Center for American Progress

GETTY IMAGES / LOGAN CYRUS

Curbing Hate Online

What Companies Should Do Now

By Henry Fernandez October 2018

Center for American Progress

Curbing Hate Online

What Companies Should Do Now

By Henry Fernandez October 2018

Contents

Author's note:

The Center for American Progress joined with partners in the development of policy recommendations found in "Appendix: Recommended Internet Company Corporate Policies and Terms of Service to Reduce Hateful Activities," available for download on this report's webpage and cited in endnote 13. The Center is a part of a coalition that developed this tool for both internet companies and for advocates who care about the issues involved. But we also have our own mission and vision, so we want to be sure to explain how CAP, as an organization committed to progressive social change, thinks about the use of technologies that are transforming our nation and world. As such, the ideas outlined in this report are CAP's views—and not necessarily those of our partners. Whether our partners would agree with our thoughts in whole or in part, we credit them for many of our ideas; they emerged from the lessons we learned while working with them as well as from their expertise.

Introduction and summary

Online tools have become essential for everyone. They make it possible to easily search for information, pay for services almost instantaneously, communicate with friends around the world, and even hail a ride or bring food to the front door. They also have helped to raise money for flood victims, to organize people to respond to public health crises, and to register people to vote. The internet opens doors for a richer, more inclusive democracy, one in which voices that have historically been locked out by discrimination and money can express the needs of their communities. We have seen activists empowered to mobilize online, grow their audiences, and tell their stories—for example through the protests for criminal justice reform in Ferguson, Missouri,[1] the emergence of the #MeToo movement against sexual harassment and sexual assault,[2] and the response to a recent white supremacist rally in Boston that led to tens of thousands of peaceful counterprotesters.[3]

But we have also seen how internet tools have empowered those driven by or capitalizing on hate. The violent and ultimately deadly neo-Nazi march in Charlottesville, Virginia, was planned, advertised, coordinated and even paid for using mainstream online tools such as PayPal,[4] Facebook,[5] and the gamer chat app Discord.[6] In a highly connected world, it is not surprising that those engaging in mass atrocities targeting minorities in Myanmar are using the internationally ubiquitous WhatsApp to spread dangerous lies.[7] The same app has been used to instigate deadly mob violence in India.[8]

Following the violence in Charlottesville, the Center for American Progress joined with the Southern Poverty Law Center and Free Press to convene experts from civil, human, and media rights groups, as well as open internet organizations, to better understand how hate organizes online and to determine what could be done about it. We sought out and listened to experts on terrorism, human rights, media manipulation, technology, and law. We heard from those working domestically, as well as experts working in Asia, Africa, and Europe. We visited with colleagues tackling the same issues in Europe and met with leaders in the United Kingdom, Germany, and at the European Union to understand how they were addressing concerns in their communities.

After months of research, our focus returned to what could be done in the United States to address hate online, and we made a key decision: We chose very early not to make a recommendation for government regulation or to advocate for legal or policy changes at the federal or state levels as part of this project. Unlike our allies in Europe, the United States has the First Amendment, reflecting an important commitment to severely limit government intervention in speech.

Instead, we focused on internet companies'[9] self-regulation of what occurs on their platforms. Virtually all internet companies already have terms of service or acceptable use policies that purport to regulate hateful activities on their services. We decided to help make those user-facing policies more effective, fair, and transparent.

It is important to note that the power of internet companies to regulate content on their own services is a right protected by the First Amendment. They may set the ground rules for how they will operate, how they will build their client base, what communications they will allow, and what they will charge to use their services—and they may do so without government interference. There are exceptions to this corporate freedom that have been carved out by Congress and the courts. For instance, in the provision of their services, internet companies cannot allow an advertiser to target advertising for home sales to white potential buyers only[10]—nor can an internet company allow or ignore the distribution of child pornography via their services.[11] These boundaries are clear. But so is the right of companies to set their own direction in how they will interact with their customers and users with regard to content on their platforms.

While user-facing policies are an important part of reducing hateful activities online, we determined that these alone are not enough. This is evident in the reality that the existence of anti-hate provisions in most companies' terms of service has not stopped hate on various platforms. Thus, we also include recommendations on enforcement as well as on other corporate policies, including how staff are trained and where authority lies within management and boards for addressing hate. We sought to respect user privacy whenever possible and worked to ensure that our recommendations would not create a new avenue for government intervention on user speech outside of legally available remedies that respect the First Amendment, such as law enforcement seeking a warrant.

We recognize as well that companies will make mistakes. In addition, there are those who try to use anti-hate provisions against human rights and racial justice activists, leveraging those provisions to kick activists off social media platforms and other

internet services. Indeed, there is evidence this occurred in Myanmar as Rohingya activists were targeted and lost the ability to speak on an online platform in condemnation against their oppression.[12] We recommend addressing this through engaging experts, training staff, and locating human content-reviewers with relevant expertise in affected regions and by employing people from those regions, while also providing an easy-to-use full appeal process by which users who have lost access to services can present evidence in their defense and have their appeal heard by a neutral party.

Finally, there is still a lot to learn about how technologies, corporate policies, and even our recommendations work to limit hate online. Thus, we recommend broad transparency that involves internet companies providing regularly updated, easily accessible data on how their policies are affecting the spread of hate on their services. This should make it possible for academics and other experts to do the necessary research to determine what is working—and what is not—as well as whether there are any unintended consequences to corporate policies, including policies which we recommend.

Even as we approach reducing hateful activities online through recommended changes to corporate policies and terms of service, as opposed to through government intervention, there are other, concurrent debates on how online tools influence our lives. These discussions include determining how to address issues such as privacy and data protection, political advertising, competition, and coordinated foreign state efforts to use technology to undermine democracy. In these spaces, there is a central role for government action through legislation, oversight, and regulation—just as there was before the internet. Our recommended corporate policies are not meant to be a substitute for a necessary debate about what that legislation and regulation should entail.

This report details how those inciting hate are using technologies to grow their audiences; to target people based on their essential characteristics, such as race, religion, gender, LGBTQ status, immigrant status, among others; and to fund their activities. It outlines our research and analysis, shares what we learned, and includes a summary of our recommended policies. In addition, we discuss some of the thorny issues that will naturally arise when attempting to balance our desire for broad, accessible speech and actions with an effective response to activities by individuals or groups driven by or capitalizing on hate.

A road map to our policies

The recommended corporate policies, including detailed explanations of the reasons why each policy reads as it does and why it is needed, are included as the Appendix in this report. We summarize here some key elements of our approach, including whom and what the policies cover.

The policies are meant to broadly encompass entities of any corporate form that perform and/or host any of the following services for internet users, whether the entity provides these services directly to the public, through intermediaries, or as an intermediary:

- Social media, video sharing, communications, marketing, or event scheduling/ticketing platforms
- Online advertising, whether directly, as a reseller, or through resellers
- Financial transactions and/or fundraising
- Public chat services or group communications
- Domain names, whether directly, as a reseller, or through resellers
- Websites, blogs, or message boards

Throughout this report and its recommended policies, we refer to these entities as "internet companies," or in the singular as "internet company." In the section "How deep to go in the stack: Internet companies covered by these policy recommendations," we discuss why we arrived at this set of companies.

This is followed by defining "hateful activities," or those that we recommend internet companies work to significantly reduce on their services. This term is defined in full in the Appendix:

> [W]e use the term 'hateful activities' to mean activities that incite or engage in violence, intimidation, harassment, threats, or defamation targeting an individual or group based on their actual or perceived race, color, religion, national origin, ethnicity, immigration status, gender, gender identity, sexual orientation, or disability.[13]

Why we developed this definition and why we think it is appropriate is described in the section "Our approach to defining hateful activities."

The remainder lays out in detail each of the specific recommendations, why they are important, and language that internet companies can adopt to be consistent with each recommendation. The recommendations include the following:

Terms of service or acceptable use policies: The internet company will clearly describe for the user that they may not use the company's services "to engage in hateful activities or use these services to facilitate hateful activities engaged in elsewhere, whether online or offline."[14]

Enforcement: The internet company will maintain the combination of technology, staffing, training, user flagging, including a trusted flagger program, and effective responses to flaggers, necessary to enforce its hateful activities rules.

Right of appeal: The internet company shall have an easy-to-use mechanism for users to appeal denial of services under the hateful activities policy.

Transparency: The internet company will provide a range of data on hateful activities on their services in a format such that the general public and researchers can determine the scope of hateful activities and what is working—and not working—to address those activities.

Evaluation and training: The internet company will establish a team of experts on hateful activities who train and support both programmers and content assessors. The internet company will routinely test any technology used to identify hateful activities to ensure it is not demonstrating bias. The internet company will locate assessment teams within affected communities and ensure they have an understanding of relevant social, political, and cultural history and context.

Governance and authority: The internet company will grant a designated member of senior management and a board committee ultimate authority for addressing hateful activities, as well as create a committee of outside experts to generate an annual report on the effectiveness of the company's efforts to curb hateful activities.

What we learned about hate online

In nine months of research and conversations with experts, we learned quite a bit about how hate groups benefit from using online platforms and how those groups organize online. We also examined how foreign state actors are using hateful messages to sow discord within pluralistic democracies such as the United States. What we learned ultimately informed our recommendations. However, a key lesson from our research is that there simply is not enough information available yet about hateful activities online. In this section, we delve a bit deeper into our findings in each of these areas.

Why online platforms matter to hate groups

Hate groups have long relied on online technology for communications and organizing purposes, because these technologies help them meet four key needs:

1. **The ability to operate anonymously:** Anonymity is important, because identifying as a member of a hate group, or participating in hateful activities, can lead to job loss, removal from community roles, general unpopularity,[15] or arrest if violence or other illegal activities are involved.
2. **The ability to reach a broad audience across a large geography:** This enables the ideologies of white supremacists and other hate groups, repugnant to most Americans, to find supporters more easily.
3. **An opportunity to indoctrinate new followers through a methodical approach:** This often involves introducing racist, misogynistic, Islamophobic, and anti-Semitic ideas to targeted individuals through humor and snark using mainstream social media platforms. Subsequently, these groups demonstrate via racist websites that there is a community that holds these same ideas, ultimately engaging in organizing efforts through message boards and chat apps.

4. **The ability to raise money:**[16] Raising funds for hate groups would be difficult without the internet, given the need for member anonymity and the geographic disbursal of members. Hate groups' strong online presence allows them to raise funds online in the same way others do: quickly, with a limited paper trail, and 24 hours a day.

Because of these benefits, hate groups have been early adopters of online technologies. This is evident in the oldest and largest member-based hate site, Stormfront.org,[17] which began as a message board for white supremacists in the early 1990s and continues today, a quarter-century later. Racists have similarly been early adopters of technologies that were never intended for hateful purposes. For example, hate groups used the gamer-oriented chat app Discord to organize for the deadly neo-Nazi Unite the Right rally in Charlottesville.[18]

Technology companies' early unintentional facilitation of hate online

Prior to the violent neo-Nazi march and murder of Heather Heyer in Charlottesville, many tech companies operating in the United States paid limited attention to how their technology might facilitate the outreach, growth, violence, and influence of hate groups.[19] For example, prior to Charlottesville and the investigations that followed, it was possible on the two largest online ad-buying platforms, Google and Facebook, to target and directly market to users who self-identified as racists and anti-Semites.[20] This lack of attention persisted for years, despite efforts by multiple civil rights organizations, including the Lawyers' Committee for Civil Rights Under Law,[21] Color of Change,[22] Southern Poverty Law Center, and the Anti-Defamation League, to encourage tech companies to more aggressively disallow hate on their platforms or services.

Internet companies' inaction is not only a result of inattention. Within a significant segment of the tech community, there is a strong commitment to a broad interpretation of free speech. This commitment is laudable in many ways and has proved important in some international democracy efforts.[23] But among many companies, it combined with a narrow focus on user growth to facilitate some large-scale, coordinated attacks against women, people of color, members of the LGBTQ community, and religious minorities. Some of the most egregious examples in the past few years include violent threats of gassing or cooking Jews, directed at Jewish reporters via Twitter.[24] Threats of murder and rape targeting women have proven a recurring problem online, including organized attacks on female video game critics and developers—known as Gamergate—via online platforms such as 4chan.[25]

White supremacists bent on indoctrination have also benefited from search algorithms that attempt to intuit what a user is searching for based on a user's prior search history and what other similar users have chosen from search results. Some social media platforms also drive content to a user based on what their algorithms predict the user wants to see constructed from their prior interests.[26] These algorithms can be gamed by hate groups using multiple techniques to raise the profiles of their websites in search results as well as their content views on social media. This is especially true when the algorithms are not weighted to remove racist or otherwise hateful lies from displayed content or to move such content off the first page of top search results.

Dylann Roof pointed to an online search as beginning his descent into the white supremacist online world of hate. One of the top results from Google when Roof typed in "black on white crime" was the website for the racist hate group Council of Conservative Citizens, where he found propaganda erroneously declaring the prevalence of black people assaulting and killing white people. This and similar inaccurate racist misinformation, as opposed to facts debunking such lies, dominated the front page of search results. Roof read and embraced this and other racist propaganda he found online, ultimately murdering nine African Americans in the basement of Mother Emanuel AME Church in Charleston, South Carolina.[27]

Hate groups' reliance on online fundraising tools

Technology has been essential for hate groups to fund their work. Donations to hate groups or purchases of racist paraphernalia—when it is used as a funding source for hateful activities—can be accomplished easily using online payment tools combined with back-end support from credit card companies. It also allows for some anonymity. As with other companies with an online fundraising presence, PayPal, given its flexibility and market share, was for many years a key tool for hate groups. Following Charlottesville, PayPal has worked to kick hate groups off its platform.[28] Other payment processors have taken some steps in this direction as well, but credit card companies have been more mixed in their responses.[29]

Eventbrite, a large platform for selling tickets to events, made changes to deny hate groups access to their services following a monthslong campaign by Color of Change, Southern Poverty Law Center, and CREDO Action.[30] When kicked off fundraising tools such as GoFundMe,[31] individuals affiliated with hate groups have taken to racist message boards and chat groups to let supporters know that they can fund their efforts

by making donations to seemingly innocuous fundraising requests on the same or similar platforms. While not ideal, when hate groups must hide their intentions in order to use these platforms, their reach and effectiveness are significantly decreased.

Governments' and political operatives' use of bots and troll armies to push hate

There is now significant evidence that Russia manipulated social media platforms to sow discord and influence elections in the United States and Europe.[32] In the United States, this effort included the use of social media and online ad purchases to deliver racist and anti-Muslim messages targeting users based on their ethnicity and prior search and social media histories.[33] U.S. intelligence officials believe that these efforts aimed to create tension among Americans by taking advantage of this nation's long history of racial discrimination and strife.[34] Whatever their purpose, these ads were racist and demeaned African Americans, Latinos, immigrants, and Muslim Americans. They were also widespread in their impact: Facebook has indicated that Russian-supported Facebook advertising may have reached 70 million Americans,[35] and similar ads appeared on Instagram.[36]

Russia's interference also included the use of thousands of bots—computer programs designed to impersonate a human on a social media platform—and coordinated groups of people, often referred to as "troll armies" or "web brigades." These efforts directed racist and anti-Muslim memes at targeted users across large social media platforms such as Twitter and Facebook.[37]

For a long time, print and broadcast advertising from political campaigns and outside groups such as political action committees has been regulated in the United States to require disclosure of who is paying for the advertising and to ensure that foreign parties are not involved.[38] More effective internet company corporate policies to reduce hateful activities are not a substitute for congressional review of how this appropriate federal regulatory role should extend to online electoral communications.

Elections outside the United States and Europe have also witnessed the use of hateful messages to exacerbate ethnic, religious, and racial tensions to encourage certain groups to turn out to vote and others to stay home out of fear for their safety. This includes the use of text messages to incite ethnic violence around elections in Kenya,[39] as well as social media posts in South Sudan.[40] In just one example of a creative local response, Kenyan civil society groups built networks of community

volunteers to respond to racist misinformation campaigns with rapid-response messaging. These efforts to inform the public countered false rumors of ethnic violence that were spread online to stoke hatred during the 2017 elections.[41]

Unfortunately, even worse examples exist outside the context of elections. The United Nations this year publicly and in strong terms called out the use of Facebook as a means by which military-affiliated groups in Myanmar spread racist disinformation targeting the Rohingya minority.[42] This use of social media was an essential part of the large-scale ethnic cleansing that has, according to the United Nations, forced more than 700,000 Rohingya to leave their homes to escape violence and discrimination in just the last year. On August 27 of this year, Facebook announced:[43]

> *The ethnic violence in Myanmar has been truly horrific. Earlier this month, we shared an update[44] on the steps we're taking to prevent the spread of hate and misinformation on Facebook. While we were too slow to act, we're now making progress – with better technology to identify hate speech, improved reporting tools, and more people to review content.*
>
> *Today, we are taking more action in Myanmar, removing a total of 18 Facebook accounts, one Instagram account and 52 Facebook Pages, followed by almost 12 million people. We are preserving data, including content, on the accounts and Pages we have removed.*
>
> *Specifically, we are banning 20 individuals and organizations from Facebook in Myanmar — including Senior General Min Aung Hlaing, commander-in-chief of the armed forces, and the military's Myawady television network. International experts, most recently in a report by the UN Human Rights Council-authorized Fact-Finding Mission on Myanmar, have found evidence that many of these individuals and organizations committed or enabled serious human rights abuses in the country. And we want to prevent them from using our service to further inflame ethnic and religious tensions.*

The need for more research

As we sought to accumulate information on these hateful activities online, we found it quite difficult to find comprehensive data on the effectiveness of internet companies' various approaches. Much of what we now know about hateful activities

online and internet companies' responses emerged only recently. Generally, this new knowledge has had three sources:

1. **Innovative research conducted jointly between companies and independent academics:** This occurs when internet companies make their data available to specific independent researchers. A good example of this is a partnership announced in April 2018 between Twitter and the Dangerous Speech Project at Harvard University, testing whether publishing clear, accessible terms of service creates positive norms for users.[45]
2. **Materials and information gleaned from leaks and undercover reporting efforts:** Examples of this include the recent leaks of Facebook moderation policies with regard to, among other things, distinctions between white supremacy, white nationalism, and white separatism.[46] Similarly, the Discord chat app was identified as a key alt-Right and neo-Nazi organizing tool based on undercover journalism by *The New York Times*[47]and Unicorn Riot.[48]
3. **Internet companies' provision of data on their websites in easy-to-understand formats:** While not yet common across internet companies, Google has an easy-to-understand transparency report[49] that shares basic information on videos removed from YouTube and the role of automated and human flagging in these removals.

These are all important tools to understand hateful activities online, but they are haphazard and do not answer a wide range of questions about most internet apps. Questions that remain as we try to tackle hateful activities online include:

- What is the full range of methods that hate groups and other hateful actors are using to indoctrinate people, especially young people, online?
- What are the points at which this indoctrination can most effectively be stopped?
- What technologies can be developed to assist companies in identifying hate online and stopping it on their platforms? How can these companies integrate these technologies into their platforms?
- What trainings and knowledge must tech company employees have in order to develop platforms that are effective at keeping hate groups from using their technologies to indoctrinate users into hate?
- How is online hate financed?

We believe that a significant change in how internet companies make data on hateful activities available could help answer these and similar questions. Simply put, we believe in transparency. We call for tech companies to regularly update and make machine- and human-readable data available online. This would include data on the volume of hateful content and whom it is targeting, the effectiveness of different responses, the impact of different kinds of flaggers, how many people effectively appeal being denied services, and several other fields. We believe that this will generate a rapid and significant boost in the kinds of research that are being completed, as well as in the lessons learned about hateful activities.

To understand what this might look like, consider the accessibility of data on U.S. voter behavior that are made publicly available, generally online, in large data sets by the U.S. Census Bureau and each of the 50 state election authorities. This has led to thousands of research projects emerging from graduate students and professors on a wide range of related topics. Research institutes and university departments regularly derive new lessons from these data. As a result, we know a lot about how media and money affect elections, how people identify with candidates, what issues matter to voters, the numerical insignificance of voter fraud, and a range of other important information that helps voters make informed choices and legislators develop good policy.

If data on how hateful activities occur online become widely available, we can expect similar new research on what works and what does not work to stop hateful activities. There will be new ideas, new technologies, and new institutions built and designed to study and make use of the data.

To see the range of data that we recommend internet companies make available, see the "Transparency" section of the Appendix.

Improvements in internet companies' responses to hate

Internet companies have taken major steps forward in multiple areas related to these issues. We should acknowledge these efforts to help curb the growth of hate online, to limit the use of internet tools to organize among hate groups, and to address the recruitment of new people into hate groups. We point out once again that all these companies already have terms of service or acceptable use policies that attempt to limit hateful activities. Our efforts are to make these policies more effective, fair, and transparent. We are sharing these improvements not to indicate that the efforts to date are sufficient; rather, we think these successes demonstrate that the recommen-

dations we are making are reasonable and achievable—even if they require companies to take new approaches and to be open to greater reflection and transparency.

It is also worth noting that internet companies have taken successful and often creative steps to curb other kinds of inappropriate content on their services, including redirecting people looking for Islamic State group terrorist recruitment videos,[50] debunking terrorism recruitment tropes, and striking comprehensive deals with record labels[51] to allow users to add popular music to the background of their videos without violating copyright rules. These past successes demonstrate that companies can develop inspired and substantial solutions to content problems that present different kinds of challenges.

This list is by no means comprehensive, but the following examples reflect significant improvements by internet companies:

- At the end of 2014, Apple removed 30 white supremacist bands[52] from its iTunes store. This was an early move by Apple that helped set a standard for other downloading and streaming services.

- Facebook has undertaken efforts to significantly increase the number of people reviewing its content. When counting full-time employees and contractors, that number is now 7,500. In a July 26, 2018, post, Facebook explained:[53]

  > *In addition to language proficiency, we also look for people who know and understand the culture. For example we want to hire Spanish speakers from Mexico — not Spain — to review reports from Mexico as it often takes a local to understand the specific meaning of a word or the political climate in which a post is shared.*

  This is consistent with our recommendation that internet companies should "locate assessment teams enforcing the hateful activities rules within affected communities to increase understanding of cultural, social, and political history and context."[54]

- Three days after the deadly neo-Nazi Unite the Right rally in Charlottesville, PayPal issued a strong statement[55] decrying the loss of life and intolerance and stating its commitment:

  > *Regardless of the individual or organization in question, we work to ensure that our services are not used to accept payments or donations for activities that promote hate, violence or racial intolerance. This includes organizations that advocate racist views, such as the KKK [Ku Klux Klan], white supremacist groups or Nazi groups.*

This followed months of advocacy by Color of Change,[56] which reported that PayPal, consistent with its statement, dropped a number of organizations that Color of Change had flagged for the company as engaging in hateful activities.

- Twitter announced on September 25 of this year that it will expand its hateful conduct policy to include:[57]

  > *content that dehumanizes others based on their membership in an identifiable group, even when the material does not include a direct target. Many scholars have examined the relationship between dehumanization and violence. For example, Susan Benesch has described dehumanizing language as a hallmark of dangerous speech, because it can make violence seem acceptable,[58] and Herbert Kelman has posited that dehumanization can reduce the strength of restraining forces against violence.*[59]

  This is an exciting new approach, because it reflects research that seeks to understand how humans are motivated to engage in mass atrocities with an understanding that an initial step is dehumanization. This dehumanization includes perpetrators denigrating victims as less than human and comparing them to vermin, insects, and viruses.[60]

- YouTube, owned by Google, has developed and used artificial intelligence to identify problematic videos. This has been necessary because of the sheer volume of user content, reportedly hundreds of hours uploaded every minute,[61] as well as the range of concerns relevant to their platform—including not only hateful activities but also terrorism and copyright infringement. Google has also become increasingly transparent about its efforts to remove content with easy-to-understand visualizations that describe flagging and content removals on YouTube.[62] In this way, it is easy to see what efforts appear to have the most impact and where additional information might be helpful.

Our approach to defining hateful activities

In the Appendix, we share in full the recommended corporate policies and terms of service, as well as a detailed accounting of our reasons for each. In the body of this report, we explain only our definition of hateful activities, which underlies and informs all our recommendations. We aimed to balance a commitment to speech with the need to reduce hateful activities online.

Our goal was to have a definition that was understandable to users and enforceable by internet companies. This starts with internet companies defining clearly what is not allowed on their services. We recommend describing disallowed actions as "hateful activities," and we define these as:

> *activities that incite or engage in violence, intimidation, harassment, threats, or defamation targeting an individual or group based on their actual or perceived race, color, religion, national origin, ethnicity, immigration status, gender, gender identity, sexual orientation, or disability.*

We believe this definition is clear and avoids murky language that can be found in some internet companies' terms of service. For example, we avoid circular language that describes hate speech as speech that involves hate. Equally important, we focus less on the idea of speech alone being violative and instead look to whether three things are present:[63]

1. Does the user "incite or engage in" a defined activity?
2. Does that defined activity constitute "violence, intimidation, harassment, threats, or defamation"?
3. Has the user targeted an individual or group based on a limited set of personal characteristics?

Only if all three occur is there a violation of the recommended policy.

Balancing the thorny issues in our recommendations

Our goal in developing these policies is to reduce the amount of hateful activities online while maintaining a strong commitment to both free speech and user privacy. To accomplish this, we brought together organizations that see these issues differently but are concerned about all of them. We also brought in experts who did not share our views on the correct approach but had thought deeply about the issues involved. They shared their thinking, thereby significantly improving our recommended policies.

Ultimately, this work is a balancing act with no one solution that trumps all others. There are some difficult challenges, and we felt it was important to explain how we think about these.

Our balancing act includes recommending that companies clearly define hateful activities and that they deny the right to engage in these on their services. But it also includes recommending key components to protect users' speech and privacy, such as ensuring a meaningful user right-to-appeal process, arbitrated by a neutral party not involved in the original decision, as well as transparency into whom these policies affect.

Addressing concerns about speech

The most obvious concern someone might have is the idea that portions of our recommended policies involve censorship and, in some way, violate the spirit of the First Amendment. The companies are not the government and thus, neither the letter, nor the spirit, of the First Amendment apply. It is important to note that internet companies already have policies about what content they will tolerate on their platforms, and almost all already say they will not allow hateful activities. We are providing them with the ability to live up to that laudable goal in a way that will be more effective, fairer to users, and more transparent to the public.

Internet companies have their own First Amendment right to facilitate what communications are allowed on their platforms. As long as they use this right to have

terms of service that allow them to kick people off the services or otherwise limit their use of the services—and virtually all of the companies do—then the question is not whether this is appropriate, but rather whether the relevant corporate policies and rules applied to users have sufficient safeguards to limit denials of services only to those who have clearly and intentionally engaged in hateful activities. We believe our recommendations accomplish this.

We did come across the argument that tools such as Facebook and Twitter are the new public commons and thus, virtually all speech should be given extra protection as would be the case in a public park.[64] This argument does not reflect a few important realities: These internet companies' platforms are not the public commons; they are not publicly owned, and not everything that is said is heard—in part because platform algorithms often determine what users see or hear to better drive ad revenue and improve the user experience. Finally, these companies' platforms function increasingly like for-profit TV and radio broadcasters, as they build the size of their audience by providing original content and content developed by others and then generating revenue through ad sales. Legacy radio and TV broadcasters have for a long time heavily regulated what viewers hear and see.

How deep to go in the stack: Internet companies covered by these policy recommendations

In simplified terms, it is possible to think of the internet as a stack with a broadband internet access service provider—a company such as Comcast or Verizon—at the bottom of the stack and apps, or tools such as Twitter, Spotify, or Amazon, at the top. These app-providing companies are sometimes referred to as edge providers. In between are companies that provide a range of services that make it possible for edge providers to deliver their services and for users to gain effective access to those services.

One question that our coalition addressed when developing our recommended corporate policies and terms of service pertains to which companies should be covered. Our policies ultimately cover companies that provide or facilitate the following services:

- Social media, video sharing, communications, marketing, or event scheduling/ticketing platforms
- Online advertising, whether directly, as a reseller, or through resellers
- Financial transactions and/or fundraising
- Public chat services or group communications

- Domain names, whether directly, as a reseller, or through resellers
- Websites, blogs, or message boards

We chose to go a bit deeper into the stack than some others might have. For example, while we chose to go beyond social media platforms to include payment processors, domain name providers, and website hosts, we specifically exclude broadband internet access service providers and end-to-end encrypted communications from our recommended policies. These were relatively easy calls for us, because the companies we include are primarily user-facing. A website-hosting company such as Wix or domain provider such as GoDaddy interact directly with their customers and are able to relatively quickly establish the nature of their content. Complaints to these companies about hateful activities on their services can be reviewed and addressed relatively easily.

While this is true for payment processors as well, there is an additional truth about any involvement of a payment processor in funding hateful activities: Payment processors generally make money by charging a percentage, often between 2 percent and 5 percent, of the cost of an item as a fee for using their service. Thus, when members of hate groups used crowdfunding tools to raise money to attend the Charlottesville rally, the companies made money off a racist, violent, and, ultimately, deadly event.

On the other hand, there are some internet companies where users have a clear expectation that the company has no access to, or will never access, their content and thus cannot make decisions about whether that content is violative of the recommended policies. It is not this user expectation alone that drives our position on this matter but also our belief that an open internet requires that certain service providers should not review content or be able to use a review of content to make decisions about access to the internet. We believe so strongly in this that we do not believe that these kinds of companies should be involved in implementing these recommended policies.

For example, broadband internet access service providers, companies such as Verizon or Comcast that provide internet access to most homes, schools, and businesses, should not review content or make any changes to service provision based on such a review. This is one element of what is commonly called net neutrality. Unfortunately, the Federal Communications Commission has recently undermined this view of net neutrality,[65] but we believe it is essential to maintaining an open internet.

Also excluded in our policies are end-to-end encrypted chat or communications services. Central to these tools is the ability to keep prying eyes and ears, including those of governments, away from content—the large majority of which has nothing to do with hateful activities—that users may want to keep confidential for any number of reasons. Protecting users' ability to rely on these services without corporate content review allows for everyone, from business people to human rights activists, to engage in meaningful communications.

This is not to say that we believe companies providing encrypted chat services can do nothing to stop hateful activities; we mean only that they should not review purportedly encrypted content when doing so. WhatsApp had to address this when extreme violence was linked to the rapid spread via their group chat app of dangerous, racist lies about minority groups in Myanmar and India. Globally, WhatsApp limited the number of people that messages could be forwarded to and further limited its quick forwarding functionality with even stricter rules in India.[66]

This leaves our policies silent on a range of other types of service providers that are deeper in the stack than the edge providers or social media apps but higher in the stack than the Verizons and Comcasts of the world. More analysis is needed before comprehensive policies of this nature can or even should be developed for these types of services and companies.

What about Black Lives Matter or #OscarsSoWhite?

In our conversations with people, organizations, and internet companies that care about getting moderation of social media platforms and communication tools right, one specific challenge was raised: How will any approach to reduce hateful activities affect organizations committed to social justice and civil rights? The concern was that in attempting to keep the Ku Klux Klan (KKK) from recruiting new members online, we might inadvertently get committed human or civil rights organizations or activists kicked off platforms that allow them to broadly educate the public and organize events to advocate for criminal justice reforms. Their concerns are based in the reality that rule enforcement in a wide variety of contexts is subject to existing biases and has a history of negatively affecting marginalized or historically disenfranchised people. This concern was often framed in the form of this question: What about Black Lives Matter[67] or #OscarsSoWhite?[68]

We were cognizant of this very real concern and its potential ramifications. In October 2017, the Congressional Black Caucus (CBC) wrote to the FBI direc-

tor about an FBI intelligence assessment titled "Black Identity Extremists Likely Motivated to Target Law Enforcement Officers." The CBC letter said the assessment concluded that these so-called extremists "are likely to target law enforcement based on 'perceptions of police brutality against African Americans.'"[69] As the CBC members noted, not only was there no evidence to support this claim, but it was also consistent with the FBI's historic targeting of African American civil rights activists such as Martin Luther King, Jr. The assessment was, in their words, "flawed because it conflates black political activists with dangerous domestic terrorist organizations that pose actual threats to law enforcement." In the current political context, we took seriously the need to ensure that efforts to address hateful activities online do not threaten civil and human rights activists.

We balanced this with the reality that hateful activities online incite hate crimes and violence; create an atmosphere of fear and distrust; and chill speech and civic participation. And we understand that hateful activities online particularly do all of these things to marginalized and historically disenfranchised people.

The KKK, for example, has a long history of engaging in violence, intimidation, harassment, and threats targeting African Americans because of their race, and Jews because of their religion—along with a range of other people. This is part of the KKK's reason for existing. New manifestations of this kind of hate are present in the Unite the Right neo-Nazis who planned violence in Charlottesville. They encouraged attendees to bring weapons and chanted "Jews will not replace us" as their marching song that was widely shared on social media.[70]

But this is not the case for Black Lives Matter, a movement committed to ending disparities in policing of people of color, to protesting the killing of black people by police officers, and to championing broader criminal justice reforms. Returning to our definition of hateful activities, we are aware of no effort by Black Lives Matter to incite or engage in any of the barred activities, namely violence, intimidation, harassment, threats, or defamation, or to do so while targeting people based on the protected categories—race, color, religion, among others. The acknowledgement of disparities in how our criminal justice system treats people of different races or ethnicities is simply not covered by our definition.

Mentioning or discussing race or racism is not sufficient to be considered a hateful activity. The activities of activists involved in #OscarsSoWhite, an effort to highlight the disparities in the nomination of films by the Academy of Motion Pictures Arts and Sciences directed by, written by, or starring people of color, thus, are also not

restricted under our recommended policies. Highlighting the failure of the academy to embrace diversity is intended to encourage a recognition of that diversity and to create opportunities for artists of color—not an effort to harm artists who are white. This movement does not engage in violence, intimidation, harassment, or threats.

This does not mean that historically marginalized groups are immune from being removed from a service or from having their access to the service limited. For instance, consider the following made-up tweet by someone allegedly concerned about lack of representation in the movie awards:

> *If more Asian, African American, and Latino people are not nominated next year, then we should start beating up white people in Los Angeles outside the ceremony—visit our group's website to learn more. #OscarsSoWhite*

This would violate our hateful activities policy, not because of the hashtag or the purported concern about diversity but because it incites violence targeting people because of their race.

To really address this concern of unequal enforcement of those fighting against discrimination, we specifically recommend that internet companies do three things:

1. Hire recognized experts who have a demonstrated expertise on hate, such as peer-reviewed publications and solid academic credentials directly relevant to germane topics, to advise programmers, develop training content, and oversee training of assessors. The training materials should then be available to the public for review.
2. Routinely test any technology used to identify hateful activities to ensure that such technology is not biased against individuals or groups based on their actual or perceived race, color, religion, national origin, ethnicity, immigration status, gender, gender identity, sexual orientation, or disability.
3. Locate assessment teams enforcing against hateful activities within affected communities to increase understanding of cultural, social, and political history and context.

Ultimately, with these safeguards in place, we believe that internet companies can appropriately reduce the risk of curbing the activism and voice of those seeking positive social change, especially those from marginalized or historically disenfranchised groups.

Conclusion

For several reasons, internet companies will be in different places as they embark on addressing the issues raised in the recommended corporate policies and terms of service. First, many companies have already undertaken significant steps, whether driven by altruism, employee concerns, a commitment to human rights, or being publicly blamed for violence or other hateful activities—or a combination of all of these. Second, organizations have different business models and reasons people use their services. For example, some companies' services are committed to user anonymity, while others require users to accurately identify who they are—neither is good nor bad. There are reasons for the range of user experiences that go into making a diverse and highly functioning internet. But user experiences in whatever form cannot be a reason to allow for hateful activities on a service. Instead, companies will need different, unique approaches. Finally, technologies are ever changing. Large platforms are introducing features, and startups are creating new ways for people to communicate, share ideas, and raise money. These new technologies will raise new challenges for addressing hateful activities.

What is important is that internet companies, at all stages of their development, prioritize reducing hateful activities on their services. We believe the recommended policies and terms of service will help them do just that.

About the author

Henry Fernandez is a senior fellow at American Progress. For more than a decade, he has researched and written on the influence of hate groups on mainstream public policy debates. He has led legislative and electoral campaigns, including large online and broadcast media efforts as well as technology-based and traditional grassroots organizing initiatives. His current work includes evaluating the uses of social media to influence voter behavior and the efficacy of various online voter registration drives. He is a graduate of Harvard College and Yale Law School.

Acknowledgements

The author would like to thank his co-chairs for the committee that came together to examine hateful activities online: Heidi Beirich of the Southern Poverty Law Center and Jessica Gonzalez of Free Press. This work would not have been nearly as thoughtful or useful without significant input from the following contributing organizations: Color of Change, Free Press, the Lawyers' Committee for Civil Rights Under Law, the National Hispanic Media Coalition, and the Southern Poverty Law Center. Finally, Jessica Cobian, Philip E. Wolgin, and Tom Jawetz on the Immigration Policy team at the Center for American Progress provided important insights and shepherded this work to completion.

Endnotes

1 Emanuella Grinberg, "What #Ferguson stands for besides Michael Brown and Darren Wilson," CNN, November 19, 2014, available at http://www.cnn.com/2014/11/19/us/ferguson-social-media-injustice/index.html.

2 Danielle Kwateng-Clark, "TIME's 'Person Of The Year' Honorees Include Black Women And Men Who Took A Stand Against Sexual Harassment," *Essence*, December 6, 2017, available at https://www.essence.com/news/time-person-year-2017-black-honorees-metoo-sexual-harassment/.

3 Ray Sanchez, "Thousands march in Boston in protest of controversial rally," CNN, August 19, 2017, available at https://www.cnn.com/2017/08/19/us/boston-free-speech-rally/index.html.

4 Ali Breland, "Charlottesville rally leaders used PayPal to organize event," *The Hill*, August 15, 2017, available at https://thehill.com/policy/technology/346661-charlottesville-rally-leaders-used-paypal-to-coordinate-and-organize.

5 Alex Heath, "Facebook removed the event page for white nationalist 'United the Right' rally in Charlottesville one day before it took place," *Business Insider*, August 14, 2017, available at https://www.businessinsider.com/facebook-removed-unite-the-right-charlottesville-rally-event-page-one-day-before-2017-8.

6 Tom McKay, "Judge Rules Discord Must Turn Over Account Data of Neo-Nazis in Charlottesville Planning Server," Gizmodo, August 7, 2018, available at https://gizmodo.com/judge-rules-discord-must-turn-over-account-data-of-neo-1828180427.

7 Kurt Wagner, "WhatsApp will drastically limit forwarding across the globe to stop the spread of fake news, following violence in India and Myanmar," Recode, July 19, 2018, available at https://www.recode.net/2018/7/19/17594156/whatsapp-limit-forwarding-fake-news-violence-india-myanmar.

8 Ibid.

9 In the next section, we give further definition of which companies are included in our policy recommendations when we use the term "internet companies."

10 Nick Statt, "Facebook continues to let advertisers racially discriminate in housing ads," The Verge, November 21, 2017, available at https://www.theverge.com/2017/11/21/16686524/facebook-housing-advertisements-discrimination-race.

11 Legal Information Institute, "18 U.S. Code § 2258A – Reporting requirements of electronic communication service providers and remote computing service providers," available at https://www.law.cornell.edu/uscode/text/18/2258A (last accessed October 2018).

12 Betsy Woodruff, "Exclusive: Facebook Silences Rohingya Reports of Ethnic Cleansing," The Daily Beast, September 18, 2017, available at https://www.thedailybeast.com/exclusive-rohingya-activists-say-facebook-silences-them.

13 Center for American Progress and others, "Appendix: Recommended Internet Company Corporate Policies and Terms of Service to Reduce Hateful Activities" (2018), available at https://cdn.americanprogress.org/content/uploads/2018/10/24111621/ModelInternetCompanies-appendix.pdf.

14 Ibid.

15 Corrine Segal, "White supremacists once wore hoods. Now, an internet mob won't let them stay anonymous," PBS, August 20, 2017, available at https://www.pbs.org/newshour/nation/white-supremacists-wore-hoods-now-internet-mob-wont-let-stay-anonymous.

16 Franz Paasche, "PayPal's AUP – Remaining Vigilant on Hate, Violence and Intolerance," Paypal, August 15, 2017, available at https://www.paypal.com/stories/us/paypals-aup-remaining-vigilant-on-hate-violence-intolerance.

17 Southern Poverty Law Center, "Stormfront," available at https://www.splcenter.org/fighting-hate/extremist-files/group/stormfront. (last accessed October 2018)

18 Unicorn Riot, "Charlottesville Violence Planned Over Discord Servers: Unicorn Riot Reports," September 5, 2017, available at https://unicornriot.ninja/2017/charlottesville-violence-planned-discord-servers-unicorn-riot-reports/; Kevin Roose, "This Was the Alt-Right's Favorite Chat App. Then Came Charlottesville," *The New York Times*, August 15, 2017, available at https://www.nytimes.com/2017/08/15/technology/discord-chat-app-alt-right.html.

19 Issie Lapowsky, "Tech Companies Have the Tools to Confront White Supremacy," *Wired*, August 14, 2017, available at https://www.wired.com/story/charlottesville-social-media-hate-speech-online/; Julia Carrie Wong, "Tech companies turn on Daily Stormer and the 'alt-right' after Charlottesville," *The Guardian*, August 14, 2017, available at https://www.theguardian.com/technology/2017/aug/14/daily-stormer-alt-right-google-go-daddy-charlottesville.

20 Sapna Maheshwari and Alexandra Stevenson, "Google and Facebook Face Criticism for Ads Targeting Racist Sentiments," *The New York Times*, September 15, 2017, available at https://www.nytimes.com/2017/09/15/business/facebook-advertising-anti-semitism.html.

21 Lawyers' Committee for Civil Rights Under Law, "Lawyers' Committee for Civil Rights Under Law Sends Letter Demanding Facebook Revise Policies Empowering White Supremacists and White Nationalists," Press release, September 20, 2018, available at https://lawyerscommittee.org/press-release/lawyers-committee-for-civil-rights-under-law-sends-letter-demanding-facebook-revise-policies-empowering-white-supremacists-and-white-nationalists/.

22 Color of Change, "Color of Change Statement on PayPal," available at https://colorofchange.org/press_release/color-change-statement-paypal/ (last accessed October 2018).

23 Heather Brown, Emily Guskin, and Amy Mitchell, "The Role of Social Media in the Arab Uprisings," Pew Research Center, November 28, 2012, available at http://www.journalism.org/2012/11/28/role-social-media-arab-uprisings/.

24 Anti-Defamation League Task Force on Harassment and Journalism, "Anti-Semitic Targeting of Journalists During the 2016 Presidential Campaign" (2016), available at https://www.adl.org/sites/default/files/documents/assets/pdf/press-center/CR_4862_Journalism-Task-Force_v2.pdf.

25 Caitlin Dewey, "The only guide to Gamergate you will ever need to read," *The Washington Post*, October 14, 2014, available at https://www.washingtonpost.com/news/the-intersect/wp/2014/10/14/the-only-guide-to-gamergate-you-will-ever-need-to-read/?noredirect=on&utm_term=.0e443ae658c2.

26 Casey Newton, "How YouTube Perfected the Feed," The Verge, August 30, 2017, available at https://www.theverge.com/2017/8/30/16222850/youtube-google-brain-algorithm-video-recommendation-personalized-feed.

27 Rebecca Hersher, "What Happened When Dylann Roof Asked Google For Information About Race?", NPR, January 10, 2017, available at https://www.npr.org/sections/thetwo-way/2017/01/10/508363607/what-happened-when-dylann-roof-asked-google-for-information-about-race.

28 Tracy Jan, "PayPal escalates the tech industry's war on white supremacy," *The Washington Post*, August 16, 2017, available at https://www.washingtonpost.com/news/wonk/wp/2017/08/16/paypal-escalates-the-tech-industrys-war-on-white-supremacy/?noredirect=on&utm_term=.d12c0293f104.

29 Kevin Dugan, "Credit cards are clamping down on payments to hate groups," *New York Post,* August 16, 2017, available at https://nypost.com/2017/08/16/credit-cards-are-clamping-down-on-payments-to-hate-groups/.

30 Eventbrite, "Gathering for Connection, Expression, and Change," Medium, August 17, 2017, available at https://medium.com/@eventbrite/gathering-for-connection-expression-and-change-635b3cab7ac6.

31 NBC News, "Silicon Valley Kicks Hate Groups Offline," YouTube, August 17, 2017, available at https://www.youtube.com/watch?v=7KnvC5RzVSg.

32 Scott Shane, "The Fake Americans Russia Created to Influence the Election," *The New York Times*, September 7, 2017, available at https://www.nytimes.com/2017/09/07/us/politics/russia-facebook-twitter-election.html?hp&action=click&pgtype=Homepage&clickSource=story-heading&module=first-column-region&region=top-news&WT.nav=top-news&_r=1.

33 Tim Lister and Clare Sebastian, "Stoking Islamophobia and secession in Texas — from an office in Russia," CNN, October 6, 2017, available at https://www.cnn.com/2017/10/05/politics/heart-of-texas-russia-event/index.html.

34 Matthew Rosenberg, Charlie Savage, and Michael Wines, "Russia Sees Midterm Elections as Chance to Sow Fresh Discord, Intelligence Chiefs Warn," *The New York Times*, February 3, 2018, available at https://www.nytimes.com/2018/02/13/us/politics/russia-sees-midterm-elections-as-chance-to-sow-fresh-discord-intelligence-chiefs-warn.html?module=inline.

35 Ben Collins, Kevin Poulsen, and Spencer Ackerman, "Russia's Facebook Fake News Could Have Reached 70 Million Americans," The Daily Beast, September 8, 2017, available at https://www.thedailybeast.com/russias-facebook-fake-news-could-have-reached-70-million-americans.

36 Alex Pasternack, "Russia's U.S. Propaganda Campaign Infiltrated Instagram, Too," *Fast Company*, October 6, 2017, available at https://www.fastcompany.com/40478430/russia-linked-instagram-facebook-posts-ads-memes-propaganda.

37 April Glaser, "What Was Russia Up To?", Slate, October 11, 2017, available at http://www.slate.com/articles/technology/future_tense/2017/10/what_we_know_about_russia_s_use_of_american_facebook_twitter_and_google.html.

38 See Legal Information Institute, "52 U.S. Code § 30121 – Contributions and donations by foreign nationals," available at https://www.law.cornell.edu/uscode/text/52/30121 (last accessed October 2018).

39 Ofeibea Quist-Arcton, "Text Messages Used to Incite Violence in Kenya," NPR, February 20, 2008, available at https://www.npr.org/templates/story/story.php?storyId=19188853.

40 Justin Lynch, "In South Sudan, Fake News Has Deadly Consequences," Slate, June 9, 2017, available at http://www.slate.com/articles/technology/future_tense/2017/06/in_south_sudan_fake_news_has_deadly_consequences.html.

41 Kira Zalan," Keeping the Peace Via Text," U.S. News and World Report, August 3, 2017, available at https://www.usnews.com/news/best-countries/articles/2017-08-03/kenyans-turn-to-technology-to-prevent-election-violence.

42 Tom Miles, "U.N. investigators cite Facebook role in Myanmar crisis," Reuters, March 12, 2018, available at https://www.reuters.com/article/us-myanmar-rohingya-facebook/u-n-investigators-cite-facebook-role-in-myanmar-crisis-idUSKCN1GO2PN.

43 Facebook, "Removing Myanmar Military Officials From Facebook," Press release, August 28, 2018, available at https://newsroom.fb.com/news/2018/08/removing-myanmar-officials/.

44 Sara Su, "Update on Myanmar," Facebook, August 15, 2018, available at https://newsroom.fb.com/news/2018/08/update-on-myanmar/.

45 Susan Benesch and J. Nathan Matias," Launching today: new collaborative study to diminish abuse on Twitter," Medium, April 6, 2018, available at https://medium.com/@susanbenesch/launching-today-new-collaborative-study-to-diminish-abuse-on-twitter-2b91837668cc.

46 Richard Lawler, "Leaked Facebook documents show its shifting hate speech policies," Engadget, May 25, 2018, available at https://www.engadget.com/2018/05/25/facebook-moderation-leak/.

47 Roose, "This was the Alt-Right's Favorite Chat App. Then Came Charlottesville."

48 Unicorn Riot, "Unite the Right," available at https://unicornriot.ninja/tag/unite-the-right/ (last accessed October 2018).

49 Google, "Google Transparency Report," available at https://transparencyreport.google.com/?hl=en (last accessed October 2018).

50 Mike Snider, "YouTube redirects ISIS recruits to anti-terrorist videos," *USA Today,* July 20, 2017, available at https://www.usatoday.com/story/tech/talkingtech/2017/07/20/youtube-redirects-isis-recruitment-searches-anti-terrorist-videos/497392001/.

51 Josh Constine, "Facebook allows videos with copyrighted music, tests Lip Sync Live," TechCrunch, June 5, 2018, available at https://techcrunch.com/2018/06/05/facebook-lip-sync-live/.

52 Southern Poverty Law Center, "Itunes pulls hate music following SPLC report, Amazon and Spotify slow to act," December 8, 2014, available at https://www.splcenter.org/news/2014/12/08/itunes-pulls-hate-music-following-splc-report-amazon-and-spotify-slow-act.

53 Ellen Silver, "Hard Questions: Who Reviews Objectionable Content on Facebook—And Is the Company Doing Enough to Support Them?", Facebook, July 26, 2018, available at https://newsroom.fb.com/news/2018/07/hard-questions-content-reviewers/.

54 Center for American Progress and others, "Appendix."

55 Paasche, "PayPal's AUP – Remaining Vigilant on Hate, Violence and Intolerance."

56 Daniel Terdiman, "After Charlottesville, PayPal says it won't do business with hate groups," *Fast Company*, August 15, 2017, available at https://www.fastcompany.com/40454274/after-charlottesville-paypal-says-it-wont-do-business-with-hate-groups.

57 Vijaya Gadde and Del Harvey, "Creating new policies together," Twitter, September 25, 2018, available at https://blog.twitter.com/official/en_us/topics/company/2018/Creating-new-policies-together.html.

58 Dangerous Speech Project, "What is Dangerous Speech?", available at https://dangerousspeech.org/the-dangerous-speech-project-preventing-mass-violence/ (last accessed October 2018).

59 Herbert C. Kelman, "Violence without moral restraint: Reflections on the dehumanization of victims and victimizers," *Journal of Social Issues* 29 (4) (1973): 25–61.

60 Anna Szilagyi, "Dangerous Metaphors: How Dehumanizing Rhetoric Works," Dangerous Speech Project, March 8, 2018, available at https://dangerousspeech.org/dangerous-metaphors-how-dehumanizing-rhetoric-works/.

61 Daisuke Wakabayashi, "YouTube Sets New Policies to Curb Extremist Videos," *The New York Times*, June 18, 2017, available at https://www.nytimes.com/2017/06/18/business/youtube-terrorism.html.

62 Google, "Google Transparency Report."

63 Center for American Progress and others, "Appendix."

64 Heather Whitney, "Does the Packingham Ruling Presage Greater Government Control Over Search Results? Or Less?", Technology and Marketing Law Blog, June 22, 2017, available at https://blog.ericgoldman.org/archives/2017/06/does-the-packingham-ruling-presage-greater-government-control-over-search-results-or-less-guest-blog-post.htm.

65 Brian Fung, "The FCC's net neutrality rules are officially repealed today. Here's what that really means," *The Washington Post*, June 11, 2018, available at https://www.washingtonpost.com/news/the-switch/wp/2018/06/11/the-fccs-net-neutrality-rules-are-officially-repealed-today-heres-what-that-really-means/?utm_term=.1d2f45ec2656.

66 Wagner, "WhatsApp will drastically limit forwarding across the globe to stop the spread of fake news, following violence in India and Myanmar."

67 Black Lives Matter, "Home," available at https://blacklivesmatter.com/ (last accessed October 2018).

68 Patrick Ryan, "#OscarSoWhite controversy: What you need to know," *USA Today*, February 2, 2016, available https://www.usatoday.com/story/life/movies/2016/02/02/oscars-academy-award-nominations-diversity/79645542/.

69 Letter from Cedric Richmond and others to Christopher Wray, October 13, 2017, available at https://cbc.house.gov/uploadedfiles/cbc_rm_thompson_cummings_conyers_letter_to_fbi_re_intel_assessment.pdf.

70 Emma Green, "Why the Charlottesville Marchers Were Obsessed With Jews," *The Atlantic*, August 15, 2017, available at https://www.theatlantic.com/politics/archive/2017/08/nazis-racism-charlottesville/536928/.

Our Mission

The Center for American Progress is an independent, nonpartisan policy institute that is dedicated to improving the lives of all Americans, through bold, progressive ideas, as well as strong leadership and concerted action. Our aim is not just to change the conversation, but to change the country.

Our Values

As progressives, we believe America should be a land of boundless opportunity, where people can climb the ladder of economic mobility. We believe we owe it to future generations to protect the planet and promote peace and shared global prosperity.

And we believe an effective government can earn the trust of the American people, champion the common good over narrow self-interest, and harness the strength of our diversity.

Our Approach

We develop new policy ideas, challenge the media to cover the issues that truly matter, and shape the national debate. With policy teams in major issue areas, American Progress can think creatively at the cross-section of traditional boundaries to develop ideas for policymakers that lead to real change. By employing an extensive communications and outreach effort that we adapt to a rapidly changing media landscape, we move our ideas aggressively in the national policy debate.

Center for American Progress

1333 H STREET, NW, 10TH FLOOR, WASHINGTON, DC 20005 • TEL: 202-682-1611 • FAX: 202-682-1867 • WWW.AMERICANPROGRESS.ORG