order to provide information to funders seeking to invest in critical efforts, including ourselves.

The United States has recently experienced an alarming rise in hate speech and hate incidents targeting virtually all ethnic, racial and religious minorities, as well as refugees, immigrants, the LGBTQ community, those with disabilities, and women. In September 2017, the Center for the Study of Hate and Extremism reported a 20% increase nationally in hate crimes, compared to the previous year. The FBI reported that hate crimes targeting Muslims doubled in 2016, while the Southern Poverty Law Center has documented a striking increase in hate groups across the country over the past two years - something that had been in decline until 2015. Yet, even these striking statistics fail to communicate the extent of the problem. Most hate crimes do not get reported, often because the victims are ashamed or afraid to do so. Further, it is difficult, if not impossible, to fully capture incidents of bigotry and harassment that do not rise to the legal level of a hate crime, or to measure the amount of hateful rhetoric online targeting people for their identity - not to mention the individual and cumulative effects these actions have.

Identity-based hatred in the U.S. is nothing new, and is well-documented throughout American history. So why the recent rise? While causality is difficult to ascribe with precision, the current spike in hate speech may be in part due to macro-trends including globalization and automation, conflicts leading to massive refugee crises and the threat of terrorism. It is likely also in some part due to the social fragmentation and political polarization underway across the US which is noted through trends of community self-segregation, the rise of partisan media, and the echo chamber effect of social media, where people are largely voicing opinions to those who already agree with them and are less exposed to analysis or critical thinking across a diversity of perspectives. These may have contributed to a climate of fear and anger, a willingness to scapegoat groups wholesale for the loss of jobs and a perceived threat to safety and available resources, leading to deepening societal divisions.

Adding fuel to the fire, politicians continue to exploit this fear, anger and division for political gain. The crux of the debate has turned to national identity: who belongs and who does not, who is American and who is not. Hateful rhetoric targeting many groups paints them as a threat, as unwelcome, as inherently un-American, or even as less than human. Throughout our interviews, a sentiment echoed that hateful beliefs and behaviors that were previously on the fringes of American society — or at least kept quiet — have now become mainstream.

Hate speech and hate crimes are also on the rise in Europe and elsewhere. Relevant violence globally shares a similar theme of ultra-nationalism: In Kansas City, an Indian-American man and his wife were gunned down by a man yelling “Go back to your country.” Similarly, British
Parliamentarian Jo Cox was murdered by a man shouting “Britain first.” In Myanmar, the day after the U.S. election, the extremist monk leading the destruction of 1.3 million Rohingya Muslims, who is known as the ‘Buddhist Bin Laden’ wrote an ode to President Trump on Facebook, alluding to his proposed Muslim ban. Following this post, violence against the Rohingya — including murder and rape of women and children — increased exponentially.

A nascent field in the U.S.

Hundreds of organizations, activists, and other actors are working across the country to address hatred at the local, state and national level — the most widely known organizations perhaps being Southern Poverty Law Center and the Anti-Defamation League. With social norms shifting away from civility and toward intolerance, these many diverse actors focus their efforts on everything from curbing hate speech online to changing negative perceptions of targeted groups.

However, it should be noted that most actors we interviewed were not established to work on addressing hate speech or hateful acts per se, but were instead organized around an identity (religion, race, immigration status, etc.) to provide support in various forms to the people they serve and represent. For these organizations, work on hate speech has been mostly a response to recent events and, for many, a completely new part of their work.

Virtually all interviewees expressed that their staff and organizations are completely overwhelmed — many used the word “shocked” — by the sheer volume of hateful incidents in their communities and the increasingly hostile landscape. One campus Rabbi dealing with threats against Jewish students, hate-filled op/eds in the campus newspaper, and swastikas painted on buildings around campus, said in his interview, “I’ve never seen anything like this. Students are afraid — some are dropping out of school. I don’t even know how to do my job anymore.” Almost all interviewees reported being vastly under-resourced in both capacity to respond and effective strategies and tools — particularly since, as noted above, most were not established to work specifically on these issues.

Defining The Problem

The definition and usage of the term ‘hate speech’ is hotly debated. In fact, many organizations and activists working on what some would call hate speech, do not use that term themselves. Globally, the atrocity prevention community has documented the tremendous power of what has been labeled Dangerous Speech: any form of expression (speech, text, or images) that can

---


increase the risk that its audience will condone or participate in violence against members of another group. Its hallmarks, as framed by Susan Benesch, founder of The Dangerous Speech Project, include dehumanization of specific groups, which can make violence seem more acceptable, and claims that members of the target group pose a mortal threat to the rest, which makes violence seem not just acceptable, but necessary. In fact, identity-based violence is often preceded and accompanied by fear-inducing and divisive rhetoric that has striking similarities across various contexts. A few organizations in the U.S. have begun to adopt this term into their work. However, most organizations — particularly at the grassroots level — are not yet familiar with this concept.

Finding one term to use throughout this report that appropriately categorizes all manner of identity-based bias, bigotry, and hateful speech mentioned by interviewees proved difficult. Among the 400+ groups and experts consulted for this scoping, there was no apparent consensus on terminology and parameters.

While definitions and terminology are important and should be addressed beyond this scoping, for the purposes of this report, we chose to be more inclusive than exclusive, and to focus more on the information and recommendations than definitions. Further, given that debate over exact definitions could distract from the very real and pressing issue at hand — at least at this moment — and given that the term ‘hate speech’ has both colloquial and legal definitions, we made the decision to use the umbrella term “Hate Speech/Dangerous Speech” (abbreviated “HS/DS” hereafter). This does not mean that there is no difference in the meanings of the various terms that we attempt, for the purpose of this report, to encapsulate with HS/DS. To be sure, they can and often do mean different things. Rather, it became a necessity to settle on one word or phrase — again, only for the purposes of this report. While we recommend that an effort be made to develop terminology that can be widely agreed upon by all actors invested in addressing these critical issues, such an effort — or any related disagreements over nomenclature — should not preclude immediate action. As Supreme Court Justice Potter Stewart said in the famous 1964 case about profanity, when it comes to hate speech and Dangerous Speech, we often know it when we see it. And we are undoubtedly seeing it across geographies, identities and socio-economic classes in the U.S. today.

A note about Dangerous Speech: When discussing Dangerous Speech and its potential consequences, it can be potentially problematic to dismiss any of these types of speech as “not dangerous enough” in terms of its ability to inspire violence in the near future. Having worked on this issue globally for several years, and in conjunction with our partners, The Nexus Fund has concluded that there are indeed alarming trends toward, as well as outright instances of, Dangerous Speech in the United States.

---

5 The Dangerous Speech Project, http://dangerousspeech.org/
Report objective

Because these trends have been escalating at an alarming rate, a large number of organizations currently working on HS/DS are doing so in addition to work on their core mission. While organizations often report “doing the best we can”, they report being overstretched, overwhelmed, and incapable of fully addressing HS/DS in their communities.

The objective of this report is to provide interested donors, including The Nexus Fund, with information to guide funding decisions in the immediate future by providing a broad view and understanding of the following: existing efforts and actors addressing HS/DS, effective interventions and approaches, key challenges, and the most pressing needs and gaps identified by affected communities and their allies.

While a lack of capacity and funding are nothing new, the sharp rise of HS/DS — and therefore the often overwhelming amount of work that needs to be done — must be acknowledged. If existing organizations that are already primed to respond to HS/DS do not receive additional funds and capacity, better results (i.e., a proportionate decline in HS/DS and related violence) simply cannot be expected.

Given that The Nexus Fund’s small research team completed this scoping in less than three months, we recognize that more remains to be learned and documented and our team strives to continue to undertake that learning and to engage our peer funders. This scoping is intended to provide a solid baseline of recommendations for funders to begin making investments now.

Tensions between freedom of expression and countering HS/DS

As with any discussion about the impacts of speech, it is important to note the tension between protecting freedom of expression and efforts to counter HS/DS. Outside of a handful of groups working on the legalities of hate speech - such as the Anti-Defamation League - the vast majority of efforts to counter hate speech focus on social, rather than legal, contexts. Many efforts focus on countering speech with more speech, rather than trying to censor the speakers. Experts including Susan Benesch cite this as being the most effective strategy when countering HS/DS. However, the debate around the legality, morality and efficacy of censorship continues to be a contentious issue, particularly as HS/DS becomes more prevalent, and as more citizens are socially, emotionally and physically threatened.

Online, HS/DS is as problematic as it is offline. The tech industry is in a difficult position when it comes to this issue because they can determine what content (speech) is allowed — and not allowed — on their platforms. Further, they are incentivized to allow all speech on their platforms in order to attract the most users and generate the most profit. Tech companies such as Twitter, Facebook and Tumblr receive enormous pressure from two opposing sides: those who advocate for no censorship whatsoever, and those who advocate for removing HS/DS as well as improving and enforcing policies to limit HS/DS on tech platforms.
Tech companies have responded with varying levels of action and differing opinions on their obligations to address this issue. Many advocates believe that companies use the cover of free speech in order to justify inaction. As public pressure increases, however, more companies appear to be taking steps to address HS/DS on their platforms. The question many advocates have is: are these companies using small efforts to merely pay lip service and placate critics, or are their intentions to truly address the issue at hand?

**Research Design and Methodology**

Our research focused primarily on organizations and community leaders working with or on behalf of racial, ethnic and religious minorities, immigrants, refugees, the LGBTQ community, women and those with disabilities.

Research began with a preliminary review of literature relevant to HS/DS in the U.S., including but not limited to studies examining online and offline HS/DS, case studies and examples of countering HS/DS, and articles reporting on recent hate-fueled incidents. Researchers then aimed to identify a large sample of organizations and individuals working on HS/DS in each targeted community; organizations were sought through expert recommendations, desk research, examining news articles, as well as referrals from interviewees. Researchers aimed to identify reactive efforts to HS/DS and hate incidents, such as rapid response measures, as well as proactive and longer-term interventions, such as working to integrate an immigrant population into a local community or creating interfaith dialogue wherein the underlying goal is to shift the attitudes, behaviors and narratives.

Researchers also strived for diversity in organizational reach (national, regional, state, and/or local focus), as well as geographic diversity (rural and urban areas, various regions of the country, etc.). Researchers also utilized online surveys with key questions, and followed up with the most relevant respondents for more formal interviews.

Targeted interviewees included relevant nonprofit organizations, technology companies, advocacy groups, coalitions, religious organizations and religious leaders, academics, journalists and prominent activists.

Interviews were semi-structured and held by phone or video conference, with most interviews lasting over an hour. A total of 431 interview requests were made (some were made to more than one person at an organization or company) and 230+ interviews were conducted. Online surveys were utilized for those unable to participate in a formal interview, and to obtain more in-depth information on key questions. For organizations that did not respond or were unable to give interviews or fill out surveys, their programs and scope of the organization were incorporated into the research more broadly.
Important considerations

This report is not meant to be an exhaustive list of every individual actor working in this space, but rather to reflect a large sample of actors whose work and insights provide a rich foundation for identifying and assessing the approaches, challenges, and needs that lead to a solid set of recommendations for funders interested in supporting efforts to understand and address HS/DS.
PART I: ADDRESSING HS/DS: AN ECOSYSTEM

Just as no single event or person shapes our individual values and opinions, no single approach to addressing HS/DS can be a silver bullet. Rather, there must be a healthy, well-resourced ecosystem of approaches, tools, organizations, efforts, and research.

For those interested in investing in and supporting the larger ecosystem, we suggest considering the following, all of which are key to a successful ecosystem to address HS/DS in the U.S. While we were unable to explore all of these in this scoping, we believe they are worth mentioning and considering for further exploration.

We included here the core areas of the ecosystem identified by interviewees, many of which we explore in in more depth throughout this report. Again, it should be noted that we do not intend this to be a comprehensive list, but rather a reflection of what we learned from desk research and speaking with a broad array of interviewees.

Approaches
While the approaches are categorized as “shaping attitudes” and “response”, there is natural overlap; something that is used to shape attitudes may also be used as part of a response effort, and response efforts are often meant to shape attitudes.

Shaping attitudes
Education at various levels, including:
- K-12
- College and university
- Public education (things like digital literacy)

Culture change and exposure
- Personal interaction with “the other”
- Interfaith dialogue
- Positive social norms
- Influential figures who provide positive messages

Response to HS/DS
Activism
- Public displays of protest or support
- Rapid response
- Strategic campaigns

Allyship
- Bystander intervention
• Cultivating effective messengers

Victim support
• Direct services
• Safety and security

Efforts to engage hateful speakers
• Having someone whose opinion they respect conduct outreach to them
• Storytelling and rehumanizing the targets of their vitriol

A strong field
A strong ecosystem must be comprised of organizations, activists, leaders, and influencers that have the resources and the training they need to ensure that all of the above happen — and are done as effectively as possible.

Healthy relevant institutions and industries
The institutions and industries that affect these issues must be engaged to serve as promoters of positive social norms, to portray targeted groups more positively, to ensure fair and just treatment, and more. These include but are not limited to:
• Government (policy and implementation of laws)
• Politics (political parties and politicians)
• Private sector (particularly media, entertainment industry and the technology sector)
• Religious institutions
PART II: KEY APPROACHES

The following are covered by this scoping, although it is important to note that the ecosystem noted above is more comprehensive and each piece merits further exploration.

Shaping attitudes

Education

The values we form as children and young adults become the values we are more likely to maintain throughout our lives. In America, the average student is in school for approximately 900 to 1,300 hours a year. This time represents an opportunity for students to interact with other students and teachers of different backgrounds, faiths and family structures. Education is critical to the prevention and shifting of hatred through the introduction and ingraining of concepts including diversity, tolerance, understanding, and empathy. It also includes the teaching of analytical thinking, including digital and media literacy. Similarly, the diversity of many college campuses presents an opportunity to shape views while students are interacting with people unlike themselves — sometimes for the first or last time in their lives, as they may have come from, and go back to, homogenous communities. Finally, hateful, alt-right groups have made significant inroads on campuses across the U.S., which is shifting social norms to be more hateful, mirroring the country’s steady decline into division.

A majority of organizations interviewed cited the importance of teaching youth about other cultures as well as about diversity, inclusion, and social justice, with 48% of interviewees citing education as one of the most critical approaches for addressing HS/DS. This must be countered with much more investment in organizations promoting diversity and inclusion.

For the purpose of this report, under this category, we cover

● Curricula and programs for K-12 schools, including anti-bullying programs
● Outreach to and engagement with youth from early childhood through college

Examples of who does this well

K-12 education

● Teaching Tolerance (SPLC): In order to effectively address bias incidents including bullying, vandalism, harassment and HS/DS, Teaching Tolerance emphasizes the necessity of incorporating ideas of diversity and tolerance through classroom materials and school culture. Recognizing that broader engagement from school leaders, not just in teachers’

---

curricula, is needed to foster an inclusive environment, Teaching Tolerance curates training materials and resources for teachers and administrators, and also provide an online forum for educators to share best practices with each other.

- **Arab Community Center for Economic and Social Services (ACCESS):** Youth and educational empowerment are key areas of focus for ACCESS. In addition to academic enrichment in core subjects, summer enrichment programs and recreational activities, ACCESS offers high school dropout prevention, technology and leadership training, community organizing and community service opportunities.
- **Erase Racism:** With the framing that separation and inequality perpetuate bias and hate, Erase Racism focuses on structural racism in Long Island, particularly in areas of housing and education. It produces training programs for teachers to incorporate diversity into their curriculum and help foster diversity in the classroom, and advocates for policies to address inequity at the local and state level. Erase Racism notes that increasing school segregation in many areas makes it more difficult for students to learn from others who are different from themselves, and hosts leadership programs across school districts to build networks and interracial friendships.

**College-specific organizations**

- **Interfaith Youth Core:** For many students, college offers a unique opportunity that expose students to various forms of diversity. The Interfaith Youth Corp capitalizes on this through its efforts to build religious pluralism guided by the belief that “interfaith cooperation is an inspiring part of the American story, and colleges are ideal environments to train leaders who write the next chapter.” By developing curriculums on the fundamentals of interfaith leadership, they provide institutional support for interfaith and religious diversity work on campuses.
- **KnowYourIX** is a survivor and youth-led organization dedicated to empowering students to reframe and end gender-based sexual violence in their communities. Considering the epidemic of sexual harassment and assault on American campuses, KnowYourIX is reframing the narrative around this issue as a civil rights issue, specifically under Title IX under the Civil Rights Act, under which colleges are federally required to report sexual violence.

**Digital literacy**

- **Simon Wiesenthal Center:** The Simon Wiesenthal Center, a museum based in Los Angeles, launched “Tools for Tolerance for Teens” program, which aims to support young people to deal with online hate, and report racism and bigotry online in the United States.

**Youth engagement — new effort worth noting**

- **Search for Common Ground — Battle for Humanity:** Combining a social media platform, mobile & web app, and real-life video game, B4H seeks to constructively engage youth living close to violence in their community. B4H examining the tactics that make video games like Call of Duty popular, B4H uses a positive psychology framework to provide users a chance to feel powerful and take positive actions to stop conflict online and offline.
Building cross-cultural understanding

When done effectively, promoting positive narratives and positive social norms can reduce bigotry and lead citizens to reject identity-based hate, mitigating HS/DS in the process. These efforts can be internally focused to reduce bias within one’s own community (or in-group), or externally focused, whereby organizations work to build relationships with and educate other groups and communities. Importance is often placed on shifting culture, building cross-cultural understanding, and normalizing diversity. Planned events to introduce or increase intergroup interactions can be simple and innocuous, making it accessible and less intimidating for less engaged community members. Shifting culture is a long-term endeavor. However, efforts to change culture can include, but are not limited to:

- Interaction with the “other”
- Interfaith dialogue
- Outreach to in-groups by allies of out-groups (see “cultivating effective messengers” below)
- Shaping identities to be more inclusive and tolerant
- Public awareness / education efforts (such as PSAs)
- Other community organizing efforts

Examples of who does this well

Facilitating interaction with “the other”

- **The People’s Supper** (a project of Hollaback!, The Dinner Party, and Faith Matters Network) — aims to repair the breach in our interpersonal relationships across political, ideological, and identity differences, leading to more civil civic discourse. They have facilitated over 900 dinners across the country to get people out of their comfort zones and join together to see each other as real people with real struggles, real fears, real hopes, and real dreams.

- **Know Your Neighbor: Multifaith Encounters Summer Campaign**: Launched by Islamic Networks Group in partnership with 70 interfaith organizations, this campaign provides tools and resources to encourage individuals and groups to interact with people of different faiths.

Public awareness / public education efforts

- **Arizona Interfaith Movement’s 'Golden Rule' initiative**: Arizona Interfaith Movement’s mission is to build bridges of understanding, respect, and support among diverse people of faith through education, dialogue, service, and the implementation of the Golden Rule. Their “Golden Rule” initiative encourages students, community leaders, residents and religious organizations in Arizona to share and apply the Golden Rule. It had led to mass adoption of the message through “Live the Golden Rule” license plates, the passage of the Golden Rule Resolution in Arizona, and an interfaith program for the Cub Scouts.

---

● **National Center for Transgender Equality:** With a mission to end discrimination and violence against transgender people, NCTE focuses on education and advocacy on key national issues for transgender communities. Its Racial and Economic Justice Initiative, for example, ensures the experiences and priorities of transgender people of color and those living in urban and rural poverty are included in the federal policy and advocacy agenda.

**Community organizing** (Note: community organizing in particular strongly overlaps with “activism and public response actions.” See below.)

● **DRUM (Desis Rising Up & Moving):** a multi-generational organization that has mobilized and built the leadership of thousands of low-income, South Asian immigrants to lead social and policy change that impacts their own lives, from immigrant rights to education reform, civil rights, and workers’ justice. It has implemented a hate-free zone in Kensington, Brooklyn.

● **ERASE Racism:** has offered race and racism dialogues in a variety of formats: film forums, workshops, discussions, and trainings. Trainings and workshops have spanned several hours to two days, and included professional development for teachers as well as the general public. Participants have been diverse, and observers have noted the dynamic and positive give and take and pledges to action that have occurred.

● **Not In Our Town:** uses community organizing, as well as documentary film and new media, to stop hate, address bullying, and build safe, inclusive communities. It has crowdsourced lessons from local communities and developed tools for schools, law enforcement and communities to take a stand against hate, intolerance, and violence. Its “Mapping the Movement Against Hate” tool is an interactive map that organizers can use to find resources by category or location.

**Response**

**Activism**

After incidents of HS/DS, events that bring community members together in condemning the speech while showing support for the target of hate are found to be helpful in allowing individuals to express, outrage, and support. This, in turn, can help promote healing, reinforce a unified identity and send a message that HS/DS is unacceptable by demonstrating positive social norms. (Note: We have separated activism / public response actions from public awareness / education and community organizing to speak to the responsive nature of public response actions. However, it should be noted that they are closely linked and often all used for both prevention and response.)

Public response actions are particularly useful to state and reset positive social norms. Efforts can include, but are not limited to:

● Protests / demonstrations and other public events

● Public statements, including public letter, op/eds
Why this approach is often effective and useful

- **It reinforces positive norms:** It is well-documented that social norms have a strong influence over our values, beliefs and behavior. In order to prevent the shift of social norms toward hatred, which happens when citizens perceive that their peers are united in their bigotry for vulnerable groups, it is critical to have ongoing public displays of support for positive norms and inclusivity — and against hatred and bigotry.

- **It helps those who are targeted feel supported:** One interfaith organization in Contra Costa County, California recalled holding a march and vigil for a mosque that was burned down, and the members of the mosque commenting they felt the whole community came out in support that day.

- **It sidelines HS/DS and refocuses on a positive message:** Organizations sometimes diverged on how to respond to an incident of HS/DS, as motivations and nature of the speech vary, but responses include condemning the content of the speech, pointing out its inaccuracies, and amplifying positive, supportive speech.

- **It can unify different communities to project a stronger collective voice:** The most recent uprising of people power against both the outcome of the election and HS/DS was the Women’s March, held the day after the inauguration. Citizens held 673 marches around the world that mobilized more than five million people. Mrinalini Chakraborty, Head of Field Operations and Strategy for the group that organized the march, cited the intersectional approach as an effective tool for countering HS/DS and violence. “By nature, by design and by intention, we are not a single-issues platform and we never wanted to be. We consider ourselves a platform for all progressive issues, but especially highlighting the issues of the most vulnerable and marginalized communities. Not just in America but across the globe.”

Examples of who does this well

- **Black Lives Matter:** With local chapters across the country, BLM has revitalized the use of public campaigns, protests and other actions to mobilize communities around the epidemic of police shootings and other instances of violence against Black communities.

- **United We Dream:** As the largest immigrant youth-led organization in the US, UWD has a nonpartisan network consisting of 100,000+ immigrant youth and allies and 55 affiliate organizations in 26 states. It organizes and advocates for the dignity and fair treatment of immigrant youth and families, regardless of immigration status, and was a leader in the creation of the Deferred Action for Childhood Arrival (DACA) program.

- **The Tennessee Immigrant and Refugee Rights Coalition** organized “We All Belong” vigils across the state following the current Administration’s first Executive Order targeting certain immigrants and refugees from Muslim-majority countries. An estimated 10,000 people participated in vigils across the state, including 3,500 people in Nashville, 2,000 in Memphis, 1,500 in Chattanooga, with several other cities each having turnout of a few hundred each. Organizers were especially pleased with the turnout in these smaller towns where pro-immigrant/refugee action is traditionally absent, and where participants and organizers were concerned there could be backlash from counter protesters and hate groups. Collective action of this scale is seen as a small but sure step to shifting narratives.
at the state and local levels, visibly challenging the assumptions that Tennesseans do not welcome immigrants and refugees.

Multiple orgs organized and participated

- **Airport protests of the Muslim ban**: a collection of local and national groups led the quick organization of airport protests following the first and second Executive Orders targeting Muslim travelers. Groups such as DC Justice for Muslims, a D.C.-based grassroots groups, and Make the Road New York, a Latino and working-class organizing group, organized demonstrations at local airports, which were replicated across the country.

- **Women’s March**: Held the day after the 2017 US Presidential inauguration, local communities held 673 marches around the world that mobilized more than five million people. The organization’s intersectional approach, particularly highlighting marginalized communities, was key in bringing together diverse crowds and organizations, allowing the March to be more than a single-issue event.

**Enlisting Allies**

Allies are a key component of any effort to support communities facing threats and stress, as they often come in the forms of the first responders - the people who speak out and take real time action in solidarity with affected persons or communities - and those who use their voice, influence and credibility to speak out on behalf of targeted communities. As the threat of violence becomes more prevalent, it creates an environment where citizens will find themselves in situations where they feel called upon to respond, but often need guidance in how to best serve as an ally to those under attack.

There are many ways individuals can create positive impact as allies. Here we highlight two of these: real-time action in situations where someone is being targeted (bystander intervention), and by serving as effective messengers on behalf of targeted communities.

**Bystander intervention** trainings empower individuals to interrupt behaviors associated with HS/DS and related violence and are designed to promote positive social norms, encourage the practice of situational awareness, and recognize community-specific vulnerabilities.

**Examples of who does this well: Bystander intervention training**

- **Collective Action for Safe Spaces**: In its efforts to build communities free from public sexual harassment and assault, CASS created innovative workshops to train anti-harassment activists to prevent public sexual harassment and assault. They also host a “Rethink Masculinity” consciousness raising group as a forum for men to work together to construct healthier masculinities, which can result in less vitriol toward others. Safe Bars, an initiative led by CASS, evolved in response to the uptick in identity-based HS/DS and now
includes: interventions/de-escalations, public awareness signage, policy and a loyalty pledge to establishments whose staff have successfully completed their curriculum.

- **Hollaback! / Heartmob:** Hollaback, a national organization dedicated to addressing harassment in across all public spaces, holds in-person bystander intervention trainings in collaboration with community organizations. Hollaback also leads HeartMob, a program and platform that provides real-time support for online harassment and empowers bystanders to speak up and intervene online.

**Cultivating effective messengers:** When hateful speakers are engaged by someone who holds credibility and influence over them, such as a friend, religious leader, or other influential figure, they are much more likely to begin the process of attitude and behavior change. Building a network of allies with influence across different sectors and groups can be an effective strategy to deploy. For example business leaders, faith leaders, popular icons and others can be engaged and cultivated to speak to their own sphere of influence, in-group and/or hateful speakers on behalf of targeted individuals and communities.

**Examples of who does this well:** Cultivating messengers

- **Veterans for American Ideals (Human Rights First):** Veterans of the armed forces are held in high esteem by virtually all Americans, making them extremely effective messengers, particularly when trying to reach conservative audiences. A small group of veterans initiated an nonpartisan effort called Veterans for American Ideals (VfAI) to train their fellow ex-soldiers to advocate for policies that protect refugees and combat Islamophobia across the country. Recruiting veterans from around the country, VfAI provides leadership and advocacy trainings to help them create or strengthen local chapters and brainstorm strategies to organize communities using personal storytelling. Using stakeholder mapping, VfAI chapter leaders identify who they are trying to influence, where they stand, and how much power/influence they have. Based on their mappings, they strategize how to leverage their individual veteran voices to reach the people or groups in the ‘moveable middle,’ particularly in more conservative circles, and advance the shared VfAI message that discrimination and xenophobic policies undermines American values and security.

- **Repairers of the Breach:** Seeks to develop effective messengers of the social gospel in places of worship, communities and workplaces who will understand the values at the heart of an anti-racism, anti-poverty, and the anti-extreme militarism movement. The effort develops leaders with a clear progressive moral vision for leadership in the 21st century and also develops local clergy who can withstand opportunism and neo-conservatism while remaining true to a principled approach to eliminating poverty and racism.

**Engaging hateful speakers**

Interviewees were split between support for efforts that engage hateful speakers and expressing that engaging these speakers is a futile exercise or that resources would best be spent elsewhere. However, several organizations have had success doing so, and given the rising level of vitriol online and offline, more needs to be done to engage these speakers to affect their
attitudes and behavior. A few organizations offer promising models for scaling up and replication.

Examples of who does this well

- **Google Jigsaw’s Redirect Method** led to over 500,000 minutes of counter-narrative video to be watched by individuals that were susceptible to ISIS recruiting. Recent research by Kevin Munger looked at groups of anti-black white male users on Twitter and found that counter speech from automated bots could influence and reduce racist speech incidents when “subjects...were sanctioned by a high follower white male.” This is an important insight into changing prejudicial behaviour online.

- **Life After Hate**, an organization created by former members of far-right extremist groups in America, leads several long-term initiatives that reach current and former members of groups like the KKK to offer them a way forward without judgment. The co-founders of Life After Hate share stories of their own transformations, emphasizing how receiving compassion from unexpected people, often targets of their hate, played a huge role in their ability to question their own intolerance and racism. Empathy and compassion are key tenets of initiating and sustaining engagement for de-radicalization efforts. Based on decades of experience, the organization understands the patience required to tackle extremist narratives, and has developed unique long-term programming that incorporates insights about the inner workings of white supremacy culture and serves to “inspire, educate, guide, and counsel” others looking to exit extremism.
PART III: FOUNDATIONS OF THE FIELD

A thriving, effective field requires that organizations and activists have the best possible resources at their disposal. Through this scoping we identified four areas that are particularly critical to laying the foundation for successful efforts moving forward. These include more and better research, network building, new and improved tools, and effective messaging.

Research

A lack of relevant, actionable and comprehensive research and data was stated as a pressing challenge by the majority of interviewees across all communities. While they are not a panacea for effectively addressing HS/DS — people are rarely moved by information alone — research and data play a critical role.

Data and research are vital for a number of reasons when addressing HS/DS; without information on the problem itself (the amount and types of hate speech proliferated online and offline and related effects), it is difficult if not impossible to drive change; solid research and data give organizations and advocates a stronger case when approaching officeholders about policy; it can lead to better and more accurate journalism; and have countless other applications. Perhaps most importantly, improved research and data are sorely needed to know what works and what doesn’t when it comes to countering HS/DS. On the flip side, poor research and incomplete data sets are certain to weaken and undermine efforts, and can even destroy the hard-won credibility of organizations working to address HS/DS.

While progress has been made in some communities to produce and obtain better research and data, it is not nearly enough and many communities are underrepresented in data that is collected, including the trans community and HS/DS targeting women online. For example, the Institute for Social and Policy Understanding (ISPU) has become one of the leading organizations providing research and polling efforts to help fill the critical knowledge gap about American Muslims, perceptions of the Muslim community, effective messaging and the impact of Islamophobia on individuals. In fact, all but one of 41 interviewees in the Muslim community named ISPU as a valuable resource that enables their work against Islamophobic HS/DS. However, this is a mere drop in the bucket compared to the well-funded industries built to propagate hatred of Muslims.

Examples of who does this well: Research & Data Analysis

- **Anti-Defamation League:** its Research and Advocacy Centers allow the organization to investigate, track and combat various forms of hate. It has two major research departments, the Center on Extremism, focused on all types of extremism, and the Center on Technology & Society, focused on cyberhate.
- **Southern Poverty Law Center**: mission is to fight hate and bigotry and to seek justice for the most vulnerable members of society. A robust investment in research and data drives their well-cited resources, educational materials, litigation and other forms of advocacy.

- **Pluralism Project**: seeks to study and document the efforts of organizations and individuals working to prevent, counter, and respond to hate in a religiously diverse society. The grassroots interfaith and civic organizations in each town, working tirelessly to promote a culture of inclusion and pluralism, play a vital, but often unrecognized role.

- **Center for New Community**: CNC is a national research and advocacy organization that has been a leader in efforts to expose and dismantle organized racism. In its current efforts to push back on anti-immigrant and anti-Muslim hate, CNC tracks organized Islamophobia bigotry in the United States and equips national and grassroots partners to mobilize.

Examples of who does this well: Connecting researchers with practitioners

- **MIT Media Lab** and **Jigsaw** organized an initial meeting of 35 researchers, advocates and platform representatives on online harassment issues in 2016 at MIT to discuss progress on online harassment and infrastructures to support research. High impact projects studying the following were discussed: estimating the chilling effects of online harassment; testing the outcomes of peer interventions against harassment; and investigating what motivates online harassing behaviors. Participants are eager to see a follow-on meeting take place.

- **Over Zero**: blends insights and approaches from diverse fields — from cognitive neuroscience to social psychology to marketing — to counter the impact of divisive and Dangerous Speech, and to increase societal resilience to hatred, division, and violence. With a focus on designing communication-based strategies, Over Zero has three main areas: providing knowledge and skills; offering strategic design and advice; and developing the field’s knowledge through pilot projects.

**Tools**

Many interviewees expressed that the strategies and tools used by most organizations in this nascent field are outdated and potentially ineffective. In fact, several organizations reported knowing their efforts were based on potentially outdated strategies — and that their tools had never been examined, revamped or improved. Many simply feel ill-equipped to research and incorporate new strategies and tools, while others expressed fear that funders would judge any change as an admission of failure.

Other fields and disciplines, such as behavioral science, neuroscience, and marketing, offer improved understanding of how to effect attitude and behavior change. However, very few organizations are utilizing lessons and tools from these other disciplines — primarily due to a lack of familiarity with them and a lack of capacity and resources to incorporate them into their efforts.
A bridgeable gap is access to and awareness of needed resources for organizations that would benefit from them. In particular, volunteer-driven, local efforts often do not have the capacity to look for resources on best practices in tackling hate or how to best engage a community. Many smaller groups we interviewed expressed frustration that while there are likely existing toolkits and programs to model and utilize, they do not know where to find them or how to implement them effectively, and often feel that they are reinventing the wheel. Further, many organizations have tools and resources that would be useful to others, but they may not be publicly available. These organizations may, however, respond positively to requests to do so. With no central information hub or any entity seeking and filling these types of requests, this gap will likely remain. If they can be made aware of relevant resources, and provided with training on how to best use those resources, their efficacy could potentially improve almost immediately.

Organizations that offer improved tools, strategies and tactics that can be shared with the field and utilized for increased efficacy

- **Over Zero**: See description in “Research” above.
- **PeaceTech Lab**, a nonprofit that supports entrepreneurs working on peace, seeks to support U.S.-based organizations to analyze, monitor and counter inflammatory language that affects their communities. Drawing on lessons learned from working in conflict affected environments globally, the Lab proposes organizing workshops to support groups to understand conflict drivers, identify partners, leverage existing data, and improve reporting on HS/DS in key U.S. cities.
- **Dangerous Speech Project** was created to test a simple, original idea: that a particular type of public speech tends to catalyze intergroup violence, and that this knowledge might be used to prevent such violence. DSP is a leader in research of online and offline DS.
- Organizations such as **Moonshot CVE** are working to counter violent extremism with big data and innovative, tech-based campaigns. These types of tools and organizations should be tapped and applied to the U.S. context.

**Messaging**

There is a significant need for improved messaging by organizations and activists across the field. Messaging is an often misunderstood and overlooked component of efforts to address HS/DS. Many organizations rely on conventional wisdom and instinct when crafting their messaging, which can do more harm than good. Messaging must take the right tone, be audience-specific, and consider a number of potential pitfalls, but most organizations and their staff are not trained to craft and deliver effective messages to the right audience(s). This category includes all messaging for efforts to address HS/DS, from public awareness campaigns to counter messaging.

**When does it work well?**

- When utilizing storytelling (see more on storytelling below)
- When consideration is given to positive messaging and humor where appropriate
● When messages are well-crafted and thoughtful about the audience
● When utilizing the right messengers
● When not being too overt with a message (show, don’t tell)
● When utilizing the right medium(s) to reach your target audience
● When a message is part of a larger, longer-term effort (not a stand alone piece). We can learn from marketing that we have to reach people multiple times before even a well-crafted message effectively reaches and influences its intended audience.

Examples of who does this well: Real-world outreach

● **Unity Productions Foundation** counts bigotry and creates peace through the media by creating high-quality, research-backed stories and films, screening them in classrooms and civic institutions accompanied with well-thought out discussion questions. They have developed trainings to help organizations learn about American Muslims and Islam, learn innovative ways to address discrimination against Muslims in America, and determine strategies to strengthen understanding while building stronger communities and reducing tensions. Additionally, UPF has successfully engaged Hollywood by providing culturally accurate information and research to script writers and producers of popular American shows. They have a decade of evaluative data to demonstrate efficacy of their films and follow up discussions, and have launched a separate initiative called MOST focused on their engagement with mainstream Hollywood.

● **Veterans for American Ideals**: See description above under “Cultivating effective allies.”

Examples of who does this well: Counter messaging online

● **ExitUSA (Life After Hate)**, a non-profit supporting former violent extremists, launched a targeted online video campaign to discredit far-right extremist groups and promote their exit program among disaffected ‘formers’ looking for a way out. Their campaign, supported in partnership with Facebook, Twitter, Google, and the Institute for Strategic Dialogue, persuaded individuals to leave the white supremacy movement.

● Muslim communities, including the **Ahmadiyya Muslim Community**, run online campaigns such as “equal entrance,” which promotes awareness around women’s rights in Islam, and “True Islam” which seeks to counter misconceptions about Muslims and extremist narratives.

● **Moonshot CVE**, a U.K.-based start-up, builds new technologies and runs guerilla-marketing campaigns in conflict zones to reduce violent extremism. They have initiated similar efforts in the United States and are planning to deepen their scope of work in the U.S. context.

**Messaging through storytelling**

When it comes to influencing attitude and behavior change, few approaches are as immediately impactful as a well-crafted story. Research shows that our values, fears and hopes are strongly shaped by stories, and that when immersed in a story, our brains physically experience anger,
Compelling stories put our brains in a highly suggestible state — just ask any ad agency. But storytelling isn’t just useful in the realm of influencing consumer behavior; when it comes to building empathy for outgroups and reducing bigotry and hatred, facts alone are simply no match for well-told stories. Of course, telling effective stories isn’t easy: studies show that in order to tap into true attitude and behavior change, people have to be transported by the story. And to influence change beyond a single room, stories need distribution such as word of mouth, social media, or through paid mediums.

**Network and movement building**

The existence of strong, broad-based networks was cited by national and regional organizations as a significant determinant of their effectiveness. Such networks are critical to building and sustaining momentum for issues affecting vulnerable communities at all levels from the federal policy stage to the frontlines of community-based work. Thus, partnerships at national, state and local levels are necessary for effective network and movement building. By creating effective distributed networks, an organization can sustain on-the-ground engagement with targeted communities, multiply programmatic impact to address HS/DS and share effective strategies that can be replicated across different communities. It’s worth noting that networks and movements are more often successful when they have dedicated coordinating staff, and they only work when the organizations involved are highly motivated to participate.

Examples of who does this well

- **Movement to End Racism and Islamophobia**: Focuses on addressing the intersectionality of racial discrimination and Islamophobia by organizing workshops and teach-ins to educate on Islamophobia and anti-racism. MERI through its partnership with other local social justice organizations, focuses their work in North Carolina.
- **Communities Against Hate**: Led by the Lawyers’ Committee for Civil Rights Under Law and The Leadership Conference Education Fund, this initiative of 11 national civil society organizations is working to address the recent spike in hate incidents across the country. They are launching a collective database that will, for the first time, aggregate data on hate incidents targeting a diverse set of impacted communities—including the Black, Latinx, LGBTQ, Muslim, Arab communities, as well as women.
- **Black Lives Matter**: Through decentralized protests and public demonstrations, BLM’s intentionally intersectional and multi-dimensional approach to expose and address forms of state violence has produced one of the most powerful and fastest growing networks in recent history.

---

PART IV: KEY INDUSTRIES AND INSTITUTIONS

Many factors contribute to how people in a society view and treat each other, including key industries and institutions that reflect and set social norms. Three of these were mentioned frequently throughout the course of our interviews: the entertainment industry, the press, and the technology sector. Whether the companies and individuals within these industries propagate or combat bigotry, feed or fight stereotypes, encourage division or inclusion is critical to shaping how Americans think about and act toward each other.

The entertainment industry

As discussed earlier, stories can be an enormously powerful for influencing attitude and behavior change, so it stands to reason that movies, television and other forms of entertainment would be among the most powerful avenues for countering HS/DS. After all, Hollywood is a storytelling industry, which has produced both positive and negative portrayals of women and minorities.

On the positive side, movements for equality and tolerance have been significantly bolstered through this industry. Will & Grace and Ellen were key to the LGBTQ community’s progress, while current programs such as The Carmichael Show, Blackish, and Fresh off the Boat are praised for effectively taking on stereotypes with humor and poignant commentary. Sesame Street has also long been a place where children can see minorities, women and those with disabilities in a positive light. In March 2017, Sesame Street introduced a puppet with autism named Julia, who is portrayed as having challenges, but also unique gifts. Through this portrayal, children can learn to recognize without judgment the behavior of those with autism, while also looking for and appreciating the special gifts they may possess.

On the negative side, stereotypical and negative portrayals of women and minorities have both reflected and contributed to bigotry. Examples include the portrayal of Muslims as terrorists on shows such as 24, while roughly half of all Latinxs on television are portrayed as criminals.

At times the propagation of bigotry is done intentionally through these mediums, but more often

---

than not it is done without knowledge of the potential consequences. Indeed, false media narratives and negative depictions in movies, television, and video games create damaging perceptions and cause feelings of inferiority in affected community members. It should also be noted that, to many, a lack of inclusion and diversity in entertainment — both on- and off-camera, sends a signal that non-whites and women of all colors are inferior.

That said, the entertainment industry is largely an untapped resource for vulnerable groups who would benefit from the creation and dissemination of these types of stories, primarily due to (1) a lack of connection between the groups and Hollywood and (2) a lack of direct advocacy and education to the entertainment industry, although there are a few notable efforts that advocate for increased visibility. Further, this is long-term work. One episode — or even one season — of a show isn’t likely to change society on its own. To have real impact on beliefs, attitudes, behaviors, actions — evaluation, data will take time, coordinated efforts, and significant resources.

Why this approach is often effective and useful

- **It breaks down misguided notions of differences:** Shows, like those mentioned above, that portray vulnerable groups as having relatable lives and problems, are also showing that all people have more in common than not: universal themes of struggle — love, money, family, friendships and work — can break down the us vs. them narrative. After all, if “they” are just like “us” there is nothing to fear. And alleviating fear can go a long way to alleviating hatred.

- **It helps targeted communities feel included in society:** By increasing and normalizing a community’s visibility, and doing so around positive depictions, it helps dispel their own feelings of isolation and otherness and instill pride and help increase self-esteem.

- **Wide distribution:** The reach of movies and television far surpasses what organizations are able to garner with their own networks. This is critical for widespread social change.

- **Expert storytelling:** The organizations working on social justice issues and HS/DS are often understaffed with people inexperienced at storytelling. The fact that stories must be well told to be effective means that expertise is not just a “nice to have” — it’s almost always a “must have.”

Example of who does this well: Entertainment Industry engagement

- **MOST (Unity Productions Foundation):** MOST, an initiative that provides the creative community with resources and information on Muslims in America, have worked with several shows to improve and diversify Muslim representation, introducing well over a dozen characters on shows such as 24, Homeland, Tyrant, The Simpsons, Bones, and Grey’s Anatomy. MOST has two key components: a resource service to provide information and facts about Islam, and seminars and special events that bring key policymakers and opinion leaders together with writers, creative executives, agents, and actors to discuss issues related to the Muslim world. MOST’s efforts are geared toward the Hollywood community (actors, writers, producers, directors, etc.) While the initiative has not faced active pushback
on their efforts, the challenges lay in beginning these relationships with writers — currently they rely primarily on backchannel relationships, often contingent on the whims of the given show.

- **BoomGen Studios**: Combats the stereotypical, racist, and outright offensive caricatures of their culture and community in popular media. BoomGen is an incubator for stories by and about the peoples and cultures of the Middle East, Central/South Asia, and North Africa aimed at the global entertainment markets. They also consult on movies, Broadway plays and television shows.

- **Color of Change**: By consulting writers rooms, showrunners and industry executives, Color Of Change seeks to reshape dominant narratives and shift cultural norms to ensure the entire entertainment industry responds with greater responsibility to the ways in which they represent Black people and other marginalized communities. Their objective is to provide an entry point for new audiences to understand our nation’s complex racial history, our present reality, and how the decisions we make moving forward will shape and reflect who we want to be as a country.

### The Press

Journalists and news outlets help to shape views through how they describe these communities and their experience, but often use words, phrases and framing that inadvertently send negative messages about targeted communities. Efforts to engage, inform and influence the media and entertainment industry are crucial to creating more positive portrayals of these communities.

A few interviewees cited the comparative news coverage of the unarmed, largely peaceful Black Lives Matter protests with the coverage of the anti-government protesters at the national wildlife refuge in Oregon. While the media focused coverage on any seemingly violent incident, such as looting, that happened at the BLM protests, coverage of the Oregon group was limited, tame and measured. This type of lopsided coverage reinforces negative stereotypes about African-Americans while providing a constant “benefit of the doubt” to white people perpetrating serious crimes.

Having a watch group to respond to those depictions is helpful, even if their message is not always heeded. For example, condemnations of whitewashing Asian characters in film shifted the discourse around films like *Ghost in the Shell* and *Aloha* — and potentially hurt their box office earnings. A recent New York Times story on *Miss Saigon* likewise demonstrated the impact of a mobilized boycott for placing a main character in yellowface, which was changed the future casting of the musical.

**Examples of who does this well:** Media engagement

- **ColorOfChange** partnered with **Media Matters for America** to study the representation of Black people in local news reporting on crime. The result is an outrageous level of distortion: while two out of every four people the NYPD arrest for murder, assault and theft are Black,
three out of every four people the news media show as responsible for those crimes are Black. The exaggerated amount of Black faces linked to crime breeds suspicion and hostility toward Black people, as does the under-reporting of white-perpetrated crime.\textsuperscript{16}

- **Asian American Journalists Association**: Efforts are aimed at both monitoring representations of Asian Americans in the media (both frequency of coverage and type of coverage) and increasing number of Asian American journalists.
- **Define American’s #WordsMatter and #FactsMatter campaigns** work to ensure that news media uses humanizing language and properly sourced facts when reporting on immigration. As a result of #WordsMatter, the Associated Press, NBC, ABC, and over a dozen other national media outlets changed their editorial guides to prohibit the use of “illegal” when referring to undocumented Americans.
- **UndocuMedia**: Media startup driving culture change, share information/rapid response and counter hate narratives against immigrant communities. The initiative is led by and for the undocumented community.

**Technology Sector**

*For more information, research and recommendations specific to the tech sector, see the Tech Appendix.*

Online, HS/DS has long been identified as a problem, both in the United States and around the world. Pew reported in 2017 that 66% percent of U.S. adult Internet users report observing harassment, which they define as six behaviors: offensive name-calling, purposeful embarrassment, physical threats, sustained harassment and sexual harassment. 41% have experienced it personally, up from 35% in 2014.\textsuperscript{17} Those surveyed by Pew observed that harassment online was often based on political, ethnic or religious differences.

Advocates and researchers argue the limited actions tech companies have taken to respond to HS/DS do not correspond with the scale of the problem. Americans seem to agree: 79% of Americans believe that online services and platforms have an obligation to address this issue. Several advocacy groups seek opportunities to engage companies on content moderation practices and to educate company representatives on how these decisions affect their work and their constituencies. However, more work is needed to widen access to backchannel engagement opportunities, particularly to data scientists and engineers.

Examples of who does this well

- **Anti-Defamation League’s** engagements to host stakeholder meetings with Silicon Valley companies, for example, provides a model for how organizations representing the interests of women, Muslims and African-Americans could engage these companies. ADL’s Online

\textsuperscript{16} ColorOfChange, “Not to be Trusted: Dangerous Levels of Inaccuracy in TV Crime Reporting in NYC” \url{https://www.colorofchange.org/newsaccuracyratings/}

\textsuperscript{17} \url{https://civic.mit.edu/sites/civic.mit.edu/files/OnlineHarassmentWorkshopReport-08.2016.pdf}
Hate Index (OHI), a project of their Center for Technology and Society and UC Berkeley’s D-Lab, is an AI/machine learning enabled way to study hate speech, from the perspective of targeted populations. Its goal is to help tech platforms better understand the growing amount of hate on social media. The OHI takes into account more than 300 different word embeddings to bring context and usage into the data analysis. It also looks at semantic, grammatical, and linguistic indicators of HS/DS, in addition to the content.

- **The Free Speech and Safety Coalition**, co-founded by WAM! board member Soraya Chemaly, mobilized 100 international organizations and helped shift content policies at Facebook. The same coalition helped WAM! amplify and strengthen direct action campaigns and related partnership work with Twitter.

- **The Online Harassment Taskforce**, a harassment focused coalition created in 2014 by Hollaback!, has organized online discussions and workshops with Facebook, Google, and Twitter. The coalition includes technologists, journalists, activists, researchers and survivors. AccessNow provides a helpline services that supports activists to deal with challenging content issues on technology platforms, including Facebook. This could be expanded to support more organizations in the U.S. who are seeking to receive swift support on content issues.

- **Nextdoor**, a private neighborhood social network for neighborhoods, deployed a successful model in 2016 around community engagement to deal with racial profiling posts on the platform that could be replicated by other companies. The tech start-up worked with community groups in Oakland such as Neighbors for Racial Justice and 100 Black Men, in addition to City Council and the Mayor’s office to collect feedback. The team then took this feedback and learnings and incorporated it into the design of a new posting flow to reduce profiling on the platform. The company has reported that the design changes lowered incidents of racial profiling by 75 percent and their work has been publicly recognized by the City of Oakland and the 100 Black Men.

- Launched at RightsCon in 2017, the **U.N. Counter-terrorism Committee** and ICT4Peace’s “Tech Against Terrorism” project provides tech start-ups tools and tactics to engage with civil society and other third party organizations (security experts, educators, and law enforcement) to respond to misuse of their products and platforms.
PART V: KEY TARGET AUDIENCES

While many key target audiences should be explored, two in particular are worth noting in depth: those in homogenous and non-urban areas, and the “moveable middle.”

Those in homogenous or non-urban areas

Americans are self-segregating — and have been for some time — into geographic areas to live near others they perceive to be like them.\(^\text{18}\) This has many negative consequences that were mentioned in our interviews. For one, studies show that people in communities with little to no immigrants or refugees report the most xenophobia, indicating that when people don’t interact with others that aren’t like themselves, they are susceptible to negative narratives about those groups. Second, people tend to become more entrenched in their views as those views are constantly reinforced by others around them who agree; people start to feel they are living in a world where it is irrational not to agree with them. (Certainly, increasingly biased media plays a role in this as well.)\(^\text{19}\) Third, it is becoming harder for organizations that promote tolerance and diversity to reach the very people who need to be reached, and easier to feel that ‘preaching to the choir’ is paramount to effective action. For these reasons and more, support for organizations based in these areas is more important than ever, as they have physical access to these valuable target audiences. Engagement of this nature should be supported for both short and long term projects, with an expectation that culture shift in these areas requires concerted efforts over many years.

Examples of who does this well

- **Dinners Across Difference** is a community-building initiative meant to engage white people and people of color in facilitated discussions about race, class, and privilege. The goal is to create new relationships through inter-racial dialogue, and to discuss local opportunities and challenges for racial justice, particularly in neighborhoods that are segregated by race and class.

- **Southerners On New Ground (SONG)** is a regional Queer Liberation organization made up of people of color, immigrants, undocumented people, people with disabilities, working class and rural and small town, LGBTQ people in the South. SONG strives to bring together these marginalized communities to work towards justice, and employ tactics such as community organizing, political education, storytelling, music, breaking bread, resistance, humor, performance, critical thinking, and celebration.

- **Montana Human Rights Network** aims to promote democratic values such as pluralism, equality and justice; challenge bigotry and intolerance; and organize communities to speak

---


out in support of democratic principles and institutions. To challenge hate groups and other extremists who use violence and intimidation as tools for political activism.

- **Rural Organizing Project**’s mission is to strengthen the skills, resources, and vision of primary leadership in local autonomous human dignity groups with a goal of keeping such groups a vibrant source for a just democracy. Local, grassroots, community organizing and advocacy.

The “moveable middle”

Two of the biggest challenges in community organizing are: (1) growing a base of engaged supporters; and (2) reaching beyond one’s own base. An audience frequently dubbed the ‘moveable middle’ are neutral or malleable individuals who can be persuaded to engage both their own group and across groups, given the right context, message and messenger. They may also be people who agree with one’s message or mission, but do not know how to engage on the issue(s).

Reaching the ‘moveable middle’ can be done using many of the approaches listed earlier in this scoping, including interfaith dialogue, facilitating interaction between people of various races and ethnicities, public education campaigns and more. Naturally, this audience overlaps with those in homogenous or non-urban areas.
PART VI: RECOMMENDATIONS

Based on our findings, the following are our top recommendations for funders seeking to support the efforts of organizations and activists working to address HS/DS across the country.

1. **Invest in research, data and measurement**
   
   In order to improve efforts, the field needs more and improved research and data to inform and measure their efforts. Key areas identified by interviewees include:

   - The link between HS/DS and violence (qualitative and quantitative research)
   - Comprehensive, data-driven analysis of strategies and tools currently being used to counter HS/DS: are they effective? What can we learn from them?
   - Tools and resources to monitor and measure efforts to address HS/DS
   - Research on effective strategies, mediums, messages, etc., including those catered to/for specific communities (API, MASA, interfaith, immigrants)
   - Research and data on HS/DS and hate crimes for communities where information is limited, including the transgender community and women.
   - Understanding the fear-mongering machinery in the US: How do they spread HS/DS? What are their best practices and mechanisms of operation, message testing, and dissemination? What can we learn from and use to inform counter efforts?
   - Research and project evaluations show that engaging victims and perpetrators of hate, as well as bystanders and others, with positive messaging, empathy and compassion is the most effective strategy. However, many organizations do not incorporate this knowledge into their efforts.
   - Efforts to learn from efforts across identity lines.

2. **Create opportunities for shared learning**
   
   Efforts to address HS/DS are naturally siloed, as most organizations working on these issues tend to be identity-based. This speaks to the nascent nature of this field in the U.S., and requires concerted efforts to connect actors to increase their ability to be effective. Further, many “gaps” cited by interviewees are resources that do exist, but which they are not familiar, which presents an opportunity to simply connect people with existing resources and make them more publicly available in one shared place. Unfortunately, there are few if any efforts to make best practices and new learnings in this field actionable for other actors in the field. Shared learning as part of a community of practice will be a critical area for the field moving forward if we are to expect any improvement.

3. **Support lesser known organizations and efforts**
   
   Often — and particularly in times of crisis — individual and institutional donors provide funds to the organizations that are the largest and most well known. These organizations also generally have by far the most funding to begin with, and the most resources dedicated to
fundraising. For example, when the immigration ban was announced, targeting majority-Muslim countries, Americans donated millions of dollars to the American Civil Liberties Union to support a legal response. While the ACLU is inarguably a key organization for addressing and fighting the travel ban, and certainly deserving of support, many other lesser-known organizations that also do exceptional work in this area remained overwhelmed and underfunded. This example speaks to two recommendations:

- Providing support for smaller and lesser-known organizations doing good work on these issues; and
- Providing them with resources and, if needed, training for brand-raising and fundraising that they most likely do not have.

Further, there is an enormous gap in funding for small, community-based projects that can affect attitude and behavior change, and whose work at the hyper-local level, in the aggregate, could have an enormous impact in creating and maintaining positive social norms across the country while addressing HS/DS. Funders that are unable to manage small grants may want to consider providing funding to re-granting organizations that are poised to do so.

4. Support organizations and projects in red states and rural areas
Hearts and minds can only be changed if you can reach them. Geographic divisions create physical barriers to reaching key target audiences, particularly in rural areas and red states. The good news is that efforts exist to do just this, but they tend to be sorely under-resourced and are often run by untrained volunteers. Key actors and organizations in these states should be engaged directly for a deeper understanding of their needs.

5. Target young people, teachers and schools

College campuses are currently ground zero for the hearts and minds of the next generation. We recommend extensive, focused support for student-led organizations addressing HS/DS (or who are poised to do so) as well as organizations working with students, and campus staff.

Grade-school students represent the most malleable of any target audience for pro-tolerance and pro-diversity messages. Initiatives working with students, teachers and/or administrators on issues related to diversity, bullying and HS/DS are key to creating the next generation of open-minded adults that celebrate difference rather than propagating fear.

6. Invest in the development of more effective training, strategies and tools
Organizations working on HS/DS – particularly those with minimal resources - are often using outdated and ineffective strategies and tools to address HS/DS, but lack the time to and familiarity with finding and implementing them into their efforts. Further, resources for training on new strategies and tools are scarce. Best practices and cutting-edge resources exist and
are constantly being improved. Organizations must have access to the best possible, most up-to-date training, strategies and tools in order to achieve the best possible outcomes.

7. Explore opportunities for cross-sectoral collaboration

Private sector: The private sector has an enormous amount of resources and expertise that could be utilized to inform and improve efforts to address HS/DS. Specifically, the marketing sector has knowledge of how to influence attitude and behavior that could be immediately and directly applied. The tech sector can be tapped to provide tools, trainings and other resources to organizations and activists working on these issues.

Academics / sciences: From both our research and experience, many relevant academics are eager to share their time and talent with those working to address HS/DS. Key academic areas include psychology, behavioral science and neuroscience. Cutting-edge work being done in these areas can vastly improve efforts.

As this is not a comprehensive list, other sectors should also be explored for potential collaboration.

8. Provide increased security for those working on these issues

The organizations and individuals addressing HS/DS often face harassment and threats of violence. This is especially true for organizations that have robust staff/volunteer hybrid structures, where a bulk of their most visible leaders come from impacted communities, such as transgender women of color or undocumented migrant workers. The public visibility of these leaders causes significant vulnerability for further targeting by HS/DS and violent threats; these leaders are literally creating the tools for their own survival. In addition, hosting public events can carry security risks, increasing event costs for organizations already working on a shoestring budget.

As it stands, there is limited capacity for these organizations to provide infrastructure and the lack of unrestricted funding further hinder abilities to address urgent security needs as they come up, without compromising volunteer recruitment, or leading to burnout and staff turnover, or other negative impacts on sustainability.

9. Provide critical funding

While all of the recommendations speak to a need for funding, the following needs are especially worth noting:

Support for long-term work
Effectively countering HS/DS cannot be done through discrete projects alone; it requires a long-term commitment of time, resources, and staff dedicated to the often difficult, slow work
of shifting social norms. Across the board, interviewees expressed a need for longer-term work on these issues that often involves community engagement over a period of many years, with staff and other needed resources dedicated to that work. This is only possible if donors provide medium to longer-term support and funding, and are realistic about the nature of social change.

**Capacity building**
- **Unrestricted funds** in multi-year grants at needed levels would allow organizations to focus on the work at hand, embark on longer-term projects, and build the best team possible.
- **Staff:** Understaffing is one of the primary barriers for organizations to address HS/DS. Whether through multi-year grants or any other type of support, organizations working on this issue are in desperate need of more staff.
- **Travel grants** can help organizations attend conferences to improve their operations, they can help rural-based organizations organize, and they can grassroots organizations bring in speakers and expand their outreach.

* * * *

_The Nexus Fund, a project of the New Venture Fund, a 501(c)(3) public charity, helps to build and support the global community to prevent mass atrocities. We act as a connector, a convener, a thought leader and a funder of small grants to fill a critical gap in the atrocity prevention space globally, with a strong focus on supporting efforts to address Dangerous Speech._

**Authors:**
Sally Smith
Arpitha Peteru
Sabrina Hersi Issa
Erin Hughes

**Technology Appendix:**
Sarah Oh (lead)
Michael Kleinman
Ross Dakin

**For more information, contact:**
Sadia Hameed
Managing Director
[sadia@nexus-fund.org](mailto:sadia@nexus-fund.org)
240-645-2642
APPENDIX:
THE ROLE OF TECHNOLOGY AND THE TECH SECTOR IN ADDRESSING HATE SPEECH AND DANGEROUS SPEECH IN THE U.S.
TABLE OF CONTENTS

Context
- Overview
- Report objective

Key Findings
- Fostering sustained engagement between platforms, civil society, and affected communities
- Funding research on trends and mitigation strategies to inform programmatic decisions and allow for more accurate determination of the most effective interventions
- Supporting key civil society groups engaging their communities on this issue, and the technology industry
- Encouraging tech incubator programs to support products and services aimed at addressing Dangerous Speech

Research process and actors

Hate Speech Online: A Continuum
- Understanding the creation of hate speech and Dangerous Speech online
- Preventing dissemination online
- Mitigating dangerous content consumption and engagement

Developing new products and services

Additional needs, gaps and recommendations
CONTEXT

In the United States, the past two years have seen an alarming rise in hate speech and hate incidents targeting virtually all ethnic, racial and religious minorities, as well as refugees, immigrants, the LGBTQ community, those with disabilities, and women. While reporting on hate crimes across the US is patchy at best - many cities do not have legal requirements to document reports of such crimes, and many victims do not report these crimes out of shame or fear of reprisal - the trends that we are able to see from what is reported give ample cause for concern. For example, in September 2017, the Center for the Study of Hate and Extremism at California State University at San Bernardino reported a 20% increase in hate crimes, compared to 2016. The Southern Poverty Law Center collected more than 1,300 “bias incidents” in just the four months following the 2016 Presidential election.

The technology sector occupies an important space in both the spread and countering of hatred online. Among other negative consequences, it can incite violence offline. This is what is known as “Dangerous Speech” - any form of expression (speech, text, or images) that can increase the risk that its audience will condone or participate in violence against members of another group. Its hallmarks, as framed by Susan Benesch, founder of The Dangerous Speech Project, include dehumanization of specific groups, which can make violence seem more acceptable, and claims that members of the target group pose a mortal threat to the rest, which makes violence seem not just acceptable, but necessary. In fact, identity-based violence is often preceded and accompanied by fear-inducing and divisive rhetoric that has striking similarities across various contexts.20 (For the purposes of this report, we use the term “hate speech / Dangerous Speech,” shortened to HS/DS hereafter. For more information on the choice of this term, please reference the introduction in the full report.)

For the purposes of this appendix focusing specifically on the tech sector, an operational definition for HS/DS includes several things:

- Harassment or attacks that are made on individuals or groups based on perceived identity;
- False or misleading information and stories that negatively depict an individual based on their perceived identity; and
- Calls for violent action against someone based on perceived identity.

During and following the U.S. presidential election in 2016, researchers collected qualitative data on a high volume of vitriol and personal attacks online across the political divide in the United States. A substantial portion of this content targets individuals based on their perceived identity. The media reported that the election cycle appeared to have emboldened trolls on social media platforms online, and/or made their actions more visible.21

---

20 The Dangerous Speech Project, http://dangerouspeech.org/
Poverty Law Center (SPLC) found in a survey of 100 hate groups operating on Twitter that these accounts amassed more “likes” in 2016 than any year since 2008. Researchers characterize a large proportion of this type of vitriol, attacks, and content produced by hate groups as HS/DS, or speech attacking an individual based on one’s identity.

The election cycle has made these issues more salient, but online abuse, a prominent facet of HS/DS, has long been identified as a problem in the United States. Pew reported in 2017 that 66% percent of U.S. adult Internet users report observing harassment, which they define as six behaviors: offensive name-calling, purposeful embarrassment, physical threats, sustained harassment and sexual harassment. 41% have experienced it personally, up from 35% in 2014. Those surveyed by Pew observed that harassment online was often based on political, ethnic or religious differences.

The tech industry is in a difficult position when it comes to this issue because of the inherent tension between Freedom of Expression and the need to effectively address – and at times censor – HS/DS. Because they can determine what is and is not allowed on their platforms, tech companies such as Twitter, Facebook and Tumblr receive enormous pressure from both sides: those who advocate for no censorship whatsoever, and those who advocate for removing HS/DS, as well as improving and enforcing policies to limit HS/DS on their platforms. Even well-intentioned efforts to limit the spread of hate speech can be controversial and difficult to do perfectly. For example, Twitter was recently in hot water for accidentally shutting down legitimate accounts in a widespread effort to shut down Russian bots set up to spread misinformation and sow discord. Further, it is not always easy to identify HS/DS online; users can rapidly adapt to new policies by using coded language and symbols.

Unsurprisingly, tech companies have responded with different levels of action and differing opinions on their obligations to address this issue. Many advocates believe that companies hide behind the issue of free speech in order to justify inaction. As public pressure increases, however, more companies appear to be taking steps (or taking more steps) to address HS/DS on their platforms.

HS/DS can turn into Dangerous Speech, or expression that can inspire people to commit or condone violence against members of another group, according to The Dangerous Speech Project. The project’s research finds that violence might be prevented by making Dangerous Speech less prevalent or less convincing. However, there are several challenges in supporting prevention efforts due to gaps of knowledge and data around the nature of HS/DS online, and strategies that can work to mitigate its effects on leading to violence.

This report provides a snapshot of promising efforts in the United States to address the prevalence of HS/DS online in the United States. It is not comprehensive; but rather, provides a

---

view into ongoing discussions and efforts. These efforts are organized in three categories:

1. Understanding the creation of HS/DS online;
2. Preventing amplification and dissemination of HS/DS; and
3. Minimizing the negative effect of online consumption of HS/DS.

**KEY FINDINGS**

The team identified four primary opportunities for funding and supporting efforts to address HS/DS online in the United States:

1. **Fostering greater sustained engagement between platforms, civil society, and affected communities.**
   a. Support groups seeking to broaden channels for constructive industry engagement on content moderation; and
   b. Support multi-stakeholder dialogues.

2. **Funding qualitative and quantitative research on trends and mitigation strategies to inform programmatic decisions and allow for more accurate determination of the most effective interventions.**
   a. Gather better baseline and trend data on overall levels of HS/DS and HS/DS reinforcing efforts to detect relevant speech and online attacks;
   b. Rigorously test the effectiveness of prevention and response interventions; and
   c. Connect the research community to practitioners designing activities to address HS/DS in the U.S.

3. **Supporting key civil society groups engaging their communities on this issue, and the technology industry.**
   a. Help organizations identify and monitor online speech trends that affect their communities;
   b. Help organizations engage more effectively with key actors within the technology industry;
   c. Support innovative messengers who aim to positively shape social norms online; and
   d. Support community educators seeking to promote digital literacy among groups in their community.

4. **Encouraging tech incubator programs to support products and services aimed at addressing dangerous speech.**
RESEARCH PROCESS AND ACTORS

The research team surveyed 53 individuals from leading advocacy organizations, research institutions, media organizations, and technology companies engaged in addressing HS/DS in the U.S. The team sought to assess engagement opportunities regarding the origination, dissemination, and consumption of dangerous speech.

We spoke with nine technology firms, including Facebook, Google, Microsoft, Yahoo, Reddit, Yelp, and Twitter, working in various functions at both leadership and operational levels related to content policy, legal support, human rights initiatives, and community management. These individuals help establish policies, or terms of service agreements that govern content on their platforms.

We also spoke with six startups that have developed new products and services that are either responding to, or seeking to monitor and address, the effects of dangerous speech.

Prominent civil society thought leaders from 14 non-profit and advocacy organizations shared their perspectives on the challenges they face when it comes to helping their communities respond to HS/DS online. Less than a quarter of the organizations represented by these thought leaders identified responding to HS/DS or harassment as part of their core mission; only recently have they had expanded their efforts to respond to community demands to respond to dangerous speech. However, nearly all organizations indicated that they were seeking support or funding for active or forthcoming projects designed to address HS/DS in the U.S.

Finally, we spoke with academic and research institutions and think tanks leading research in this area, including the MIT Media Lab, Harvard Berkman Klein Center, Stanford University, New York University, and Jigsaw shared data and insights into current research. They provided an overview of research questions that needed to be addressed, and barriers to investigating those questions.

Our team intended to investigate efforts focused on addressing dangerous speech. However, very few companies and organizations that we interviewed used this term. In fact, we observed a commingling of terms to describe content that we would categorize as dangerous speech. Interviewees used the terms HS/DS, online harassment or abuse, “dangerous content” and discriminatory practices against minority or vulnerable groups. In addition, the team discussed online misinformation and disinformation to the extent false information has lead to increased tensions between different groups.

Finally, given the sensitive nature of these issues for the technology sector, the research team faced challenges in receiving data and insights from companies on this topic.
THE CONTINUUM

Developing a comprehensive understanding of HS/DS online, how and why it happens, what is being done about it – and what can be done – can be daunting. With this in mind, this paper lays out the most relevant challenges and opportunities in terms of a chronological continuum: the initial creation of speech, the dissemination of the speech, and the consumption/engagement.

Understanding the Creation of HS/DS Online

HS/DS commonly appears on social media and gaming platforms in the form of text comments, posts, videos, or messages between users.24 It can also take the form of offensive icons, images, and memes. The Anti-Defamation League identified Pepe the Frog, for example, as a hate symbol for its use by the far right in 2016.25 The media has reported on the visibility of HS/DS on prominent public Facebook pages and Twitter accounts, including those of political leaders and elected officials.26

Studies have observed that supporters of Donald Trump’s Presidential campaign and men affiliated with “Gamergate,” an online movement objecting to progressive bias in the gaming industry, share HS/DS content, especially on Twitter. In a recent survey of 100 hate groups on Twitter, the Southern Poverty Law Center (SPLC) identified anti-immigrant and anti-Muslim groups to have the largest presence, following by anti-LGBT, black separatists, white supremacists, and anti-Semtic groups. Worryingly, data scientists such as Jonathan Morgan, who analyzes behaviors of the “alt-right” online, believes there has been an increase in the rate of radicalization of violent extremists, or those calling for genocide against Jews, the killing of Muslims, and African-Americans, on Twitter since the 2016 election period.27

In addition to social media, low-quality hyper partisan news websites such as TrueTrumpers.com produce and distribute false stories about minority groups in the U.S. as part of larger information dissemination campaigns. These sites, for instance, publish false stories about Muslims committing crimes, and make references to taking violent actions towards Muslims.28 An emerging and frightening application of technology known as ‘Deepfakes’ is a face-swapping technology that can make it appear as if someone is doing or saying something, often so realistically that it is difficult to discern any manipulation. The technology is named for the Reddit user that popularized its use switching celebrity faces into pornography. It is easy to

24 Other popular content sharing platforms where HS/DS could be posted include Instagram, Pinterest, Wordpress, Google+, and messaging applications such as Telegram.
28 https://www.buzzfeed.com/craigsilverman/anti-muslim-traffic-arbitrage-is-a-thing?utm_term_=nvbPkxEoDm#.xm5E76QJL
see how dangerous this technology could be if it were to be used to create video and audio content of political leaders to say and do offensive or illegal things, particularly on the heels of the so-called “fake news epidemic” that has shown how easy it can be to cast doubt on any news based on one’s cognitive biases and political persuasion.

The research team has identified three important drivers with respect to the creation and spread of HS/DS online:

1. **Political and social polarization**, or the separation of the country into identity-based groups, has seemingly led to greater amounts of polarized speech and HS/DS online.\(^{29}\) Buzzfeed OpenLabs research finds online responses to new stories on social media are split into factions and appear to mirror offline political divisions.\(^{30}\)

2. Individual users now have **the ability to amplify and promote** their views on social media at an unprecedented scale: organized individuals and organizations can use a variety of strategies, including spreading deliberately inaccurate information, as well as using networks of bots and fake accounts, to share their content and gain influence through real networks.

3. **Programmatic advertising**, or the use of software to purchase digital advertising allows for the execution of sophisticated and targeted advertising campaigns online.

There is no baseline data that allows us to measure the increase in HS/DS on the web or within specific platforms like Facebook or Twitter, much less across such platforms. Moreover, there is a lack of qualitative data to help researchers and advocacy groups to understand the nature of HS/DS and how it manifests online.

Technology companies do not share data or trends on online dangerous speech, abuse or the spread of discriminatory information and its impact on society.\(^{31}\) The companies consulted for this scan indicated that they often do not gather data on dangerous or HS/DS outside of ‘terms of service’ violations, particularly because it is difficult to identify accurately ex ante, or at scale. These companies are opaque to researchers and generally do not share any data that can be studied. Existing analyses about behaviors or trends have been limited to Twitter, which allows data to be extracted through its API, and a handful of other companies, such as Reddit, which has reportedly opened data for researchers.

**Current Efforts**

A challenge for determining opportunities to respond to HS/DS is the lack of concrete data and research on the nature and scale of the issue. There is no comprehensive data on overall levels

---

30 https://www.buzzfeed.com/lamvo/this-is-what-facebooks-filter-bubble-actually-looks-like?utm_term=.gdVXQoDKz#ie5E3gyLi
31 An uptick in terms of terms of service violations may not necessarily be an accurate gauge of the overall volume of HS/DS as: terms of service vary from company to company; terms of service takedowns usually rely on third party reporting and as a result, do not provide a comprehensive overview of the volume of HS/DS; and overall numbers can reflect changes in company policy or willingness to enforce policy.
of HS/DS within specific platforms, but much less across platforms. The Anti-Defamation League’s database on hate symbols, which catalogs the most frequently used symbols by hate groups, and the Southern Poverty Law Center’s database on hate organizations provides an entry point for understanding the problem, but there is no consensus on definition(s) of dangerous speech, a problem made all the more complicated by the fact that terms can change and evolve rapidly. This also poses a challenge of unclear parameters of what companies should and could be addressing.

There are early efforts focused on developing a shared standard for what constitutes HS/DS in practice, especially given how rapidly such speech can change and evolve, and collecting quantitative data to validate even basic trends around HS/DS.

Qualitative and quantitative research to track overall trends and better understand how HS/DS is generated (and can best be mitigated) is urgently needed to inform programmatic decisions, help platforms understand the impact of HS/DS on their business models, and allow for more accurate determination of what is effective. Large-scale interventions to curb HS/DS have failed and in some cases, has had an adverse effect, exposing people to new problems.\(^\text{32}\)

**Strategic Opportunities for Support**

**Need: Detect HS/DS and Online Attacks\(^\text{33}\)**

A significant challenge in collecting and assessing data on this issue is the context-specific nature of HS/DS. For example, anti-Semitic groups and individuals are known to put three parentheses around the names of Jewish people and organizations as a (previously) coded way of identifying them as Jewish. HS/DS cannot always be detected through keyword matches or image recognition. Facebook, Google, and others are exploring and actively deploying the use of artificial intelligence and machine learning approaches to detect harmful content such as child pornography or terrorist content, but are cautious to deploy these tools.

In fact, third-party efforts to detect HS/DS at scale relying on keyword searches are controversial. Hate Index, an effort to crowdsource hate terms and search for them at scale, and Geography of Hate, a project to identify geographic origins of hate, both rely on searches for HS/DS based on keyword matches. Some researchers worry which may result in a high volume of false positives. One notable new effort is to compile data on HS/DS online, including the *Anti-Defamation League Center for Technology and Society*’s Online Hate Index, an AI/machine learning enabled way to study hate speech from the perspective of targeted populations. Its goal is to help tech platforms better understand the growing amount of hate on social media. The OHI takes into account more than 300 different word embeddings to bring context and usage


\(^{33}\) There is a discussion around whether HS/DS detection is a worthwhile endeavor. One group of academics and practitioners emphasized the need to improve detection and analysis of HS/DS data online; others noted that any detection exercise would be imperfect and simply validate the hypothesis that HS/DS is a problem and requires sustained response.
into the data analysis. It also looks at semantic, grammatical, and linguistic indicators of HS/DS, in addition to the content. Other AI/machine learning projects that would scan online content for HS/DS based on keywords related to different communities would benefit from learning from the challenges these projects have faced.

One model for successfully detecting HS/DS is to scrape data, analyze and publish online articles on community-specific dangerous content trends on Facebook, as online media organizations in other countries have done. There are successful efforts internationally that have empowered ordinary citizens to monitor and detect online HS/DS that could serve as models for potential replication. Another model is to collect and analyze flagged content requests from tech companies at scale, as academic institutions have done around copyright issues. Projects seeking to address HS/DS detection are listed below.

- **Jigsaw**, Google’s technology incubator, has piloted tools, *Conversation AI and Perspective AI*, which use artificial intelligence to code comments that could better inform content decisions for platform providers. An initial pilot with The New York Times and Wikimedia offers promise to help publishers score comments to detect harmful content. They are seeking opportunities for additional testing and the development of more use cases. The company is developing use cases that can be used by publishers to manage comment sections.
- The **U.C. Berkeley Human Rights Center**, through its open source investigation lab, supports students to verify instances of HS/DS on social media. They have collaborated with Amnesty International, and more recently with *ProPublica*, a nonprofit news organization, which will house the project data that journalists can use for their investigations. *ProPublica* collects stories about people who have witnessed or were a victim of a hate crime for journalists, researchers, and civil rights organizations through their *Documenting Hate* project.
- The **Southern Poverty Law Center** launched a #reporthate project that collects incidents of hateful intimidation and harassment.
- **MentionMapp**, a social media application, makes following key hashtags and discussions on Twitter, including those around HS/DS, easier to find, measure and understand. They seek to expose “weapons of mass amplification” through their tool.
- The **Quilliam Foundation**, a London-based think tank is designing a program to apply the methodologies used to determine ISIS and terrorist networking strategies to white supremacy networks in the United States.

### Preventing the Dissemination of HS/DS Online

HS/DS is commonly disseminated and discovered online through social media. A May 2016 Pew Research Center report showed that a majority (62%) of Americans rely on social media

---

34 [http://mandola-project.eu](http://mandola-project.eu)
35 [https://www.lumendatabase.org](https://www.lumendatabase.org)
platforms for news. Some technology platforms have reluctantly acknowledged that their products have spread and amplified HS/DS through the (1) misuse of their platform by bad actors and (2) the algorithms that serve content to their users. Facebook has publicly shared how politically motivated actors have disparaged groups or causes, using ethnicity or religion as a hook to amplify dangerous narratives on Facebook.36 The company in April 2017 published information on coordinated efforts to spread false and manipulative information about different groups to achieve political goals. 37 Tactics such as creating racist memes, coordinating harassing comments, coordinating liking and sharing have been documented.

A force that contributes to the spread of HS/DS online is the use of programmatic advertising, or the use of software to purchase digital advertising that allows individuals and organizations to execute sophisticated and targeted advertising campaigns online. Programmatic advertising creates incentives for these websites to exist, often with no content on their website, because they are able to generate revenue through the ad impressions it receives. A Buzzfeed investigation found that a Eastern European run website called TrueTrumpers.com was the source of false stories about Muslims on a Facebook page “American President Donald J. Trump” with 400,000 fans.

On the web, groups have expressed significant concern about platforms such as Google serving results through algorithms, which govern what content is prioritized for viewers through their platforms. Private social networks, such as NextDoor, an application to connect neighborhoods, have also faced a rise in racist comments on their platforms.

Researchers note that solutions to address the dissemination of HS/DS online will not only have to think about how to remove it, but also how quickly: researchers believe social media posts usually reach their widest audience within the first few hours after posting.38

Corporate Response

Given the size and reach of their platforms, a few prominent technology companies naturally bear more responsibility to prevent the dissemination of HS/DS online. Numerous technology industry leaders have spoken out in support of groups that have been the target of online HS/DS since the 2016 election cycle.39 They have, however, been slower to take related actions affecting their policies and products that would address these issues.

Since 2016, companies have slowly changed their policies and identified operation gaps, and indicated a willingness to work with journalists and civil society and in some cases, have sought their counsel on issues related to violence, harassment, sexual exploitation and HS/DS.

38 http://www.newstalk.com/OPINION:-Theres-no-easy-fix-for-tackling-hate-speech-online
AppNexus, operator of one of the largest digital advertising services, made news in November 2016 when it barred Breitbart News from using its ad-serving tools because the publisher violated its HS/DS rules. Several months later, in the face of strong public criticism and requests from their shareholders, Facebook, Google and Twitter slowly responded to take action as well. Many of these actions are specific to address problems associated with the spread of false information; however, for the reasons stated above, they have implications for addressing the spread of HS/DS content online. A summary of major actions taken are listed below:

- **Facebook.** In April 2017, Facebook announced an expansion of its security focus from traditional abusive behavior, which covers spam and hacking, to include other forms of misuse of their platform, including efforts to manipulate and distort civic discourse, mislead users, and affect political outcomes. The company publicly made a commitment to disrupt economic incentives to undermine financially motivated operations to spread misleading information, build new products to improve information diversity and help users make more informed decisions when they encounter low quality content. They also committed to improving their content review process by hiring more 3,000 more employees to monitor violent video content.41

- **Google.** In response to criticism that the company spreads hateful content through its ad products, particularly on YouTube, the company has updated its policies to prohibit a wider range of content that could be discriminatory; the new policy not only prohibits content that advocates against specific groups, but also any dangerous or derogatory content that has any “characteristic that is associated with systemic discrimination or marginalization.” This means content that disparages immigrants and refugees, denies the Holocaust, or advocates for the exclusion of certain minority groups are no longer allowed.42 Google has also added new features to their search product allowing users to report hateful, dangerous or other objectionable search results or content, including their auto-complete function product that provides search predictions. By using artificial intelligence, the company has also been able to review more videos and ad-related content to aid in the removal of bad content.43 In early 2017, the company made a landmark announcement regarding adjustments of their algorithm to demote low grade content, which includes offensive content and false information about specific groups.44

- **Twitter.** Several actions have been taken to respond to abusive accounts that spread dangerous content, including automatically blocking accounts and restricting users who misuse the platform to share HS/DS.45 The company reported removing networks of

40 https://www.nytimes.com/2017/04/12/opinion/online-advertising.html?_r=0
41 https://www.wsj.com/articles/zuckerberg-says-facebook-will-add-3-000-people-to-review-content-after-violent-posts-1493822842
42 https://www.recode.net/2017/4/26/15426130/google-expanding-hate-speech-policy-adsense
43 https://www.wired.com/2017/04/zerochaos-google-ads-quality-raters/
45 https://www.engadget.com/2017/04/26/twitter-is-gaining-more-users-losing-less-money/
users who were advancing political or religious violence through messages on the platform by using advanced natural language processing software.46 To engage with third-party organizations on online abuse issues, Twitter manages a Trust and Safety Council.

Civil Society Efforts

Civil society has been seeking to identify pathways to advocacy when it comes to informing technology companies about content policy issues. Groups like Color of Change push technology companies to take responsibility for the content these companies host through outside advocacy strategies. They support advocacy efforts to help ensure the removal of discriminatory content, which can take the form of a post on a public page on Facebook, or a crowdfunding campaign to support an officer responsible for the death of an unarmed black man.

Open MIC, a nonprofit promoting shareholder engagement in technology companies, works with groups focused on civil rights, media reform and media justice to support shareholder advocacy by helping convey investor interests on media justice and diversity to company leadership. Their shareholder advocacy efforts at Facebook and Google provide a model for engaging the companies on content policies and other operational asks to mitigate the effects of HS/DS.47

Other organizations have advocated for more procedural, rather than normative, solutions. For instance, The Dangerous Speech Project and the Center for Democracy and Technology (CDT) advocate for companies to help users better understand terms of service by raising awareness of the platform guidelines. CDT is working with the Harvard Berkman Center on analysis and recommendations for how technology companies enforce content policies and how this can be improved. The CDT also advocates for the establishment of a remedy process for users to appeal content decisions.

Strategic Opportunities for Support

Need: Foster Greater Sustained Engagement Between Platforms, Civil Society and Affected Communities

Advocates and researchers argue the limited actions companies have taken to respond to HS/DS do not correspond with the scale of the problem. Several advocacy groups seek opportunities to engage the company on their content moderation practices and to educate company representatives on how these decisions affect their work and their constituencies. Advocacy, media, and community-based organizations including WAMI, Sum Of Us, Color of Change, Center for Media Justice, Daily Kos, and Jewish Voice are on the front lines of

46 http://www.dailymail.co.uk/sciencetech/article-4358312/Twitter-block-certain-words-search-engine.html
responding to HS/DS and HS/DS that affects their communities. Some of them respond to community demands to flag trolls and HS/DS content, and directly bring high-risk or high-profile instances to tech companies. They have expanded these efforts to help the companies respond more quickly to flagged content, correct content decision errors, and review high-risk content review requests. However, they say more work could be done to widen access to backchannel engagement opportunities, particularly to data scientists and engineers.

Practical solutions presented by civil society include supporting companies to generate higher-grade review or content removal requests on a volunteer basis. Some groups, for example, take it upon themselves to communicate urgent takedown requests directly to the companies for expedited review through back channels.

- There is an opportunity to replicate or broaden successful civil society efforts to engage companies. The Anti-Defamation League’s engagements to host stakeholder meetings with Silicon Valley companies, for example, provides a model for how organizations representing the interests of women, Muslims and African-Americans could engage these companies. The Free Speech and Safety Coalition, co-founded by WAM! board member Soraya Chemaly, mobilized 100 international organizations and helped shift content policies at Facebook. The same coalition helped WAM! amplify and strengthen direct action campaigns and related partnership work with Twitter. Key organizations include Women’s Media Center, Community Red, Take Back the Tech, YWCA Canada, and NNDEV.

- The Online Harassment Taskforce, a harassment focused coalition created in 2014 by Hollaback!, has organized online discussions and workshops with Facebook, Google, and Twitter. The coalition includes technologists, journalists, activists, researchers and survivors. AccessNow provides a helpline service that supports activists to deal with challenging content issues on technology platforms, including Facebook. This could be expanded to support more organizations in the U.S. who are seeking to receive swift support on content issues.

- Nextdoor, a private neighborhood social network for neighborhoods, deployed a successful model in 2016 around community engagement to deal with racial profiling posts on the platform that could be replicated by other companies. The tech start-up worked with community groups in Oakland such as Neighbors for Racial Justice and 100 Black Men, in addition to City Council and the Mayor’s office to collect feedback. The team then took this feedback and learnings and incorporated it into the design of a new posting flow to reduce profiling on the platform. The company has reported that the design changes lowered incidents of racial profiling by 75 percent and their work has been publicly recognized by the City of Oakland and the 100 Black Men.

**Need: Test Prevention and Response Strategies on Platforms**

There are opportunities to identify ways to work within a platform for more beneficial outcomes
even without direct platform access through “successor systems,” says Ethan Zuckerman.

Examples of these strategies include: (1) **Peer-to-peer sanctioning**, or how online users can better police dangerous content for themselves and their communities. Kevin Munger, a researcher at NYU, deploys bots to respond to racist content on Twitter and identifies trends in which bots and comments secure positive outcomes. The Dangerous Speech Project also seeks to identify instances where Twitter users have successfully rebuked other users to understand the factors that lead to positive outcomes where HS/DS may be removed by the original user who posted the hateful content. Another example is (2) **altering identity markers**. Based on his bot research, Kevin Munger hypothesizes that expanding the scope of a user’s identity online, by including information beyond one’s profile photo and name, could decrease the likelihood of negative identity-driven interactions and thus increase civility online. Projects to explore and share additional strategies are listed below.

- **Civilservant.io** and **CivilServant**, a project of Dr. J. Nathan Mathias, formerly of the MIT Media lab and now a postdoc at Princeton, has created software to support online communities to run experiments on the effects of community moderation practices. By optimizing the crowd and empowering online moderators and volunteer activists who design and contribute to public response on platforms, these projects allow larger groups of people to run more tests. As a result, more knowledge can be generated at a quicker rate.

- **Data for Democracy**, a community project of data scientists, is partnering with the Southern Poverty Law Center to identify HS/DS and corresponding counter speech, or social media content that tries to positively influence online behaviors in response to HS/DS. Data for Democracy supports groups to learn how to use new open source products to do this work to increase their capacity to continue this work in the future. They seek to support other journalists, activists and community groups.

- Susan Benesch’s **Dangerous Speech Project** intends to commission a literature review of engagement strategies to reduce hate and vitriol online and offline. The organization’s new website, which will launch in 2017, will gather effective and promising strategies that can be of use to the practitioners working to address hate and dangerous speech.

- **The Oxford Internet Institute** is looking at the potential application of “good” bots, which could counter “bad” bots.

### Need: Connect the Research Community to Practitioners

There are several issues that require urgent attention in this category. Researchers argue that inadequate definitions of online HS/DS and harassment prevent collaboration opportunities. In addition, given the speed at which platforms evolve, and the relative small number of researchers studying online behaviors and activities, there is a need to produce research that can be distributed quickly enough to be actionable. However, there have been initial efforts among researchers to bring academics together across technology, academia, and practitioners to collaborate and share knowledge:
MIT Media Lab and Jigsaw organized an initial meeting of 35 researchers, advocates and platform representatives on online harassment issues in 2016 at MIT to discuss progress on online harassment and infrastructures to support research. High impact projects studying the following were discussed: estimating the chilling effects of online harassment; testing the outcomes of peer interventions against harassment; and investigating what motivates online harassing behaviors. Participants are eager to see a follow-on meeting take place.

**Need: Support Multi-Stakeholder Dialogues**

Despite ongoing engagement, advocacy and media groups strongly believe existing channels for engagement with technology companies should be expanded. Previous stakeholder convenings (led by the Obama Administration) were successful in galvanizing action from the technology companies to address extremist and ISIS content on their platforms.

- Launched at RightsCon in 2017, the U.N. Counter-terrorism Committee and ICT4Peace’s “Tech Against Terrorism” project provides tech start-ups tools and tactics to engage with civil society and other third party organizations (security experts, educators, and law enforcement) to respond to misuse of their products and platforms.
- One civil society organization is pursuing a model to prepare technology companies for potential and future abuse of their platform to promote hate content through a high-level convening of key stakeholders from the tech sector.

**Mitigating Dangerous Content Consumption and Engagement**

Many groups we spoke with worried that online HS/DS could negatively shape public discourse on important issues offline. #Pizzagate, a conspiracy theory during the 2016 election cycle that led to a Washington, D.C. shooting in 2016, showed how false information and rumors could shape a larger public discussion and even lead to dangerous behavior. Researchers who study radicalization among jihadist extremists have documented how Islamic State recruiters and propagandists exploit vulnerabilities to move potential extremists to action.  

Compounding the harm, according to technology companies and advocacy groups, is the lack of digital media literacy and education among Americans. The interviews conducted for this scan indicate that the passive nature by which users discover and engage with content needs to be tackled. A 2016 Pew Research Center survey of 2,000 Americans showed that only 56% of respondents could recall the name of a news outlet when following provided news links. This calls into question whether online media consumers are aware of the sources of the information they consume.

Several advocacy groups and researchers believe social norms are a powerful way to regulate dangerous speech. To this end, branded and unbranded counter-messaging campaigns seek to counter hateful messages, or to promote alternative messages and views. These campaigns often take the form of promoting original content (e.g., a Facebook campaign with original content debunking myths about Islam) or amplifying existing content (e.g., ADL’s project to provide resources on anti-Semitism alongside search returns of hateful terms). Counter messaging can be powerful, especially group norms-based campaigns that aim to promote and proclaim specific norms or behaviors, says The Dangerous Speech Project.

Finally, campaigns such as #MuslimWomensDay have experienced significant backlash online in the form of harassment that make conversations about online security important and relevant. Many organizations seek resources and support to learn strategies and methods for conducting this work.

Corporate Response

Corporate-led programs, including Microsoft’s Digital Civility Program and Google’s Internet Citizen program, as well as U.S.-focused programs such as The Family Online Safety Initiative that offer frameworks and content that could be used by third-party organizations focused on online HS/DS. Institute for Strategic Dialogue, a UK-based organization, has developed curriculum for Google’s Creators for Change program, including a digital tolerance program for teens, and advised Facebook on counter-narrative initiatives.

Social media companies have been slowly supporting news integrity initiatives. Facebook is working with journalists, civil society and others through the Facebook Journalism Project and News Integrity Coalition. Facebook’s News Integrity Initiative, a consortium managed by CUNY Graduate School of Journalism, will fund research and projects to advance news literacy. Google News Lab, is an effort that seeks to support journalists conduct better reporting and improve the news experience for the public.

Online counter-messaging is a strategy that Facebook, Google, Yahoo, Microsoft and other companies support as a response to HS/DS. Facebook provides ad credits and other in-kind support in the form of training, workshops, and access to marketing tools to organizations running counter-messaging campaigns. They also run a peer-to-peer program that supports 59 universities to support projects to address extremist content. Google, through its Creators for Change program, has been supporting workshops and training on counter-content creation on YouTube to address HS/DS and extremist content in the United States for community based

---

49 http://www.huffingtonpost.com/entry/muslim-womens-day-takes-off-on-twitter-amid-celebration-and-backlash_us_58d93b5ce4b02a2eaab62db4
51 https://media.fb.com/2017/01/11/facebook-journalism-project/
organizations and will continue this program.

Strategic Areas for Support

Need: Reinforce Norms Online through Counter-Messaging

Efforts to promote counter and alternative messages to HS/DS and narratives are urgently needed and should be prioritized for support, say advocacy groups and researchers. Many are worried by the speed and strength with which extremist groups have organized and executed wide-reaching online message campaigns to promote HS/DS content. The mobilization of far-right users in the U.S. to bombard French social media with propaganda supporting far-right candidate Marie Le Pen around the April 2017 French presidential election, for example, provides evidence of this. Interestingly, the campaign of Le Pen’s rival reportedly attempted to “blur” false and negative narratives by putting out their own through secret backchannels. This was seemingly a successful tactic, given the decisive outcome of the election in Macron’s favor.

Visible work is being led by vulnerable groups and minority groups who are seeking to counter this speech; however, many observers worry these groups are too few and lack resources to execute successful campaigns. Others have highlighted the challenges groups supporting minority rights causes (i.e. progressive organizations) face in reaching audiences beyond their supporter and donor bases. Practitioners observe current messaging efforts run the risk of being viewed as partisan, or “preaching to the choir.” Eli Pariser, author of “The Filter Bubble” and founder of Moveon.org, notes the need to better understand the epistemology of non-left groups in order to reach them. Other actors, under cover of anonymity, seek to amplify alternative messages to audiences who are sympathetic to the views of extremists.

Discussions around counter messaging strategies appear to focus on two approaches: 1) directly countering false information or dangerous narratives with real information and compelling counter-narratives or 2) promoting alternative messages. Many groups hypothesize that effective campaigns will target and provide opportunities for disaffected groups, such as young men in rural towns without jobs, to receive support and help.

A summary of active efforts are listed below:

- **ExitUSA**, a non-profit supporting former violent extremists, launched a targeted online video campaign to discredit far-right extremist groups and promote their exit program among disaffected ‘formers’ looking for a way out. Their campaign, supported in partnership with Facebook, Twitter, Google, and the Institute for Strategic Dialogue, persuaded individuals to leave the white supremacy movement.

- Muslim interest communities, including the Ahmadiyya Muslim Community, run online

---

campaigns such as “equal entrance,” which promotes awareness around women’s rights in Islam, and “True Islam” which seeks to counter misconceptions about Muslims and extremist narratives.

- PeaceTech Lab, a nonprofit that supports entrepreneurs working on peace, seeks to support U.S.-based organizations to analyze, monitor and counter inflammatory language that affects their communities. Drawing on lessons learned from working in conflict affected environments globally, the Lab proposes organizing workshops to support groups to understand conflict drivers, identify partners, leverage existing data, and improve reporting on HS/DS in key U.S. cities.

**Need: Strengthen Data and Security Skills and Survivor Support**

Organizations, particularly those working at the community level, have identified several needs that would prepare them to understand and respond to online HS/DS that affects the communities they serve and support. These include building capacity to better track and understand the nature of HS/DS, supporting those who are harassed or affected by HS/DS, and protecting their organizations and individuals who design or implement interventions to respond to online HS/DS. Identified projects include:

- The National Network to End Domestic Violence, a national membership and advocacy organization, helps women navigate online threats to their safety including how to thwart abusers through training and advocacy through its Safety Net Technology Project. Its resources and support services to survivors who have encountered harassment, hate and abuse online could benefit other communities.

- Services offered by harassment prevention and support focused nonprofits such as Hollaback! and Online SOS serve and support people experiencing online harassment through free and direct services that could also be made more widely available to vulnerable and minority communities. Heart Mob, a product of Hollaback! allows users who are targets of HS/DS or harassment to recruit online bystanders to help them. Grassroots feminist activists have also developed online resources such as “Speak Up & Stay Safe(r)” that could be made available for more audiences.

- Digital Security Exchange (DSX), a new U.S.-based project founded by a former Access Now staffer, supports organizations to develop stronger security practices to help them respond to organizational threats and other online security challenges they might face in responding to HS/DS. The project is expanding a community of security and technology experts who support non-profits facing significant online threats.

- Benetech, a social enterprise that develops social impact software, in partnership with the Southern Poverty Law Center, seeks to build more digital security products to support vulnerable communities, based on Benetech’s Martus software, which is used internationally by NGOs to document human rights violations, and conducting threat modeling for Black Lives Matter to discuss building trusted spaces online.

---

53 [https://www.techsafety.org](https://www.techsafety.org)
● Existing efforts such as Data Kind’s open call to support data analysis and the use of predictive technologies on HS/DS projects could be expanded to support groups working at the community level.

Need: Improve Digital Media Literacy

Groups are beginning to respond to an urgent need to increase digital media literacy and digital citizenship education to foster greater understanding of what it means to create, share and consume content online. Many advocacy groups noted that there is a lack of resources for promoting meaningful education programs around digital media literacy for children and young adults. Identifying digital literacy projects and programs was not a focus of this scan; however, there are several new programs that were mentioned to our team, including those listed below:

● The Simon Wiesenthal Center, a museum based in Los Angeles, launched “Tools for Tolerance for Teens” program, which aims to support young people to deal with online hate, and report racism and bigotry online in the United States.

DEVELOPING NEW PRODUCTS AND SERVICES TO ADDRESS DANGEROUS SPEECH

Several media thought leaders note the opportunity for product and service development. Entrepreneurs are beginning to “productize trust and safety.” Brand-safety firms such as GIPEC work on deep web attribution, and Moat, a digital advertising search engine support large companies with their content challenges. Media observers also note the opportunity for predictive technologies and computational journalism to support civil society and media to respond to dangerous speech. Bay Area-based incubators Matter VC and YCombinator have both expressed interest in prioritizing teams that seek to address issues that increase democracy and promote civic engagement through broader narratives. A list of active efforts follows below.

● Newknowledge.io, a start-up, builds machine-learning technologies and software applications to identify and combat online extremism and disinformation, particularly far right extremism and associated HS/DS. They are seeking to further develop and scale the use of their technologies that use machine learning to identify HS/DS, use natural language processing to identify communities becoming radicalized, hate symbols, and influencers.

● New America Foundation’s new initiative in public interest technology, announced in early 2017, will support technologists to develop products and services supporting populations that may be negatively impacted by new policies related to immigration, refugee assistance, and criminal justice. This work is expected to include direct service and product development and is likely to address online HS/DS.
- **Matter VC**, a media accelerator, is prioritizing engaging communities and start-ups that are using new distribution models on new issue areas such as: promoting vulnerable or underrepresented voices (e.g., feminism) and social impact content; supporting vulnerable groups to own their identities online; and helping users break away from their filter bubbles. Examples of companies they are supporting include: *News Deeply*, which covers important, underreported single-issue stories pertaining to refugees, and women and girls.

- **The Tempest**, an online media company, builds, amplifies and connects diverse millennial women through innovative content and products.  

- **Imzy**, a new social media community platform with tens of thousands of users that seeks to be inclusive. Content is monitored, and harassment or HS/DS is removed.

### ADDITIONAL NEEDS, GAPS, AND RECOMMENDATIONS

In addition to the efforts and opportunities identified above, the research team identified several gaps of knowledge that warrant further attention and discussion in the future. Initial recommendations are provided.

### Tech Company Engagement on HS/DS Issues in the U.S.

Tech companies do not appear to have identified U.S. HS/DS as a priority issue for response from an operations perspective, despite increased advocacy on this front from those representing minority groups in the U.S. There is an opportunity to include the perspectives of vulnerable and minority groups in more active discussions about “fake” news, information operations online, and counter-terrorism to highlight the implications of these trends on the work of national and community-based organizations representing the interests of minority and other vulnerable groups.

### Building a Community of Practice

There is an overall lack of centrality or a hub for conversation among advocacy organizations, researchers, and technology industry representatives to discuss matters around HS/DS both within and across their communities; discussions around these issues have coalesced around small hubs that are limited to a handful of funders and academics. The community of researchers studying and publishing research on online HS/DS, harassment, and false information in the U.S., in particular, appear to be far more distributed and disjointed compared to communities addressing counter-terrorism or open data, for example. Groups implementing the programs above, in particularly, noted that they could benefit from ideation and collaboration support from and engagement with experts and other community organizations who have implemented messaging and community outreach programs.

---

55 [https://thetempest.co](https://thetempest.co)
Tapping International Experience
Work on hate and HS/DS in other geographies present rich learning opportunities. Nonprofit thought leaders note the gap of knowledge between the international media development, counter-terrorism, and peace-building communities with those developing programs and interventions in the United States. The countering violent extremism (CVE) community, for example, offers insights that are relevant to groups in the U.S. focused on counter-messaging interventions. Think tanks and community-based organizations in the international CVE community, such as the Quilliam Foundation, Hedayah Foundation, Demos and the Institute for Strategic Dialogue could offer expertise to U.S.-based organizations.

Civic Tech Community’s Engagement on HS/DS
There is an opportunity for figures that have led civic engagement or other projects at the intersection of technology and public engagement that have credibility with the technology industry to support civil society in their efforts to develop new approaches and projects to address dangerous speech.

Supporting Online Moderators
Online moderators of conversations on platforms such as Reddit and private Facebook groups are actively engaged in responding to HS/DS in their communities. Some have developed novel approaches to responding to these issues through progressive community standards. Research from the MIT Media Lab indicates these online communities and their moderators are a valuable constituency to engage in this discussion. The Lab recently hosted a workshop by Civil Servant with Reddit moderators doing work to support improved content moderation. These and other relevant actors could be better engaged by the research community.

Impact Evaluation
There are several practical challenges around measuring the problem of HS/DS and how interventions might have a measured impact. The lack of baseline data, for example, makes the design of a possible intervention and evaluating its success challenging. At the project level, the evaluation of online counter-messaging initiatives is still in an experimental phase. More support and strategies are needed to support organizations to evaluate the success and impact of their interventions. Improving social media data analysis methods is a related area that should be explored.