ENLISTING USEFUL IDIOTS: THE TIES BETWEEN ONLINE HARASSMENT AND DISINFORMATION

BRITTAN HELLER*

Online harassment has been a blind spot for major platforms for many years. The problem became mainstream with Gamergate in 2014, the first public reckoning with intimidation of women in the online gaming community. This problem is still plaguing social media, with progress being made in fits and starts after publicized incidents of bullying or silencing of minority voices.[1] *However, the problem has grown beyond these applications to new harms. Online harassment has created serious policy, technical, and structural vulnerabilities that have been exploited by malign actors and gone largely unnoticed—or unprioritized—by defenders. Trolling has become the vocabulary and testing ground of digital authoritarians. Understanding how online harassment works is integral to combatting State-based disinformation and efforts to undermine faith in both democracy and the internet.*[2]

This Article describes the threat posed by online harassment and then outlines how lessons learned from combatting online harassment can be used to counter a wide range of disinformation actors. To do this, the Article will define "digital authoritarians" to show how disinformation and online harassment are connected. It will then use two examples of online harassment to demonstrate how

[*] Brittan Heller is Counsel in the Global Business and Human Rights practice at Foley Hoag LLP. She was an inaugural 2019-2020 Technology and Human Rights Fellow at Harvard Kennedy School. Heller was the founding director of the Center for Technology and Society and prosecuted human rights and technology-related crime at the U.S. Department of Justice. She received her J.D. from Yale Law School and her M.A./B.A. from Stanford University.

This Article is based on an address, "Tech Pessimism," given at Silicon Flatirons Conference on February 9, 2020. Special thanks to Amie Stepanovich, Ben Nimmo, Camille Francois, Nathaniel Gleicher, Danielle Citron, Ben Wittes, Mark Lemley, and Gare Smith for their input and support.

[1] Caitlin Dewey, *The Only Guide to Gamergate You Will Ever Need to Read*, WASH. POST (Oct. 14, 2014), https://www.washingtonpost.com/news/the-intersect/wp/2014/10/14/the-only-guide-to-gamergate-you-will-ever-need-to-read/ [https://perma.cc/7VNQ-AYJW].

[2] *See, e.g.*, Adrian Shahbaz, *The Rise of Digital Authoritarianism*, FREEDOM HOUSE, https://freedomhouse.org/report/freedom-net/2018/rise-digital-authoritarianism [https://perma.cc/MWU3-XRM8] (last visited Oct. 12, 2020) (discussing digital authoritarianism).

the phenomenon has been a continual problem for online platforms, using post-Gamergate examples. These stories show how harassment has grown more complex—allowing malign actors to go beyond intimidating individuals to targeting entire communities via systemic platform vulnerabilities. The weaknesses are created by missing or underenforced content moderation policies. While many articles and essays discuss the harm of online harassment, few scholars and advocates discuss the influence of harassment on elections or democracy in general.

This Article concludes with three concrete suggestions of how industry can counter harassment-based disinformation and the targeting of vulnerable persons and groups. First, companies should commit to a risk-based allocation of resources. This is especially important when it comes to addressing harassment and disinformation. Second, companies should also follow best practices in other industries by conducting human rights impact assessments. Harassment should be included in the topics surveyed. And third, companies should take a systemic, and not a piecemeal, approach to online harassment by looking at behaviors instead of actors and content. If these are accomplished, online harassment can be taken out of the digital authoritarians' toolbox as a means to influence elections, undermine democracy, and eliminate political critics—which means it is more critical than ever to take online harassment seriously.

INTRODUCTION

Maria Ressa is a world-renowned journalist with a problem.[3] She is the founder and CEO of Rappler, an independent investigative news outlet founded in the Philippines in 2012.[4] Her background consists of over thirty years of experience, including running CNN's Manila and winning the 2018 Person of the Year Award from Time Magazine.[5]

Despite her renown and expertise, in recent years Ressa has found both Rappler—and herself—the targets of online smear campaigns. Ressa's prominence grew with her coverage of the populist and authoritarian government in the Philippines. After the 2016 election ushered in Rodrigo Duterte's administration, a series of widespread disinformation campaigns targeted the president's critics, his political opponents, and independent media.[6]

Ressa documented these systematic campaigns of online harassment and mapped them to government sources. The attacks included networks of pro-Duterte bloggers and bot-driven social media accounts designed to malign the independent press and increase support for the president.[7] Ressa described the ferocity of the harassment and its origin in her own words:

> Rappler and I became a target after we did a series on the "propaganda wars." We released stories that showed how this hate was being used to create doubt in institutions and in journalists… That triggered a wave of attacks against me and against Rappler that reached as many as ninety hate messages per hour… ninety messages a week, you can handle, but an hour? That becomes a whole different ball game, and our response to it was to do what we do as

[3] *See Maria Ressa on Digital Disinformation and Philippine Democracy in the Balance*, NAT'L ENDOWMENT FOR DEMOCRACY, https://www.ned.org/wp-content/uploads/2018/02/Maria-Ressa-on-Digital-Disinformation-and-Philippine-Democracy-in-the-Balance.pdf [https://perma.cc/P8W8-GHRC] (last visited Oct. 12, 2020).

[4] *The People Behind Rappler*, RAPPLER (June 16, 2012, 4:19 PM), https://www.rappler.com/about/the-people-behind-rappler [https://perma.cc/UH6M-APWG]; *About Rappler*, RAPPLER (Dec. 14, 2011, 8:00 AM), https://www.rappler.com/about/about-rappler [https://perma.cc/2ASK-5XD2].

[5] *Maria Ressa on Digital Disinformation and Philippine Democracy in the Balance*, *supra* note 3; Karl Vick, *2018: The Guardians: Maria Ressa*, TIME MAG., (Mar. 5, 2020, 6:50 AM), https://time.com/5793800/maria-ressa-the-guardians-100-women-of-the-year/ [https://perma.cc/FQ56-G4MF].

[6] *Maria Ressa on Digital Disinformation and Philippine Democracy in the Balance*, *supra* note 3.

[7] *Id.*

> journalists: to shine a light and tell people that these attacks were happening, that journalists were being targeted. . .
>
> After journalists were targeted, opposition politicians were next, and the one who I think really bore the brunt of the propaganda machine's attacks was Senator Leila de Lima, the former Commission on Human Rights chief and justice secretary who had been investigating Duterte and then became a Senator; President Duterte began targeting her and within a few months, she was jailed. . . . The harbinger of the attacks against her in the real world was a social media campaign. What we saw with these attacks is not just an attempt to tear down the credibility of anyone questioning or perceived to be a critic of government, but also to seed doubt in truth, and this is where you can see the disinformation campaign that continues today.[8]

Ressa's story, unfortunately, is not unique. Online harassment, including the targeting of journalists, is an old problem.[9] However, a recent development is the harnessing of online harassment as a weapon by governments around the world to silence critics and bend the truth to their will.[10] As this Article will show, in countries around the world—from the Philippines to China, Russia and elsewhere—governments have adopted techniques from trolls, exploited platform vulnerabilities, and used harassment to spread disinformation via unwitting accomplices.

I. ONLINE HARASSMENT IS A CONTINUAL PROBLEM FOR ONLINE PLATFORMS.

The story of Maria Ressa exemplifies how governments have invested in the online harassment game. The term "digital authoritarians" describes primarily State actors who use the internet to enable mass manipulation and intimidation to meet their political ends and silence dissent.[11] As this Article will

[8] *Id.*

[9] *See, e.g.*, *Anti-Semitic Targeting of Journalists During the 2016 Presidential Campaign*, ANTI-DEFAMATION LEAGUE (Oct. 19, 2016), https://www.adl.org/sites/default/files/documents/assets/pdf/press-center/CR_4862_Journalism-Task-Force_v2.pdf [https://perma.cc/PJ8G-KJYS]; Caroline Sinders, *An Incomplete (but growing) History of Harassment Campaigns since 2003*, MEDIUM (Nov. 25, 2018), https://medium.com/digitalhks/an-incomplete-but-growing-history-of-harassment-campaigns-since-2003-db0649522fa8 [https://perma.cc/SQT5-BGPK].

[10] *See Maria Ressa on Digital Disinformation and Philippine Democracy in the Balance*, *supra* note 3 (discussing Chinese and Russian state-sponsored social media accounts).

[11] Shahbaz, *supra* note 2.

describe, malign actors often use online harassment to fuel their own agendas.[12]

We can best understand how this threat has evolved, and why it is so effective, by looking at the experiences of those who have been targeted by online violence and harassment, whether or not a government agent was responsible for the harassment. As two of these examples show, harassment is an effective hammer to silence voices because the experiences of those targeted go beyond disruptive to being frightening, or even dangerous—and they cannot be solved by simply turning off a computer.

Noor is a young woman who lives in the Washington D.C. metro area.[13] Around January 2017, at the time of President Trump's inauguration celebration, Noor was targeted by online accusations.[14] She was accused of setting a Trump supporter's hair on fire, and supposed video evidence was posted online and viewed by 1.5 million people.[15] However, the truth was that the woman in the video was not Noor.[16] In fact, it was a woman with only a passing resemblance to her.[17] When the incident happened, Noor was actually visiting her gravely ill father.[18]

The targeting of Noor was racist, sexist, and anti-Muslim, and paired offline threats with online harassment to terrorize Noor and her family. Noor was misidentified—most likely intentionally by the original poster—by crowdsourcing on forums like 4Chan, WeSearchr, and right-wing political forums.[19] The cybermob combined its fixation on her with calls to make her pay for her supposed transgressions.[20] Anonymous calls were placed to Noor's work.[21] Her email inbox and Facebook page began to fill with angry and violent communications from strangers.[22] After Noor's family noticed cars circling her house, they moved into a hotel for a week.[23]

[12] Lisa Reppell & Erica Shein, *Disinformation Campaigns and Hate Speech: Exploring the Relationship and Programming Interventions*, INT'L FOUND. FOR ELECTORAL SYS. (Apr. 2019), https://www.ifes.org/sites/default/files/2019_ifes_disinformation_campaigns_and_hate_speech_briefing_paper.pdf [https://perma.cc/6SQ7-K5ZY].

[13] Terry Collins, *Here's the Brutal Reality of Online Hate: Attack of the Trolls*, CNET (Nov. 27, 2017, 3:20 PM), https://www.cnet.com/news/the-brutal-reality-of-online-hate-neo-nazis/ [https://perma.cc/ZR3L-X8XF] (name changed for privacy considerations).

[14] *Id.*

[15] *Id.*

[16] *Id.*

[17] *Id.*

[18] *Id.*

[19] *Id.*

[20] *Id.*

[21] *Id.*

[22] *Id.*

[23] *Id.*

Noor sought to get the misidentifying posts taken down, but at least one website that played host to the amateur sleuths demanded a statement from police indicating that Noor was not a subject or target in the investigation of the alleged assault on the protestor.[24] The people behind the website saw themselves as vigilante online investigators who were unmasking the perpetrator of an assault, and were suspicious of efforts to remove what they considered to be their good detective work.[25] This Article posits that disinformation is often a component of online harassment—threats are often founded on a mixture of true and false statements, weaving a tapestry around the victim that leaves them struggling to separate truth from fiction.[26]

Initially, local law enforcement was unsympathetic to Noor's plight. Her experience was common to many harassment victims[27]—local police told her that all she could do was to disengage from the internet and turn off her devices. Law enforcement protocol prevented the department from publicly clearing her before the investigation was positively concluded.[28] However, once they saw the extent of the ongoing harassment, after being presented with a dossier proving that Noor was with her father and his hospice nurse at the time of the assault, the police issued a statement exonerating her, which helped dampen the cybermob.[29]

Julie is a journalist, who must engage with the public on Twitter in order to do her job.[30] During coverage of the 2016 U.S. presidential campaign, Julie was targeted by masses of anti-Semitic trolls.[31] For example, for nineteen hours straight, Julie received a series of Tweets: herself superimposed into a concentration camp wearing a yellow Jewish star; her face pasted on a string of victims' bodies; images of herself in a gas chamber;

[24] The author personally knows this from her background in law enforcement, particularly from working as a prosecutor in the District of Columbia where this incident occurred.

[25] The author personally knows this from her background in law enforcement, particularly from working as a prosecutor in the District of Columbia where this incident occurred.

[26] *See*, *e.g.*, Reppell & Shein, *supra* note 12.

[27] *See* Julie Zeilinger, *An interview with Cynthia Lowen, director of online harassment documentary 'Netizens'*, WOMEN'S MEDIA CTR. (May 30, 2018), https://womensmediacenter.com/fbomb/an-interview-with-cynthia-lowen-director-of-online-harassment-documentary-netizens [https://perma.cc/PQM4-UNLY] (explaining that disengaging from the internet is common advice given to victims of cyber abuse).

[28] The author personally knows this from her background in law enforcement, particularly from working as a prosecutor in the District of Columbia where this incident occurred.

[29] Collins, *supra* note 13.

[30] Name has been changed for privacy considerations.

[31] *Anti-Semitic Targeting of Journalists During the 2016 Presidential Campaign*, *supra* note 9, at 3, 8.

and images of the gates of Auschwitz altered to read "Machen Amerika Great."[32] After receiving thousands of messages like these, Julie bought herself a firearm.[33] Julie's fellow political journalist Elizabeth also received similar negative attention, going beyond social media to include a phone call from "Overnight Caskets."[34]

This carefully seeded falsehood was enough to make both Elizabeth and the business owner part of this effort—one as a victim, the other as an unwitting tool. It is telling that there is a term from Russian disinformation operations—a "useful idiot"—that describes exactly this technique, where an unwitting bystander is used to perpetuate a disinformation campaign.[35]

Unfortunately, these stories are not unusual. Each could be taken from the headlines in the mid-2010s, when Gamergate and other controversies like the harassment of female online personalities like Kathy Sierra introduced online harassment to the public, and would not be out of place today.[36] While most cases of online harassment are not as extreme as the examples listed above, what begins online often moves to offline consequences.[37] Outside authorities—be it platform content reviewers, school officials, or police officers—are enlisted as useful idiots, squandering their resources while further victimizing the targets.[38] Furthermore, in many cases, the perpetrators are intimately familiar with the contours and weaknesses of the law.[39] They use this knowledge to

[32] *Id.* at 11–14. Note that while the original Tweets were deleted by Twitter, copies of the tweets using this imagery are available within the report.

[33] Matt Katz, *Trump-Inspired Anti-Semitism Prompts Fear, Police Reports…and a Gun Purchase*, WNYC (June 28, 2016), https://www.wnyc.org/story/trump-inspired-anti-semitism-spikes-prompting-conservative-writer-protect-herself-gun/ [https://perma.cc/X5DY-TJYJ].

[34] Lauren Gambino, *Journalist who profiled Melania Trump hit with barrage of antisemitic abuse*, THE GUARDIAN (Apr. 28, 2016), https://www.theguardian.com/us-news/2016/apr/28/julia-ioffe-journalist-melania-trump-antisemitic-abuse [https://perma.cc/DR5N-5MTC] (name changed for privacy considerations).

[35] RSA Conference, *Your Democracy Needs You: Taking On Digital Dictatorships*, YOUTUBE (Feb. 28, 2020), https://www.youtube.com/watch?v=5jy4saeW5Ho#action=share&ab_channel=RSAConference [https://perma.cc/6S62-7UDB].

[36] Dewey, *supra* note 1; Kathy Sierra, *Why the Trolls Will Always Win*, WIRED, (Oct. 8, 2014, 4:49 PM), https://www.wired.com/2014/10/trolls-will-always-win/ [https://perma.cc/AYH4-56SU].

[37] *See*, *e.g.*, Jacqueline Beauchere, *Digital civility at lowest level in 4 Years, new Microsoft research shows*, MICROSOFT ON THE ISSUES (Feb. 10, 2020), https://blogs.microsoft.com/on-the-issues/2020/02/10/digital-civility-lowest/ [https://perma.cc/D6SC-UZX2].

[38] *See*, *e.g.*, Robert Salonga, *Facebook exec targeted by hoax call*, MERCURY NEWS (Jan. 9. 2019), https://www.mercurynews.com/2019/01/09/facebook-exec-targeted-by-hoax-call-drawing-heavy-police-response/ [https://perma.cc/S4SF-FA9L].

[39] Luke O'Brien, *The Making of an American Nazi*, THE ATLANTIC (Dec. 2017), https://www.theatlantic.com/magazine/archive/2017/12/the-making-of-an-american-

their advantage to weaken enforcement or enable their own ends. For example, it is common to see coded language in calls to harass people, or reminders to harassers to let targets "know your opinion" to make it more difficult for online platforms to take action against it by wrapping themselves in the mantle of freedom of expression and the First Amendment.[40]

However, online harassment has not been static for the past decade—in fact, the techniques behind it have evolved and the actors using it have evolved. As people's offline and online selves grow closer together, advocates have seen more aggressive blends of online and offline harassment. And in addition to angry individuals, politically-motivated and state-backed actors are now leveraging these techniques.

II. MODERN ONLINE HARASSMENT HAS EVOLVED IN THE LAST DECADE, NOW TARGETING ENTIRE COMMUNITIES, RATHER THAN ONLY INDIVIDUALS.

Since 2016, researchers and advocates have learned more about how online harassment functions as a tool, sometimes for ends other than solely individual intimidation. Increasingly, online harassment is being used to target entire communities—everyone who looks like or identifies with a particular target—and to exacerbate social divisions and societal fissures.

Today's harassment is often multivariable and intersectional. A study that the author conducted with U.C. Berkeley's D-Lab[41] and the Center for Technology and Society used machine learning to study xenophobic language on Reddit in the build-up to the 2016 election.[42] Going into the study, the researchers expected to see primarily anti-immigrant slurs and targeting of particular communities. The team's expectation was that harassers would be provoked by specific animus, like the way hate crimes target specific people based on their protected class, like race, religion, or ethnicity. But what came back was far more diverse. Top phrases

nazi/544119/ [https://perma.cc/WT4V-FWLG]. *But cf.* Alexia Fernández Campbell, *The limits of free speech for white supremacists marching at Unite the Right 2*, VOX (Aug. 12, 2018), https://www.vox.com/policy-and-politics/2018/8/10/17670554/unite-the-right-dc-free-speech-first-amendment [https://perma.cc/FZ7W-WZ4T].

[40] *Through Conspiracies and Coded Language, White Supremacists Use Social Media Networks to Aid and Abet Terror, New Study Finds*, ANTI-DEFAMATION LEAGUE, https://www.adl.org/news/press-releases/through-conspiracies-and-coded-language-white-supremacists-use-social-media [https://perma.cc/2KJ5-4624] (last visited Oct. 12, 2020).

[41] *Online Hate Index: Scalable Detection of Online Hate Speech*, D-LAB, https://dlab.berkeley.edu/landing-page/online-hate-index-scalable-detection-online-hate-speech [https://perma.cc/7P24-BT6F] (last visited Oct. 12, 2020).

[42] *The Online Hate Index*, ANTI-DEFAMATION LEAGUE, https://www.adl.org/resources/reports/the-online-hate-index#implications [https://perma.cc/HFZ9-5VXN] (last visited Oct. 12, 2020).

ranged as broadly as "Jew," "Woman," "Black," "White," and "Hate," and the campaigns seemed to be disturbingly equal-opportunity.[43] Harassers targeted many communities, making these efforts look less like targeted hate crimes and more like efforts to systematically exacerbate social fissures in American communities.[44]

Harassment remains constant, but its application and nuances are not fixed; other assumptions about online harassment may not play out as one may presume based on past experiences. Advocates—the author included—sounded alarms about rising hate speech in the lead-up to the 2016 election, but new studies show that online hate speech was not more prevalent during the 2016 election than it is normally.[45] Does this mean that online harassment is normally much worse than we think, but has been historically underreported? How deeply is hate speech now embedded in American political expression because of increased polarization and what does this look like? More study is needed to understand this phenomenon and how online harassment evolves.

III. HARASSMENT IS A VITAL TACTIC USED IN THE SPREAD OF DISINFORMATION AND ELECTORAL INTERFERENCE.

What we do know, thanks to the Senate Intelligence Report about Russia's actions in the 2016 election, is that online harassment has been tied into other vectors, like disinformation and electoral interference.[46] The report described how Russian actors, using bots and accounts impersonating Americans, sought to magnify divisions in the United States, particularly in flashpoint areas like race relations.[47] This targeting of minority groups is most effective when it is used to provoke domestic actors, fusing expressions of aggression and hatred into legitimate political expression by these unwitting proxies. An example of this occurred

[43] *Id.*

[44] *See id.*

[45] Alexandra A. Siegel, Evgenii Nikitin, Pablo Barberá, Joanna Sterling, Bethany Pullen, Richard Bonneau, Jonathan Nagler & Joshua A. Tucker, *Trumping Hate on Twitter? Online Hate Speech in the 2016 US Election Campaign and its Aftermath* (Mar. 6, 2019) at 1, https://alexandra-siegel.com/wp-content/uploads/2019/05/Siegel_et_al_election_hatespeech_qjps.pdf [https://perma.cc/GU4U-3556]; *cf. Anti-Semitic Targeting of Journalists During the 2016 Presidential Campaign*, *supra* note 9.

[46] *See* Philip Ewing, *Report: Russian Election Trolling Becoming Subtler, Tougher To Detect*, NPR, (Mar. 5, 2020, 3:05 PM), https://www.npr.org/2020/03/05/812497423/report-russian-election-trolling-becoming-subtler-tougher-to-detect [https://perma.cc/295Q-KAVK].

[47] *See* Tim Mak, *Senate Report: Russians Used Social Media Mostly to Target Race in 2016*,
NPR, (Oct. 8, 2019, 2:50 PM), https://www.npr.org/2019/10/08/768319934/senate-report-russians-used-used-social-media-mostly-to-target-race-in-2016 [https://perma.cc/V3HW-PVTR].

in 2016 in Houston Texas, where Russian-linked Facebook groups organized two offline protests—one anti-Muslim and one pro-Muslim—to occur at the same time, in front of the same mosque, on the same day.[48] The event appeared to be protest and counter-protest to American observers and resulted in confrontations and exchanges of verbal abuse.[49] The atmosphere was tense and the potential for violence was high.[50] Notably, the instigators, safe in Russia, never had to set foot in the United States.[51]

Harassment is also leveraged by state actors and politically-motivated actors for reasons that go beyond the personal targeting of their immediate victims. This has been intricately tied to politics. For example, in Brazil, the Supreme Court is investigating the so-called "office of hate," a group of online influencers, bots, and digital marketing firms allegedly hired by President Bolsonaro and his allies to disparage critics of the regime and create manufactured waves of pro-government public opinion.[52] Observers have noticed the professional nature and narrow scope of the attacks, like individual harassment of Bolsonaro's political opponents and multiple daily pro-Bolsonaro issue-based campaigns.[53] Facebook did an investigation and took down networks of fake accounts tied to employees in the office of Bolsonaro, his sons, and other conservative allies of the president.[54] These accounts commented on "domestic politics and elections" and posted "criticism of the political opposition, media organizations and journalists."[55]

Sometimes masses of people will organize to harass political opponents and abuse the platform architecture designed to protect people from harassment. For example, in Vietnam, cybermobs reportedly coordinated in 2014 to participate in a tactic known as brigading: coordinated groups targeting critics online; in this case, the brigade hit the "report" button en masse on their target's content, resulting in Facebook kicking political activists or

[48] Claire Allbright, *A Russian Facebook page organized a protest in Texas and the counterprotest.* TEXAS TRIBUNE, (Nov. 1, 2017, 4:00 PM), https://www.texastribune.org/2017/11/01/russian-facebook-page-organized-protest-texas-different-russian-page-l/ [https://perma.cc/L357-46XA].

[49] *Id.*

[50] *See id.*

[51] *See id.*

[52] Andrew Rosati & Mario Sergio Lima, *In Hunt for "Office of Hate," Brazil's Supreme Court Closes In*, BLOOMBERG, (June 22, 2020, 2:00 AM), https://www.bloomberg.com/news/articles/2020-06-22/in-hunt-for-office-of-hate-brazil-s-supreme-court-closes-in [https://perma.cc/PM7U-D3RT].

[53] *Id.*

[54] Chandler Thornton & Rodrigo Pedroso, *Facebook Shuts Down Network of Fake Accounts Tied to Employees of Brazil's Bolsonaro and Sons*, CNN, (July 9, 2020, 2:55 PM), https://www.cnn.com/2020/07/09/americas/brazil-bolsonaro-facebook-fake-accounts-intl/index.html [https://perma.cc/6A48-72E7].

[55] *Id.*

independent journalists off the platform.[56] The mechanisms intended to mitigate abuse resulted in at least forty-four independent Vietnamese journalists being removed from Facebook:

> For journalist Pham Doan Trang, it's meant that the free space of Facebook has become effectively controlled by her political opponents. "They have been escalating since mid-June (2014)," Trang says, "by now, it's hundreds of pages that have been knocked down," both from individuals and larger publications. Generally, the pages were taken offline through Facebook raids like the one that targeted Trang—a large group of people all pressing the Report Abuse button at once. Anything they were trying to say is effectively silenced, whether it's breaking news or reports from a protest.[57]

While attribution is difficult, the pattern is clear: pro-government forces report critics and manipulate platform mechanisms to silence critiques of the Vietnamese government.

Furthermore, tactics for mass media manipulation adapt according to digital authoritarians' evaluation of the political environment and the best way to spread their intended message. The groups comprising useful idiots are not always the usual suspects—and are increasingly volunteer conscripts. For example, in China, it was commonly known that pro-government forces would pay individuals to post positive content or to disparage opponents.[58] Starting in 2016, an organic movement emerged, starting on an online soccer forum called Diba that became the center of nationalistic online attacks:

> [Diba is] known for its highly organised nationalist "battle missions". Its troops are divided into groups and assigned different tasks for its actions, which are always advertised in advance on their social networks. . . . During the attacks, some Diba members translate poems and pro-China slogans into various languages or create memes – the "ammunition" as it is known – while others are administrators, directing

[56] Russel Brandom, *Facebook's Report Abuse Button has become a Tool of Global Oppression*, THE VERGE (Sept. 2, 2014, 10:16 AM), https://www.theverge.com/2014/9/2/6083647/facebook-s-report-abuse-button-has-become-a-tool-of-global-oppression [https://perma.cc/3F4Q-V5KG].

[57] *Id.*

[58] Phoebe Zhang & Laurie Chen, *The Emergence and Evolution of China's Internet Warriors Going to Battle over Hong Kong Protests*, S. CHINA MORNING POST (Sept. 4, 2019, 7:15 AM), https://www.scmp.com/news/china/society/article/3024223/emergence-and-evolution-chinas-internet-warriors [https://perma.cc/A47Q-RSET].

> the "battle" from headquarters, sending out links and new materials for the troops to copy, paste and spam.[59]

On Diba, a forum like Reddit, users publicly brag of their successes targeting Taiwanese pro-independence celebrities and politicians, including Taiwanese President Tsai Ing-wen; hundreds of thousands of Diba users inundated the comments of their targets' Facebook pages, leading to Diba accounts being banned by Twitter, Facebook, and YouTube.[60]

Current Diba efforts target other non-Chinese interests, like pro-Hong Kong democracy activists, and misuse online architecture to stifle debate.[61] Reports state that fans of "A-zhong" (a fictional persona standing in metaphorically for the Chinese state, who is often depicted like a pop star) "are encouraged to refrain from engagement with pro-Hong Kong commenters, who may try to 'brainwash' them; instead, team leaders suggest, it's better to immediately report those Instagram users who appear to be spreading propaganda."[62]

The prevalence of these digital brigades has led to an online expression, *wumao* or "50 cent trolls," a term alluding to the fact that previous online armies were known to be paid by the Chinese government; the term denigrates a poster, assuming they are trolling-for-hire and insulting their expression as cheap.[63] China's online citizen armies are a newer technique that enlist patriotic youth who volunteer to troll and spread Chinese nationalistic memes and pro-government content out of a sense of civic pride.[64] This organic brand of troll is diverse and has included Chinese overseas expatriates, rappers, and teenage fangirls—all of whom "refuse to be labelled as nationalists, saying they simply want to offer counterpoints to 'misinformation' about China, or to protect China from being smeared. Like Diba, they find strength in numbers and work in a similar, coordinated fashion."[65]

The relationship between these trolls and the Chinese government officials is very close, suggesting it may be sanctioning the efforts.[66] Politicians have been supportive of some of these groups, and the Chinese Communist Youth League has endorsed

[59] *Id.*

[60] *Id.*

[61] *See* Lauren Teixeira, *China is Sending Keyboard Warriors Over the Firewall*, FOREIGN POL'Y (Aug. 26, 2019, 9:25 AM), https://foreignpolicy.com/2019/08/26/china-is-sending-keyboard-warriors-over-the-firewall/ [https://perma.cc/66DV-BT7C].

[62] *Id.*

[63] *See* Zhang & Chen, *supra* note 58.

[64] *See id.*

[65] *Id.*

[66] Teixeira, s*upra* note 61.

the movement.[67] These newer efforts contrast with paid astroturfing by government or private parties, and instead weaponize internet subcultures to harass political opponents.[68] China may have learned that the best useful idiots may not be idiots at all—but willing authentic parties eager to harass others.

IV. BECAUSE ONLINE HARASSMENT IS HARD TO DEFINE, AND RULES PROHIBITING IT ARE UNDERENFORCED, IT IS SYSTEMATICALLY EXPLOITED BY MALIGN ACTORS.

Advocates have been complaining about online harassment for almost as long as the internet has existed.[69] While the problem itself is undeniable, it is notoriously hard to get a handle around the scope, scale, and shifting content of online harassment.[70] Because of this volatility, it should be no surprise that digital authoritarians and malign online actors with ulterior motives would seek to exploit it.

Harassment is extremely hard to quantify for any single actor because it often spans multiple platforms.[71] Perpetrators exploit this weakness, often transferring abuse from platform to platform, so no single company can fully grasp, or mitigate, the entirety of the abuse that a target is facing.[72] Restrictions on companies sharing information about their users are designed to protect privacy, and while this is an important goal, one unfortunate consequence is to limit platform coordination and increase the burden on targets themselves to combat multi-platform harassment.[73]

Different platforms have their own unique terms of service, and their users have distinct understandings of what constitutes

[67] *See id.*

[68] *See* Zheng & Chen, *supra* note 58.

[69] *See, e.g.*, *About ADL's Work Combating Cyberhate and Countering Violent Extremists Online*, ANTI-DEFAMATION LEAGUE (Feb. 29, 2016), https://www.adl.org/news/article/about-adls-work-combating-cyberhate-and-countering-violent-extremists-online [https://perma.cc/5R9M-7PPV] (ADL's first report on online abuse was in 1985).

[70] *See generally*, Maeve Duggan, *Online Harassment 2017*, PEW RES. CTR. (July 11, 2017), https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/ [https://perma.cc/7CY7-NNPW].

[71] *See Online Abuse 101*, WOMEN'S MEDIA CTR., https://www.womensmediacenter.com/speech-project/online-abuse-101/#crossPlatformHarassment [https://perma.cc/N2RU-VUHK] (last visited Oct. 12, 2020).

[72] *See* Jordan Kraemer & Danya Glabau, *The Trolls are Teaming up—and Tech Platforms Aren't Doing Enough to Stop Them*, FAST COMPANY (Dec. 10, 2019), https://www.fastcompany.com/90440915/the-trolls-are-teaming-up-and-tech-platforms-arent-doing-enough-to-stop-them [https://perma.cc/6F8L-YLN8].

[73] *See generally id.*

harassing behavior on different platforms.[74] This means that even though online harassers use multiple platforms, the platforms themselves are not created equal. Different types of social media and policy approaches to online harassment lend themselves to campaigns of online harassment in different ways.

For this analysis, it is useful to divide up social media platforms in two dimensions: (1) based on the type of social media they host (images, video, message boards); and (2) based on the approach their content policies take to addressing online harassment (permissive and mostly unmoderated; mixed tolerance of harassing content; or restrictive against harassment). This second distinction is particularly important. Some platform policies clearly condemn online harassment, while others take a far more permissive approach, often rooted in theories of free speech or libertarianism.[75] As will later be discussed, whether a platform has taken an aggressive stance against online harassment—demonstrated through its policies, product features, and enforcement record—should be a significant factor for legislators, regulators, and civil society groups assessing how to treat that platform.

[74] *See* Jessica A. Pater, Moon K. Kim, Elizabeth D. Mynatt & Casey Fiesler, *Characterizations of Online Harassment: Comparing Policies across Social Media*, PROC. OF THE 19TH INT'L CONF. ON SUPPORTING GRP. WORK 373 (Nov. 2016), https://dl.acm.org/doi/pdf/10.1145/2957276.2957297 [https://perma.cc/9W68-FAZ7] (describing that platforms, too, have distinct terms of service or community standards that characterize content as "harassing" in different ways; there is no one standard definition of harassment).

[75] *See id.* at 370.

TABLE: PLATFORM APPROACHES TO ONLINE HARASSMENT

		Type of Social Media Platform				
		Image Sharing	**Video**	**Multiple Formats**	**Discussion Board**	**Messaging**
Approach to online harassment	**Restrictive**	Instagram, Snapchat	TikTok	Facebook, Twitter, LinkedIn	Reddit	
	Mixed					WhatsApp, Telegram
	Permissive			Parler, VKontakte	Gab, 4Chan, 8Chan	

Obtaining inside information to evaluate the prevalence of online harassment is not easy. Scholars have a difficult time getting back-end information from platforms for researching online activity, although this is starting to change.[76] For example, Facebook now issues requests for proposals from academics to partner with them to study online phenomena, like hate speech and disinformation.[77] Twitter has also worked with researchers and issued requests for proposals to study how to increase the health of online conversations.[78] But currently, the most detailed research comes from internal teams at platforms or market research firms. Microsoft has conducted an annual global survey of digital civility since 2016.[79] Pew Research has also done two landmark surveys of abusive behaviors throughout the internet ecosystem.[80]

This combined body of work has shown interesting trends. Notably, different demographics have distinct thresholds for acceptable versus harassing behaviors. While there is more online harassment reported than ever, younger generations have grown up expecting abuse or "flaming" to be part of their online experience, and have lower baseline expectations about user safety, privacy, and security.[81] This can also be seen manifesting in China's Diba community, who see their trolling as an expression of national pride, and not as a demonstration of harmful online behavior.[82]

Perhaps due in part to this realignment of user expectations, combined with the lack of transparency into platform enforcement

[76] *See*, *e.g.*, *Content Policy Research on Social Media Platforms Request for Proposals*, FACEBOOK RES. (Jan. 30, 2019), https://research.fb.com/programs/research-awards/proposals/content-policy-research-on-social-media-platforms-request-for-proposals/ [https://perma.cc/37N6-JFR8]; Vijaya Gadde & Yoel Roth, *Enabling Further Research of Information Operations on Twitter*, TWITTER (Oct. 17, 2018), https://blog.twitter.com/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.html [https://perma.cc/24SD-5UR4]; *see also* Jeffrey Mervis, *Researchers finally get access to data on Facebook's role in political discourse*, SCIENCE (Feb. 13, 2020, 5:10 PM) https://www.sciencemag.org/news/2020/02/researchers-finally-get-access-data-facebook-s-role-political-discourse [https://perma.cc/X9GZ-4J3K].

[77] Pater et al., *supra* note 74.

[78] *Twitter Health Metrics Proposal Submission*, TWITTER (Mar. 1, 2018), https://blog.twitter.com/en_us/topics/company/2018/twitter-health-metrics-proposal-submission.html [https://perma.cc/L5EG-M4LL].

[79] *See* Beauchere, *supra* note 37.

[80] *See* Duggan, *supra* note 70 (Pew's prior study was done in 2014).

[81] *See generally Online Abuse 101*, *supra* note 71; *see generally* Amanda Lenhart, Michelle Ybarra, Kathryn Zickuhr & Myeshia Price-Feeney, *Online Harassment, Digital Abuse, and Cyberstalking in America*, DATA & SOC'Y, (Nov. 21, 2016), https://www.datasociety.net/pubs/oh/Online_Harassment_2016.pdf [https://perma.cc/6ZPJ-TA9P].

[82] *See* Zhang & Chen, *supra* note 58.

and challenges with information sharing, online harassment presents a strategic vulnerability for many platforms. This is substantial, just like a cybersecurity breach, even though it is a failure of company policy and enforcement. It is no surprise, given the consistent complaints of inaction by victims, that harassment has become a tool used by bad actors who want to use the internet to influence elections, undermine democracy, and eliminate critics.[83] Oftentimes, as seen with Russia, China, Vietnam, Brazil, and the Philippines, harassment is integrated into influence campaigns, meaning "any actions taken. . . to distort political sentiment… to achieve a strategic gain and/or geopolitical outcome."[84]

V. BECAUSE ONLINE HARASSMENT IS A FREQUENT PART OF INFLUENCE CAMPAIGNS, WE NEED TO UNDERSTAND HOW IT WORKS TO GRAPPLE WITH EMERGING ONLINE THREATS TO ELECTIONS AND INFORMATION INTEGRITY.

Examples from around the world show that trolling is the template used by digital authoritarians—particularly related to the targeting of minority groups.[85] While journalists and political activists are particularly at-risk, it is also religious, ethnic, and racial minorities in a society that can bear the brunt of harassment from useful idiots.[86]

As previously mentioned, within studies of disinformation and foreign influence operations, we see other types of harmful online activity, like hate speech and online harassment, being used to target minority groups.[87] This transition is how offline phenomenon, like racial tensions, become translated into online environments. In other words, disinformation can leverage social pressures and prejudice that explode into online harassment; oftentimes the harassers are useful idiots, especially when influence operations provide them with an impetus for action. According to the Institute for the Future, eight different minority groups experienced targeted disinformation before the 2016 U.S.

[83] *See generally* Ewing, *supra* note 46.

[84] *See* Reppell & Shein, *supra* note 12, at 10 (an earlier term is also "information operations").

[85] *See id.*

[86] RSA Conference, *supra* note 35.

[87] *Id.*

Presidential election.[88] The Oxford Internet Institute has similarly studied the gender-based facets of disinformation.[89]

Examples of disinformation leading to harassment include situations with potentially dangerous outcomes. As previously mentioned, Russian trolls organized a protest against the "Islamization" of a Houston community along with a counter protest, across the street, supporting Muslim neighbors in 2016.[90] The foreign actors never had to set foot in Texas, yet they created a situation with a real potential for violence.

Those who broker disinformation know that harassment can be a bridging mechanism to offline violence and apply it accordingly. How does this work? According to Ben Nimmo, one of the foremost experts on information operations in the world, the most common tactic used by digital authoritarians is to dismiss the speaker.[91] In a talk given at RSA Conference 2020, the author and Nimmo describe how Morgan Freeman, who had previously espoused anti-Russia sentiments for its interference in American political processes, was targeted in Russian troll tweets for marijuana usage.[92] This has nothing to do with Freeman's credibility about politics, but is akin to an *ad hominem* attack, insulting him to attempt to stop people from listening to his message. The dismissal tactic is designed to prevent audiences from examining evidence or taking critique seriously.[93] If you combine dismissal with the aim of disinformation (creating chaos by amplifying preexisting social fissures in a society), it is easy to understand how online harassment can be deployed against a speaker. Since harassers often attack using social biases against the target, this can easily be deployed in the service of information operations to amplify existing prejudices and undercut pluralistic societies.[94]

Many platforms have begun taking action to prepare for the 2020 U.S. elections, but this varies from company-to-company and

88. The Human Consequences of Componential Propaganda, Eight Case Studies from the 2018 US Midterm Elections, Institute for the Future, (Katie Josef & Samuel Woolley eds., 2019), https://www.iftf.org/fileadmin/user_upload/downloads/ourwork/IFTF_Executive_Summary_comp.prop_W_05.07.19_01.pdf [https://perma.cc/Y7F8-6A36].

[88] AtlanticCouncil, *360/OS London – The Digital Authoritarian's Toolbox*, YouTube (June 21, 2019), https://www.youtube.com/watch?v=3Dj2DE9VzGo [https://perma.cc/HW32-Q3LG].

90. Donie O'Sullivan, *Russian trolls created Facebook events seen by more than 300,000 users*, CNN (Jan. 26, 2018, 5:13 PM), https://money.cnn.com/2018/01/26/media/russia-trolls-facebook-events/index.html [https://perma.cc/9PTD-47LN].

91. RSA Conference, *supra* note 35.

92. *Id.*

93. *Id.*

94. *Id.*

on each platform's perceived level of risk.[95] Furthermore, while disinformation campaigns often begin with elite hackers and carefully seeded dumps of stolen documents, the end game is to engage average citizens to spread the message organically.[96] As previously mentioned, a successful operation will engage brigades of useful idiots, who drive home the message via attacking the speakers. In the United States, this can create a dilemma, as platforms do not want to squelch the legitimate (even if odious, hateful, or aggressive) expression of their users, within their terms of service. This is especially true when companies are under increasing scrutiny from political constituencies, who allege their political viewpoints have been censored.[97] It can create the perfect storm: harassment that looks like typical online hate and vitriol that some audiences have become accustomed to online, but it is more than meets the eye. This content moves the Overton window, primes audiences with disinformation, and—as Russia's actions demonstrate—can potentially impact democratic processes.

VI. POTENTIAL SOLUTIONS FOR ONLINE HARASSMENT THAT PROMOTES DISINFORMATION

What can be done about the new application of online harassment in service of disinformation? First, companies should apply a risk-based allocation of resources.[98] When previously confronted with harassment, some companies did not have the staff, the resources, or the expertise to handle such a problem.[99] However, given the intersection with other platform integrity issues, this Article recommends that companies treat harassment with the same gravity that applies to other security-focused threats. As part of this analysis, companies should expand their

[95] For a comprehensive survey of platform policies, *see* Jared Newman, *Tech Platforms Screwed Up the Last Election: Here's How They're Prepping for 2020*, FAST COMPANY (Mar. 4, 2020), https://www.fastcompany.com/90467733/tech-platforms-screwed-up-the-last-election-heres-how-theyre-prepping-for-2020 [https://perma.cc/6YCY-UY74].

[96] *Id.*; *Exposing Russia's Effort to Sow Discord Online: The Internet Research Agency and Advertisements*, U.S. H.R. PERMANENT SELECT COMM. ON INTELLIGENCE (last visited Oct. 12, 2020), https://intelligence.house.gov/social-media-content/ [https://perma.cc/8WWA-TTGF].

[97] Emily A. Vogels, Andrew Perrin & Monica Anderson, *Most Americans Think Social Media Sites Censor Political Viewpoints*, PEW RESEARCH CTR. (Aug. 19, 2020), https://www.pewresearch.org/internet/2020/08/19/most-americans-think-social-media-sites-censor-political-viewpoints/ [https://perma.cc/55QD-DZXC].

[98] Brittan Heller, *Is this Frog a Hate Symbol or Not?*, N.Y. TIMES (Dec. 24, 2019), https://www.nytimes.com/2019/12/24/opinion/pepe-frog-hate-speech.html [https://perma.cc/FF2T-DGDL].

[99] *See* Jessica Valenti, *If Tech Companies Wanted to End Online Harassment, They Could Do It Tomorrow*, THE GUARDIAN (Dec. 1, 2014), https://www.theguardian.com/commentisfree/2014/dec/01/tech-companies-online-harassment-courts-social-media [https://perma.cc/6GY7-XRJJ].

focus to other regional markets and categories of vulnerable users, like journalists and political candidates, during especially critical times. Specifically, elections and periods of civil unrest are indicators of times when platform manipulation is more likely to occur.[100] For example, companies should look to users or customer bases in Africa and southeast Asia around their elections as the testing grounds for new types of media manipulations. This is where the most sophisticated threat actors are testing out new theories, and where harassment and hate speech have been gaining ground.[101]

Companies can also learn from other industries' best practices and commit to independent human rights impact assessments (HRIAs), which should include examining the prevalence, means, and likelihood of online harassment.[102] HRIAs are a mainstay of corporate social responsibility, and have been conducted by other volatile industries like oil, mining, and gas for years.[103] Internal assessments are part of some companies' commitment to multi-stakeholder processes, like the Global Network Initiative, and are part of the policies, standards, and procedures they undertake to adhere to the U.N. Guiding Principles on Business and Human Rights; however, these efforts do not incorporate online harassment as a serious user safety threat.[104] The benefit of looking at harassment as part of a HRIA would be that it would provide a feedback loop for companies to hear from impacted populations and allow them to anticipate risks before they hit the breaking point. As situations around the world erupt into ethnic or political violence—like the targeting of the Rohingya in Myanmar—online harassment of minorities can serve as a litmus test showing that tensions may be reaching the level of mass violence.[105]

[100] RSA Conference, *supra* note 35 (While major platforms are starting to realize this, the majority of focus has been U.S.-based, when the audience of companies like Facebook is primarily international).

[101] Researchers are just beginning to look into countries in the Global South and Asia. *See*, *e.g.*, Shelby Grossman, Daniel Bush & Renée DiResta, *Evidence of Russia-Linked Influence Operations in Africa*, STAN. INTERNET OBSERVATORY (Oct. 29, 2019), https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/29oct2019_sio_-_russia_linked_influence_operations_in_africa.final_.pdf [https://perma.cc/4LYU-RYC9]; *Taiwan Election: Disinformation as a Partisan Issue*, STAN. INTERNET OBSERVATORY (Jan. 21, 2020), https://cyber.fsi.stanford.edu/io/news/taiwan-disinformation-partisan-issue [https://perma.cc/P8NH-Y6JJ].

[102] *See* Heller, *supra* note 98.

[103] *See*, *e.g.*, PRINCIPLES FOR RESPONSIBLE INVESTMENT, DIGGING DEEPER: HUMAN RIGHTS AND THE EXTRACTIVES SECTOR 9–10 (2018), https://www.unpri.org/download?ac=5081 [https://perma.cc/U62V-XPB6].

[104] *See The GNI Principles*, GLOBAL NETWORK INITIATIVE, https://globalnetworkinitiative.org/gni-principles/ [https://perma.cc/748A-WU7M] (last visited Oct. 12, 2020).

[105] Facebook did commission a HRIA for Myanmar, which detailed the role that social media played in inciting mass violence against religious and ethnic minorities.

Companies should understand that regulatory sands may be shifting, and legislators and the courts may pay increased attention to how platforms confront online harassment. Section 230 of the Communications Decency Act of 1996 (CDA) provides online platforms and websites broad immunity from liability for content posted by their users.[106] This broad protection has come under increasing fire in recent years, with lawmakers, activists, and advocates calling for it to be repealed or revised.[107]

The intermediary liability protections afforded by Section 230 are an important enough component of the internet's legal framework that limiting its protections should be taken carefully and with due consideration for secondary consequences. One of the most promising methods to draw a line like this was proposed by Danielle Citron and Benjamin Wittes, who recommend conditioning the protections of Section 230 to apply only to sites that "take[] reasonable steps to prevent or address unlawful uses of [their] services."[108]

This Article has argued that online harassment is both harmful in its own right, and is increasingly a tool of digital authoritarians, who use it to silence critics and fuel aggressive disinformation campaigns. It has noted further that social media platforms have taken different approaches to combatting online harassment. While the scope and scale of social media means that gaps (some significant) will always exist, platforms can be categorized based on how seriously they have taken their responsibility to confront online harassment among their users. Given this interplay it would be worth legislators or the courts considering whether they should apply Citron and Wittes' "Good Samaritan" proposal to combat online harassment. Perhaps platforms that knowingly encourage, or "do not take reasonable steps to prevent" systemic online harassment should incur liability for the individual and societal consequences of that endemic harassment.

However, this was after the violence had already swept through the country and the company had been criticized for its inaction for several years. While Facebook has begun to do HRIAs, this is not yet widespread for analyzing the risk of emergent situations of violence instigated or exacerbated by social media platforms. *See* BSR, HUMAN RIGHTS IMPACT ASSESSMENT: FACEBOOK IN MYANMAR 24 (2018), https://about.fb.com/wp-content/uploads/2018/11/bsr-facebook-myanmar-hria_final.pdf [https://perma.cc/VL7N-2AT5].

[106] *See* 47 U.S.C. § 230 (2020).

[107] Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans Section 230 Immunity*, 86 FORDHAM L. REV. 404 (2017); Lauren Feiner, *Republican Bill Seeks to Limit Social Media Liability Protections Without Getting Rid of Them*, CNBC (Sept. 10, 2020, 10:07 AM), https://www.cnbc.com/2020/09/10/republican-bill-seeks-to-limit-section-230-protection-for-tech-platforms.html [https://perma.cc/Z7WF-W97W].

[108] Citron & Wittes, *supra* note 107, at 419.

Assessing the boundaries of this definition will be particularly delicate. An overbroad imposition of liability could burden innovation and even become a tool of the very same governments that leverage online harassment today. However, as Citron and Wittes argue, a carefully construed limitation on the broad immunity currently provided to social media platforms can avoid given "platforms a free pass to ignore destructive activities" while still defending free speech online.[109] Given how online harassment is being wielded by repressive governments, this is an essential balance to strike not only for individual rights, but for society.

Finally, companies should move to systemic approaches to combat harassment.[110] Patchwork efforts tend to focus only on harassing content, after it has already been posted, and thus are already too late. Best practices, derived from studies of disinformation, would include looking for harassing behaviors, instead of just content or actors.[111] This is a better indicator due to the rapidly shifting context behind much harassment, and the prevalence and volatility of the useful idiot problem.

Companies can also build awareness of harassment into their product development pipeline, along with mechanisms that minimize the burden on targets.[112] If this were to occur, it is likely we would see companies as more than triage teams, and a decrease in the impact of harassment on individuals, communities, and nations—and a lessening of the impact of any useful idiots or related disinformation.

CONCLUSION

With the onset of disinformation as a tool of statecraft, it behooves companies, regulators, and citizens to view online harassment in a new light. It is no longer just an issue of personal online safety. The examples in this paper have shown how governments across the globe and their affiliated actors have used online harassment to terrorize journalists, activists, and political opponents; to widen social fissures in societies around the world based on race, religion, and ethnicity; and to disrupt democratic

[109] *Id.* at 413.

[110] *See* Heller, *supra* note 98 (arguing that platforms must embrace multiple methods to effectively fight online harassment, extremism, and hate).

[111] *See* Camille Francois, *Actors, Behaviors, Content: a Disinformation ABC*, TRANSATLANTIC WORKING GRP. (Sep. 20, 2019), https://science.house.gov/imo/media/doc/Francois%20Addendum%20to%20Testimony%20-%20ABC_Framework_2019_Sept_2019.pdf [https://perma.cc/3GWB-EJ7P] (describing the ABC's of viral deception—actors, behaviors, and content).

[112] *See* Stine Eckert, *Fighting Online Abuse Shouldn't Be Up To The Victims*, THE CONVERSATION (Nov. 26, 2017, 6:45 PM), https://theconversation.com/fighting-online-abuse-shouldnt-be-up-to-the-victims-87426 [https://perma.cc/ABC4-C662] (advocating that software companies should step up efforts to combat online harassment).

processes. Because of this, platforms should view online harassment as a strategic vulnerability—and take immediate steps to close the policy gap, lest their products and services become playgrounds for the useful idiot problem.