

ADL REPORT:

CONTROL-ALT-DELETE

Recommendations of the ADL Task Force
on the Online Harassment of Journalists



TABLE OF CONTENTS

Foreword	1	Crash Override Network	16
Introduction	1	The Women’s Media Center Speech Project	16
Why ADL Is Focused On Cyberhate	2	Producing New Tools for Journalists	17
SECTION ONE: SUMMARY OF KEY RECOMMENDATIONS	2	The Legal Framework	17
Recommendations for Industry	2	The Communications Decency Act	17
Recommendations for Journalists, Targets, and Advocates	3	The First Amendment	17
Recommendations for the Legal System and Policymakers	4	Exceptions to the First Amendment	17
SECTION TWO: THE PROBLEM	4	Practical Difficulties in Bringing Criminal Charges	18
Fighting Cyberhate and Upholding the First Amendment	4	Doxxing and Swatting	18
Journalists as Targets	5	Hate Crime Laws	19
Fighting Cyberhate While Upholding the First Amendment	5	Challenges for Law Enforcement	20
Difficulty of Policing Online Hate	5	Civil Lawsuits and Copyright Remedies	20
Why Harassment Is a Particularly Serious Problem for Twitter	6	CONCLUSION	21
SECTION THREE: DEEPER CONTEXT	7	APPENDIX A: METHODOLOGY	22
The Social Media Ecosystem	7	APPENDIX B: TASK FORCE MEMBERS	22
How the Major Platforms Mitigate Abuse	8	APPENDIX C: ADL’S HISTORY OF RESPONDING TO RECURRING PATTERNS AND DISTURBING TRENDS	23
Anti-Harassment by Design	12	APPENDIX D: MORE OF THE LAW REGARDING “TRUE THREATS”	24
Privileging Accountable Online Identities	13	APPENDIX E: ANTI-SEMITISM IN THE U.S.	25
New Platforms and Services Designed to Help Mitigate Abuse	14	APPENDIX F: THE FIRST REPORT	27
Heartmob	15		
Trollbusters	15		
Hate Speech Blocker	16		

FOREWORD

Sometime during the 2016 presidential campaign, it became dangerous to be a journalist.

Throughout 2016, ADL received complaints from journalists covering the presidential campaign that they were being serially harassed online merely for doing their jobs.¹ Journalists received altered images of themselves as concentration camp inmates, wearing yellow “Juden” stars, and of Auschwitz’s infamous entry gates emblazoned with the slogan, “Machen Amerika Great.” They were directed to go “back to the ovens” and tarred with anti-Semitic slurs. This abuse had real-life consequences, as one journalist discovered when he saw white supremacist images embedded in a video that was designed to trigger his epilepsy,² and as other reporters found when they were repeatedly harassed at campaign events.

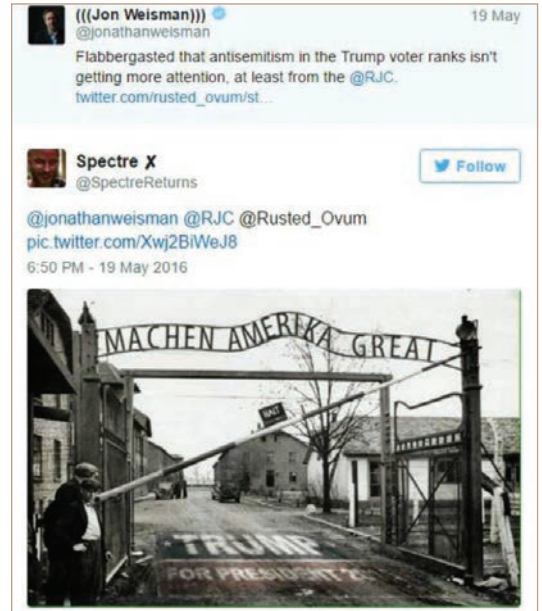
In response to this deluge of hate, ADL embarked on a groundbreaking project to quantify the scope of the abuse, to evaluate how, when, and where it occurred, and to create a profile of the abusers. ADL’s report,³ the first of its kind, found that anti-Semitic language on social media was shockingly pervasive: A total of **2.6 million** tweets containing anti-Semitic language were posted on Twitter between August 2015 and July 2016, with an estimated potential reach of **10 billion impressions**. ADL believes this outpouring of hate has contributed to the reinforcement and normalization of anti-Semitic language on a massive scale.

We’re following our initial report with a second installment: ADL’s recommendations for addressing internet harassment, which we hope will help policy makers, law enforcement, the internet industry, the targets of online harassment, and the public. ADL believes we have a collective obligation to develop the tools and strategies to confront online hate in order to ensure that the internet remains a medium of free and open communication for all people.

INTRODUCTION

This report is organized in three main sections.

- Section One provides a summary of our key recommendations.
- Section Two provides background on how we arrived where we are today.
- Section Three sets out some context that is key to understanding the social media ecosystem, the relevant legal framework, and innovative research and new tools that are already providing creative responses to abuse.



Dana Schwartz (Observer)
This meme is repeated with various journalists pictured inside the gas chamber.

Finally, we include several appendices providing background on the Task Force, our methodology, the initial report, and some additional analysis and context that may be helpful.

WHY ADL IS FOCUSED ON CYBERHATE

From its inception, ADL has understood that the fight against one form of prejudice requires battling hate in all forms.⁴

ADL took an early lead against those who use new technologies to foment hatred, undermine democratic values, and tear our society apart. From anti-Semitic images in mass-circulation newspapers to stereotypical depictions of Jews and African-Americans in the entertainment industry, ADL has consistently recognized the power of media and communications technology to shape public attitudes on issues of prejudice, hatred, and discrimination – both for ill and for good.⁶ With every major advancement in technology in the past century, ADL has fought those who would use new platforms to spread hate, and emphasized the importance of promoting acceptance, inclusiveness, and the protection of civil and human rights.

ABOUT ADL

The Anti-Defamation League (ADL) was founded over one hundred years ago “to stop the defamation of the Jewish people, and to secure justice and fair treatment to all.”⁵ From its inception, ADL has understood that the fight against one form of prejudice requires battling hate in all forms.

Today ADL is one of the nation’s premier civil rights and human relations agencies. ADL fights anti-Semitism and all forms of bigotry, defends democratic ideals, and protects civil rights for all through information, education, legislation and advocacy.

SECTION ONE: SUMMARY OF KEY RECOMMENDATIONS

RECOMMENDATION FOR INDUSTRY

IMPROVE MECHANISMS FOR REPORTING HATE SPEECH AND HARASSMENT

Reporting mechanisms should be thorough, quick, and responsive – Allow complainants to explain why content is offensive and to show the cultural, social, or personal context that makes them feel targeted.

Reviewers need cultural context and continual training – Ensure that reviewers understand specific cultural contexts and receive continual updates to their training.

Internet companies should be transparent – Explain review processes and moderation guidelines to users, so they can better participate in the system.

Complaint mechanisms should be efficient – Permit complainants to report instances when the same abusive material is posted multiple times or on multiple channels.

Targets should not be re-victimized – Reduce the need for prohibited or suspended material to be re-flagged every time it appears or is reported.

RESPOND TO BYSTANDER REPORTING OF HARASSMENT

Accept third-party harassment reports – Update platform systems to respond not only to reports from harassment victims, but also to reports by observers to harassment.

Encourage user reporting – Encourage social media users to use existing complaint mechanisms and reporting solutions.

SIMPLIFY THE APPEALS PROCESS FOR DENIALS OF REQUESTS TO ADDRESS HATE SPEECH OR HARASSMENT

Simplify and clarify appeals processes – Include the original complaint in all responses to complainants, especially since targets may have multiple active complaints. Include a link within the complaint response to appeal the company’s decision.

ENCOURAGE CROSS-PLATFORM COMMUNICATION BETWEEN COMPANIES

Disrupt online harassment campaigns – Determine ways to track and respond to cross-platform abuse when harassers employ multiple platforms to harass others.

Social media platforms should collaborate on, or co-invest in, anti-harassment tools – Pool resources and combine expertise to develop new approaches to combat online harassment.

INVEST IN INNOVATIVE RESEARCH

Develop programming solutions to address harassment – Explore promising avenues to combat harassment, like natural language processing and machine learning, and technical solutions that may preserve anonymous speech but permit accountability.

Encourage “anti-harassment by design” – Encourage innovators to consider harassment problems when designing new platforms and build in protections from the start.

Privilege accountable online identities – Propose platform structures that privilege “verified” online identities, while still carving out spaces for free expression and the potential need for anonymity.

RECOMMENDATIONS FOR JOURNALISTS, TARGETS, AND ADVOCATES

HARNESS POSITIVE COUNTER-SPEECH

Beat trolls at their own game – Emulate groups like HeartMob and Trollbusters that combat harassment in real-time by unleashing supportive positive content on a target’s social media feeds.

Use more speech to educate – Learn from new services, like Hate Speech Blocker, that issue real-time alerts to educate users when they use hate speech or abusive language, making them pause and consider their choices before sending the content.

PRODUCE NEW TOOLS FOR JOURNALISTS AND TARGETS

Give writers control of their online space – Develop innovative solutions for journalists who must engage with social media, like the Coral Project has done, by improving online commenting forums and building out tools for newsrooms to customize.

Silence the noise – Look to platform features or third-party applications that allow users to block content, keywords, or abusers from social media.

Support targets of online abuse – Support advocacy groups, like Crash Override Network, that mitigate the effects of online abuse while providing legal and emotional support to targets of online hate.

DEPLOY EASY TO USE AND ACCESSIBLE CYBERSECURITY TOOLS

Apply smart online security solutions – Follow best practices in online safety, including using a password manager and multifactor authentication, regularly patching and updating devices and software, separating work and personal services, and only accepting social networking requests from people you know.

RECOMMENDATIONS FOR THE LEGAL SYSTEM AND POLICYMAKERS

MODERNIZE STATE LAWS TO FULLY COVER CYBERSTALKING

Close gaps in state laws – State cyberstalking statutes should be updated to prohibit online abuse as it frequently occurs, like indirect harms committed by virtual mobs, since these behaviors may not be covered by existing laws.

LEGAL REDRESS FOR ACTIONABLE DIRECT THREATS AND HARASSMENT

Dedicate more funding for better law enforcement responses – Cyberharassment cases can be time consuming and challenging to investigate and prosecute. Government should direct more resources towards this growing problem, enabling law enforcement respond more quickly.

GOVERNMENT-SPONSORED STUDY OF THE SCOPE OF CYBERHATE

Update federal studies on cyberhate – In 1993, the National Telecommunications and Information Administration issued a report on the use of broadcast and internet media to encourage harassment and abuse. Today Congress should direct the NTIA to revisit its study on cyberhate, the way it spreads, the damage it causes, and how it can be countered.

IMPROVE RESPONSES TO ONLINE ABUSE

Provide better training for state and local law enforcement – State and local law enforcement are often the first responders to online hate. Government should ensure that these departments have the expertise and resources to effectively investigate and prosecute cyber cases.

Centralize the reporting of online abuse – A single national reporting center (along the lines of the Internet Crime Complaint Center) would help establish more consistent reporting of these crimes, and ensure that the right authorities are always informed and able to assist with the investigation.

NEW LAWS AGAINST NEW FORMS OF ONLINE ABUSE

Criminalize new forms of online abuse – New federal laws addressing “doxxing” and “swatting” as forms of cyberharassment would provide law enforcement officials more tools to respond to these dangerous practices, which use online activity to harm victims in the physical world. As a starting point, we recommend a conference that brings specialists in online harassment, the First Amendment, law enforcement and privacy experts together to identify new avenues of improving responses to online harassment that moves off-line.

SECTION TWO: THE PROBLEM

FIGHTING CYBERHATE AND UPHOLDING THE FIRST AMENDMENT

Cyberspace can be dangerous. This is particularly the case for journalists, who find themselves in the public eye as they report the news. Oftentimes, these individuals find themselves targeted with serious harms, like private disclosure of personal information, identity theft, constant harassment with hate speech, harm to professional reputations, and threats of physical and sexual violence —and even extending to members of their family.



JOURNALISTS AS TARGETS

Responding to a disturbing upswing in online, anti-Semitic abuse of reporters in 2016, ADL formed a Task Force on Harassment and Journalism to assess the scope and source of anti-Semitic, racist, and other harassment of journalists in order to better understand how this harassment impacts the democratic process and free speech.⁷

The Task Force issued its first report in October 2016. This pioneering report presented findings based on a broad set of keywords (and keyword combinations) designed by ADL to capture anti-Semitic language on social media. Using this metric, a total of 2.6 million tweets containing language frequently found in anti-Semitic speech were posted across Twitter between August 2015 and July 2016. Those tweets had an estimated 10 billion potential impressions (reach), which ADL believes contributes to reinforcing and normalizing anti-Semitic language – particularly racial slurs and anti-Israel statements – on a massive scale.

FIGHTING CYBERHATE WHILE UPHOLDING THE FIRST AMENDMENT

The key problem identified in this Report is the massive explosion of cyberhate through Twitter, Facebook, and other social media platforms. These platforms have made it far easier to communicate and have brought people together on a global scale, while simultaneously paving the way for racists, white supremacists, and anonymous individuals to spread hate and target individual users for abuse. The key tension in the United States is to empower technology companies, law enforcement, and users to fight back against cyberhate while maintaining the internet as a free and open channel of communication.

Cyberhate takes many forms. Cyberbullying involves the use of the internet to bully other individuals, bringing the schoolyard bully into the virtual playground. Cyberharassment is the use of the internet to harass others, such as by repeated postings or hurtful comments on a target's social media channels ("trolling"), the use of social media to expose private information, or the use of technology to threaten a target virtually, often with very real consequences. Cyberhate also includes the use of the internet and social media by individuals or organized hate groups to promote their messages of bigotry.

Cyberhate can target both precisely and broadly. Often, particular individuals are targeted based on their race, religion, ethnic background, gender, or sexual orientation. Sometimes the perpetrator is known to the target; sometimes the harasser is anonymous. The targeting of an individual can have dire consequences for any friend or contact of the target who shares a similar background; like a hate crime, targeting can intimidate and silence entire communities. Cyberhate can also broadly target members of particular groups, as evidenced by white nationalist groups that use the Internet to spread their hateful messages.

DIFFICULTY OF POLICING ONLINE HATE

Policing online hate is challenging both technologically and legally. From a technology perspective, it is difficult for targets to respond to cyberhate in real-time. Reporting is often delayed, reporting mechanisms are inefficient, and perpetrators may employ methods to shield their identities. Among the challenges companies face are:

- The vast amount of materials on their sites so that they cannot proactively police it;
- The vast number of complaints sites receive make serious discussions, personalized responses and transparency difficult;

TARGETED HATE

Often it is minorities, women, and people in the public sphere who are disproportionately targeted for cyberharassment. A prominent example is Leslie Jones, a comedian on Saturday Night Live. She was brutally trolled on Twitter, following a film review of her new film, *Ghostbusters*, by alt-right provocateur Milo Yiannopoulos. Her private accounts were also hacked, resulting in the distribution of private photographs on the internet. Jones responded by signing off of Twitter. The next day, Twitter permanently banned Yiannopoulos from the platform.⁸

-

Accordingly, companies often respond too slowly to complaints, when they respond at all.

Legally, it is challenging to prosecute these cases. The First Amendment shields a great deal of speech, no matter how hateful, from sanction, although exceptions exist for true threats and certain other categories of speech. Law enforcement, including police and prosecutors, often lack the skills to trace and log cyberharassment and cyberbullying, and so depend on targets to preserve evidence. Companies don't always respond quickly and efficiently to complaints or law enforcement requests.

WHY HARASSMENT IS A PARTICULARLY SERIOUS PROBLEM FOR TWITTER

Trolls disproportionately target minority groups, religious groups, and women.¹⁰ Failure to effectively address these attacks may result in fewer viewpoints, as the targets disengage from online spaces. Speech is not a zero-sum game; when action is not taken to stop abuse, the speech of harassers is privileged while the voices of targets are driven off the platform.¹¹

6

SECTION THREE: DEEPER CONTEXT

THE SOCIAL MEDIA ECOSYSTEM

In the 21st century, many of our media and social interactions take place online. Sixty-five percent of American adults use social networking sites – from Facebook and Twitter to Snapchat and Pinterest, and a dozen others – up from just seven percent in 2005.¹⁵ As use of social networks has increased, so has the amount of time that users spend on those sites – in 2015, an average of nearly two hours a day.¹⁶

As the time spent on social media has increased, harassment on social media has also increased. In fact, according to the Pew Research Center, 70 percent of 18-24 year olds who use the internet have experienced harassment online, and 26 percent of women in that age group reported being stalked online. Social networks appear to be particularly fertile ground for online harassment – 66 percent of users who experienced online harassment said the most recent incident occurred on a social network.¹⁸

“Cyberharassment” covers a range of activities, from name calling to sustained sexual harassment, and it takes place throughout the online world – on websites, in games, and on online dating services. According to Professor Danielle Citron, cyberharassment “involves threats of violence, privacy invasions, reputation-harming lies, calls for strangers to physically harm victims, and technological attacks.”¹⁹

The online nature of cyberharassment differentiates it from other forms of harassment, she explains, because of “...the different way the Internet exacerbates the injuries suffered . . . by extend[ing] the life of destructive posts.”²⁰ It also typically involves a “course of conduct,” meaning that attacks are sustained campaigns, not one-off experiences, and may involve “cyber mobs” of strangers acting in concert to heighten the damage caused by their outrageous and abusive targeting of their victim.²¹

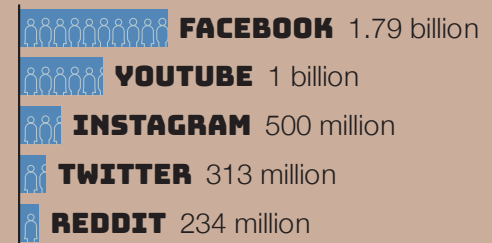
Social media platforms have been criticized for being tight-lipped about the logic underlying their policies, procedures, and capabilities for tracking and responding to online abuse. While these companies have terms of service, targets of online abuse claim that flagging or reporting content is not always effective in getting that content removed. Many individuals consulted for our study specifically noted that reporting mechanisms are challenging to understand and use, and they offer limited explanation of what a user can expect after the report has been submitted.

This lack of transparency may be motivated by efforts to limit the ability of online harassers to do end-runs around known anti-harassment programs. Unfortunately, it also undermines the ability of victims to effectively complain and obtain redress.

Social media platforms should collectively face the common challenge of online harassment. It is an opportunity for companies to collaborate to the extent allowed by user privacy considerations, to pool resources, and to innovate new technical solutions and community mechanisms. Jointly-developing mechanisms to disrupt cross-platform abuse would help the industry eliminate the “whack-a-mole” problem that plagues those experiencing online abuse.

TOP SOCIAL NETWORKS

(by number of users)¹⁷



FUTURE INDUSTRY TRENDS

Innovation will influence the strategies that companies will deploy to combat future cyberharassers. Social media companies want user growth and more adaptable features, like live video. As new ventures emerge – coverage of breaking news, development of e-commerce applications, integration with virtual and augmented reality, and creation of new hardware partnerships – companies will need to be increasingly vigilant against online abuse.

HOW THE MAJOR PLATFORMS MITIGATE ABUSE

Many of the major internet platforms today rely on trust and safety teams to assess online abuse and respond to user concerns. The data these teams use can come from leveraging crowd-sourced moderation tools or enabling users to vote – either explicitly or through proxies (such as “likes”) on the appropriateness of comments.²² When comments are flagged as inappropriate, or otherwise brought to the attention of trust and safety teams, those responsible for the offensive content may be subject to actions ranging from temporary restrictions to permanent bans. Some companies – particularly in the gaming and educational industries – also rely on lists of prohibited words or user-controlled, word-based filters to identify violators.²³

If companies have not yet done so, they should conduct outreach and create educational materials to ensure that users understand how removal processes and procedures work. This would address a complaint that ADL hears repeatedly from advocates and targets of abuse – that the internal review processes and moderation guidelines need to be more transparent and applied more consistently within the same platform. Furthermore, reporting mechanisms should be thorough, quick, and responsive.

Content reviewers are often overseas contract workers or new college graduates.²⁵ Because these individuals may not come from the same background as the user, companies should allow complainants to explain why the content they are reporting is offensive and allow them to show the cultural, social, or personal context that makes them feel targeted. The place for this in a review process should be clearly marked and explained. Furthermore, to ensure that reviewers understand specific cultural contexts, companies should require that these reviewers receive continual updates to their training.

By and large, companies are able to choose what scope of response to online hate is most appropriate for their community, and what mechanisms they should rely on to accomplish this goal. Although certain markets compel action in limited circumstances,²⁶ in most cases U.S. companies can decide for themselves the scope of moderation that they feel is appropriate, under the law. In other words, in the United States, the internal policies of internet companies primarily govern how, if, and when content is moderated. Furthermore, there are no legal requirements that the review and removal process be transparent.

LEADING THE WAY IN COLLABORATION

Microsoft has set an example of collaboration to combat shared harms, in the realm of child exploitation. It has made a version of its anti-child pornography PhotoDNA available to social media companies at no cost. This sharing takes away potential hurdles for smaller companies and other organizations that want to give users the freedom to upload content while ensuring the integrity of their platforms.²⁴

TAILORING STRATEGIES TO HARMS

Different online abuse causes different harms and calls for different responses. For example, an anonymous mob targeting an individual because of their race can have both a specific impact – that individual may be silenced – and a more general impact – others that identify with that individual may be silenced. Different solutions might address different aspects of these harms. For instance, shielding the target but not removing the speech might mitigate the specific harm but not the general one. In evaluating responses to online abuse, it is essential to think about the specific types of harm at issue and to deploy specific mechanisms designed to address that harm. There is no one-size-fits-all solution.

BASIC INDUSTRY PRACTICES RELATED TO CYBERHATE AND HOW THEY HAVE CHANGED IN THE LAST THREE YEARS

PRACTICES	2013	2016
Hate Speech Policies	The Terms of Service for many platforms did not address hate speech directly or used vague terminology in policies	Multiple platforms, including Facebook, Google, Twitter, Amazon, Microsoft gaming, and Yahoo, now include specific prohibition of hate speech
User-Friendly Reporting	Complaint mechanisms or contact details were often buried or limited in functionality	Virtually every major service and platform uses post, profile and image flagging. Now standard practice to send receipt of complaint acknowledgements and provide links to further policy/process information.
Enforcement Mechanisms	In cases where hate speech was prohibited, penalties were mostly delineated	Google, Facebook, Twitter have instituted flagging for specific posts and partial content removal. Several social media platforms have implemented “stop and think before sending” messages and campaigns.
Transparency	Pervasive tendency for companies not to explain why content allowed to remain after a complaint; little explanation offered to users whose material was deleted	Most platforms offer explanations to users whose content has been deleted and provide an appeals process. Complainants on Facebook and YouTube are advised if content has been removed. Public disclosure of rationales for removals is limited.
Counter-speech	Counter-speech education by only limited number of companies, and un-coordinated between companies	Counter-speech projects are being studied and changes implemented by major platforms.

CHALLENGES THAT THE INDUSTRY AS A WHOLE CONFRONTS WHEN DEALING WITH CYBERHATE

INTERNAL INDUSTRY CHALLENGES	2013	2016
Industry Realities	No effort to broadly explain the challenges created by evolving technology, unintended consequences and the volume of content	Industry platforms are sharing more data on traffic, members' complaints and responses than ever before - but still falling short in adequately illuminating the enormous and ever-growing volume of content and the challenge of addressing issues that require human evaluation and intervention
Anonymity	Anonymous participation on many platforms tolerated despite policies to the contrary	Anonymity continues to pose challenges for enforcement of Terms of Service. New technologies are better at detecting users with multiple accounts being used to evade website policy.
Industry Coordination	No coordinated industry statements or projects obvious to the public	The Anti-Cyberhate Working Group has become a major venue for the industry to coordinate anti-cyberhate activity. Major breakthroughs: publications of ADL's "Best Practices for Responding to Cyberhate" and well-received Cyber-Safety Action Guide. There is more dialogue between companies on hate related issues than ever before.
Hate speech links and linked material	Platforms took no substantial responsibility for third party or linked content	Ongoing debate and discussion regarding platform as publisher and impact of link distribution
Corporate Voices	Few if any corporate voices spoke about online hate	Anti-hate speech voices in industry now led by Facebook, Microsoft, and Google with recent important statements by Twitter

EXTERNAL CHALLENGES THAT IMPACT THE INDUSTRY'S ABILITY TO ADDRESS CYBERHATE

EXTERNAL INDUSTRY CHALLENGES	2013	2016
Cross Border	Limited coordination of cross border issues	In the borderless environment of the Internet, almost all initiatives and resolution programs remain geographically based
Government Intervention	Uncoordinated or unenforceable regulations	Increasing disconnect between online ideals and achievable targets for action compared to laws under consideration and being enacted to curb online hate
Cyber-Terror/Hacking	Hacking (website defacement) mainly performed on an opportunistic basis without consistent political motivation or targeting	Sharp increase in politically motivated hacking targeting Jewish institutions and Western interests

ANTI-HARASSMENT BY DESIGN

Because of intermediary liability protections for internet service providers, ADL suggests working with internet companies to develop creative solutions to address online hate – from the way websites are built to the way users interact with interfaces and with each other. We call this philosophy “anti-harassment by design.” Designers and user experience specialists, who craft the interfaces that make up the internet, can help to quell harassment. ADL has looked to news websites as examples of platforms that have discouraged online hate by harnessing design and new innovation.

Designers’ interfaces influence how online users communicate. *The New York Times*, for example, has changed its comments section over the past few years, creating a more civil environment. Prior to October 2007, the *Times* allowed commenters to post on blogs, but not on dedicated news stories.²⁷ On October 30, 2007, an online Science Times article and an online editorial added comment sections, marking the first time users could comment on more traditional news stories.²⁸ But in November 2011, the paper redesigned its comment section, allowing users to post directly below the article text rather than requiring them to go to a separate webpage.²⁹ Researchers from the Engaging Media Project examined how this redesign affected behaviors within the comment section and how the moderating team’s interactions affected comments.³⁰ They found that the redesign resulted in an increased number of comments left on the website and a decline in the use of abuse flags.³¹

Maintaining online comments sections is complicated, however. While these sections provide opportunities to contribute to public debates, they also require constant maintenance to remain free of trolling. In 2015, many news sites gave up on online comment sections altogether, including *Recode*, *Reuters*, *Popular Science*, *The Week*, *Mic*, *The Verge*, and *USA Today’s FTW*.³²

These decisions are thanks in part to online abuse. According to the Global Report on Online Commenting, 65 percent of the organizations questioned reported that their journalists were subject to cyberharassment, with opinion pieces often generating the most comments.³³ Sensitive topics were most often the target of trolling.

Good comment moderation requires time and money. *The New York Times* is often touted as a success story, but moderation is very labor-intensive, and only a small percentage of Times stories are open to comments. Other outlets outsource evaluation of comments or attempt to moderate reader contributions on a smaller scale, such as by closing sections to comment after an initial window.³⁴

To effectively combat harassment, ADL believes media and technology companies must think about the architecture of the internet and harness innovation. Companies should consider how social media functions, from a structural point of view, and generate solutions to meet different models. According to the Committee to Protect Journalists, there are two distinct kinds of social media platforms: “those like Facebook, where each person is presented with a curated section of material based on preferences defined by the user, and those like Twitter and instant messenger, where information displayed is not informed by the platform or its algorithms.”³⁵ Therefore, to make headway in addressing online abuse, new solutions for existing platforms should take these distinctions into account.

It is easier to account for such distinctions at the beginning of development. Innovators within and outside the industry are now starting to consider harassment problems when designing new platforms and building in protections early on.

The research now underway on designing platforms to discourage and prevent harassment is a start, but more needs to be done, both within existing platforms and externally. Social media companies should be aggressive in pushing this research and innovation, providing financial support, partnering with researchers, and sponsoring hackathons to develop new talent. There are promising avenues to consider, like natural language processing and machine learning, and technical solutions that may preserve anonymous speech but permit accountability for abuse.³⁶

Some media companies are rising to this innovation challenge. Jigsaw, a Google subsidiary, recently introduced an artificial intelligence-related solution to content moderation called Conversation AI. Conversation AI can “automatically flag insults, scold harassers, or even auto-delete toxic language.”³⁷ This algorithm studies and instantly flags abusive language, and then rates it with an “attack score” of 0-100. If the score is 0, then it means that there was no abusive language detected; if the score is 100, then the language showed some form of harassment or abuse.³⁸ Conversation AI has been trained to spot toxic language with a reported 92 percent certainty and a 10 percent false-positive rate.³⁹ The technology will be beta-tested in the comments sections of The New York Times, and Wikipedia also plans to use it. Eventually, Conversation AI will be open-source, so websites or social media platforms can apply it to catch and deflect abuse in real-time.

Because haters quickly find avenues around anti-harassment measures, however, those seeking to counter cyberhate must accelerate the pace of innovation. Within days of Jigsaw’s announcement of its Conversation AI online content moderation system, for instance, internet trolls had developed a “secret code” to circumvent it. The trolls attempted to associate the names of prominent companies with racial slurs; for example, Jewish people were termed “Skypes.” Until Twitter removed it, a “Skype Directory” profiled Jewish professionals and celebrities, marking them as targets for further online abuse.⁴⁰ Just as the potential for advancement is embodied by new technology, so is the capacity for new forms of hate.

PRIVILEGING ACCOUNTABLE ONLINE IDENTITIES

Twitter is taking its obligation to create safe online spaces seriously. In addition to the anti-hate policies it has implemented since 2013 – and in light of its recent problems with hate speech and trolls – Twitter will announce changes to its terms of service and user interface in November 2016.⁴¹ The new changes to Twitter’s safety policy may give users more control over their experience.

Twitter’s approach to cyberhate has evolved through experience. At its founding, Twitter encouraged users to tweet as much as possible with little to no moderation.⁴² But as Twitter learned that online abuse was impacting users’ experience, it introduced tools like the “Report Abuse” button and modifications to its terms of service to prohibit indirect threats and non-consensual nude images.⁴³ After high-profile targeting of celebrities in summer 2016, Twitter took additional steps to mitigate harassment.⁴⁴ For example, in July 2016, the verified account function, putting a blue check next to a Twitter handle that is certified as authentic, was opened to all users.⁴⁵ In August, new “quality” spam filters and a “notifications” filter – disabling users from being mentioned by tweets and replies of people they did not follow – became available.⁴⁶ However, ADL’s research indicates that many journalists considered these features “clunky,” and the company is still considered to be slow to respond to complaints of abuse, when it responds at all.

ADL recommends that Twitter streamline its complaint mechanisms, especially for targets of repeated online abuse. Allowing targets to report instances of the same abusive material that are posted and reposted multiple times, or within multiple channels, would prevent targets from being re-victimized in the reporting process. Systems should eliminate the need for targets to flag prohibited or suspended material every time it reappears or is reported.

Simple changes could also provide significant relief to targets of mass abuse. For example, including the original complaint in all responses to a complainant would be immensely helpful to targets of wide-scale abuse, who may have several complaints being processed at the same time. Additionally, the inclusion of a button or mechanism to simplify the appeals process – like a link within the company’s response to a complaint that would allow the user to appeal the company’s decision – would improve the user’s experience throughout the process.

ADL also recommends that Twitter take specific steps to privilege online engagement that uses people’s real identities, while simultaneously carving out a space for anonymous speech, with both governed by enforceable terms of service. This could be done by allowing customers to choose between a premium level of service for verified accounts, perhaps with comparatively faster service, and a general level of use, which would be open to any account.

RECOMMENDATIONS FROM JOURNALISTS THEMSELVES

Some journalists have recommended steps that Twitter could take to rein in trolling.

These steps might include:

- Popularizing the platform's user verification feature, which is open to all Twitter users who are willing to use their real names, but is not required by the platform. Fake accounts and bots would be ferreted out by user reporting. Verification could be incorporated into Twitter's quality filter feature, and machine learning could improve the filter until it was able to remove abusive tweets from other users. Eventually, Twitter could offer users the option to view tweets from people they follow, verified strangers, or unverified strangers that a user has manually whitelisted, or agreed to accept messages from. If this was made the default setting, Twitter would see a large reduction in unverified users with a small number of followers – a user profile that fits many who use the platform to peddle hate. Twitter would then have the bandwidth to devote more attention to handling reports of online abuse in a more active and nuanced way.⁴⁷
- Improving the verification feature, which users have complained is slow and awkward. The problems with this feature provide an additional challenge for journalists. By improving the user experience, more people might sign up and take advantage of built-in anti-harassment features. The features help users control what they see on Twitter. The first filter hides notifications from accounts that the user does not follow. The second filters out accounts from notifications and timelines based on factors like account origin and behavior.⁴⁸

To the extent that they are not already doing so, companies like Twitter should maximize the use of in-house resources to analyze and track online abuse in order to determine if their policies and procedures are effective. They should also report publicly on their analysis of what features, products, or services were effective in diminishing online abuse. If they find their current policies and procedures inadequate or insufficiently ineffective, companies should invest in sufficient manpower or new tools to reduce online abuse in a measureable and transparent fashion.

Furthermore, to make sure they accurately understand the scope of online abuse on their platforms, companies like Twitter should take new measures to encourage user reporting of online abuse through existing complaint mechanisms and reporting solutions. Additional user education and proactive outreach may be required. Journalists are often reticent to report abuse, due to their heightened awareness of speech issues, and other users have accepted such abuse as part of their user experience, even when it violates terms of service.

NEW PLATFORMS AND SERVICES DESIGNED TO HELP MITIGATE ABUSE

A number of important new platforms and services designed to address and mitigate abuse have emerged recently. These services are cutting-edge because they excel at beating trolls at their own game. While not all new developments can be highlighted here, several noteworthy initiatives deserve mention.

SUCCESSFUL ONLINE CIVILITY

Wikipedia has been touted by some as a success in encouraging commentators to engage civilly, despite disagreement. From this example, we can see factors that might be applicable to journalism: the tone of encyclopedic-neutrality, the frequency with which commentators are confronted by opposing viewpoints, and the necessity of backing up factual assertions with citations.⁴⁹

Some combat harassment by harnessing positive counterspeech. Targets receive real-time positive support on their social media feeds. Others use more speech to educate, asking users to stop and consider the implications of their actions before posting hate speech. Still others have digitized newsrooms, giving writers more control over their online environment and building out tools for newsrooms to customize. In addition to silencing the noise of online abuse, these approaches recognize that the effects of abuse go deeper than social media feeds, and they support targets by providing legal advice, cybersecurity best practices, and emotional support resources.

HEARTMOB

Heartmob is a new platform that provides crowd-based positive reinforcement to targets of harassment in real-time.⁵³ It flips the group dynamics typically exploited by trolls on its head, and empowers bystanders to act to stop abuse that they observe. Heartmob users can report harassment, call on others to defend targets or send them support, and publicly “out” trolls.⁵⁴

HeartMob is a project by Hollaback, an organization dedicated to fighting real-life harassment on the streets. With Heartmob, it applies that expertise online and addresses several challenges of cyberharassment. For example, reporting harassment typically involves monitoring multiple platforms and flagging individual messages for review by social media companies. Since companies do not regularly engage in cross-platform analysis, each only sees a piece of the abusive behavior that targets experience. Using Heartmob, a target is able to create a more accurate portrait of the online abuse in order to create a more compelling case if they turn to companies or law enforcement for redress.

When HeartMob users report harassment, they have the option of (1) keeping their report private and saving it for later if the abuse escalates, or (2) making the report public and choosing from options that indicate how they would like bystanders to support them, take action, or intervene. Heartmob further allows bystanders to provide support to targets of abuse by receiving public requests for assistance, along with the target’s chosen actions of support. HeartMob staff moderates the messages and reports closely to ensure their platform remains safe and supportive.

Heartmob provides additional resources to targets of harassment, including information on safety planning; how to identify real threats online harassment laws; reporting online harassment to authorities; and referrals to other organizations that provide counseling and legal services.⁵⁵

TROLLBUSTERS

Trollbusters is a positive messaging service which took top-honors at a 2015 hackathon sponsored by the International Women’s Media Foundation.⁵⁶ The event was designed to address trolling and cyber harassment that against women. Trollbusters describes itself as “just-in-time rescue services to support women journalists, bloggers and publishers who are targets of cyberharassment.”⁵⁷ It has three elements: identifying troll networks; rapid response counterspeech teams; and the provision of emotion, legal, and technical support to targets.

Trollbusters’ service counters cyberattacks in real-time with online community support and positive messaging. When a journalist arrives on the Trollbusters website, she can enter her Twitter handle or someone else’s Twitter handle and a link to an offending tweet. TrollBusters then emails the target, asking for her consent to receive tweets from them. After

COMBATTING ONLINE ABUSE IN GAMING

Riot Games makes multiplayer online games, including the popular League of Legends (LoL) – 100 million players every month by last count.⁵⁰ In 2014, LoL also had a harassment problem – players were quitting the game and citing noxious behavior as the reason. Riot made a series of small tweaks – from turning off chat between opposing players, to providing better notifications to players being sanctioned for harassing behavior.⁵¹ These initiatives didn’t “solve” harassment on LoL, but they did substantially shrink the problem.⁵²

that, Trollbusters will tweet an announcement from their account, putting the troll on notice, and will continue monitoring activity from the offending user to address the situation if it escalates.

Trollbusters relies on directed supportive counterspeech via positive and affirming messages that are sent to a target of abuse at the point of attack, including memes, endorsements, and testimonials. This positive wave drowns out the harassment, while providing emotional support and reputation management to targeted writers. Seeding troll-invaded discussions with supportive comments may derail the intended abuse. Since it interrupts abuse in progress, it neutralizes one of the stronger critiques of counterspeech – that the damage has already been done. This strategy may potentially deter trolls – and create more space and support for voices that trolls seek to silence. Such affirmative counterspeech may help journalists protect their voices, websites, and businesses from trolling. Trollbusters’ volunteers also advise targets on managing their online reputations using techniques like search engine optimization and provide legal resources, online safety tips, and emotional support.

Trollbusters is developing software to identify groups of trolls using natural language processing.⁵⁸ This network analysis technology aggregates information on organized networks of trolls. Tracking networks of cyberharassers makes it easier to filter them out of the conversation, identify them, and hold them accountable. Currently, Trollbusters helps guard against attacks on Twitter, but it may expand to other social media as the product develops.

HATE SPEECH BLOCKER

Another innovative solution lies not in mitigating harassment, but in asking speakers to pause and reflect on their words before they post online. In October 2016, a U.K.-based NGO called International Alert won a “peace-building hack-a-thon” with a “spellchecker” for hate.⁵⁹ The team developed Hate Speech Blocker, a Chrome browser plugin that analyzes text in real-time before it is sent. The text is compared against Hatebase API, a nonprofit online service that collects data about hate zones and derogatory terms in different parts of the world, covering a wide swath of slurs, from religious-based intolerance to online bullying.⁶⁰ If particular terminology is recognized, Hate Speech Blocker issues an alert to the user, asking him or her to pause before posting. While users are not stopped from posting, the alert suggests why particular language may be offensive or construed as hate speech in that specific context. It also provides further information about the targeted group, in an effort to educate the user about other communities. The developers hope that Hate Speech Blocker will curb online hate speech while simultaneously protecting freedom of speech.

CRASH OVERRIDE NETWORK

Crash Override Network is an organization that emerged in the wake of Gamergate to directly help targets of online abuse and to work toward eliminating the online abuse.⁶¹ They provide many resources, including guides to protecting users and a hotline to email if confronted with online abuse. Other survivors and experts provide users with information about how to deal with the harassment technologically, emotionally, and legally. The service also monitors and reports abuse so that a victim need not read it himself or herself. They are a trusted Twitter resource.⁶²

THE WOMEN’S MEDIA CENTER SPEECH PROJECT

This organization provides comprehensive resources for individuals dealing with online harassment.⁶³ Resources include a glossary of terms to help targets identify what they might be facing, information about possible legal measures, statistics about women and online abuse, and updates on pending legislation.

PERSONALIZING YOUR ONLINE EXPERIENCE

Platforms like Instagram have been introducing tools to help users personalize their online experiences by filtering out specific hashtags. Twitter also allows users to export their block lists and share them with others, as a way to crowdsource against harassment. While this is applicable to journalists, it is also widely available to the general public, and can help reduce online abuse.⁶⁴

PRODUCING NEW TOOLS FOR JOURNALISTS

Technology companies are developing new tools to enable journalists to fight harassment in order to do their critically important work reporting, investigating, and documenting events. Mozilla, for example, has developed the Coral Project, which is creating new open-source resources to allow journalists to design their own level of media engagement.⁶⁵ With the Coral Project, journalists can build communities of trusted users, control and moderate comments, and develop new ways to engage users with the newsroom.

New tools and resources are also available through advocacy groups like Crash Override Network, mentioned above, that provide walkthroughs and security check-ups.⁶⁶ Among the key advice for journalists is use of a password manager and multifactor authentication, regular patching and updating of electronic devices, use of separate services for work and personal accounts, and limiting contacts in social networks to people one personally knows.⁶⁷

THE LEGAL FRAMEWORK

There are two important things to understand about the legal environment governing social media: (1) intermediary liability protections mean that companies are generally not legally responsible for the content that users post on their websites; and (2) freedom of speech protections mean that most hate speech is also legal speech, unless it falls into specific exceptions to the First Amendment.

THE COMMUNICATIONS DECECY ACT

American law insulates most internet service providers from liability for much of the content that appears on their platforms.⁶⁸ The Telecommunications Act of 1996 (the “Communications Decency Act”) provides that “no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.”⁶⁹ This applies to all social media platforms, search engines, and message boards. This law was designed to encourage innovation and shield internet service providers from legal responsibility for hosting content on the internet.

THE FIRST AMENDMENT

The First Amendment to the U.S. Constitution states, in relevant part, that “Congress shall make no law...abridging the freedom of speech, or of the press....” The First Amendment guarantees all Americans freedom of speech, even those whose opinions are abhorrent.

The U.S. Supreme Court has reaffirmed that the government may not regulate the content of speech by Americans online to any greater extent than it may regulate that speech in the public square or the media. As a legal matter, any governmental effort to restrict expression based on content must further a compelling government interest and be the least restrictive means of meeting that compelling interest. That is a very high burden for the government to satisfy.⁷⁰

The effect of the First Amendment and the Communications Decency Act is to protect intermediaries from legal liability for the content posted on their platforms. The actions that these companies take to regulate their spaces, through terms of service and community guidelines, is largely voluntary.

EXCEPTIONS TO THE FIRST AMENDMENT

There are certain exceptions to the First Amendment, however, and certain kinds of speech that can be legally sanctioned. These include expression that constitutes a “true threat” against an identifiable individual or institution; libel; expression that meets the legal test for harassment, stalking, or cyberbullying; and expression that constitutes incitement to imminent lawless action likely to result in such action.⁷¹

For a **true threat** to be legally actionable, the person making the threat must understand that it would be interpreted by the recipient or target as a serious expression of intent to harm.⁷² The specific circumstances are also important in determining whether the threatening language was intended to be transmitted to the target, which is a key element of the analysis. Courts examine a variety of factors and closely analyze language, intent, context, intended audience, and usage. In some cases, therefore, language that makes someone fearful – that is unquestionably obnoxious, vile, and hateful – may nevertheless be protected.

Libel can be actionable, but when the target is considered a public figure, the victim would have to show “actual malice.”⁷³ This is a high standard that goes beyond disliking someone and wishing them harm. Actual malice means that the individual acted with “knowledge of falsity or reckless disregard for the truth.” Prominent journalists, who would often be considered public figures, would have to meet this standard to succeed in any libel action.

For online **harassment** to be actionable, it must inflict significant emotional or physical harm, and be directed at specific individuals – meaning that blanket statements of hate directed toward Jews, Muslims, African-Americans, Latinos, gays, or other groups are legally protected in the United States.⁷⁵ Finally, the exception regarding **incitement to violence** will rarely, apply to hate online because of the need to demonstrate a direct connection to imminent violence.⁷⁶

PRACTICAL DIFFICULTIES IN BRINGING CRIMINAL CHARGES

Even if the content does not constitute protected speech, targets face considerable practical challenges in bringing criminal charges. For example, law enforcement officials must investigate the harassment or threat, identify the perpetrator, and connect him or her to the conduct beyond a reasonable doubt. Many state and local police departments are ill-equipped, inexperienced, and under-resourced to conduct this sort of investigation when the threats or harassment are communicated online.

ADL recommends that legislators close gaps in state laws to make sure that law enforcement has a full criminal toolkit. This includes updating state cybersecurity laws to reflect cyberstalking behavior.⁷⁷

Electronic harassment differs from physical harassment in several important ways. What begins as a state or local matter may turn into a federal matter if the abuser and the target are in different states. Physical analogues to crimes do not often work. For example, stalking laws often require the perpetrator and the victim to be located in the same geographic area. However, cyberstalkers may be located in another state or country. Lack of direct contact between the perpetrator and the target can make it difficult for law enforcement to identify, locate, and arrest the perpetrator.

Furthermore, the pace of cyberharassment exceeds the pace of government investment. New methods to harass emerge frequently, and these new developments are often not covered by state, local, or federal laws. In practice, many harassers employ tactics like distributing private information (telephone numbers, addresses, emails) so others can harass the target, or sending messages to a target’s family, friends, or employer. In certain states, this behavior is not considered cyberstalking.

DOXXING AND SWATTING

Many statutes covering acts of violence, intrusion, intimidation, or hate crimes in the physical world do not yet have online equivalents or represent untested areas of law. There are many ways to communicate online. A journalist can be intimidated through techniques that are not one-on-one communications, like public release of the journalist’s private documents or personal records.⁷⁸ Targets can be silenced through tactics including intrusion into their private accounts, posting altered photographs of them online, enabling third parties to harass or threaten them, or impersonating them. These tactics – referred to as doxxing – may not fall under the auspices of an online “communication,” as is required by some harassment and cyberstalking laws.⁷⁹ Congress should consider a new law prohibiting this practice.

Swatting is another disturbing practice that warrants legal attention. Swatting refers to calling 911 anonymously with a fake emergency that sends SWAT teams and/or first responders to another person's address.⁸⁰ One such example is a call reporting falsely that someone has a gun or is threatening another person. The law enforcement response to such a call could obviously produce serious and perhaps deadly results. Some see swatting as an extension of doxxing, and the practice probably already violates laws regarding false reports or unauthorized use of telecommunications. A law addressing both doxxing and swatting as forms of cyber harassment, however, would call attention to these dangerous practices and provide law enforcement officials with important new tools to respond.⁸¹

HATE CRIME LAWS

Hate crime laws typically increase a criminal's sentence if the prosecution proves that the perpetrator of a crime intentionally selected a victim based on the victim's real or perceived race, ethnicity, religion, gender, sexual orientation, or other similar status.⁸² There must be an underlying crime, such as an assault or vandalism; hate speech by itself is legally protected in the United States.⁸³ While in principle some state hate crime laws could apply to online content, such as a true threat that falls outside the protection of the First Amendment, the relevant federal criminal statute would not, because it requires a violent act or attempt to commit a violent act. In general hate crime laws seem an unlikely recourse, complicated further by the difficulty in determining what state law would apply when, as in most cases, the perpetrator and the target are in different states or countries.

Just because hate crimes statutes may not apply does not mean cyberhate is not a serious phenomenon. Law enforcement's basic recommendation to anyone subjected to online harassment, and who feels he or she is in danger, is to call local police. Since the late 1990s, a majority of states have enacted cyberstalking and harassment laws.⁸⁵ Still, for someone facing what is felt to be an "imminent threat," (or trying to determine if something rises to the level of an "imminent threat"), determining if there are applicable statutes is not a realistic option. Moreover, even in states where such laws have been enacted, there is a very small likelihood that an officer, accustomed to crimes committed in person, by mail, or over the telephone, will have the training or understanding to handle online threats and harassment.

Without training for state and local law enforcement, targets of online harassment are not likely to receive adequate assistance or protection. Nor will those who engage in cyber-attacks be held accountable. Training should occur on all levels: federal law enforcement should develop national trainings for U.S. Attorney's Offices, and state attorney generals should develop statewide programs and resources to support local law enforcement. It should include guidelines on how to deal with victims, who will often be vulnerable due to their age or the nature of the harassment. Departments should dedicate victim-witness resources to these type of crimes and consider treating cyberharassment with the same resources and sensitivity as departments treat sex crimes, instead of treating cyberharassment like hacking or other computer-based crimes.

A GLOBAL PROBLEM?

While the focus of this report is the United States – and the overview provided here logically focuses on American law because most major internet companies are U.S.-based – the international nature of the internet means that content protected here is not protected everywhere. Indeed, hate speech violates the law in many other countries.⁸⁴ Consequently, it is quite possible that a victim of online hate or harassment might have a stronger legal claim against a perpetrator located outside the United States if that perpetrator's message appears and is accessible to users in a country that prohibits hate speech.

CHALLENGES FOR LAW ENFORCEMENT

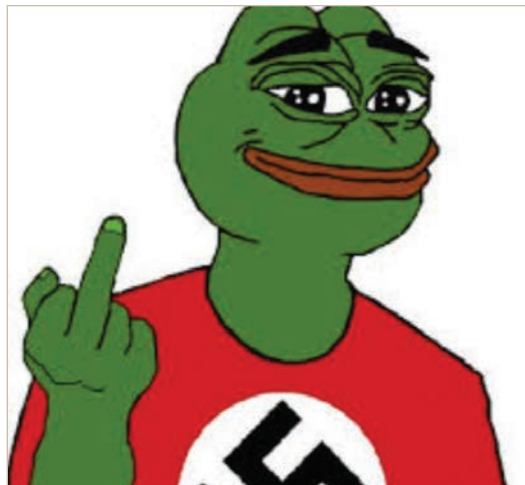
Law enforcement faces substantial challenges in fighting cybercrime, and online harassment is no exception. Overall, law enforcement faces resource constraints that make addressing cyber harassment even more challenging than responding to similar crimes committed in the real world. These investigations are often technically challenging, and made even more so by the pace and fluidity of the internet.

First, the pace of cyberharassment exceeds the pace of government investment. New methods to harass emerge frequently, and often these developments are often not covered by state, local, or federal laws.

Second, there is no consensus about the scope of the problem and where government responsibility lies. The last government study on the use of broadcast and internet media to encourage cyberharassment and abuse was completed in 1993 by the National Telecommunications and Information Administration.⁸⁶ ADL recommends that this report be updated given the breathtaking evolution of technology since 1993.

Additionally, a centralized repository for reporting online abuse would funnel cases to the appropriate authorities and provide more realistic portrait of the extent of online abuse. While law enforcement would recommend that a journalist seeking assistance or protection contact the local FBI office or the Bureau's Internet Crime Complaint Center (IC3), IC3's expertise is in white collar crime.⁸⁷ It would be valuable to have a center dedicated specifically to dealing with cyberharassment.

Finally, new symbols to communicate hate are constantly being developed. ADL, which identifies and tracks these symbols, recently added Pepe the Frog to its database.⁸⁸ Pepe is a green cartoon frog that was originally a popular internet meme, but beginning in 2015, he was co-opted by the alt-right movement as a mascot who came to stand for white supremacy. Law enforcement needs to be aware of a constantly-shifting online environment, where new symbols are used to intimidate targets of abuse and meanings are often opaque or coded.



Pepe the Frog, an internet meme that became a mascot for white supremacy.

CIVIL LAWSUITS AND COPYRIGHT REMEDIES

Finally, targets of online abuse may try to sue their attackers in civil suits or use notice-and takedown procedures from copyright law to remove images.

The applicable torts depend on the specific facts, but often include defamation, intentional infliction of emotional distress, harassment, and public disclosure of private facts.⁸⁹

In order to gain redress through a civil suit, online harassment victims must locate and identify the original speaker.⁹⁰ But potential civil plaintiffs are often stymied by an inability to unmask defendants who have engaged in the harassment. Unmasking may be facilitated, in some instances, with a subpoena directed to an internet service provider seeking information that could identify the speaker. However, some websites choose not to store identifying information about their users, and perpetrators may insulate themselves by using "anonymizers" that make it extremely difficult to identify them.

Copyright is a promising avenue for individuals targeted by images or video. Individuals can file complaints with the Internet Crime Complaint Center or send notice to internet service providers hosting any copyrighted images. However, this is challenging for several reasons.⁹¹ First, targets may not own the copyright of pictures taken of them, since copyright resides with the photographer, so they may have to get ownership of the images transferred to them. Second, copyright remedies often take time, which prolongs the experience of being abused online. Finally, this approach requires the victim to track and trace the spread of harassing images, and forward them on to authorities – which places additional burdens on a target of harassment.

CONCLUSION

JOIN ADL EFFORTS TO COMBAT CYBERHARASSMENT.

We need advocates who want to create safer and more civil online spaces and who are willing to take a stand for real change. We invite you to join with ADL and similar organizations who are working towards solutions to the problem. Addressing hate online is an ADL priority, and the primary focus of the League's Center on Technology and Society.

APPENDIX A: METHODOLOGY

To generate recommendations on how to best combat cyberharassment, ADL convened its Task Force on Hate Speech and Journalism. Building on ADL's decades of experience in monitoring and exposing hate and hate groups, as well as its central role in working with the Internet industry to address online hate, the Task Force provided insight into how to approach the problem.

Additionally, with the help of the Task Force, ADL identified a select group of over 150 outside experts and representatives of journalism, law enforcement, academia, Silicon Valley, and nongovernmental organizations. Participants included leading internet companies, like Google, Facebook, Microsoft, and Twitter. Using a combination of oral and written interviews, the participants were asked to define cyberharassment, to explain what strategies were effective in combatting cyberharassment, and to evaluate what problems currently exist in addressing online harassment. They were then asked to discuss technical, legal, policy, and pragmatic solutions to those problems.

With this advice and counsel, ADL proposed solutions and/or countermeasures that can prevent journalists – and other individuals – from becoming targets for hate speech and harassment on social media.

APPENDIX B: TASK FORCE MEMBERS

TASK FORCE MEMBERS AND ADVISOR TEAM

ADVISORS:

Isaiah Berlin, Senior Fellow in Culture and Policy at The Brookings Institution, Danielle Citron, Professor of Law at the University of Maryland; Steve Coll, Dean of the Columbia University Graduate School of Journalism; Todd Gitlin, Professor and Chair, Ph.D. Program, Columbia Journalism School; Brad Hamm, Dean of the Medill School of Journalism at Northwestern University; Shawn Henry, retired Executive Assistant Director of the Federal Bureau of Investigation; Bethany Mandel, *New York Post*; Leon Wieseltier, Contributing Editor at *The Atlantic* and Christopher Wolf, Partner at Hogan Lovells LLP.

PROJECT TEAM:

Marvin D. Nathan, National Chair
Jonathan A. Greenblatt, CEO & National Director
Glen S. Lewy, President, Anti-Defamation League Foundation
Deborah M. Lauter, Senior Vice President, Policy and Programs
Steven M. Freeman, Deputy Director, Policy and Programs
David Friedman, Vice President, Law Enforcement, Extremism and Community Security
Todd Gutnick, Vice President, Marketing and Communications
Brittan Heller, Director, Technology and Society
Oren Segal, Director, Center on Extremism
Jonathan Vick, Assistant Director, Cyberhate Response
Marilyn Mayo, Research Fellow, Center on Extremism
Jessica Reaves, Content Specialist, Center on Extremism
Daniel Kelley, Assistant Director, Policy and Programs

APPENDIX C: ADL'S HISTORY OF RESPONDING TO RECURRING PATTERNS AND DISTURBING TRENDS

- In the 1930s, when Father Charles Coughlin used the new technology of radio to spew anti-Semitic diatribes and pro-German propaganda over the airwaves, ADL monitored his actions and forcefully opposed him.⁹²
- In the 1950s, at the dawn of television, President Eisenhower used ADL's televised 40th anniversary celebration as a platform to denounce Senator Joseph McCarthy and the anti-communist witch-hunts and conspiracies sweeping the country.⁹³
- In the 1980s, ADL published *Computerized Networks of Hate*, a prescient report raising concern about the spread of hate on new technology platforms, including the use of dial-up computer bulletin boards as a communications tool for any white supremacist with a modem and a home computer.⁹⁴
- ADL was on the forefront of responding to the growing threat of cyberbullying, having been alerted to increased incidents of online harassment through its tracking of anti-Semitic incidents and its anti-bias educational programs. The term cyberbullying emerged online around 2004, with the phenomenon affecting children and sexual minorities.⁹⁵
- In May 2012, ADL convened a Working Group on Cyberhate to develop recommendations for the most effective responses to online hate and bigotry.⁹⁶ Among the Working Group's members are leading representatives of the Internet industry, including Facebook, Google/YouTube, Microsoft, Twitter, and Yahoo, as well as civil society groups, the legal community, law enforcement, and academia. In 2014, with significant input from the Working Group, ADL released a set of Best Practices for Responding to Cyberhate, an initiative that established guideposts for the industry and the Internet community to help prevent the spread of online hate speech.⁹⁷ Industry leaders welcomed these best practices: Facebook said they provide "valuable ways for all members of the internet community to engage on this issue," and Twitter encouraged users to "keep these best practices in mind when dealing with difficult situations online."⁹⁸

BEST PRACTICES FOR RESPONDING TO CYBERHATE

The Working Group on Cyberhate's 2014 Best Practices for Responding to Cyberhate call on internet providers to:

1. Take reports about cyberhate seriously, mindful of the fundamental principles of free expression, human dignity, personal safety and respect for the rule of law.
2. Providers that feature user-generated content should offer users a clear explanation of their approach to evaluating and resolving reports of hateful content, highlighting their relevant terms of service.
3. Offer user-friendly mechanisms and procedures for reporting hateful content.
4. Respond to user reports in a timely manner.
5. Enforce whatever sanctions their terms of service contemplate in a consistent and fair manner.

At the same time, the best practices called on the internet community to:

1. Work together to address the harmful consequences of online hatred.
2. Identify, implement and/or encourage effective strategies of counter-speech — including direct response; comedy and satire when appropriate; or simply setting the record straight.
3. Share knowledge and help develop educational materials and programs that encourage critical thinking in both proactive and reactive online activity.
4. Encourage other interested parties to help raise awareness of the problem of cyberhate and the urgent need to address it.
5. Welcome new thinking and new initiatives to promote a civil online environment.⁹⁹

- In 2016, ADL became an inaugural member of the Twitter Trust & Safety Council, which the company formed to develop strategies to combat hate on their platform while maintaining the ability of Twitter users to freely share their views.¹⁰⁰ Twitter asked ADL and other Council members to provide input on the safety of their products, policies, and programs, and the Council has brought Twitter together with safety advocates, academics, and researchers working to prevent abuse.

APPENDIX D: MORE OF THE LAW REGARDING “TRUE THREATS”

This report provides an overview of the legal framework to help understand the challenges posed in seeking legal redress for cyberhate – and particularly for online threats. The purpose of this Appendix is to provide some additional context by referring to a few particularly important legal cases.

When courts have evaluated whether threatening statements are unprotected speech subject to criminal prosecution, the specifics of the language have mattered more than the effect of the speech. In *Virginia v. Black*, for example, the Supreme Court defined true threats as “statements where the speaker means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals.”¹⁰¹ However, what constitutes a true threat online remains an open question.

The intention of the speaker may matter. In *Elonis v. U.S.*, the defendant was originally convicted of violating a law requiring: “that a communication was transmitted,” “that it contained a threat,” and that the accused individual knew this.¹⁰² The defendant made a series of Facebook comments, which he characterized as “self-styled rap lyrics,” that included a desire to kill his wife, to blow up police officers, to shoot children in a kindergarten class, and to cut the throat of an investigating FBI agent. He argued his writing was “therapeutic” and that he was merely expressing his First Amendment rights. At Elonis’ trial, the jury convicted him on grounds that what he wrote in his posts would be understood by “a reasonable person” as a threat to his estranged spouse and to others who were the targets of his posts.¹⁰³ However, the Supreme Court overturned Elonis’ conviction because he was convicted without sufficient proof that he knew what he was writing and that the ordinary meaning of his words would be a threat.

Other cases, like *State v. Locke*, have parsed the historical and temporal context of threats.¹⁰⁴ In this case, the defendant appealed his conviction for threatening the governor of Washington State. He sent a series of threatening messages to the governor via her website. His first communication read: “I hope you have the opportunity to see one of your family members raped and murdered by a sexual predator. Thank you for putting this state in the toilet. Do us a favor and pull the lever to send us down before you leave Olympia,” which the court found to be hyperbolic political speech rather than a true threat.¹⁰⁵ The second email to the governor told her “you should be burned at the stake like any heretic,” but the court found that this was also not a true threat, because “the ancient political or religious pedigree of burning at the stake” was not a realistic modern threat. However, the third communication was not an email, but an event request, titled “[Governor’s] public execution,” with the location designated as the governor’s residence. Because this communication was sent within seventeen days of the shooting of Congresswoman Gabrielle Giffords, it was found to be a true threat.

Additionally, evidence that the threatening language was intended to be transmitted to the target is an important element of the analysis. In *United States v. Alkhabaz*, the defendant posted several stories to the Usenet group alt.sex.stories which involved the rape, torture, and murder of young women, including one of his university classmates.¹⁰⁶ Investigators later found emails describing the author’s plan to kidnap the young woman and carry out the violent fantasies. The defendant was arrested and charged under 18 U.S.C. § 875(c), which makes it a federal crime to transmit any communication in interstate or foreign commerce containing a threat to injure the person of another. But the Sixth Circuit found that the defendant’s stories did not constitute a true threat, and were therefore protected speech. A key factor in the ruling was that the defendant apparently never intended his classmate to see the emails, and he was not emailing his

correspondent to threaten his classmate or to attempt to intimidate her. Therefore, his emails and stories did not, to the Sixth Circuit, constitute a threat.

What is clear from this line of cases is that courts will examine a variety of factors and closely analyze language, intent, context, intended audience, and usage. This means that language that makes a subject fearful may not constitute a true threat under the law. It may be obnoxious, vile, and hateful — but protected — speech.

In addition, case law suggests that courts may be unwilling to convict defendants for criminal acts that constitute “only” harassment on social media. In *United States v. Cassidy*, a defendant posted on Twitter and a blog targeting a local religious leader.¹⁰⁷ The leader claimed the harassment made her fear for her safety. However, the indictment was dismissed because the speech in question was deemed to be protected speech (in part because the leader was a public figure). Additionally, the court stated the leader had the ability to block content or divert her eyes. Therefore, courts may be wary of cases based solely on online targeting, especially if they follow *Cassidy* and advise that the victims should shield themselves from the harassment.

APPENDIX E: ANTI-SEMITISM IN THE U.S.

In June 2016, ADL released its annual Audit of Anti-Semitic Incidents.¹⁰⁸ The Audit identifies both criminal and non-criminal acts of harassment and intimidation, including distribution of hate propaganda, threats and slurs. It is compiled using information provided by victims, law enforcement, and community leaders and evaluated by ADL’s professional staff. The Audit provides an annual snapshot of one specific aspect of a nationwide problem while identifying possible trends or changes in the types of activity reported. This information assists ADL in developing and enhancing its programs to counter and prevent the spread of anti-Semitism and other forms of hatred.

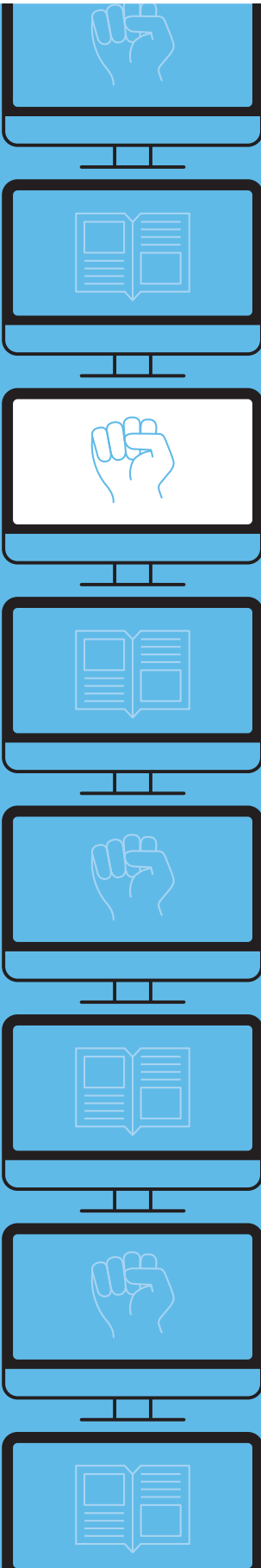
ADL recorded a total of 941 incidents in the U.S. in 2015. This was an increase of about 3 percent from the 912 incidents recorded in 2014. Anti-Semitic incidents were reported in 39 states and the District of Columbia.

Here is the summary of the 2015 findings:

- ADL reported a total of 56 anti-Semitic assaults on Jewish individuals (or individuals perceived as Jewish) in 2015, up from 36 in 2014. Incidents involved the use of physical force and/or violence, spitting and thrown objects. Forty-four of the 56 assault incidents (79 percent) were reported in New York State.
- The ADL Audit reported a dramatic increase in anti-Semitic incidents on campus in 2015. A total of 90 incidents were reported on 60 college campuses, compared with 47 such incidents reported on 43 campuses in 2014.
- The ADL Audit recorded 377 cases of anti-Semitic vandalism in 2015, up slightly from 363 in 2014. Vandalism incidents are individually evaluated by ADL and are categorized as anti-Semitic based on the presence of anti-Semitic symbols or language; the identity of the perpetrator(s), if known; and the target of the vandalism and its proximity to Jewish homes, communities and institutions.
- The ADL Audit recorded 508 cases of anti-Semitic harassment in 2015, down slightly from 513 in 2014. Incidents included verbal attacks and slurs against Jewish individuals (or individuals perceived to be Jewish); anti-Semitism conveyed in written or electronic communications, including anti-Semitic cyberbullying; and anti-Semitic speeches, picketing or events.

Overall, anti-Semitic incident totals in the U.S. were historically low. During the last decade, the number of reported anti-Semitic incidents peaked at 1,554 in 2006 and has been mostly on the decline since.

APPENDIX F: FIRST REPORT



ADL REPORT:

ANTI-SEMITIC TARGETING OF JOURNALISTS DURING THE 2016 PRESIDENTIAL CAMPAIGN

A report from ADL's Task Force on Harassment and Journalism

October 19, 2016



INDEX

Key Findings

**Background: ADL Task Force on
Harassment and Journalism**

Introduction

Methodology

Why Twitter?

Detailed Findings

Spike Analysis

Impact

**White Supremacists Encourage Online
Harassment of Jewish Journalists**

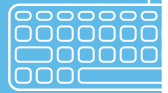
Tweets

Memes

KEY FINDINGS

- Based on a broad set of keywords (and keyword combinations) designed by ADL to capture anti-Semitic language, there were 2.6 million tweets containing language frequently found in anti-Semitic speech between August 2015 – July 2016.
- These tweets had an estimated 10 billion impressions (reach), which may contribute to reinforcing and normalizing anti-Semitic language on a massive scale.
- At least 800 journalists received anti-Semitic tweets with an estimated reach of 45 million impressions. The top 10 most targeted journalists (all of whom are Jewish) received 83 percent of these anti-Semitic tweets.
- 1,600 Twitter accounts generated 68% of the anti-Semitic tweets targeting journalists. 21% of these 1,600 accounts have been suspended in the study period, amounting to 16% of the anti-Semitic tweets.
- Sixty percent of the anti-Semitic tweets were replies to journalists' posts (11% were regular Tweets and 29% re-tweets). In other words, anti-Semitism more often than not occurred in response to journalists' initial posts.
- There was a significant uptick in anti-Semitic tweets in the second half (January-July 2016) of this study period. This correlates to intensifying coverage of the presidential campaign, the candidates and their positions on a range of issues.
- There is evidence that a considerable number of the anti-Semitic tweets targeting journalists originate with people identifying themselves as Trump supporters, "conservatives" or extreme right-wing elements. The words that show up most in the bios of Twitter users sending anti-Semitic tweets to journalists are "Trump," "nationalist," "conservative," "American" and "white." This finding does not imply that Mr. Trump supported these tweets, or that conservatives are more prone to anti-Semitism. It does show that the individuals directing anti-Semitism toward journalists self-identified as Trump supporters and conservatives.
- While anti-Semitic tweets tended to spike in the wake of election-related news coverage, the language used in the anti-Semitic tweets was not solely election-related. Many tweets referenced classic anti-Semitic tropes (Jews control the media, Jews control global finance, Jews perpetrated 9/11, etc.). This suggests that while the initial provocation for anti-Semitic tweets may have been at least nominally election-related, the Twitter users generating targeted anti-Semitism may have used news events as an excuse to unleash anti-Semitic memes, harassment, etc.
- The words most frequently used in anti-Semitic tweets directed at journalists included "kike," "Israel," "Zionist," and "white" etc., an indication that the harassment may have been prompted by the perceived religious identity of the journalist.

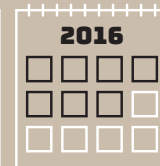
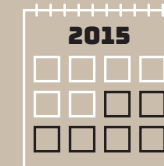
OVERALL DATA PULL



based on keywords correlating
with anti-Semitism

2,641,072

TOTAL MENTIONS



from August 1,
2015 through
July 31, 2016
contained these
keywords

10,000,000,000

Number of estimated impressions generated



66%

Percentage of tweets posted by male
users, based on user-disclosed details

- While anti-Semitism was primarily directed at journalists who are Jewish (or perceived to be Jewish), non-Jewish journalists also received anti-Semitic tweets following criticism of Mr. Trump – presumably intended to be either an insult or threat. This is likely connected to the anti-Semitic tropes related to Jews “controlling” the media, and the media “controlling” the government.
- As previously stated, there is no evidence suggesting these attacks were explicitly encouraged by any campaign or candidate. In fact, ADL has been able to identify individuals and websites in the white supremacist world that have played a role in encouraging these attacks.
- While this report did not investigate whether social media attacks have a chilling effect on journalists, it does show that targeted anti-Semitic tweets raised the cost of entry into (and staying in) the marketplace of ideas for journalists, particularly Jewish journalists.

Please note that this is the first stage of a two-stage reporting process. This data gathering and analysis phase will be followed by a series of recommendations, to be released on November 19, 2016.

NOTE ABOUT THE REPORT AND THE PRESIDENTIAL ELECTION

ADL is a nonprofit organization and does not take sides for or against any candidate for elective office, so it is crucial to be perfectly clear about what this report says and what it does not say.

This report identifies some self-styled followers of presidential candidate Donald Trump to be the source of a viciously anti-Semitic Twitter attack against reporters. Accordingly, we wish to make it clear that based on the statistical work we have performed, we cannot and do not attribute causation to Mr. Trump, and thus we cannot and do not assign blame to Mr. Trump for these ugly tweets. While candidates can and do affect the environment in which social media operates as well as the tenor of its messages, the individuals who tweet hateful words are solely responsible for their messages.

BACKGROUND: ADL TASK FORCE ON HARASSMENT AND JOURNALISM

In June 2016, in the wake of a series of disturbing incidents in which journalists covering the 2016 presidential campaign were targeted with anti-Semitic harassment and even death threats on social media, the Anti-Defamation League (ADL) announced the creation a Task Force on Harassment and Journalism.

Building on ADL's decades of experience in monitoring and exposing hate and hate groups, as well as its critical work with the tech industry in efforts to address online harassment, the Task Force sought insights from a group of experts from the world of journalism, law enforcement, academia, Silicon Valley and nongovernmental organizations. Their advice and counsel will help ADL to do the following:

- Assess the scope and source of anti-Semitic, racist and other harassment of journalists, commentators and others on social media;
- Determine whether and how this harassment is having an impact on the electorate or if it has a chilling effect on free speech;
- Propose solutions and/or countermeasures that can prevent journalists from becoming targets for hate speech and harassment on social media in the future.

With the release of this landmark report, ADL has unveiled the extent to which the 2016 presidential election cycle has exposed journalists to anti-Semitic abuse on Twitter. Our first-of-its-kind investigation included wide-ranging surveys of journalists as well as a quantitative analysis of anti-Semitic Twitter messages and memes directed at reporters.

This initial report, produced by ADL's Center on Extremism, which has worked closely with social media and internet providers for more than two decades in responding to anti-Semitism and online hatred, will be followed by a final report, which will incorporate a broad range of recommended responses to bigotry on social media. The final report will be released at ADL's Never is Now Summit on anti-Semitism on November 17, 2016.

*Participation in the Task Force does not imply agreement with, or assent to, the findings of this report.

INTRODUCTION

Over the course of the 2016 Presidential campaign, an execrable trend has emerged: reporters who voiced even slightly negative opinions about presidential candidate Donald Trump have been targeted relentlessly on social media by the candidate's self-styled supporters; reporters who are Jewish (or are perceived to be Jewish) have borne the brunt of these attacks.

There is evidence that Mr. Trump himself may have contributed to an environment in which reporters were targeted. Indeed, he repeatedly denounced reporters as "absolute scum," and said of "most journalists" in December 2015, "I would never kill them, but I do hate them. And some of them are such lying, disgusting people. It's true." Accordingly, while we cannot (and do not) say that the candidate caused the targeting of reporters, we can say that he may have created an atmosphere in which such targeting arose.

The social media attacks on journalists were brutal.

When journalist Julia Ioffe wrote a profile of Melania Trump for the May 2016 issue of *GQ* magazine, a firestorm of virulently anti-Semitic (and misogynistic) responses on social media followed. One tweet called Ioffe a "filthy Russian kike," while others sent her photos of concentration camps with captions like "Back to the Ovens!"

On May 19, *New York Times* editor Jonathan Weisman tweeted about casino magnate Sheldon Adelson's support for Trump, and the anti-Semitic response to Ioffe's article. The reaction was immediate, with Twitter user CyberTrump leading the charge against Weisman: "Do you wish to remain hidden, to be thought of one of the goyim by the masses?" As other racists and anti-Semites piled on, Weisman received images of ovens, of himself wearing Nazi "Juden" stars, and of Auschwitz's infamous entry gates, the path painted over with the Trump logo, and the iron letters refashioned to read "Machen Amerika Great."

After criticizing Mr. Trump, conservative writer Ben Shapiro became the target of a wave of anti-Semitic tweets calling him a "Christ-Killer" and a "kike." Jake Tapper, John Podhoretz and Noah Rothman have all received similar messages after voicing opinions perceived to be critical of Mr. Trump. In the midst of the attacks, Rothman tweeted: "It never ends. Blocking doesn't help either. They have lists, on which I seem to find myself."

While much of the online harassment of journalists is at the hands of anonymous trolls, there are known individuals and



websites in the white supremacist world that have played a role in encouraging these attacks (see “White Supremacists Encourage the Online Harassment of Jewish Journalists” section).

METHODOLOGY

This report covers the time period of August 2015 through July 2016.

To capture the vast sweep of anti-Semitic Tweets directed at journalists, ADL utilized the latest in “big data” techniques. There were four phases to the report.

Phase one: ADL interviewed journalists impacted by the anti-Semitic harassment and they provided critical background information and described their experiences as targets of harassment on Twitter. They also described the effect the attacks had on their work and personal sense of safety.

Phase two: ADL conducted a search of tweets using a broad set keywords (and keyword combinations) designed by ADL to capture anti-Semitic language. These keywords did not include any terms associated directly with the 2016 presidential campaign. This yielded 2.6 million results.

Phase three: We focused our search on tweets received by a list of 50,000 journalists and compared those with the 2.6 million results.

Phase four: We manually reviewed each of these tweets and narrowed the results to 19,253 overtly anti-Semitic tweets, which we found were directed at 800 journalists.

Note 1: One can never include all of the words that might be used in an anti-Semitic attack, and you can’t predict the ways in which anti-Semites will create “codes” to avoid censure and potential exclusion by social media platforms. (In October 2016, for example, after this analysis was complete, white supremacists attempted to avoid tech-based approaches to isolate online harassment. To do so they assigned tech-oriented code words to their favorite slurs, referring to “kikes” as “Skypes,” among many others).

Note 2: It is impossible to capture all of the anti-Semitic tweets or identify all of the anti-Semitic Twitter users, and because 21 percent of the accounts responsible for tweets containing anti-Semitic language have been deleted (either by Twitter or by the users), there is reason to conclude that the numbers in this report – especially the number of anti-Semitic Tweets received by individual journalists – are conservative.

WHY TWITTER?

This report is focused on Twitter because it is the primary social media platform used to perpetrate these attacks on journalists, according to the journalists themselves.

TWEETS TO JOURNALISTS

based on keywords correlating with anti-Semitism

19,253

ANTI-SEMITIC TWEETS



at US journalists from a pool of 50,000 journalist Twitter handles

45,000,000

Number of estimated impressions generated



CLASSES OF TWEETS

to US journalists

60%

29%

11

Replies Re-Tweets Regular Tweets

We are also not attributing the abuse to the Twitter platform: as with all of the major social media companies, Twitter does not proactively monitor and regulate speech, but like other platforms, claims to respond when hate speech is reported.

ADL conducted a search of tweets using a broad set of keywords (and keyword combinations) designed by ADL to capture anti-Semitic language. These keywords did not include any terms associated directly with the 2016 presidential campaign. This yielded 2.6 million results.

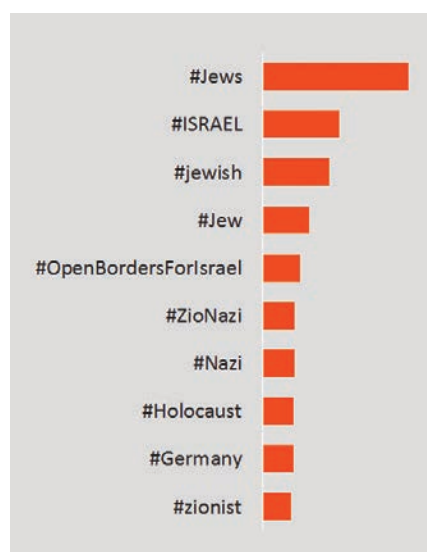
Sixty-eight percent of the 19,253 Tweets were sent by 1,600 Twitter users, confirming that these were persistent attacks on journalists by a relatively small cohort of Twitter users.



* The above word cloud is based on the 2.6 million tweets.



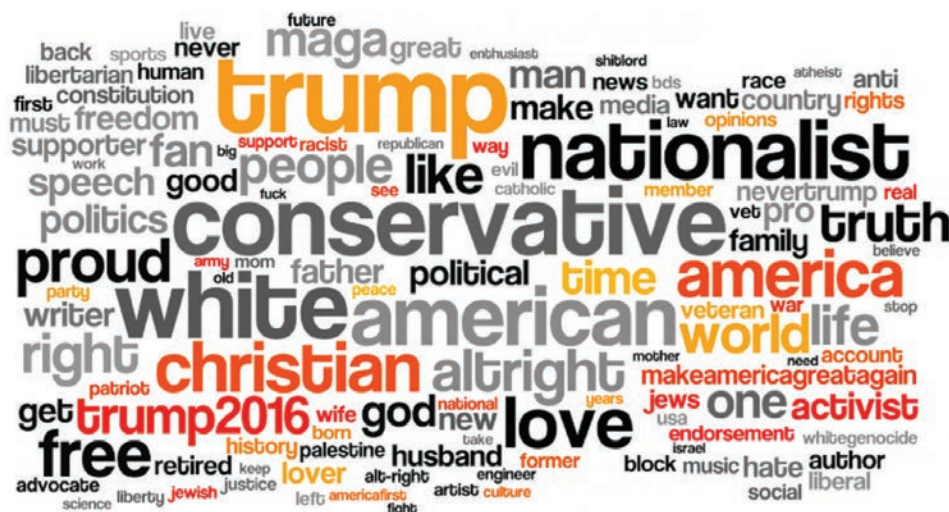
** The above word cloud is based on the 19,253 anti-Semitic Tweets directed at 800 journalists.*



* To the left are the most common Twitter hashtags appearing in the anti-Semitic tweets

There was a significant increase in the volume of anti-Semitic tweets in the second half of the reporting period. Seventy-six percent of Tweets at journalists were posted between February to July 2016. This corresponds with intensifying coverage of the presidential campaign, the candidates, and their positions on a range of issues.

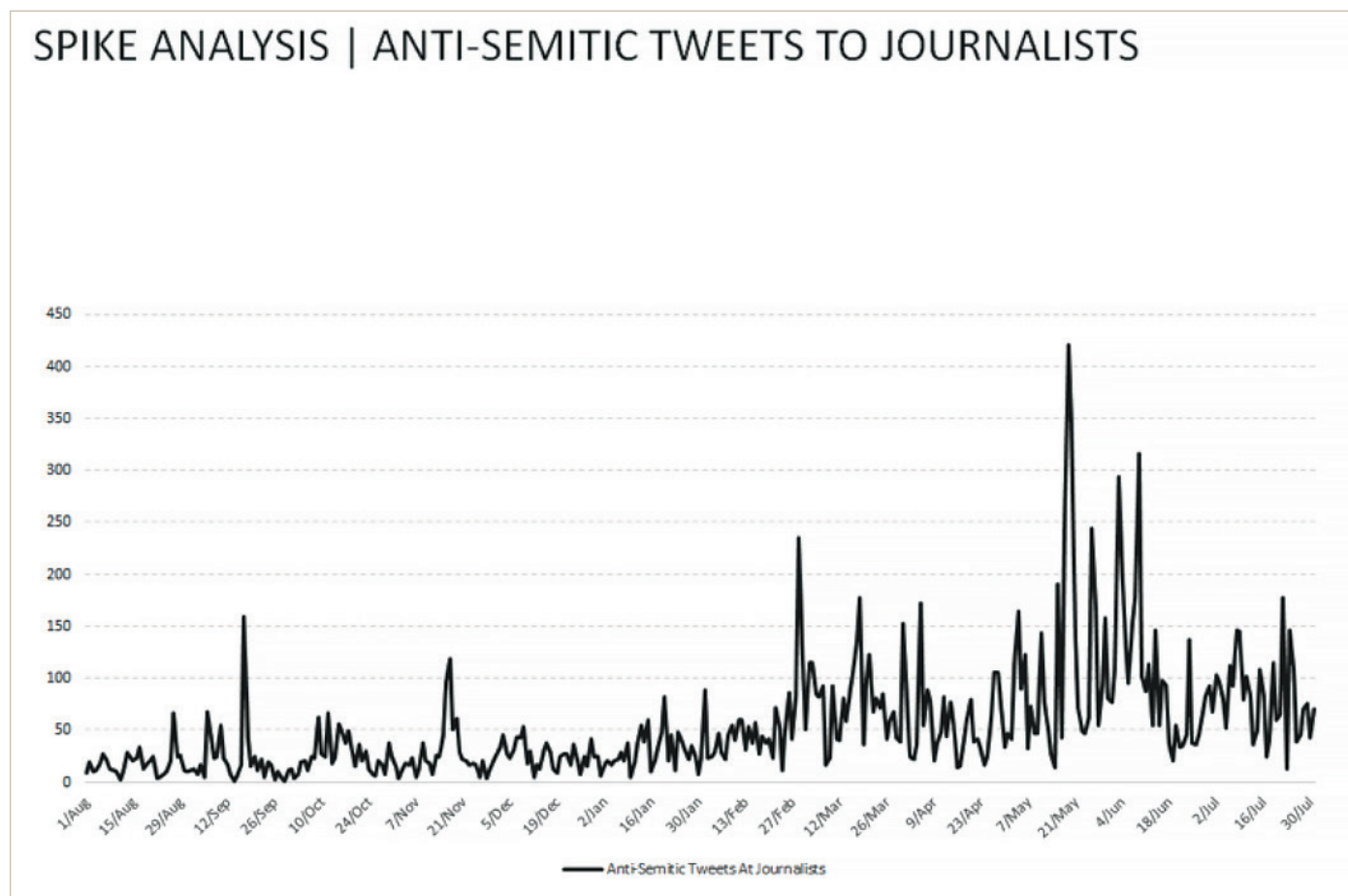
The percentage of these 1,600 users whose accounts were suspended, or 20% of the 19K-plus total tweets removed.



Bethany Mandel

34

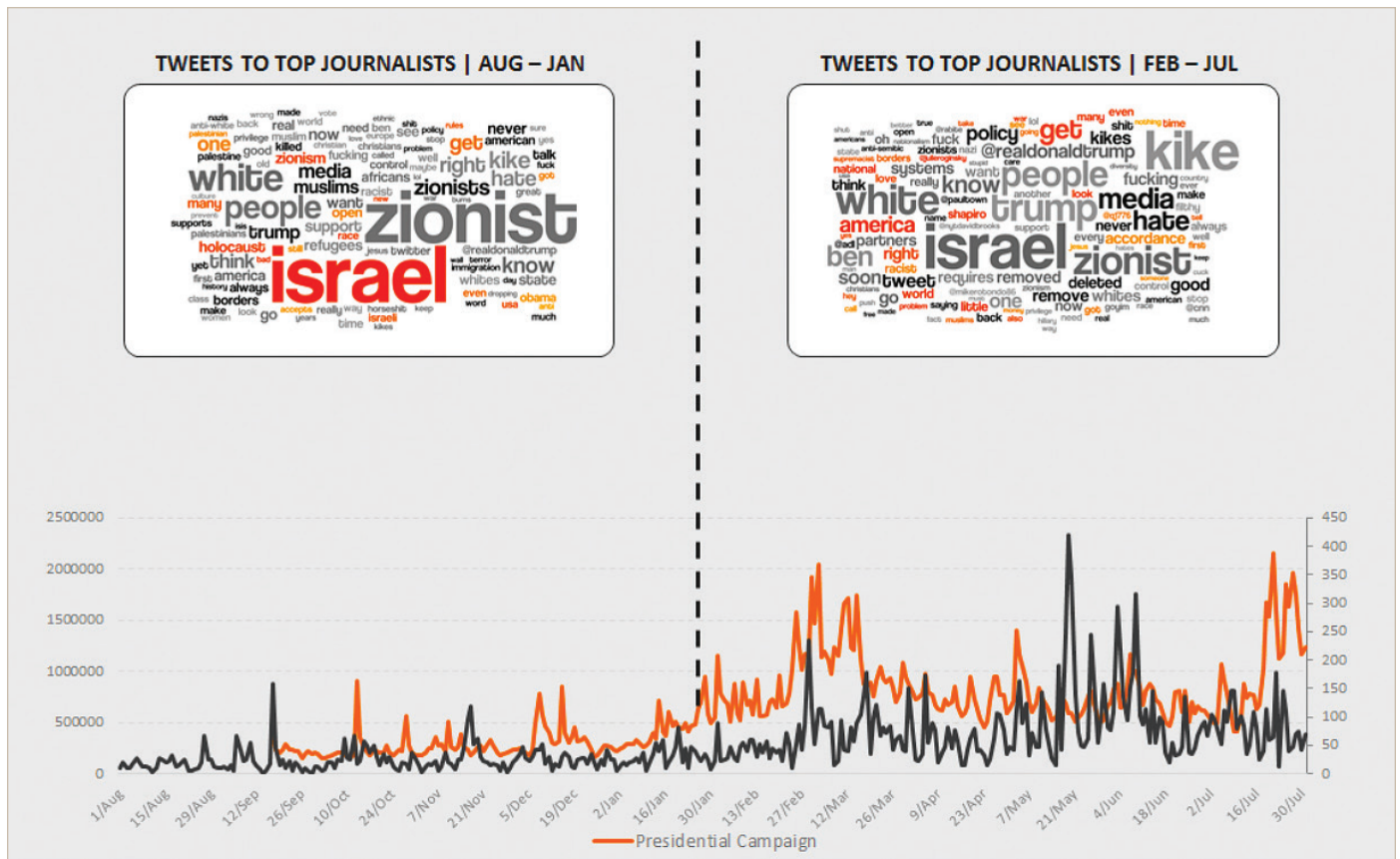
SPIKE ANALYSIS



As stated, there is no known causal relationship between Mr. Trump or his campaign and the wave of anti-Semitic attacks against journalists. However, these self-appointed Trump surrogates used events in the campaign, especially actions by Mr. Trump, as a justification for attacking journalists.

Examples:

- One of the most significant spike in anti-Semitic Tweets occurred on/around March 13, 2016, when Mr. Trump blamed Bernie Sanders for violence at a Trump rally.
- There was a similar spike in anti-Semitic Twitter activity on February 29, 2016, during peak coverage of Trump's refusal to "disavow" the Ku Klux Klan
- Another spike occurred on May 17, 2016, when Melania Trump asserted that Julia Ioffe "provoked" the anti-Semitic attacks against her;
- A similar spike occurred May 25, 2016, when Trump verbally attacked a federal judge whose parents emigrated from Mexico.



But while anti-Semitic tweets demonstrably spiked following election-related news events, the language used in anti-Semitic tweets was not solely election-related. Many tweets referenced classic anti-Semitic tropes (Jews control the media, Jews control global finance, Jews perpetrated 9/11, et cetera).

Racial slurs and anti-Israel statements were the top two manifestations of anti-Semitism. This suggests that while the initial provocation for anti-Semitic tweets may have been related to the election, the Twitter attackers may have used news events - as well as the public airing of these anti-Semitic tweets - as an excuse to unleash more general anti-Semitic memes and attacks. When Jonathan Weisman tweeted about the racist reaction to his comments about Trump, he was inundated by a wave of anti-Semitic Twitter responses.

In February and March 2016, as the so-called #NeverTrump movement took hold, self-styled Trump supporters from the alt-right attacked. (Alt-right is short for "alternative right," a range of people on the extreme right who reject mainstream conservatism in favor of forms that embrace implicit or explicit racism or white supremacy). This is when the Twitter attacks on Ben Shapiro, an originator of the #NeverTrump movement, began in earnest.

"It's amazing what's been unleashed," Shapiro told ADL. "I honestly didn't realize they were out there. It's every day, every single day." Despite Shapiro's efforts to shield his family from the abuse, his wife and baby were targeted as well. "When my child was born there were lots of anti-Semitic responses talking about cockroaches."

Bethany Mandel, a freelance reporter who wrote critically about Trump, was also viciously harassed on Twitter. One user tweeted about her for 19 hours straight, and she received messages containing incendiary language about her family, and images with her face superimposed on photos of Nazi concentration camps. Mandel, like the other Jewish journalists interviewed by ADL, has been targeted by anti-Semitic language before, but these attacks stood out, she said, for their "volume and the imagery. It also seemed coordinated - they would come in waves and 50 percent of the time I couldn't identify the source."

IMPACT

A landmark 2014 Pew Research Center study shows that only five percent of people who are harassed online report the problem to law enforcement. Many more – a combined 31 percent – withdraw, either by changing their username, deleting their account, bowing out of an online forum, or simply not attending certain offline events. When people stop talking because they're afraid, that's evidence of a chilling effect.

But for a lot of people, including journalists, quitting social media simply isn't an option – and the Pew data reflects that. Forty-seven percent of those who are harassed online stood their ground and confronted their tormenter online. Forty-four percent blocked the person responsible, and 22 percent reported the person to the website or online service hosting the exchange.

Half of the journalists we interviewed decided not to report the harassing tweets, some because they believed people should have a right to say whatever they want, and others because they weren't confident Twitter would do anything to address the issue. Across the board, the criticisms of Twitter were consistent: The company doesn't do enough to enforce its terms of service.

Jonathan Weisman told us, "I think suspending or deleting [attackers'] accounts is pointless, because they just come back on under a different name. Twitter has to decide if they are going to stand by their terms of service or not. If they decide tomorrow, 'Look, we don't have the capacity to monitor all of this, and we want it to be a free exchange of ideas,' – then fine, we would know what it was. But they want to have it both ways – the halo of having terms of service, but not enforcing them. Or enforcing them only sporadically."

Some of the journalists, including Weisman, stepped away from Twitter, at least for a while, while others stuck with the platform, hoping for a respite even as they braced for more abuse.

While this particular report did not test whether there was a chilling effect on journalists, it does show that targeted anti-Semitic on Twitter undoubtedly raised the cost of entry into (and staying in) the marketplace of ideas for journalists, particularly Jewish journalists.

WHITE SUPREMACISTS ENCOURAGE ONLINE HARASSMENT OF JEWISH JOURNALISTS

While much of the online harassment of journalists is at the hands of anonymous trolls, there are known individuals and websites in the white supremacist world that have played a role in encouraging these attacks. These people and websites represent a sampling of the people and sites engaged in this activity, and have been on ADL's radar for some time.

Two of the neo-Nazis responsible for some of the attacks on Jewish journalists are Andrew Anglin, founder of the extremely popular white supremacist website The Daily Stormer and Lee Rogers of Infostormer (formerly The Daily Slave). While both Anglin and Rogers are banned from Twitter, they have encouraged their followers to Tweet anti-Semitic language and memes at Jewish journalists, including Julia Ioffe and Jonathan Weisman.

Ioffe wrote a profile of Donald Trump's wife, Melania, for the May 2016 issue of GQ. Anglin and Rogers (self-identified Trump supporters) felt the piece was unflattering. Anglin wrote to his supporters on April 28, "Please go ahead and send her a tweet and let her know what you think of her dirty kike trickery. Make sure to identify her as a Jew working against White interests, or send her the picture with the Jude star from the top of the article." Anglin provided Ioffe's Twitter address and the anti-Semitic picture he mentioned. Rogers followed a similar path a few days later, telling his supporters, "I would encourage a continued trolling effort against this evil Jewish bitch." He then provided Ioffe's Twitter address.

The situation with Jonathan Weisman was somewhat different. After Weisman tweeted out an article by Robert Kagan on the emergence of fascism in the United States and Donald Trump, he was bombarded by anti-Semitic Tweets and memes. Anglin attacked Weisman on May 25, 2016, for publicizing the hateful tweets directed at him. But Anglin went much further. Writing about Weisman and Ioffe, “You’ve all provoked us. You’ve been doing it for decades—and centuries even—and we’ve finally had enough. Challenge has been accepted.”

A couple of days later, Anglin, echoed by “Marcus Cicero” on Infostormer, urged supporters to Tweet anti-Semitic questions at Weisman, including, “Why do Jews demand that White Christians go fight and die in wars for them?”

White supremacist Andrew Auernheimer, an associate of Anglin and an Internet hacker also known as “Weev,” also tweeted at Weisman, “Get used to it you fucking kike. You people will be made to pay for the violence and fraud you’ve committed against us.”

Weisman was one of the first journalists, in the *New York Times*, to publicize another form of harassment – the use of the echo symbol (multiple parentheses) around names to identify that person as Jewish in an article. In his May 26, 2016 article, Weisman noted that some of the anti-Semitic tweets included his name in parentheses. He asked one of the tweeters why, and that person responded, “It’s a dog whistle, fool. Belling the cat for my fellow goyim.”

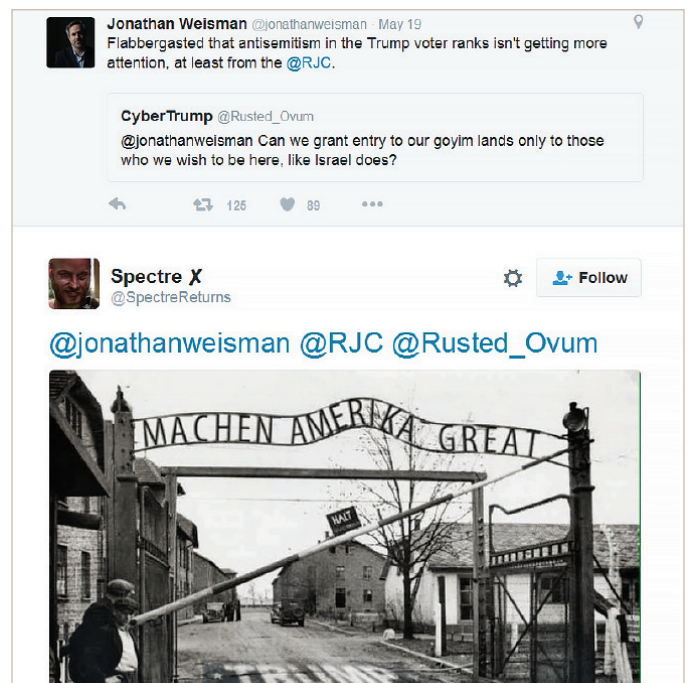
A few days later, two journalists at *Mic* traced the origins of this anti-Semitic typographical symbol to a 2014 podcast “The Daily Shoah” on The Right Stuff (TRS), a racist and anti-Semitic website. The podcast used an echo sound effect when someone on the podcast mentioned a Jewish name. According to TRS, “all Jewish surnames echo throughout history. The echoes repeat the sad tale as they communicate the emotional lessons of our great white sins, imploring us to Never Forget the 6 GoRillion.” Other anti-Semites translated the audio echo into a typographical symbol used primarily on social media sites, including Twitter. TRS was also behind the “Coincidence Detector” app, a Google Chrome plugin (removed on June 2, 2016 by Google) whose purpose was, according to Mic, “compiling and exposing the identities of Jews and others who are perceived as ‘anti-white.’” According to the creators of the app, it “can help you detect total coincidences about who has been involved in certain political movements and political empires.” It was, of course, referring to Jews. Users of the app would then put the echo around a Jewish name.

The publicity generated by the echo symbol resulted in a more widespread, defiant counter-use of the echo, as thousands of Twitter users, including Jewish journalists, changed their Twitter screen names to echo themselves.

TWEETS

Of the 19,253 Tweets sent to 800 journalists, 79 percent were text only, while 12 percent contained links and 8 percent contained images.

A (small) sampling of anti-Semitic tweets sent to journalists:





Julia loffe

MEMES

A few of the most frequently employed memes in the anti-Semitic online (Twitter) harassment of journalists:



Dana Schwartz (Observer)
This meme is repeated with various journalists pictured inside the gas chamber.

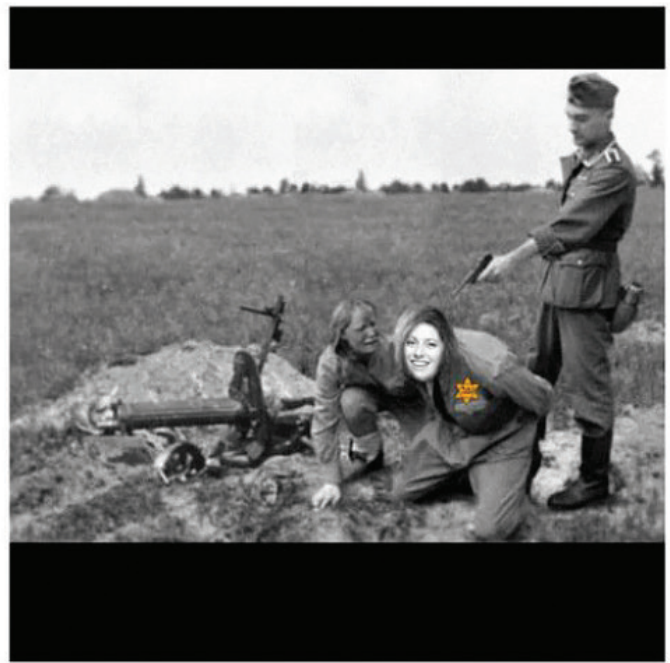




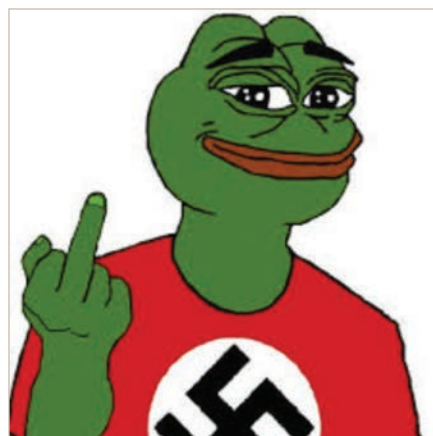
AmeriKKKanHistoryX™
@KingAntiochus



@bethanyshondark Shalom Bethany!



AmeriKKKanHistoryX™ @KingAn... 1h
@bethanyshondark Disgusting, filthy, and treacherous Jewess gets a



Bethany Mandel

ANTI-DEFAMATION LEAGUE TASK FORCE REPORT PROJECT TEAM

Marvin D. Nathan
National Chair

Jonathan A. Greenblatt
CEO & National Director

Glen S. Lewy
President, Anti-Defamation League Foundation

Deborah M. Lauter
Senior Vice President, Policy and Programs

Steven M. Freeman
Deputy Director, Policy and Programs

David Friedman
*Vice President, Law Enforcement,
Extremism and Community Security*

Todd Gutnick
Vice President, Marketing and Communications

Brittan Heller
Director, Technology and Society

Oren Segal
Director, Center on Extremism

Jonathan Vick
Assistant Director, Cyberhate Response

Marilyn Mayo
Research Fellow, Center on Extremism

Jessica Reaves
Writer, Content Specialist, Center on Extremism

Daniel Kelley
Assistant Director, Policy and Programs

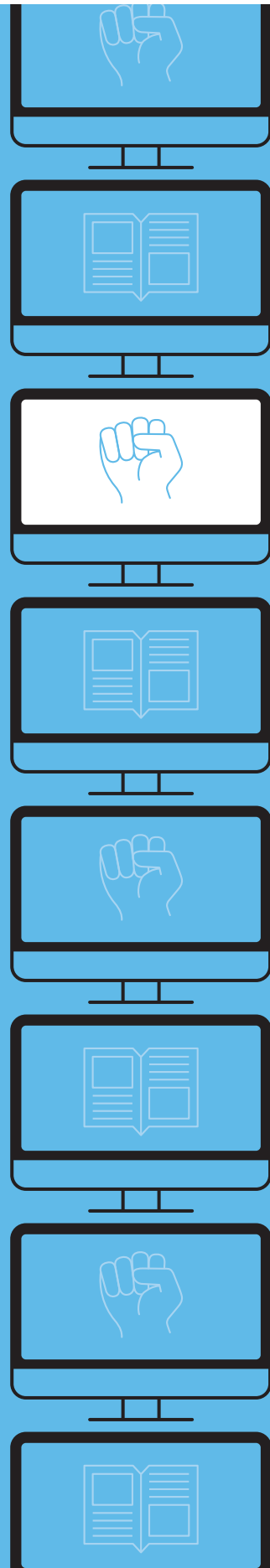
For additional and updated resources please see: www.adl.org

Copies of this publication are available in the Rita and Leo Greenland Library and Research Center.

©2016 Anti-Defamation League | Printed in the United States of America | All Rights Reserved



Anti-Defamation League
605 Third Avenue, New York, NY 10158-3560
www.adl.org



1 See ADL's first report at http://www.adl.org/assets/pdf/press-center/CR_4862_Journalism-Task-Force_v2.pdf

2 <http://www.newsweek.com/epileptogenic-pepe-video-507417>

3 <http://www.adl.org/press-center/press-releases/anti-semitism-usa/task-force-report-anti-semitic-harassment-journalists-twitter-2016-campaign.html>

4 <http://www.adl.org/about-adl/>

5 Id.

6 <http://www.adl.org/education-outreach/bullying-cyberbullying/>

7 <http://www.adl.org/press-center/press-releases/discrimination-racism-bigoity/adl-forms-task-force-to-address-anti-semitic-racist-harassment-journalists-social-media-1.html#.WB9nkrldU>

8 <http://fusion.net/story/327103/leslie-jones-twitter-racism/>

9 <http://www.dailydot.com/via/phillips-dont-feed-trolls-antisocial-web/>

10 <http://www.nytimes.com/roomfordebate/2014/08/19/the-war-against-online-trolls/women-and-minorities-as-targets-of-attack-online>

11 Danielle Keats Citron (2014). *Hate Crimes in Cyberspace*. Harvard University Press. ISBN 978-0-674-36829-3. <http://www.hup.harvard.edu/catalog.php?isbn=9780674368293>

12 <https://blog.twitter.com/2016/announcing-the-twitter-trust-safety-council>

13 <https://www.bloomberg.com/news/articles/2016-10-17/disney-said-to-have-dropped-twitter-pursuit-partly-over-image>

14 <https://techcrunch.com/2016/10/27/twitter-lays-off-9-of-its-workforce-as-it-posts-a-much-needed-positive-q3/>

15 <http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/>

16 <http://www.globalwebindex.net/blog/social-media-captures-30-of-online-time>

17 <http://www.dreamgrow.com/top-15-most-popular-social-networking-sites/>

18 Pew, online harassment., <http://www.pewinternet.org/2014/10/22/online-harassment/>

19 Danielle Keats Citron (2014). *Hate Crimes in Cyberspace*. Harvard University Press. ISBN 978-0-674-36829-3. <http://www.hup.harvard.edu/catalog.php?isbn=9780674368293>

20 Id.

21 Id.

22 See Facebook's explanation of their process here: <http://wersm.com/how-does-facebook-moderate-content-infographic/>

23 <http://www.theatlantic.com/technology/archive/2016/08/the-social-media-invisibles/497729/>

24 <http://news.microsoft.com/features/microsofts-photodna-protecting-children-and-businesses-in-the-cloud>

25 <https://www.wired.com/2014/10/content-moderation/>

26 For example, certain E.U. countries impose strict legal bans on the use of Nazi imagery and Holocaust denial.

27 <http://www.niemanlab.org/2016/10/when-are-comments-sections-of-news-sites-worth-keeping-alive-what-are-some-options-for-taming-them/>;
<https://engagingnewsproject.org/research/10-things-we-learned-by-analyzing-9-million-comments-from-the-new-york-times>

28 Id.

29 Id.

30 Id.

31 Id.

32 <http://www.niemanlab.org/2015/09/what-happened-after-7-news-sites-got-rid-of-reader-comments/>; <https://www.wired.com/2015/10/brief-history-of-the-demise-of-the-comments-timeline/>

33 <http://www.wan-ifra.org/reports/2016/10/06/the-2016-global-report-on-online-commenting>

34 <http://www.niemanlab.org/2016/10/when-are-comments-sections-of-news-sites-worth-keeping-alive-what-are-some-options-for-taming-them/>

35 <http://www.osce.org/fom/220411>

36 See, e.g., <http://andreaforte.net/ForteCSCW17-Anonymity.pdf>

37 <http://www.businessinsider.com/google-jigsaw-anti-harassment-tool-conversation-ai-2016-9>

38 <http://takingnote.blogs.nytimes.com/2016/09/21/can-bots-fight-bullying>

39 <https://www.wired.com/2016/09/inside-googles-internet-justice-league-ai-powered-war-trolls/>

40 https://www.buzzfeed.com/alexkantrowitz/racist-social-media-users-have-a-new-code-to-avoid-censorshi?utm_term=.rro3vRz40#.ekkD1K8Ga

41 <http://www.businessinsider.com/twitter-meaningful-changes-tackle-abuse-harassment-november-2016-10>

42 <http://www.forbes.com/sites/kalevleetaru/2016/01/15/is-the-internet-evolving-away-from-freedom-of-speech/#3622f6236770>; <http://motherboard.vice.com/read/the-history-of-twitlers-rules>

43 <http://www.pcworld.com/article/2045904/twitter-introduces-intweet-abuse-button-after-complaints.html>

44 <http://fusion.net/story/327103/leslie-jones-twitter-racism/>

45 <https://blog.twitter.com/2016/announcing-an-application-process-for-verified-accounts-0>

46 <https://blog.twitter.com/2016/new-ways-to-control-your-experience-on-twitter>

47 http://www.slate.com/articles/technology/technology/2016/10/how_twitter_s_verification_tool_could_help_solve_its_abuse_problem.html

48 http://www.slate.com/articles/technology/bitwise/2016/01/twitter_needs_a_drastic_plan_to_save_itself_here_it_is.html

49 <http://www.independent.co.uk/life-style/gadgets-and-tech/can-wikipedia-save-the-internet-a7380786.html>

50 <https://www.wired.com/2014/05/fighting-online-harassment/>

51 Id.

52 Id.

53 <https://iheartmob.org>

54 Id.

55 Id.

56 <http://www.troll-busters.com/>

57 Id.

58 <http://alldigitocracy.org/combating-hate-speech-against-women-on-twitter/>

59 <http://www.international-alert.org/>

60 <http://blogs.voanews.com/techtonics/2016/10/21/hate-speech-plugin-gives-internet-trolls-a-chance-to-pause/>

61 <http://www.crashoverridenetwork.com/>

62 <https://twitter.com/crashoverridenw/status/601111469904662528>

63 <http://wmcspeechproject.com/>

64 <https://www.hashtags.org/platforms/instagram/instagram-update-introduces-ban-on-certain-hashtags/>; <https://www.wired.com/2015/06/twitter-block-list/>

65 <https://coralproject.net/>

66 <http://www.crashoverridenetwork.com/resources.html>

67 Id.

68 <https://www.eff.org/issues/bloggers/legal/liability/230>

69 <https://www.law.cornell.edu/uscode/text/47/230>

70 <http://www.fas.org/sgp/crs/misc/95-815.pdf>

71 Id.

72 See Appendix D for a further discussion of true threats jurisprudence.

73 <http://injury.findlaw.com/torts-and-personal-injuries/defamation-law-the-basics.html>

74 Id.

75 <http://www.nytimes.com/2008/06/11/world/americas/11iht-hate.4.13645369.html>

76 <https://cdt.org/files/pdfs/MorrisTerrorRecruitTestimonyFinal.pdf>

77 Danielle Keats Citron (2014). *Hate Crimes in Cyberspace*. Harvard University Press. ISBN 978-0-674-36829-3. <http://www.hup.harvard.edu/catalog.php?isbn=9780674368293>

78 <http://computer.howstuffworks.com/what-is-doxing.htm>

79 <http://scholarship.law.unc.edu/nclr/vol94/iss1/3>

80 <http://www.ibtimes.com/what-swatting-celebrities-gamers-now-congresswoman-have-all-been-targeted-2289880>

81 For practical tips of how journalists can protect themselves from doxing, see here: <http://niemanreports.org/articles/how-to-deter-doxing/>

82 <http://www.adl.org/combating-hate/hate-crimes-law/>

83 Id.

84 <http://fortune.com/2016/09/30/german-facebooks-zuckerberg-hate-speech-complaint/>

85 <https://nobullying.com/cyber-harassment-laws/>

86 <https://www.ntia.doc.gov/legacy/reports/1993/TelecomHateCrimes1993.pdf>

87 <https://www.ic3.gov/default.aspx>

88 <http://www.adl.org/combating-hate/hate-on-display/c/pepe-the-frog.html>

89 <http://www.theatlantic.com/technology/archive/2014/11/what-the-law-can-and-cant-do-about-online-harassment/382638/>

90 See the history of filings in the Autoadmit case, <http://www.dmlp.org/threats/autoadmit>

91 <http://www.theatlantic.com/technology/archive/2014/11/what-the-law-can-and-cant-do-about-online-harassment/382638/>

92 <http://www.adl.org/assets/pdf/press-center/TEN-DECADES-OF-ADL-IMPACT.pdf>

93 Id.

94 http://archive.adl.org/poisoning_web/introduction.html

95 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4276384/#b1-pch-19-527>

96 http://www.adl.org/combating-hate/cyber-safety/best-practices/#.WB9kR_krIdV

97 http://www.adl.org/combating-hate/cyber-safety/best-practices/#.WB9ev_krIdU

98 Id.

99 <http://www.adl.org/combating-hate/cyber-safety/best-practices/#.WCF96xlrKu6>

100 <https://blog.twitter.com/2016/announcing-the-twitter-trust-safety-council>

101 <http://caselaw.findlaw.com/us-supreme-court/537/465.html>

102 <http://www.scotusblog.com/case-files/cases/elonis-v-united-states/>

103 Id.

104 <http://caselaw.findlaw.com/wa-court-of-appeals/1641379.html>

105 Id.

106 <http://caselaw.findlaw.com/us-6th-circuit/1279576.html>

107 <http://www.nytimes.com/2011/12/16/technology/judge-dismisses-case-of-accused-twitter-stalker.html>

108 <http://www.adl.org/press-center/press-releases/anti-semitism-usa/2015-audit-anti-semitic-incidents.html#.WCF63xlrKu4>

109 http://www.adl.org/assets/pdf/press-center/CR_4862_Journalism-Task-Force_v2.pdf

ANTI-DEFAMATION LEAGUE TASK FORCE REPORT PROJECT TEAM

Marvin D. Nathan
National Chair

Jonathan A. Greenblatt
CEO & National Director

Glen S. Lewy
President, Anti-Defamation League Foundation

Deborah M. Lauter
Senior Vice President, Policy and Programs

Steven M. Freeman
Deputy Director, Policy and Programs

David Friedman
*Vice President, Law Enforcement,
Extremism and Community Security*

Todd Gutnick
Vice President, Marketing and Communications

Brittan Heller
Director, Technology and Society

Oren Segal
Director, Center on Extremism

Jonathan Vick
Assistant Director, Cyberhate Response

Marilyn Mayo
Research Fellow, Center on Extremism

Jessica Reaves
Writer, Content Specialist, Center on Extremism

Daniel Kelley
Assistant Director, Policy and Programs

For additional and updated resources please see: www.adl.org

Copies of this publication are available in the Rita and Leo Greenland Library and Research Center.

©2016 Anti-Defamation League | Printed in the United States of America | All Rights Reserved



Anti-Defamation League
605 Third Avenue, New York, NY 10158-3560
www.adl.org

