Table of Contents

Executive Summary

This report strives to build awareness of the direct and indirect impacts of gendered and sexualized disinformation on women in public life, as well as its corresponding impacts on national security and democratic participation. In an

» This presents a democratic and national security challenge; as adversaries attempt to exploit widespread misogyny, women may be less likely to choose to participate in public life.

- ***A be fcna e ge c ca e dee ec g a d e d g ege de ed a d e a ed ab e a dd f a :***

  » Malign creativity is perhaps the greatest challenge to detecting, challenging, and denormalizing online abuse because it is less likely to trigger automated detection and often requires moderate-to-deep situational knowledge to understand.

  » "Platform policies lack a coherent definition of 'targeted harassment,' meaning" much of the abuse women face is not violative of platforms' terms of service, leaving abusers to continue their activities without facing consequences. There is also a lack of intersectional expertise in content moderation, which results in abuse toward women, people of color (POC), and other marginalized communities going unaddressed.

  » Targets bear the onus of detection and reporting. Managing an onslaught of abuse on social media requires time to block, report, and mute abusers. These burdens are discounted and affect their daily lives offline.

- ***S ca ed a af , a a e , a d e e na e a ge g ed n n ea de c ac a d a a ec .*** In order to mitigate the threat, the researchers recommend:

  » **Social media platforms** should introduce incident reports that allow women to report multiple abusive posts at once to provide more context and a more holistic view of the abuse they are experiencing.

  » They should also regularly update platform classifiers or keywords to reflect and root out malign creativity, improve automated detection methods, and introduce nudges or friction to discourage users from posting abusive content.

  » Finally, they should create a cross-platform consortium to track and respond to online misogyny, similar to existing consortiums which counter terrorism and extremism.

  » **Lawmakers** should include content moderation transparency reporting requirements in social media regulation bills to improve understanding of the problem and introduce accountability for women's online protection.

  » They should create clear standards that prohibit the use of gendered and sexualized insults and disinformation in official business.

  » Critically, US lawmakers should reauthorize the Violence Against Women Act (VAWA) and include provisions against online gender-based harassment.

»419 0 T57ew 419Mo3.8.45< 19Mo3elo.349men2.tac /T1»

and textual memes, and other tactics to avoid detection—which compounds the problem and makes responding more difficult. In addition to the data gathered from the social media platforms, the analysis also draws upon in-depth interviews with targets of state-sponsored gendered and sexualized disinformation, as well as two focus groups of female scholars and analysts in information operations, internet governance, human rights, and communications, many of whom have experienced online misogyny themselves.

Over half of the research subjects were targeted with gendered or sexualized disinformation narratives, with women of color subjected to compounded, intersectional narratives also targeting their race or ethnicity.

More broadly, this report strives to build awareness of the direct and indirect impacts of gendered and sexualized disinformation upon women engaged in or aspiring to engage in public affairs, as well as its corresponding impacts on national security and democratic participation. Though this research necessarily has a small sample size, it demonstrates the high level of vitriol to which women in public life are subject, with over 336,000 individual pieces of gendered and sexualized abuse posted by over 190,000 users directed at 13 research subjects during the two-month data collection period. Over half of the research subjects were targeted with gendered or sexualized disinformation narratives, with women of color subjected to compounded, intersectional narratives also targeting their race or ethnicity. While the report demonstrates that this problem is diffuse and, thanks to the use of malign creativity, difficult to detect, it is a problem that demands action. The report outlines the most urgently needed solutions within social media platforms, government, and at the employer or organizational level.

This report is by no means easy reading, but it offers a glimpse into the realities of being a woman online, and a step toward making our public spaces more equitable, democratic, and secure.

Existing Scholarship on Gendered and Sexualized Disinformation

There is a burgeoning field of scholarship focused on harassment and abuse in online spaces. However, academic and policy discussions around gendered disinformation as a distinct type of disinformation remain fairly new. In order to learn from previous work and to effectively position our study, the research team selected and reviewed literature from the disinformation, online abuse, and human rights scholarship. This included scholarship on online abuse and harassment directed at women generally, abuse directed at women active in public life, state-backed influence operations with gendered tactics, and disinformation that employs gendered or sexual language. Engaging with the existing body of work helped the team to better understand the unique online threats facing women, in particular, female political leaders, journalists, and activists.

Women in public service use social media as a tool to gain exposure, to connect with constituents, and to advance their messages on their own terms outside the medium of the news media.[7] Activists and journalists use social media as a part of their jobs: to report on evolving stories, to connect with sources, and to publicize their work.[8] While maintaining a social media presence is now necessary for success in these and other careers, an online presence can be a "double-edged sword" that can open the door to online harassment.[9]

Women are uniquely affected by gendered abuse online. In , a sociological survey of women in public life, Sarah Sobieraj characterizes online digital abuse against women as expressions of "digital misogyny" and "patterned resistance" against women's full and equal participation in public life. She argues that the attacks are "aimed at protecting and reinforcing a gender system in which women exist primarily as bodies for male evaluation and pleasure."[10] Drawing upon qualitative interviews with victims of gender-based attacks online and professionals in content moderation and internet safety, Sobieraj observes that

rather than a specific kind of disinformation.[20] Others have adopted the term "gendered disinformation," expanding on the broad description of the term set by Nina Jankowicz, the lead author of this study, in her 2017 reporting: "[a mix of] old ingrained sexist attitudes with the anonymity and reach of social media in an effort to destroy women's reputations and push them out of public life."[21] However, follow-on studies employ a distinct definition. Di Meco describes

Methodology

This exploratory research takes a sequential mixed methods approach, drawing upon quantitative and qualitative data sources. Key among these were: cross-platform social media data gathered from six social media platforms; existing

formation campaigns. Over the course of the project, a small number of candidates were excluded or added based on current events or whether sufficient data was available for analysis.

The team collected data on six US House of Representatives candidates and two Senate candidates—representing three Republicans and five Democrats—as well as Senator Kamala Harris during her successful Vice-Presidential bid. The team also included Michigan Governor Gretchen Whitmer in data collection, given the sustained threats against her during the coronavirus pandemic.[43] Three other international politicians across the political spectrum in English-speaking countries were also included in order to provide a comparative assessment of the gendered disinformation environment.

Subjects were selected to represent diversity in political affiliation, race, ethnicity, and levels of visibility. In order to explore the intersection between gender and race, the research team included women from diverse backgrounds; however, the team recognizes that future research will require a wider look at diversity and integration of an intersectional approach.

The following individuals comprised the final list of subjects for the project:

International Politicians (3)

- Prime Minister Jacinda Ardern (New Zealand)
- Secretary of State for the Home Department Priti Patel (UK)
- Deputy Prime Minister Chrystia Freeland (Canada)

US Politicians (10)

- Senator Susan Collins (R)
- Senator Kirsten Gillibrand (D)
- Senator and Vice-President-Elect Kamala Harris (D)
- Rep. Jaime Herrera Beutler (R)
- Rep. Alexandria Ocasio-Cortez (D)
- Rep. Ilhan Omar (D)
- Rep. Elissa Slotkin (D)
- Rep. Elise Stefanik (R)
- Rep. Lauren Underwood (D)
- Governor Gretchen Whitmer (D)

Quantitative Data Collection

The research team collected data from six social media platforms, selected based on size of user base and ideological variation in users: Twitter, Reddit, Gab, 4chan, 8kun, and Parler. Facebook and Instagram were also considered for collection, but were not included as both platforms employed data collection limitations, which in turn limited our ability to collect data in a structured, sustainable manner.

For all platforms, a list of keywords reflecting abuse or disinformation was built to inform data collection. These lists

(ii) The secondary group consists of the Tweets in which the Tweet body does not include the collection criteria, and is dependent on the Quoted Tweets for proper matching. These are secondary users not necessarily directly involved in the conversation, though they may share similar sentiments. While these Tweets have been excluded from the current analysis, the behavior trend is indicative of the wider engagement in these conversations.

Network Visualization Methodology

The visualization of cross-platform networks enables better understanding and representation of the content sharing behaviors. Data in the network visualizations corresponds to data points, posts containing abusive or disinformation narratives according to keyword lists (see Appendix C), as described above. Zooming into the most abusive content or narratives within the networks allowed exploration of user behavior, including whether and how users disseminated their own messages individually and alongside the messages of others. Network visualizations were created to capture two different types of interactions: user-to-abusive keyword relationships (captured on all social media platforms, such as when a number of users publish the word "tranny" on both Twitter and 4cg5cgversation.5 a-to (hg5cgto-ab9 (e

to conduct, transcribe, and analyze the interviews. The interviews were conducted on-the-record, lasted between 45-60 minutes, and were recorded and fully transcribed with interviewee consent. In addition to describing their experiences, each interviewee was asked a series of questions (see Appendix D) about characteristics of gendered disinformation they have experienced and observed. These responses informed the research team's definition of gendered and sexualized disinformation.

Research team members Alexandra Pavliuc and Nina Jankowicz also conducted two focus groups with a total of eight female scholars and analysts who study disinformation, including two women of color and one representative of a marginalized group. The researchers issued a broad invitation to 30 women in media, academia, and policy analysis who have focused their work and engagement on disinformation or issues adjacent to it; the eight respondents participated in one of two 60-minute focus groups. They discussed their personal experiences with online misogyny, attempted to define their experiences, and workshopped solutions to the problem (see Focus Group Guide, Appendix E). Contributions to the focus groups were not for attribution. The focus groups were recorded and fully transcribed, with participant consent.

Data Analysis

Overview

The data collected can be broadly separated into categories of gendered abuse, uncoordinated disinformation, and coordinated disinformation. Gendered abuse involves the often casual use of derogatory terms aimed at degrading or insulting women based on gender. The gendered abuse recorded throughout this project ranged from name-calling to sexually violent threats. One widespread example of such abuse is the frequent reference to Alexandria Ocasio-Cortez's former job as a bartender, which abusers used in attempts to undermine her political qualifications and express misogynistic views. For example, in response to Ocasio-Cortez's attempts to block Trump from picking a new Supreme Court Justice, one user wrote: "Suddenly the slut bartender is now a constitutional scholar." In the data collected, gendered abuse was more widespread across all of the subjects than disinformation. While this type of language is undeniably problematic and deeply harmful, this analysis will focus primarily on coordinated or uncoordinated disinformation directed at the research subjects.

In the context of gender, disinformation involves the spreading of rumors or alleged "facts," often of a sexual nature, in order to humiliate, discredit, or disempower the subjects. These campaigns could be either coordinated or uncoordinated. A coordinated disinformation campaign is one which is intentionally conducted by a person or group of people, who may or may not believe in the narrative. Though not of a sexual nature, one example of coordinated disinformation was a campaign orchestrated in September 2020 by Project Veritas, which claimed to have uncovered evidence that Congresswoman Ilhan Omar was orchestrating a widespread ballot harvesting scheme. Whether or not Project Veritas or their collaborators truly believed the story, their coordinated efforts to spread it had a significant impact on Omar's public reputation and made her a target of increased abuse online. The day the story was released, abuse against Omar rose 1,871 percent.[44]

By comparison, uncoordinated disinformation is unplanned, but can be equally detrimental. Part of its impact comes from the difficulty of identifying the originating source: whereas coordinated disinformation may have a clear origin

was openly calling to have Trump killed. While the source of this narrative is unknown and does not appear to have been coordinated, it was nonetheless extremely powerful: mentions of "8645" reached over 6,000 instances on the day of Whitmer's interview and were echoed by the Trump campaign.[46] Importantly, while it is theoretically possible to define the difference between a coordinated and uncoordinated disinformation campaign, it is not always possible to distinguish between the two in practice. Arguably, the most successful coordinated disinformation campaigns are those which appear organic.

The following analysis examines the primary disinformation narratives identified by the research team, key findings from the platforms examined, and unanticipated results.

Top-Level Quantitative Findings

- Between September 1 and November 9 2020, over 336,000 data points of gendered abuse and disinformation were posted on the platforms monitored by the research team. Of the 13 research subjects, the overwhelming majority of recorded keywords relating to abuse and disinformation were directed toward Kamala Harris, accounting for 78% of the total number of recorded instances.
- The research team identified three overarching types of disinformation narratives that impacted multiple subjects: **sexual, transphobic, and racist**

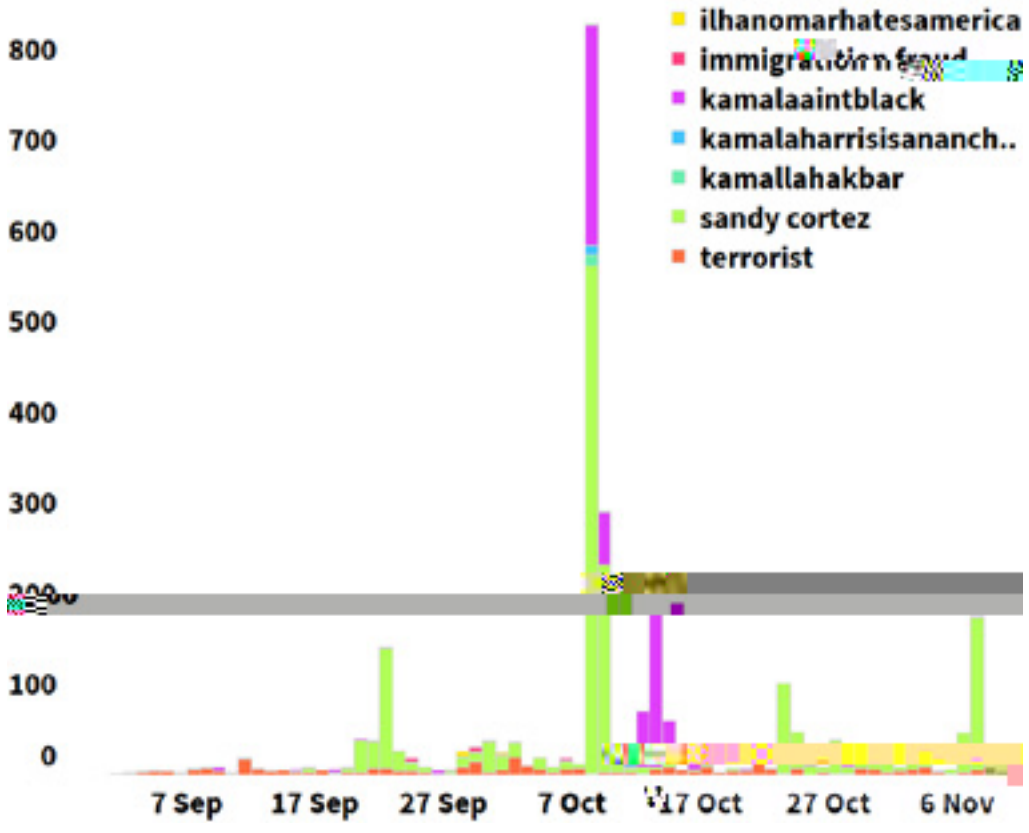
female Prime Minister who gave birth while in office, and Whitmer as a female Governor who imposed significant restrictions during the COVID-19 crisis. Considering these trends, the "secretly transgender" narrative may serve two possible goals. The first is to strip these women of their power and attractiveness, since the transphobia inherent in this narrative dictates that transgender people can be neither attractive nor powerful. The second is to justify their political success, as the misogyny inherent in the narrative dictates that women, particularly young and attractive women, cannot rise to power without deception or male characteristics. It is also possible that this narrative is a byproduct of the sheer volume of abuse received by high-profile political women.

Gendered abuse online also manifested in racist or racialized disinformation. Of the 13 subjects examined, five are women of color who were subjected to racist abuse. Three of these women were targeted with racist and racialized disinformation.

- This was most clearly observed in reference to **Ilhan Omar**, whose Black and Muslim identities were weaponized to portray her as a dangerous foreign "other." Abuse targeting Omar's ethnicity and religion has manifested in a multitude of ways since she entered the political sphere. In 2019, a photoshopped image circulated purporting to show her without her hijab, revealing a balding head with unkempt hair. The photo aimed to humiliate Omar based on her appearance, religion, and ethnicity. Narratives identified in the data collected, however, appear to have shifted from attempts at humiliation towards a portrayal of Omar as a terrorist and political saboteur. These are grounded in Omar's identity as a refugee and her connections with other Somali immigrants and attempt to

latter narrative plays on the taboo of incest to portray Omar as a foreigner who does not adhere to American cultural norms and by extension, cannot represent the American people. As depicted in the network visualization below, a small but vocal group of users on Parler posted a number of racist hashtags such as #DeportIlhanOmar (with one user responsible for 91 uses of the hashtag). A wider array of users posted keywords relating to the false narrative about Omar marrying her brother on Twitter, Reddit, 4chan,and 8kun (with one user engaging extensively with this narrative on 8kun).

- A similar narrative targeted **Kamala Harris**, whose citizenship status was called into question throughout her political campaigns. Users focused on the fact that Harris' parents are both immigrants, incorrectly arguing that Harris is not a natural-born citizen and is therefore ineligible to run for office. Additionally, the minority identities of both Harris and Ocasio-Cortez were used as sources of criticism and delegitimization in several instances. One narrative insisted that as a Black and South Asian woman, Harris could speak for neither group, while others online accused her of exaggerating her racial identities in order to further her political goals. This resulted in the hashtag #KamalaAintBlack, which was recorded 657 times throughout the collection period. **Alexandria Ocasio-Cortez** was targeted with similar criticism when it was revealed that she used to go by the nickname "Sandy." The hashtag "Sandy Cortez" subsequently spread with the aim of delegitimizing Ocasio-Cortez's identity as a working-class woman of color and reframing her as a privileged politician.



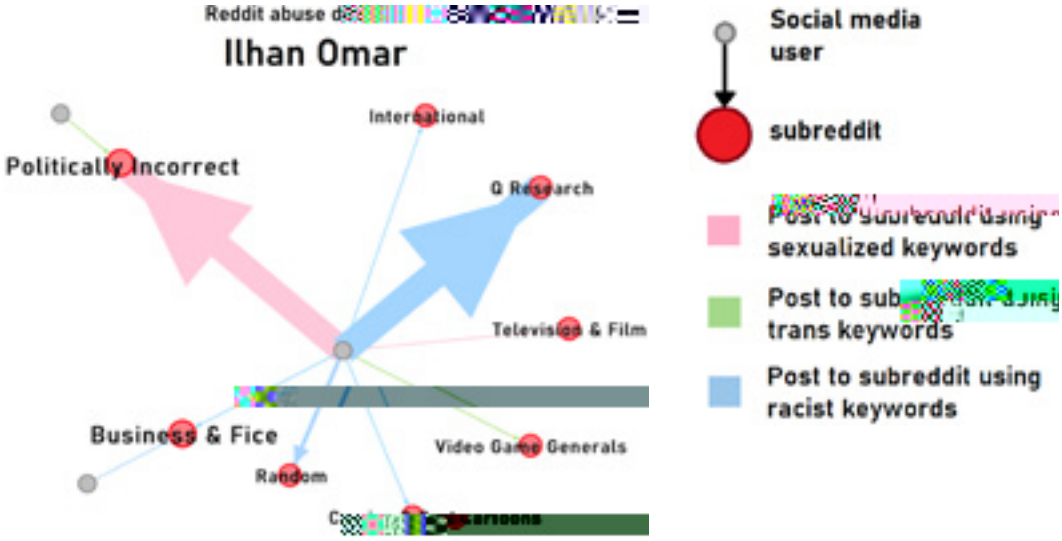
As comparative data was not collected on male politicians of color, it is difficult to know whether these racialized narratives occur with more frequency when directed at women leaders. However, racialized disinformation provides an additional avenue by which female politicians of color receive abuse online and is often compounded with other modes of harassment.

and 8kun, while Parler's top three keywords were #PhonyKamala, #RecallWhitmer, and #RecallGovWhitmer. Addi-

Exploring Behavior and Coordination

In order to evaluate user behavior and assess potential coordination, the research team visualized user relationships (i) between the abusive keywords users employed across social media platforms (captured on all monitored social media platforms), and (ii) between subreddits and boards users posted to (captured on Reddit, 4Chan, 8kun, and Gab).This overview of user interactions with keywords across platforms illustrates how the themes of sexual, racist, and transphobic disinformation narratives and broader abuse overlapped across platforms. Overall, **we observed patterns of intersectionality between generally abusive and sexually abusive keywords or narratives with which users engaged.We also observed instances of individual users who exhibited repetitive abusive posting patterns**.

While the three themes of gendered disinformation recorded during this data collection period - transphobic, racist, and sexual narratives - were presented separately in this report, it is important to emphasize that they do not operate in silos. Keyword networks of user interaction with #KamalaAintBlack and "Sandy Cortez" showed that some users whodT1krra -1.rks



Unique Findings

Certain keywords recorded were found to have both positive and negative connotations, depending on the intent and context of the post. Two notable examples of this were "Big Gretch," a nickname developed for Gretchen Whitmer, and the use of the word "bitch." Whitmer's critics initially employed "Big Gretch" in relation to Whitmer's COVID-19 policies. Subsequently, however, the nickname was also adopted as a term of endearment, becoming the subject of a rap song publicly embraced by Whitmer herself.[49] A similarly ambiguous term used in relation to other research subjects is "bitch." Initially considered an explicitly abusive term, the research team found that it is frequently used in admiration of the subjects. For example, one user writes "AOC is a that bitch [sic]. I love this woman!." These findings indicate that a certain percentage of false positives exist among terms that would otherwise be considered clearly abusive. The duality of meaning, as well as the ambiguity of intent, can also lead to challenges in automatic detection of abusive and false narratives, a critical consideration for platform content moderation responses.

Among the research subjects, Priti Patel is the only high-profile politician against whom coordinated disinformation campaigns were not recorded. This is particularly interesting given Patel's long history of controversial policies. Online abusers appeared to target her directly with misogynistic abuse; "bitch," "fat" and "witch" were the most frequent abusive keywords directed towards Patel. It is unclear what accounts for this apparent lack of targeted disinformation during the data collection period. As the project's only British research subject, it is possible that the difference is cultural; however, this conclusion cannot be drawn without further research to enable a comparison to other British political

figures. It is also possible that the lack of a major electoral event in the UK during the project's collection period can account for this. This finding raises questions regarding the frequency and intensity of disinformation cross-culturally.[50]

While examining the data collected, the research team noted a unique passage that appeared 31 times in reference to Kamala Harris. The passage is an example of a "copypasta," internet slang for a body of text that is copied and pasted across different websites. It begins with the phrase: "I see Kamala Harris as a challenge, more than anything. Here is a woman who, in every single aspect, is absolutely revolting - her exterior AND her personality - yet I can't help but wonder what would be like, to plunge balls-deep into her repeatedly." The passage continues by describing different sexual acts in explicit detail, referring to Harris as a "thing" that the user allegedly despises but wishes to conquer sexually. Upon investigation, the research team found that this exact passage has been circulating online across different websites since at least 2012, and has been used for several different women and girls including Amy Schumer, Chloe Moretz, Emma Gonzalez, Greta Thunberg, and Maisie Williams.[51] Given that the passage is too long to be posted on Twitter where the user could tag the subject directly, it appears to be a way for male or male-identifying users online to assert their masculinity over influential women, using the fantasy of violent sex as a proxy for domination. Kamala Harris was the only research subject targeted by this passage during the data collection period.

The data collected precludes any definitive conclusions regarding differences in abusive language directed at Democratic and Republican subjects. Subjects including Kamala Harris or Alexandria Ocasio-Cortez do not have Republican equivalents that are of comparable political stature, preventing a reliable comparison of keywords and sentiments. Additionally, the time frame during which data was collected saw primarily Democratic or Liberal women centered in historic events, leading to their increased prominence online. While some data suggests that the abuse directed towards

State-sponsored Gendered and Sexualized Disinformation

Gender and sexuality has long been used to argue that women are ill-suited to positions of power. The domestic deployment and amplification of these narratives is undoubtedly a threat to democracy, but when this societal vulnerability is weaponized by adversaries, it becomes a threat to national security. This section details three overt state-sponsored harassment campaigns aimed at female journalists targeted by Iran, China, and Russia. In all three case studies, state-sponsored entities deployed false or misleading narratives that played on gendered or sexualized

Hong Fincher believes it was Zha's Tweet that was the trigger for the harassment campaign against her. "It was something very personalized, but it wasn't over the line harassment," she noted, referencing platform policies about "targeted harassment" that rarely lead to enforcement. Hong Fincher's experience is a classic example of dogpiling, which many of this project's focus group participants underlined as a loose type of coordination between online users, with devastating effects. Hong Fincher blocked and reported hundreds of accounts in the aftermath of Zha's Tweet, including the accounts impersonating her. Cockerell, the journalist from Coda Story, also uncovered hundreds of seemingly-automated pro-China accounts that amplify Zha and other pro-China voices like him.[57] Some of these accounts, as well as those impersonating Hong Fincher, were removed, as evidenced by notes about deleted accounts or deleted Tweets among the replies to Zha's Tweet and Hong Fincher's original thread:

The harassment targeting Hong Fincher did not stop at inauthentic amplification, however. Others associated with state-run Chinese media expanded on the false narrative that she was "against mixed-race marriage" despite being the product of and party to a mixed race marriage. For instance, Tom Fowdy, a blogger who is a freelance author

with efforts to ban Carl Zha from Twitter over "allegations of [her] targeted harassment." His blog did not address the substance of the "allegations," but suggested criticism of Zha was a "part of the broader culture amongst China analysts and opponents of the country to discredit everyone who contravenes their views." "Shortly after that," she recalled, "there was a Xinhua correspondent who was based in Belgium," who suddenly began to criticize Hong Fincher's PhD research, conducted at Tsinghua University in Beijing. "I have never been [criticized] by Chinese state media officially," Hong Fincher noted, nor had her work in China been the target of Chinese state derision. At the same time, she noticed that "all these Chinese ambassadors around the world who are on Twitter started dismissing the report about mass sterilization of Muslim women."

Finally, trolls took to other platforms, including Hong Fincher's Amazon page, where they left inauthentic reviews of her work and promoted the false narrative about her views on mixed-race marriages. Hong Fincher says "there were people calling me a multitude of sexualized insults, misogynistic insults...there have been people threatening to gang rape me and rape me and referring to my children." She wondered on Twitter, "Is it any wonder that most women prefer not to call out harassers publicly?."[58] She does credit Twitter with some response; after an email exchange with a Twitter employee and a public awareness campaign led by the Coalition for Women in Journalism, they began taking action on some of the content in the campaign and verified Hong Fincher's account. But women without the profile, resources, or volition to escalate evidence of abuse may not have been able to achieve this result.

Despite Twitter's action, when asked to describe the disinformation campaign against her in one word, Hong Fincher called it a "tsunami." But Hong Fincher weathered that storm. "It was clearly an attempt to intimidate me and shut me up and exhaust me," she said. "I didn't want to engage, obviously. But I just thought 'I can't let them get the upper hand.' I pinned my thread for quite a few weeks just out of defiance and I didn't want to let all these trolls know that their intimidation was working. But it was utterly exhausting and extremely unpleasant."

This case study demonstrates all three characteristics of gendered and sexualized disinformation: malicious state

marriages, used sexualized threats and insults, created fake accounts impersonating her, and unleashed state media the Chinese state and denigrate Hong Fincher.

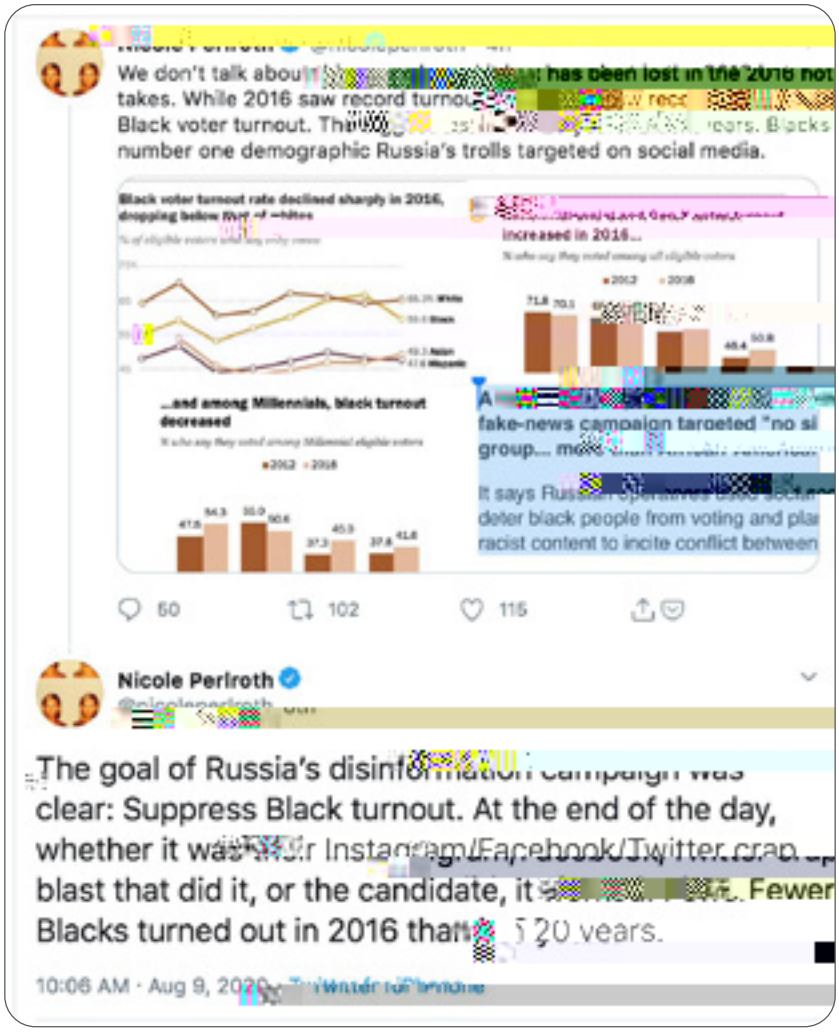
In a 2019 study, Dr. Samantha Bradshaw identifies the ways foreign influence operations rely on traditional gender stereotypes in their messaging strategies through an analysis of over 300,000 English Tweets about gender and politics culled from Twitter's Election Integrity Initiative dataset.

to only 13 percent of all discussions by observed foreign state-operated accounts—including a large proportion from Russia's Internet Research Agency—Bradshaw concluded that "gender was a cross-cutting theme that intersected"

with discussions pushed by foreign state-operated accounts on other topics like race and religion, employing gendered stereotypes "to engender fear, spread skepticism, and foment distrust."[60]

Finnish journalist Jessikka Aro, who identified and reported on pro-Kremlin trolls long before they were a household

in 2016, and what affect, if any, Russian influence campaigns had on the African American vote—when Russian trolls pretended to be Black Lives Matter activists and spread anti-Clinton memes and narratives like 'Killary,'" she said, citing findings by the Senate Intelligence Committee, which concluded the majority of Russian influence operations in 2016



has finally gotten to the bottom of why black Americans didn't get excited about Hillary Clinton's 2016 candidacy for president: Russian trolls."[64] On Twitter, the broadcaster wrote: "#NYT reporter Nicole Perlroth has finally worked out why #Trump won the #2016Election, and it's an original theory. Actually it's not..."[65]

Additionally, RT described Perlroth's Tweet as "whitesplaining," playing on the racial themes that run through both modern and historical Russian operations.[66] During a summer of racial unrest in the United States, throughout which RT fanned the flames of American polarization and endemic racism, the Russian broadcaster alleged that Perlroth's legitimate questions about the impact of 2016 Russian information operations on their target audiences was racist. RT chose to run unattractive pictures of Hillary Clinton in what appear to be aggressive poses with the article and its corresponding

social media content, playing into narratives that Clinton was "nasty" or "shrill," and that older women do not have a place in society. Eventually, the Perlroth narrative was laundered into the American fringe media; in content repackaged from the original RT article, an American website used an unflattering image of Perlroth alongside text that claimed she was "whining" about Hillary Clinton's 2016 election loss.

Even for Perlroth, who is no stranger to online harassment, "the volume of the Twitter blowback and online chatter around my Tweet was way higher than anything I had tweeted in the last two years. So I imagine it was because of RT." In response, Perlroth deleted the tweet—a strategy used by many women experiencing online abuse to lessen the amount of harassment directed at them—which caused her abusers to insinuate that she tacitly admitted that she was wrong. Perlroth's experience demonstrates how women experiencing online abuse have no good choices; by speaking their mind and standing their ground, they may subject themselves to further abuse. The decision to delete the content engendering abuse or lock down accounts can also lead to harassment.

RT's targeting of Perlroth does not include explicitly gendered or sexualized tropes; the elements of this campaign would not be unearthed by an automated tool searching for gendered terms among the state-sponsored broadcaster's articles and social media properties. But using the Russian Federation's online media ecosystem, it advanced a far more insidious narrative—that women are stupid, "whiny," or, in the case of Hillary Clinton, unnecessarily power-hungry—in order to drive online engagement and further societal division that was later amplified and replicated within homegrown American publications and audiences.

Challenges of Detecting and Enforcing Against Gendered and Sexualized Abuse and Disinformation

There are a number of challenges in detecting and responding to online gendered and sexualized abuse and disinformation. Based on the data collected for this study and the responses of interview and focus group participants, the research team has identified the following difficulties in order to aid future responses.

Malign Creativity in Online Misogyny

While this research identified a list of keywords that might signal or be used in conjunction with gendered abuse or disinformation, we found that "malign creativity"—the use of coded language, iterative, context-based visual and textual memes, and other tactics to avoid detection—is perhaps the greatest challenge to detecting, challenging, and

Memes targeting Kamala Harris as part of the sexualized disinformation campaign alleging she "slept her way to the top" were edited, cropped, animated, or otherwise altered in order to evade detection by platforms' automated and human content moderation efforts, and significantly slowed the process of taking action against the content.

Inadequate De□nitions of "Targeted Harassment"

As they stand, platforms' policies on "targeted harassment" often do not adequately address the online abuse and

testimonials explain the drain that self-reporting–as well as responding to abuse more broadly–can have on targets.

In the data collected for this study, the highest volumes of abuse directed at the research subjects were recorded on Twitter. It is heartening to note that Twitter can sometimes successfully intercept this abuse, as occurred after Leta Hong Fincher reached out to the platform. In many cases, the spread of disinformation and abuse occurs indirectly via retweets, allowing the narratives to gather momentum "under the radar," without the need for a great number of direct abusers. With less visibility and a smaller pool of potential abusers, subjects such as Fincher were able to harness the power of blocking/reporting features to cut off their abusers' access. Focus group participants also echoed that blocking and reporting features have been helpful to them in protecting their online presence, offline safety, and their psychological health. One participant noted that "the limiting comments function on Twitter," which allows Tweet authors to decide which users can reply to their Tweets, "while not perfect, feels like a step in the right direction." Despite these successes, the burden of reporting this abuse and advocating for its removal still falls on the subjects.

Abuse Occurs Outside of Highly Visible Areas

Many women who are the subject of online abuse or disinformation campaigns find that abusive content can spread in ways that are not easily monitored outside the scope of dedicated research, such as the current study. For example, on Twitter, rather than abuse being sent solely in reply to a target's Tweet, or as a Quote Tweet or screenshot of the Tweet, abuse can be sent in reply to other content that may or may not tag the target. On Facebook, abuse occurs in the comments of posts, often on a page or group that the target does not administer or may not even know exists. Perpetrators of abuse may also refer to targets by nicknames or employ malign creativity to make it difficult to track the campaigns and narratives, often causing targets to feel overwhelmed when attempting to assess the abuse against them.

The architecture of other platforms also contributes to the effect of hard-to-detect abuse. On TikTok, abuse comes in the forms of video clips in 60 seconds or less, which replay the harassment on loop, and can also occur as text in comment sections. Abusive clips on the platform layer different forms of multimedia to produce content that compounds the severity of abuse. Harassment of public figures is not typically sent directly to them if they do not have a presence on the platform, but the videos are recommended to like-minded users by TikTok's algorithm, helping them reach wider audiences.[76]

Offline Burden on Targets Discounted

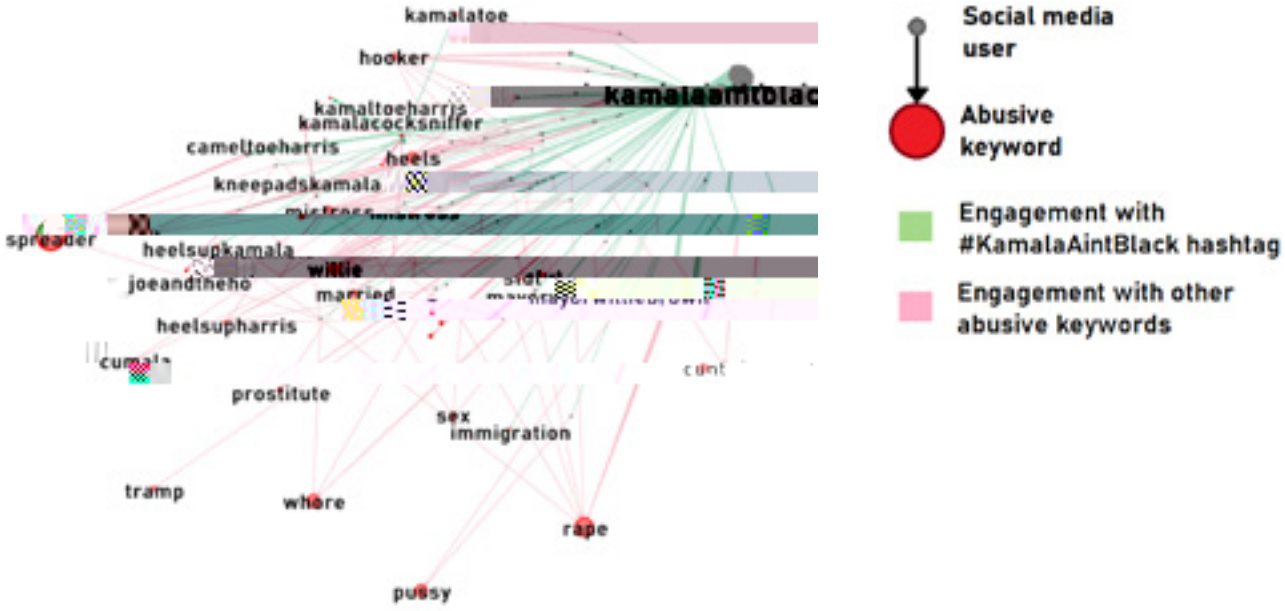
Blocking, reporting, muting, and restricting one's account are ways to manage during an abusive episode or disinformation campaign. However, these mitigating features offered by platforms do not account for the psychological and physical effects on targets. One focus group participant noted that these campaigns are "obviously...designed to push [you] down and it has the effect of this grinding away at your resistance, your ability to get through something psychologically. So, I almost see it as a psychological warfare technique."

These efforts also often have effects on women's physical security as well as the ways they participate in public discourse. Like Yeganeh Rezaian, women can be targets of hacking attempts. One focus group participant noted that in addition

Nicole Perlroth described an incident in which a colleague had been subject to gendered harassment on Twitter but received no response from the platform. It took her personal intervention to get the content removed, per Twitter's policy:

Women of Color Face Far Greater Threats

Both the quantitative and qualitative data collected in this sample underscore a trend established by other studies: gendered harassment and disinformation campaigns against women of color online are greater in volume and more serious in tone than those that white women face.[77] [78] Several of the women of color in the sample faced multiple gendered or sexualized disinformation narratives in addition to a high level of gendered abuse. White research subjects were the targets of fewer disinformation narratives and received less harassment during the collection period. The



network visualizations below demonstrate the intersectionality of the attacks against Kamala Harris and Alexandra Ocasio-Cotrez. Abusers who used the hashtag #KamalaAintBlack across social media also engaged with sexualized narratives when discussing Kamala Harris, such as her relationship with Willie Brown and false narratives around her sexuality with terms such as "HeelsUpHarris," "whore," and "superspreader." Similarly, users who demeaned Alexandra Oc12 352.484 TmitT9ssKt7r1.07 TKt7224 SimilaOrx20.4 ex20.4 z.5 (r)0bd rr Oc1l.5 (c1l.5 6.484 Tmitn4 Tmitg.5 (r)0hx20.4 ex2

Policy Recommendations

The online misogyny women face—and with it, the gendered and sexualized disinformation campaigns against them—is not a problem any one group or sector can solve alone. Currently, there are weaknesses in every adjacent function, from platform policies and content moderation, to political recognition and employer support. The recommendations in this paper focus on critical changes that could have an immediate impact on women's experiences on social media platforms and in public life.

Social Media Platforms

-

this connection and provide adequate training for those managing the public presence of women under attack. Incident reports, as described above, can also help inform updates to classifiers and keywords.

It is important to recognize, however, that updating classifiers and lists of abusive keywords is to some degree a never-ending task; abusers will continue to find ways to employ malign creativity in their campaigns. As with traditional disinformation campaigns, removing the offending content and playing "whack-a-troll" will only go so far toward creating a more equitable, democratic online environment.[81]

- ***I ea aed de ec .*** Automated and machine learning models are only as accurate as their training data. This is particularly true for gendered abuse and disinformation; this research faced challenges with collecting the many forms of misogyny, hate speech, and gendered disinformation women face online due to the varied nature of language and malign creativity of perpetrators. Machine learning techniques require carefully gathered and processed training data for the unique racial, cultural, and linguistic environments in which they operate. Any automated method must also incorporate a feedback loop, as recommended by Cathy O'Neal in .[82] A feedback loop should also continuously update machine learning models using successful reports that individual users have submitted to the platform, as these are ideal examples of human-identified problematic content. Platforms might share these systems in order to eradicate cross-platform abuse, perhaps through a global consortium, as described below.

- ***T ge ne n bec a e e e , da e de f a ge ed na a e e c c de ge de -ba ed ab e a d d f a .*** As discussed above, current definitions of targeted harassment are inadequate, and do not offer subjects of gender-based abuse and disinformation campaigns protection or recourse. After updating their Terms of Service to reflect this widespread problem, platforms should train content moderators to recognize and act against gender-based abuse and employ and consult with more subject-matter experts with intersectional background to assist in policy development, training for content moderations, and oversight of policy enforcement.

- ***C ea e a d e f ce a a e a d ea gf c e e ce f e e ab e .*** Not a single focus group participant was satisfied with social media platforms' enforcement against the online misogyny they had experienced and observed. Much of this dissatisfaction stems from clear incidents of abuse and targeted harassment that meet no or little consequence. The changes outlined above would help highlight persistent abusers; for these individuals, platforms should consider heftier consequences than removal of an isolated piece of content or locking of account features until the offending content is deleted. Platforms might consider imposing—and import9.9638 0 0 1o84.438,7datinu-0.001 rassment

tics specific to other marginalized communities should be viewed as distinct from legitimate political criticism.

Similar standards should apply to Members in their official behavior, both on and offline. Elected officials should lead by example, calling out gender-based abuse and harassment when they see it, as well as not engaging in it themselves. Members should not share or employ gendered disinformation or gender-based slurs. Members of the House of Representatives are already prohibited from sharing "visual misrepresentations of other people, including but not limited to deep fake technology," and in official communications may not "disparage" other Members, including through ad hominem attacks.[88] Given the widespread nature of attacks against women in politics, Congress should develop more detailed standards for decorum around gender issues and their intersection with trends such as disinformation.[89]

- ***Rea n e ne V e ce Aga W e Ac (VAWA) a d c de aga e ge de - ba ed na a e .*** The 2019 VAWA Reauthorization Act has not received a vote in the Senate, leaving crucial protections for victims of gender-based violence lapsed. When the next Congress considers VAWA reauthorization, lawmakers should add provisions to support targets of online gender-based harassment, including budgetary allocations to build law enforcement awareness about and increase investigation of online gender-based threats. We concur with Sarah Sobieraj that "even in contexts such as the United States, where much of the [online] abuse will not be legally actionable, responding [law enforcement] officers can play an important role in supporting victims. Helping officers become well versed on attackers' tactics and the primary venues for online hostility is an important first step."[90] Lawmakers could also consider including platform transparency and disclosure requirements about gender-based abuse and disinformation (described above) in the VAWA reauthorization bill.

Employers

- ***De e c e f e ee fac g e na a e a d ab e.*** For many public-facing industries, including the media, academia, think tanks, and government, employee engagement on social media is now critical to both brand and individual success. Many employers have policies relating to employees' or affiliates' use of social media, but far fewer have support mechanisms for those undergoing online abuse as a result of their work-related social media engagement. Employers should consider providing mental health services, support for employees or affiliates' legal fees and other expenses (such as anti-doxxing service subscriptions), as well as outlining clear mechanisms for targets to report such campaigns against them to official communications and human resources staff. Organizations with non traditional arrangements with affiliates—such as think tanks that associate with fellows who are not technically employees of the institutes they represent—must recognize that these individuals may also require support and benefits beyond institutional affiliation.

  Organizations can also engage with social media platforms on behalf of targets of abuse, adding additional legitimacy and urgency to targets' reports; in instances in which professional organizations highlight the campaigns against their members, such as when the Coalition for Women in Journalism released a statement in support of Leta Hong Fincher, platforms seem to take action more quickly. Above all, employers should recognize that

About the Authors

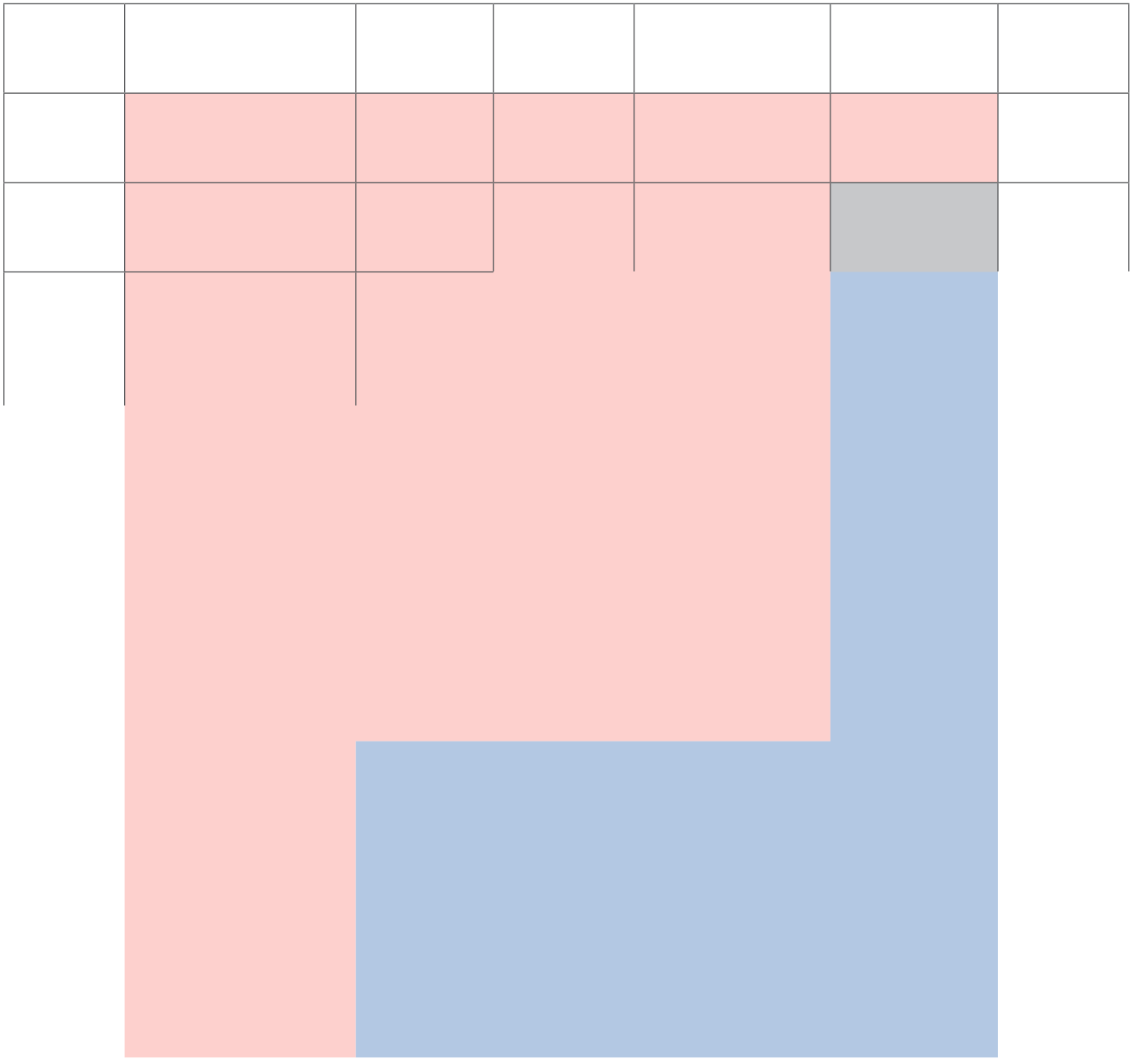
Nina Jankowicz serves as the Disinformation Fellow within the Wilson Center's Science and Technology Innovation Program. She is the author of . (Bloomsbury/IBTauris 2020). She has worked as a strategic communications adviser to the Ukrainian Foreign Ministry under the auspices of the Fulbright Public Policy Program and managed democracy assistance programs to Russia and Belarus at the National Democratic Institute. Ms. Jankowicz received her MA in Russian, Eurasian and East European Studies from Georgetown University and her BA in Russian and Political Science from Bryn Mawr College.

Jillian Hunchak is a research analyst who specializes in far-right extremism in French- and English-language contexts. She currently carries out research at Moonshot CVE and Insight Threat Intelligence, and has completed internships at the Organization for the Prevention of Violence (OPV) and the International Centre for the Study of Radicalisation (ICSR). She holds an MA from King's College London in Terrorism, Security and Society.

Alexandra Pavliuc is a PhD student at the Oxford Internet Institute who studies the spread of disinformation online. She is a researcher on the Alternative News Networks Project (part of the Computational Propaganda Project at the OII). Her work has beesihn6eshed by 'rfext.2002 Tm3C8e002 Tm3m bs Scie/Titor rbeeWPa0.345 0uCegre'olitical22s

Appendix A: Platform Policy Quick Reference

- Forbidden
- Unclear
- Permitted where not otherwise forbidden (i.e. an action violates another policy like breaking the law)



Appendix C: Keywords Used in Data Collection and Analysis

The following keywords were used to identify social media posts that contained transphobic, racist, and sexualized content that might lead to the discovery of disinformation narratives. Variations of some of these keywords were used in parts of the analysis (for example, only the word "Aroush" was used to identify all posts containing the "Kamal Aroush" narrative). Many of these subject specific keywords were discovered throughout the analysis process and added to the final data collection (such as "married her brother" for Ilhan Omar, and "HeelsUpHarris" for Kamala Harris). This process further underscores evidence of malign creativity and coded language to target and abuse women online.

Generalized Gender-Based Abuse	Sexual Disinformation	Transphobic Disinformation	Racist Disinformation
rt r	rr	t y r	rr r y
t	r t r r	r	r t r
t		tr	r
t r	D f	y y	t rr r t
rr		tr y	t
t rr	r rr	tr r	r t fr
y	rr	tr t t	y rt
t	t		
Dy	r t rr		
t	t		
y	/ 2020		
r	ff r		
r	tr		
r			
t			
t rr	rr		
t rr	rr r r t r		
t rr			
t	tr t r t		
tr	r r r		
	t		
yf	tr f		

Appendix D: Semi-Structured Interview Questions

- What have been the main characteristics of gendered disinformation that has been targeted at you online?
- What are other characteristics that you feel women generally face, but that you have not necessarily faced?
- How would you define what you face online? Please provide a term, and a longer statement.
- Have you felt that there was a level of coordination between your attackers when they attacked you online?
- Do you feel that there is a difference between the disinformation targeted against you, and other disinformation you are aware of?

Appendix E: Focus Group Script

We have invited you here today because we are interested in your unique perspectives on the challenges women in public life face online. Each of you study Internet trends in some capacity and as women, you may have faced harassment and misogyny online yourselves. We are grateful for the opportunity to tap into the deep knowledge in this virtual room and workshop policy solutions for platforms and governments.

We will be recording this focus group for note taking purposes. But, in the final report, no comments will be personally attributed to anyone. We will begin the recording now.

1. What are the main characteristics of the misogyny that has been targeted at you, or that you have seen be targeted at other women online? (we're aiming for types of content, ways that women are approached, etc)
    a. Do you think there is a level of coordination in these attacks? If yes, how so?
    b. Do you think these types of attacks have an element of disinformation (or manipulated information) to them?
    c. How would you define what you're seeing?

1. Can you think of an example of a social media platform's regulation to counter misogyny on their platform that you think was successful? Let's go around and each give an example, if we have one.
    a. Let's discuss the main characteristics of your successful examples.
2. Can you think of an example of a social media platform's regulation to counter misogyny on their platform that you think was unsuccessful? Let's go around and each give an example, if we have one.
    a. Let's discuss the main characteristics of your unsuccessful examples.

Endnotes

1 Lucina Di Meco, "#SHEPERSISTED: Women, Politics & Power in the New Media World," The Wilson Center, Fall 2019.

2 "Naked untruth; sexualised disinformation." , November 7, 2019, 47 (US).

3 Nina Jankowicz, "How Disinformation Became a New Threat to Women," , December 11, 2017.

4 Aja Romano, "Deepfakes are a real political threat. For now, though, they're mainly used to degrade women," , October 7, 2019.

5 Jane Lytvynenko and Scott Lucas, "Thousands Of Women Have No Idea A Telegram Network Is Sharing Fake Nude Images Of Them," , October 20, 2020.

6 Frances Perraudin, "Alarm Over Number of Female MPs Stepping Down After Abuse," October 31, 2019.

7 Di Meco, 2019.

8 Sarah Sobieraj, : (Oxford University Press, 2020).

9 Di Meco, 2019.

10 Sobieraj 2020, 5.

11 Ibid., 10.

12 Ibid.

13 Amnesty International, "Toxic Twitter," March 2018.

14 Amanda Lenhart, Michele Ybarra, Kathryn Zickuhr, and Myeshia Price-Feeney, "Online Harassment, Digital Abuse, and Cyberstalking in America," Data & Society, 2016.

15 Kirsten Zeiter, Sandra Pepera, and Molly Middlehurst, "Tweets that Chill," National Democratic Institute for International Affairs (NDI), 2019.

16 Ibid.

17 "#NotTheCost: A Call to Action," National Democratic Institute, 2016.

18 Sobieraj 2020, 119.

19 Ibid., 112.

20 Samantha Bradshaw, "The Gender Dimensions of Foreign Influence Operations," Global Affairs Canada, 2019.

21 Jankowicz 2017.

22 Ellen Judson, Asli Atay, Alex Krasodomski-Jones, Rose Lasko-Skinner, and Josh Smith, "Engendering Hate: The contours of state-aligned gendered disinformation online," Demos, October 2020.

23 Ibid.

24 Evelyn Douek, "What is Coordinated Inauthentic Behavior on Social Media?," , July 2, 2020.

25 Sobieraj 2020, 4.

26 Sarah Sobieraj, "Bitch, slut, skank, cunt: patterned resistance to women's visibility in digital publics," & 21, Volume 11 (2018).

27 Ibid.

28 Judson, Atay, Krasodomski-Jones, Lasko-Skinner, and Smith, "Engendering Hate," 2020.

29 See Appendix A for more detail on different platforms' user policies.

30 Twitter Support, "Notices on Twitter and what they mean," Twitter Help.

31 Twitter Support, "Synthetic and manipulated media policy," Twitter Help.

32 Facebook, "Regulated Goods," Facebook Community Standards.

33 Platforms are less likely to enforce against harassment of public citizens than harassment of private citizens. For example, users cannot "purposefully expose" public citizens to calls for their own death or threaten "severe violence", but public figures are not protected by Facebook's Violence and Incitement policy, which only applies to "minor public figures" and private citizens. Facebook, "Bullying," Facebook Community Standards. Facebook, "Violence and Incitement," Facebook Community Standards.

34 Facebook, "Bullying," Facebook Community Standards.

35 "YouTube Community Guidelines enforcement," Google Transparency Report, Google, September 2020.

36 Kaitlyn Tiffany, "Reddit Squashed QAnon by Accident," September 23, 2020.

37 Kaitlyn Tiffany, "Reddit Is Finally Facing Its Legacy of Racism," , June 12, 2020.

38 Rachel Lerman, "The conservative alternative to Twitter wants to be a place for free speech for all. It turns out, rules still apply," , July 15, 2020.

39 Nick Hopkins, "Revealed: Facebook's internal rulebook on sex, terrorism and violence," , January 3, 2018.

40 Andrew Marantz, "Why Facebook Can't Fix Itself," , October 12, 2020. Casey Newton, "The Trauma Floor: The secret lives of Facebook moderators in America," February 25, 2019. Zach Whittaker, "Facebook to pay $52 million to content moderators suffering from PTSD," May 12, 2020. Casey Newton, "The Terror Queue: These moderators help keep Google and YouTube free of violent extremism—and now some of them have PTSD," December 16, 2019.

41 Isaac Stanley-Becker and Elizabeth Dwoskin, “

81 Nina Jankowicz, “The Only Way to Defend Against Russia’s Information War

so