

The background of the entire page is a repeating pattern of white speech bubbles on a dark red background. The bubbles are scattered across the page, with a higher density on the left side.

ISD

Powering solutions  
to extremism  
and polarisation

# Public Figures, Public Rage

Candidate abuse  
on social media

Cécile Guerin

Eisha Maharasingam-Shah



## About This Report

This report presents the findings of a research project measuring the scale of online abuse targeting a variety of Congressional candidates in the 2020 US election. It found that women and candidates from an ethnic minority background are more likely than men and those who do not have an ethnic minority background to receive abusive content on mainstream social media platforms (Facebook and Twitter). It provides some recommendations and next steps which should be taken by technology companies and policymakers to protect candidates who are more likely to be targeted online and receive abusive content.

## Acknowledgements

We would like to express our gratitude to members of the Institute for Strategic Dialogue (ISD), especially Chloe Colliver, Jennie King and Natasha Bach, for their helpful feedback and revisions, and Zoé Fourel for her support in the research and writing process of this report. Any mistakes or omissions are the authors' own.

Beirut | Berlin | London | Paris | Washington DC

Copyright © Institute for Strategic Dialogue (2020). Institute for Strategic Dialogue (ISD) is a company limited by guarantee, registered office address PO Box 75769, London, SW1P 9ER. ISD is registered in England with company registration number 06581421 and registered charity number 1141069. All Rights Reserved.

[www.isdglobal.org](http://www.isdglobal.org)

## Executive Summary

This report outlines initial findings from research undertaken to analyse the scale and nature of online abuse targeting Congressional candidates during the 2020 US presidential campaign. It assesses whether gender, ethnicity and party affiliation might play a part in determining the level and type of abuse candidates receive on mainstream social media platforms (Facebook and Twitter). The report shows that online abuse continues to be widespread on social media despite tech companies' commitment to making their platforms safe for all, and that women and candidates from ethnic minority backgrounds are especially likely to be targeted systemically with abusive content.

### Key findings

- **Women were far more likely than men to be abused on Twitter, with abusive messages making up more than 15% of the messages directed at every female<sup>1</sup> candidate analysed, compared with around 5–10% for male candidates.<sup>2</sup>** The only exception to this trend was Republican Senator Mitch McConnell: 27% of the messages he received were classified as abusive.
- **Women of ethnic minority backgrounds were particularly likely to face online abuse.** On Twitter, Democratic Congresswoman Ilhan Omar (Somali–American) received the highest proportion of abusive messages of all candidates reviewed, at 39%. Democratic Congresswoman Alexandria Ocasio-Cortez (Puerto-Rican–American) received the highest proportion of abusive comments on Facebook.
- **Conversely, male politicians of ethnic minority backgrounds did not appear more vulnerable to abuse than their white counterparts.** Democratic Senator Cory Booker and Republican Senator Tim Scott (both African–American) received levels of abuse similar to white male candidates, although the abuse was more likely to be racialised. Language targeting Tim Scott in particular frequently included racist slurs and tropes.
- Mitch McConnell received the most abuse of all male candidates both on Twitter and Facebook, with levels approaching or sometimes higher than female counterparts.
- **Both on Twitter and Facebook, abuse targeting women was more likely to be related to their gender than that directed at men,** with abuse attacking female candidates based on their physical appearance and perceived lack of competence. Conversely, abuse targeting men was more generalised, often attacking their political stances.
- There was more abuse on Twitter than on Facebook; however, these results may be deceptive and relate in part to there being limited public access to Facebook data. While the overall proportions of abusive content as a share of content mentioning each candidate on Facebook were significantly lower than those

found on Twitter, female politicians still received nearly 12% more abusive comments than their male counterparts on Facebook, on average.

- **On Facebook, female Democrats received ten times more abusive comments than their male counterparts, while Republican women received twice as many abusive comments as Republican male peers.** These trends were mirrored on Twitter, with the exception of those directed at Mitch McConnell.

## Recommendations

Online abuse is just one of the many faces of harm prevalent on sites like Facebook and Twitter. For a number of years, these platforms have attempted to make voluntary changes in policy and practice to address the enormous scale of illegal and potentially harmful content and behaviour on their platforms, from disinformation to hate speech and terrorist content. Under pressure from media, policymakers and advertisers, platforms have taken steps to broaden and detail their terms of service, improve their transparency reporting and build content moderation teams. Yet this has failed to stem the tide of hate and abuse across social media sites.

The findings of this report, combined with those of ISD's consistent reporting on the prevalence of hate, extremism and disinformation on social media platforms, underscore the limits of self-regulation and self-reporting by social media companies in dealing with harmful content online. Democratic governments must act swiftly and proportionately to enforce responsibilities on companies to mitigate the risks of users and societies using their products and platforms. Instead of focusing on determining the legitimacy or not of individual pieces of content, they should concentrate on the systems, design choices and decisions that are in place to govern the online information environment, which is currently so hospitable to hatred and harassment.

The recommendations below propose two approaches. One is a set of voluntary steps for platforms to take to enable more effective and sustainable protections for all users affected by harassment and abuse on their sites; the other is a regulatory approach for governments that

seek to use their democratic mandate to curb the worst excesses of companies that host and curate potentially harmful content online.

Other stakeholders have a role to play in protecting democracies from the damage of systemic harassment of public figures. Civil society organisations can continue to highlight and expose the scale of online abuse taking place on social media platforms and the extent to which it affects high-profile figures in politics, including women and minorities.

While this research report provides some insights into the nature and scale of abuse online, we were not able to cover the full spectrum of manifestations of online abuse and the communities that are often heavily targeted by it. There is much more to be done to understand and therefore respond effectively to this broad set of issues in order to protect a robust democratic discourse and debate online and offline.

These are our recommendations for social media platforms:

- **Social media platforms need to provide greater transparency about their content moderation policies, processes and enforcement outcomes relating to harassment and abuse, specifically:**
  - o The type of content that falls inside and outside their relevant policies
  - o The financial and human resources allocated to content moderation, including the linguistic and cultural contextual expertise of those teams
  - o The appeals and redress processes in place for wrongful removals
  - o Regular reporting on the type of content removed for harassment, bullying and abuse, including the proportion of content that is removed under these policies that is image-based or video-based.
- As part of transparency efforts, **social media platforms should archive and preserve all data relating to content removals undertaken during the COVID-19 pandemic through artificial intelligence (AI)**, in order to enable future study on the successes, failures and risks of reliance on these processes for decision-making about

content moderation. ISD signed an open letter in April 2020, along with 45 other digital rights and research organisations, to call for social media platforms to enable the preservation of and access to content removed by algorithmic decision-making during the pandemic.<sup>3</sup> The letter also called for additional transparency reporting to provide further information to users and researchers on the nature and potential impacts of those currently opaque systems. These systems play an important role in enabling the identification of content violating company terms of service. It is therefore imperative that there is greater understanding of their potential inaccuracies or biases, as well as their strengths. AI moderation systems can disproportionately mislabel or wrongly categorise content from minorities, yet companies have a growing reliance on AI to keep costs low and scale high. These systems should undergo periodic, independent review to stress-test these flaws, and ensure there is an appropriate balance between technical solutions and human oversight (see our recommendations for democratic governments, below).

- **Human moderators need to receive appropriate training and guidance on detecting and assessing abuse targeting high-profile individuals, including gender-based and intersectional abuse.** To understand the scale of and adequately address gender-based and intersectional abuse on their platforms, social media companies should formalise training from expert organisations on issues of harassment and identity-based abuse for their workforces, at the policy design and enforcement levels. Company managers should regularly update policies to reflect new patterns of illegal and harmful behaviour and evolution in abusive language informed by the latest research. They should have precise policies on the use of abusive or false images and videos.
- **Social media platforms should put in place measures to address and minimise the abuse and harassment of politicians and high-profile figures on their platforms.** They should provide guidance and support to political candidates on steps to remain safe online while campaigning. Facebook's introduction of new platform features that protect prominent figures could be an example

of good practice in addressing this type of abuse, if enforced appropriately and comprehensively on the platform, which has yet to be seen. Such initiatives should be replicated across platforms that currently lack sufficient policies to address systematised abuse against public figures. The scale of abuse highlighted in this research shows the problem is widespread on social media and requires companies to take adequate measures to address it. Measures put in place by platforms should emphasise greater control for users over their settings and online experiences. The development of new features and options on platforms should be informed by existing and new research into the impact of abuse on victims, and through consultation with those victims or their representatives. One example of such a change to platform features is the introduction of a new Twitter setting to allow a user to limit replies only to their own followers, which provides a user with greater control over the content they receive on the platform.

These are our recommendations for democratic governments:

- **Democratic governments should pursue regulation that moves away from siloed, content-based rules attempting to address each type of illegal activity or breach of rights separately and should instead pursue regulation requiring transparency and accountability for the processes and systems that order, curate, promote, target, amplify or, in many cases, profit from user-generated content.** Transparency is a prerequisite for users to understand how technology platforms and products are involved in promoting illegal or harmful activity such as abuse and harassment online, and a necessary first step in designing effective and proportionate mitigation for these negative effects. Democratic governments must move away from attempting to dictate the rights and wrongs of specific content removals and towards a joined-up approach to dealing with the means of distribution and decision-making of these content-hosting sites. The UK's Online Harms White Paper has done this by charting how platforms have a 'duty of care' to deal proactively and systemically with possible risks that users might encounter when using them.<sup>4</sup> The EU also has the opportunity to

design responsible and sustainable digital regulation through the Digital Services Act.

- **Transparency and mandatory access to algorithmic systems must be the core of any regulatory set-up.**<sup>5</sup> Transparency from technology companies has improved over the last five years, with the emergence of self-reported statistics on content moderation efforts and associated policy decisions. However, this leaves companies to mark their own homework, while ignoring the bigger picture questions over how algorithmic systems and design decisions are themselves affecting outcomes relating to abuse or harassment, among other harms. It is of central importance that governments, civil society and the public are able to understand better the impact the internet has on society and democracy in order to encourage its positive effects and curb negative ones. There are four areas in which democratic governments should seek to institute mandatory, regulated transparency from content-hosting platforms: content moderation, advertising, complaints and redress, and algorithmic decision-making.<sup>6</sup>
-

## Introduction

Online abuse in politics has become a significant issue in recent years, likely both a signal and symptom of polarised political landscapes across the globe, with research showing that politicians are the target of attacks, harassment and threats on social media. During the 2019 UK general election, political candidates received an estimated four times the amount of online abuse they had done two years before.<sup>7</sup> A study analysing ‘uncivil’<sup>8</sup> messages received by Canadian politicians and US senators found that 11% of public messages sent to Canadian politicians and 15% of public messages sent to US politicians could be categorised as such.<sup>9</sup> While political figures have always received increased attention and criticism because of their public position, social media has lowered the barriers to entry for abusive and threatening behaviour towards this group, as many others.

High-profile cases of online abuse against prominent politicians have made headlines, and a number of US politicians, including Alexandria Ocasio-Cortez, Ilhan Omar and Rashida Tlaib, have spoken out about their experiences of online abuse. In the summer of 2019, Alexandria Ocasio-Cortez faced vicious attacks after blocking troll accounts on Twitter;<sup>10</sup> in April the same year, calls for violence against Ilhan Omar emerged on Twitter after a video posted by President Trump juxtaposed images of Omar with footage of the twin towers falling on 9/11.<sup>11</sup>

Media reports and existing literature have shown that women involved in politics are a particular target for online abuse. While the increase in women’s participation in the 2018 US midterm elections is to be celebrated, it was accompanied by an alarming increase in online harassment towards them.<sup>12</sup> Rheault, Rayment and Musulan’s study into the abuse of Canadian and US politicians showed that while men faced significant abuse, ‘women who achieve a high status in politics are more likely to receive uncivil messages than their male counterparts’.<sup>13</sup> Research by AI company Max Kelsen showed that Hillary Clinton received nearly twice as much abuse as Bernie Sanders, her main opponent in the 2016 Democratic primary.<sup>14</sup>

The phenomenon is widespread across the globe. A 2016 report by the Inter-Parliamentary Union profiled 55 women parliamentarians from 39 countries and found that 65.5% had received ‘humiliating sexual or

sexist remarks’, while 41.8% reported the distribution of ‘extremely humiliating or sexually charged images’ and 32.7% stated they had been harassed in some form. Among those who had experienced gender-based violence, 61.5% believed the acts had been ‘intended primarily to dissuade them and their female colleagues from continuing in politics’.<sup>15</sup>

Aside from the profound effect on victims, such abuse is indeed an impediment to democratic societies, threatening progress on diversity and representation in politics. Abuse has been shown to deter women and minorities from pursuing careers in politics, or even encourage those already engaged to step down from political life.<sup>16</sup> In the most extreme cases, online abuse can have real-world consequences and create a climate in which offline abuse thrives.

While causality between online and offline abuse is difficult to establish, offline violence often finds its origins in online conversations. Following the UK’s decision to leave the EU in 2016, pro-European MP Anna Soubry, who faced considerable online abuse,<sup>17</sup> was abused outside Parliament.<sup>18</sup> In 2010, British MP Stephen Timms was stabbed by an Islamic extremist who claimed to have been influenced by online sermons.<sup>19</sup> In 2016, the Labour MP Jo Cox was brutally murdered outside her home by a white supremacist sympathiser amid a rise in far-right sentiment in the UK.<sup>20</sup> While such cases are extreme, we have observed a mainstreaming of divisive and polarising language against candidates, including at the highest level of political office. In 2019, President Trump told four congresswomen of colour – Alexandria Ocasio-Cortez, Ilhan Omar, Rashida Tlaib and Ayanna S. Pressley – ‘to go back’ to their countries.<sup>21</sup>

Despite multiple media reports of online abuse targeting politicians and some initial research into the phenomenon, there is little data-driven research studying the scale and nature of online abuse directed at political candidates in the US. This research has combined machine learning and natural language processing technology (NLP) with qualitative analysis to analyse the scale and nature of online abuse during the 2020 presidential campaign. By studying a sample of candidates it provides initial insights into the impact of gender, race and party affiliation on online abuse.



## Methodology

### Defining the Research Scope

Online abuse can take many forms and encompass a variety of tactics and behaviours. The Women's Media Center defines it as follows: 'Online abuse includes a diversity of tactics and malicious behaviours ranging from sharing embarrassing or cruel content about a person to impersonation, doxing, stalking and electronic surveillance to the non-consensual use of photography and violent threats'.<sup>22</sup> For the purpose of this research, we examined online abuse through the lens of language and content, adopting a broad definition which does not simply encompass illegal behaviour. We therefore defined abuse as: **content which includes any kind of direct or indirect threat towards or referring to an individual; content which promotes violence against an individual based on any part of their identity including gender, race, ethnicity, religion or age ('protected characteristics'); content which aims to demean and belittle an individual based on any part of their identity.**

### Research Objectives

ISD conducted a study into abuse targeting political candidates in the US, focusing on Twitter and Facebook content from 25 June to 6 July 2020. Facebook has 190 active million users in the US,<sup>23</sup> while Twitter (over 81 million users)<sup>24</sup> is a key platform for political discussion and engagement, frequently used by political candidates to communicate publicly about their activities.

The aim of this research was to:

- Determine the scale and proportion of online abuse targeting political candidates in the US on mainstream social media platforms, ahead of the elections in November 2020
- Highlight the significance of online abuse in the US presidential campaign and its impact on democratic debate
- Understand the relevance of party affiliation, gender and ethnicity to the scale and nature of abuse directed at candidates

- Provide recommendations to tech companies, victims of online abuse and civil society partners on how to tackle the proliferation of harassment directed at public figures online.

### Choice of Research Subjects

In order to determine how gender and ethnicity impact online abuse, we selected ten US candidates according to various criteria. We wanted:

- Gender diversity (self-ascribed female versus male gender identity)
- Ethnic diversity
- Party balance (an equal split between Republicans and Democrats)
- Candidates to have a social media presence (candidates with an official presence on Facebook and Twitter).

We recognised that candidates with a large social media presence or known to be active on these platforms could be greater targets for abuse than those with a smaller online footprint. We prioritised gender and ethnic diversity of candidates above their level of social media presence. While we aimed to select candidates from a range of backgrounds, this research was limited in time and we were not able to include transgender candidates in the study as we could not identify a Republican transgender candidate. We aimed to balance party affiliation across all the candidates selected for study. Further research is needed to understand how online abuse is used to target transgender candidates. Other factors that may influence the nature and level of abuse directed at candidates also fell outside of the scope of this study, including sexuality, age and disability.

We selected the following candidates (D = Democrat; R = Republican):

- D Cory Booker (male, African–American)
  - D Dick Durbin (male, white American)
  - D Alexandria Ocasio-Cortez (female, Puerto-Rican–American)
-



- D Ilhan Omar (female, Somali–American)
- D Nancy Pelosi (female, white American)
- R Marsha Blackburn (female, white American)
- R Susan Collins (female, white American)
- R Jennifer González (female, Puerto-Rican–American)
- R Mitch McConnell (male, white American)
- R Tim Scott (male, African–American).

## Part 1: Twitter

### Methodology

We collected **all tweets sent directly to candidates in the list over an 11-day period (25 June to 6 July 2020) using Twitter’s API**. To ensure that we only analysed messages directly targeting the candidates, we excluded tweets that merely mentioned or tagged them with ‘@’, keeping only direct replies. This broader set of data, including mentions, could provide an even more comprehensive understanding of the prevalence of generalised abusive content on social media platforms, but fell outside the scope of this study.

Over this time period we collected 234,269 tweets (Figure 1)

**Figure 1** *The number of tweets sent to candidates between 25 June and 6 July 2020*

Twitter Address	Volume of Tweets
@senatemajldr	76,259
@SpeakerPelosi	65,640
@AOC	34,233
@SenatorCollins	20,140
@MarshaBlackburn	12,671
@SenatorTimScott	9,265
@IlhanMN	8,969
@CoryBooker	3,390
@SenatorDurbin	3,454
@RepJennifer	248*

\*As González received a very low number of tweets, we excluded her from the Twitter analysis.

### Automated vs manual classification

We classified the tweets manually in order to achieve accuracy and consistency, with three human annotators reviewing each tweet. This process was supported by an automated system designed to pre-sort the data into two subsets, deemed by the system as either more or less likely to be abusive.

The automated system looked for a list of keywords<sup>25</sup> in a tweet, determined manually by researchers to indicate

that a message was likely to be abusive, alongside a chain of two @ Bayes classifiers (trained with 400 and 550 manually annotated tweets from the dataset), which attempted to identify abusive or non-abusive tweets.

We used manual classification to establish the proportion of abuse targeting each candidate and to observe the patterns of abuse they were subjected to during the period of study. This was executed in two steps:

- We used the automated system to pre-sort all the messages into two sets (likely abusive, likely non-abusive). For each set, we subdivided messages into those addressed to each candidate.
- For each candidate, we then randomly selected 100 tweets from the abusive and non-abusive datasets and scored them manually as abusive or non-abusive. Finally, we used the known proportions of messages in the samples to correct these counts and provide estimates of the absolute overall number and proportion of tweets to each candidate that were abusive or non-abusive.

This approach allowed us to code and verify a sample of tweets per candidate manually to ensure greater accuracy, while still observing and scoring large data samples and obtaining a good estimate of the percentage of abusive content.

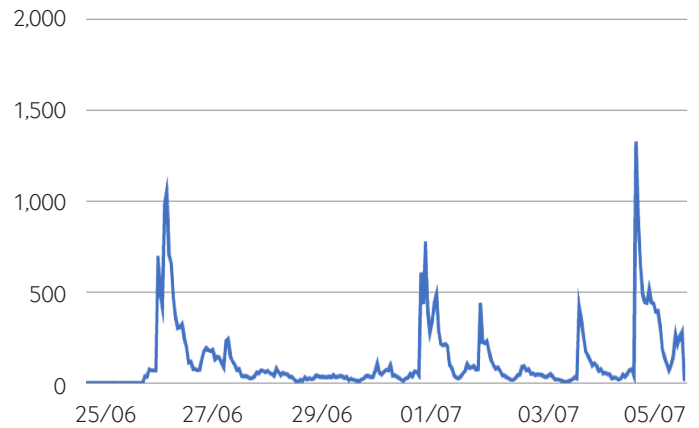
## Findings

### General volume over time

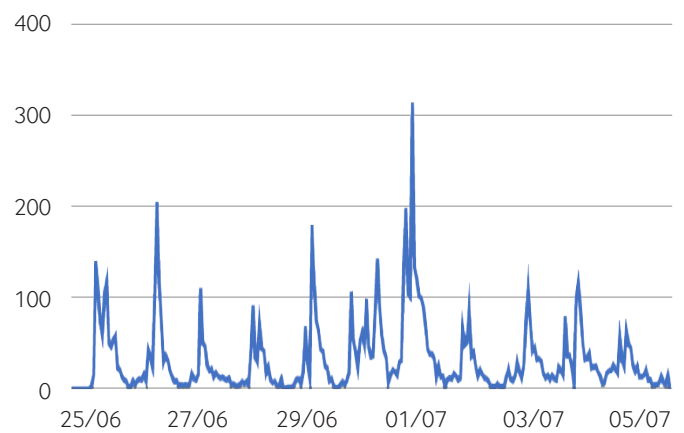
Figure 2 shows the volume of tweets candidates received during our survey period.

**Figure 2** The number of tweets received by the candidates studied between 25 June and 6 July 2020

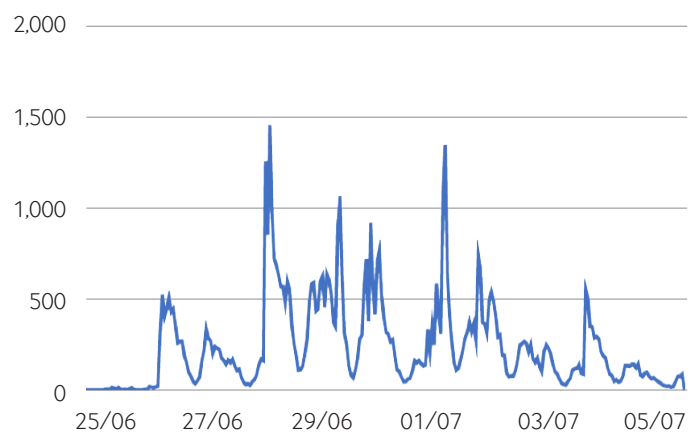
Alexandria Ocasio-Cortez @aoc



Ilhan Omar @ilhan

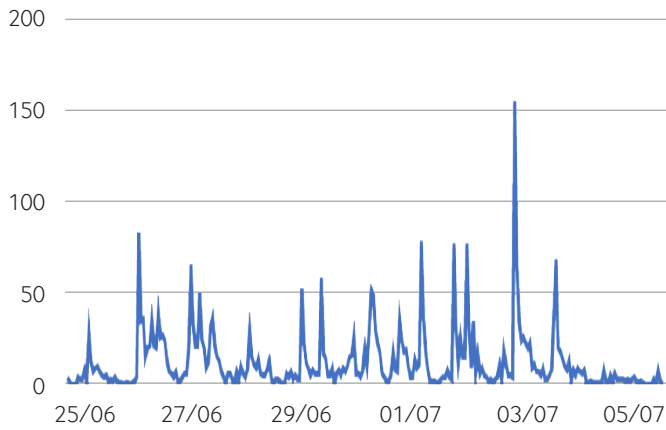


Nancy Pelosi @speakerpelosi

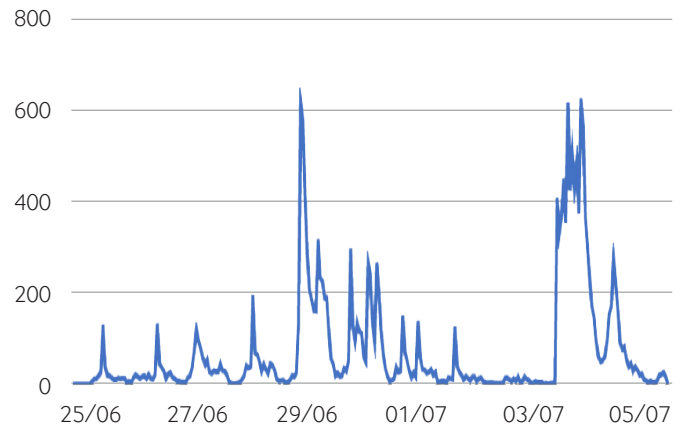


**Figure 2** *The number of tweets received by the candidates studied between 25 June and 6 July 2020*

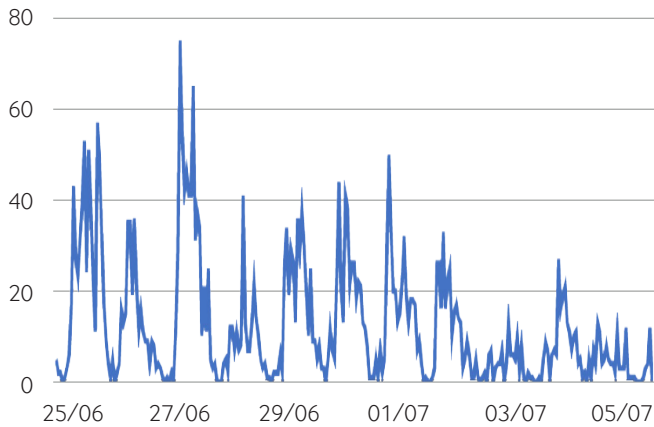
**Cory Booker** @corybooker



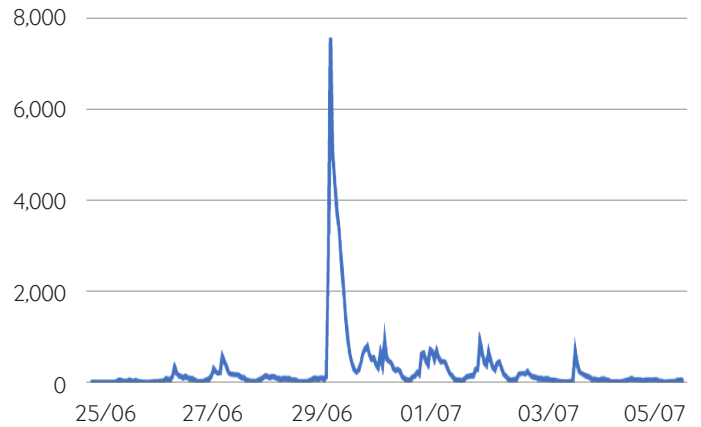
**Susan Collins** @senatorcollins



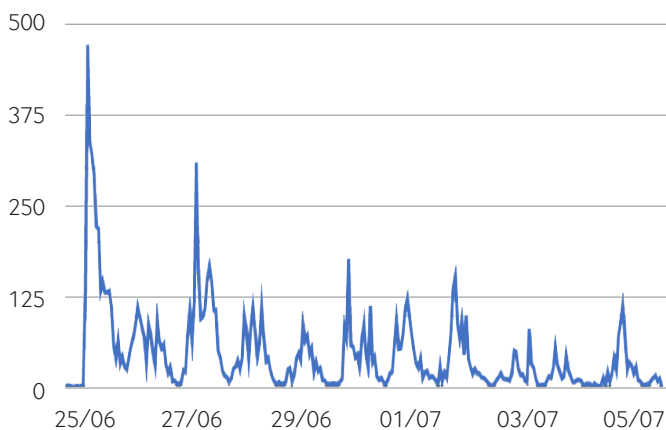
**Dick Durbin** @senatordurbin



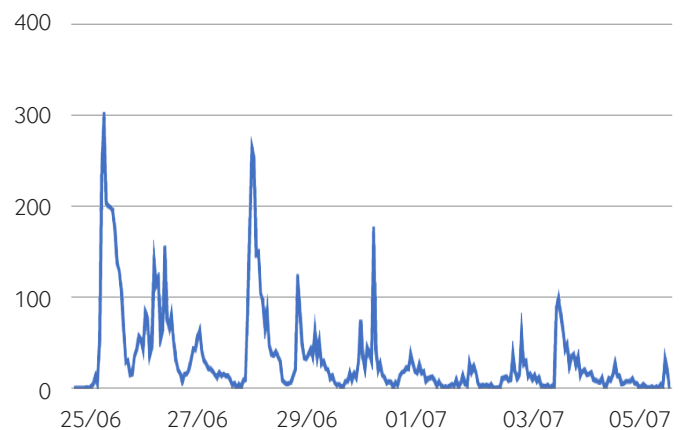
**Mitch McConnell** @senatemajldr



**Marsha Blackburn** @marshablackburn



**Tim Scott** @senatortimscott



### Estimation of scale of abuse

The manual classification gives an estimate of the proportion of tweets that are abusive for each representative (Figure 3).

**Figure 3** *An estimate of the proportion of tweets candidates received between 25 June and 6 July 2020 that were abusive*

Ilhan Omar (D)	39.1%
Susan Collins (R)	31.2%
Mitch McConnell (R)	27.8%
Nancy Pelosi (D)	21.9%
Marsha Blackburn (R)	17.0%
Alexandria Ocasio-Cortez (D)	16.5%
Tim Scott (R)	10.5%
Cory Booker (D)	10.4%
Dick Durbin (D)	5.8%

We were not able to obtain significant results for Jennifer González as the sample of direct replies to her Twitter handle was too small to draw meaningful conclusions. We therefore excluded this candidate from the analysis. From the remaining data, we found the following:

- **Women were far more likely than men to be abused on Twitter, with 15–39% of all female candidates' tweets being abusive.** Ilhan Omar received the highest proportion by a clear margin and this included abuse based on her gender (female), ethnic and national background (naturalised Somali–American) and religion (Muslim). Men received 5–10% abuse on their respective Twitter accounts.
- **The only exception to this trend was Mitch McConnell 27.8% of whose messages were abusive in the period of study,** often for his so-called 'treason' and 'support for Russia'.
- **Representatives with high-profile roles in Congress were likely to receive high proportions of abuse:** McConnell (27.8%) and Nancy Pelosi, Speaker of the House of Representatives (nearly 25%).

### Patterns of abuse and uses of language

We analysed the language most often used to target candidates over the 11-day period of study and extracted a list of 'surprising keywords and phrases' for each individual in a three-step process:

- We obtained 20 keywords by contrasting the data from tweets to a sample of standard English Wikipedia data, helping identify uncommon vocabulary.
- We extracted 20 keywords by contrasting the individual of interest's Twitter data with that of other candidates to identify language specifically targeting that person.
- We filtered these 40 keywords out through a 'blacklist' likely to generate background noise. For this research we blacklisted the terms: Trump, COVID, the individual's name and common aliases, political party names and common state names.

We ordered keywords and phrases by their degree of unexpectedness and selected the top 20 (Table 1).

Table 1 The top 20 keywords and phrases used to target candidates

<b>Alexandria Ocasio-Cortez</b>	<b>Ilhan Omar</b>	<b>Nancy Pelosi</b>	<b>Cory Booker</b>	<b>Dick Durbin</b>
defund the police	shameful	bounties on	heartbreaking	forever and ever s386 forever and
tiktok	defund the police	a hypocrite	spartacus	daca
defunding the police	a hypocrite	briefed	a hypocrite	h1b
a hypocrite	non oromos	antifa	try harder bwahaha	have to self-deport
the dnc	oromo	ccp	bounties	gc
@aoc congrats antifa	the looters	treasonous	bro	a hypocrite
@aoc lmao	oromia region	defund the police	hickenlooper	#daca stop deporting h4 age out
moron	livable wage	george floyd's name	bozo	illegals
progressives	antifa and	you and your cronies are	moron	300k legal immigrants
joe biden's husban	breonna taylor	botox	tweeting	greecard
the dumbest	deport	looters	ketamine	h4 kids
george soros	hachalu	the dnc	nibbling	unblock #s386 and support freedom
go back to bartending	illegals	george kirby	treasonous	killinois
bootleg	minnesotans	go eat your icecream	on colbert	forever and ever s386 forever and
@aoc go back to waitressing	defunding the police	unverified	ketamine	600k illegals
internships	a bailout	to chinatown	nibbling	your inaction
a waitress	taxpays	gourmet icecream	any yt ppl wanting to reparate	dickie
kpop	non oromos	on the sidewalks	hey famo i'm being targeted by	eb
outspent 10 1	somalia	botox	home murders do we need easer	killinois
	hachalu	invocation of emergency powers by	newark	s 386

<b>Marsha Blackburn</b>	<b>Susan Collins</b>	<b>Mitch McConnell</b>	<b>Tim Scott</b>
bounties on american soldiers	bounties on our troops	bounties on	shameful
ccp	kavanaugh's vote	a hypocrite	ban chokeholds
gaslighting	spineless	defund the police	bounties on
a hypocrite	you voted to acquit	treasonous	sowell is
tweeting about	enabler	antifa	a white supremacist
antifa	treasonous	ccp	chokehold
defund the police	you are a hypocrite	shameful	indefensible
treasonous	enablers	briefed	retweet
vlad the impaler is paying money	furrowed brow	moron	toothless
cowards	gullible	you and your cronies	tweeted
hypocrites	shameful	tweeting	tweeting
hkers	tweeting	declaring war on the rulebook	retweeting a video
handwritten notes	inaction	@senatemajldr we must have no stigma	retweeted a video
knoxville 901 527 9199 memphis	cowards	turtle	white supremacists
423 541 2939 chattanooga 731	the lobster industry	mcturtle	guns in greenville today what will
423 541 2939 chattanooga 731 660	kavanagh	@senatemajldr too little too late	coon
660 3971 jackson 865 540 3781	suzie	tr45son tr45son tr45son tr45son tr45son tr45son	dir clawback fees
865 540 3781 knoxville 901	kavenaugh	merrick garland	surely the token of the republikkkons
chattanooga 731 660 3971 jackson 865	furrow your brow	judgeships	voicemails
knoxville 901 527 9199 memphis 629	suze	tortoise	i'm frustrated that you



Our analysis of ‘surprising keywords’ highlighted a number of trends:

- **Language related to treason is present across candidates and political parties** and appears to be a common feature of US political debate online. In the case of non-US-born or minority candidates, this language often goes hand in hand with claims they are somehow un-American. Keywords associated with Ilhan Omar are ‘deport’ and ‘Somalia’. Meanwhile, Alexandria Ocasio-Cortez is associated with (Hungarian-born) philanthropist George Soros who has become a regular target of conspiracy theories.<sup>26</sup>
- **Women are more vulnerable to demeaning and belittling language than men, as well as being attacked for their appearance – regardless of political affiliation.** While terms such as ‘suzie’ and ‘suze’ were frequently associated with Susan Collins and used in abusive contexts, Nancy Pelosi faced attacks on her age and appearance with the keyword ‘botox’ appearing prominently in the list. **Women are also more susceptible to having their competence questioned**, with Alexandria Ocasio-Cortez frequently being referred to as a ‘waitress’ or ‘bartender’ and urged to ‘go back to bartending’.
- **Men and women of ethnic minority backgrounds were attacked with racialised or discriminatory language.** While attacks on Ilhan Omar included calls for deportation, our analysis of surprising phrases also surfaced the racist term ‘coon’, which was frequently used to target Tim Scott.

## Part 2: Facebook

### Methodology

We used the Facebook data aggregator CrowdTangle<sup>27</sup> to analyse content from public pages and groups, and ran search queries to collect all mentions of the ten candidates selected for this study.<sup>28</sup> The time frame used was the same as for Twitter (25 June to 6 July 2020).

It is important to note that given the nature of accessibility to Facebook data, the data used for this study only provides a very partial view of what is taking place on the platform. We were not able to access comments on public Facebook posts, or any data from private Facebook groups (even anonymised). This research may therefore miss abuse taking place on areas of the platform not accessible to researchers at scale. Twitter, on the other hand, provides extremely comprehensive access to data across its platform, enabling more rigorous and accurate conclusions to be drawn.

While Twitter’s API enabled us to collect direct replies to the candidate’s tweets, thereby ensuring their relevance from the outset, we collected Facebook data via a query search with the candidate’s full first name or their title and surname. These queries produced a number of irrelevant posts, including mentions of candidates in commercial spam, advertising or posts not related to them. We therefore used an NLP classifier to eliminate irrelevant posts from the dataset. Irrelevant posts were defined as posts completely unrelated to the individuals, such as mentions in spamming, advertising and commercial content.

We collected 26,584 relevant Facebook posts (Figure 4).<sup>29</sup>

**Figure 4** *The number of posts about candidates on Facebook that were abusive between 25 June and 6 July 2020*

Candidate	Posts
Nancy Pelosi (D)	10,937
Mitch McConnell (R)	5,425
Tim Scott (R)	3,421
Alexandria Ocasio-Cortez (D)	2,901
Ilhan Omar (D)	1,383
Cory Booker (D)	917
Susan Collins (R)	896
Dick Durbin (D)	434
Marsha Blackburn (R)	270

**Automated vs manual classification**

We applied the same methodology used for Twitter to pre-sort posts into two datasets (abusive, non-abusive), which were deemed by our automated system as more or less likely to be abusive. In order to analyse posts targeting candidates directly, we excluded posts that mentioned more than one of the individuals on our list.

For each set (abusive, non-abusive), we manually classified a random sample of 50 posts per candidate. As noted above, data accessed via CrowdTangle does not offer a comprehensive view of the platform.

We classified less than 2% of posts as abusive across all candidates so we shifted our research approach to establish and assess the level of abuse in content reaching the widest audiences on Facebook, creating datasets of publicly available posts with high levels of engagement. This was done in two steps:

1. We identified the five posts with the highest comment count per candidate. For each of these posts, we ordered the comments by Facebook’s ‘All comments’ option, where the comments the platform algorithm identified to be most relevant appeared first.
2. We selected the first ten comments that appeared on these posts, manually classifying them as abusive

or non-abusive. We then qualitatively analysed these comments to gain insight into the nature of conversation and abusive language.

**Estimation of scale of abuse**

Figures 5 and 6 show the cumulative comment count per candidate and the percentage of sampled comments that were deemed abusive through manual classification. A more in-depth analysis of the nature and language of abuse targeting candidates is provided in the next section.

**Figure 5** *The cumulative comment count for top 5 posts on Facebook per candidate between 25 June and 6 July 2020*

Nancy Pelosi (D)	146,140
Tim Scott (R)	42,060
Mitch McConnell (R)	25,943
Alexandria Ocasio-Cortez (D)	15,730
Dick Durbin (D)	14,471
Susan Collins (R)	9,167
Ilhan Omar (D)	7,517
Cory Booker (D)	2,349
Marsha Blackburn (R)	1,209

**Figure 6** *The percentage of comments on Facebook that were abusive per candidate between 25 June and 6 July 2020*

Alexandria Ocasio-Cortez (D)	28%
Nancy Pelosi (D)	22%
Susan Collins (R)	16%
Ilhan Omar (D)	10%
Mitch McConnell(R)	8%
Cory Booker (D)	2%
Dick Durbin (D)	0%
Marsha Blackburn (R)	0%
Tim Scott (R)	0%

## Findings

These figures highlight a number of key findings:

- Popular posts mentioning Alexandria Ocasio-Cortez were most likely to receive abusive comments, with some 28% of content classified as abusive.** The nature of the abuse she received spanned references to her gender, ethnicity and appearance. This high level of abuse reinforces the findings from Twitter, demonstrating that candidates with intersectional identities are more susceptible to receiving abusive content online.
- Women were much more likely to receive abuse in Facebook comments than their male counterparts.** Top posts about all female candidates but one (Marsha Blackburn) received at least 10% abusive comments. Consistent with our findings on Twitter, Mitch McConnell received more abuse in comments than other male candidates.
- For some candidates, posts with the highest comment count all related to the same (often high-profile) incident. This affected our analysis as many of the comments under these posts made no reference to the candidate. To ensure consistency in our manual classification, we did not select alternate posts and comment samples, but this could form the basis of a useful follow-up study to analyse a broader range of Facebook posts and subsequent comments.<sup>30</sup>
- While the overall levels of abusive content were significantly lower than those found on Twitter, **women received on average nearly 12% more abusive content on Facebook than their male counterparts.**
- Women were **more likely to suffer high levels of online abuse than male politicians across party lines.** Female Democrats received ten times more abusive comments than male Democrats, and Republican women received twice as many abusive comments as Republican men on average (Figure 7).
- Party affiliation significantly impacted the level of abuse received by women.** The proportion of

abusive comments towards Democratic female candidates was 12% higher than those targeting Republican female candidates. However, the small sample size (three Democratic women and two Republicans) limits the conclusions that can be drawn from such findings.

**Figure 7** *The percentage of overall comments on Facebook that were abusive, by gender and party affiliation*

Men	3.3%
Democrat	2%
Republican	4%
Women	15.2%
Democrat	20%
Republican	8%

### The nature of discussion and abusive language

As when examining the abuse directed at candidates on Twitter, we found that content directed at women was more targeted at candidates' personal characteristics and identities (including gender) than abuse directed at men.

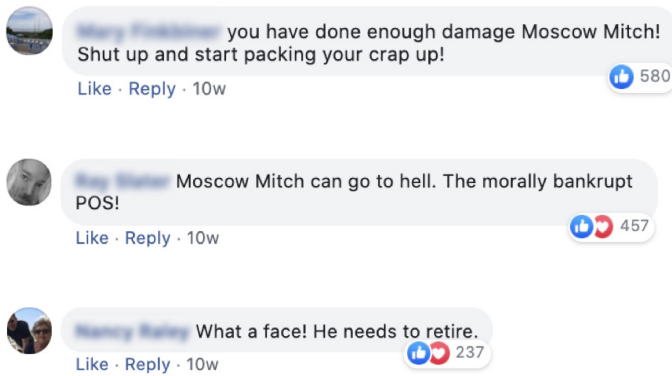
The abuse attempted to undermine women's intelligence, physical appearance or ability to hold office. This trend was most notable in comments targeting Alexandria Ocasio-Cortez and Nancy Pelosi. We found comments calling Alexandria Ocasio-Cortez a 'nutcase', 'trainwreck', 'bartender' and 'stripper', and posts describing Nancy Pelosi as a 'drunk' and a 'nan'.

Conversely, abusive comments directed at male politicians were more generalised. Examples of this were 'POS' ('piece of shit'), 'shut up' (directed towards Mitch McConnell) and 'bozo' (directed towards Cory Booker). While a large proportion of comments did not meet the threshold of abuse, they contained highly critical language. We provide an overview of key themes found in the comments below.

### Mitch McConnell

Mitch McConnell received the highest proportion (8%) of abusive comments among male candidates in Facebook posts. Posts with the highest comment count appeared between 29 June and 1 July 2020. They referenced McConnell’s comment that there should be ‘no stigma’ about wearing masks amid the coronavirus pandemic and other remarks he made on the Senate floor. The abuse directed at McConnell either attacked his appearance or was more general (Figure 8).

Figure 8 Examples of abusive comments directed at Mitch McConnell

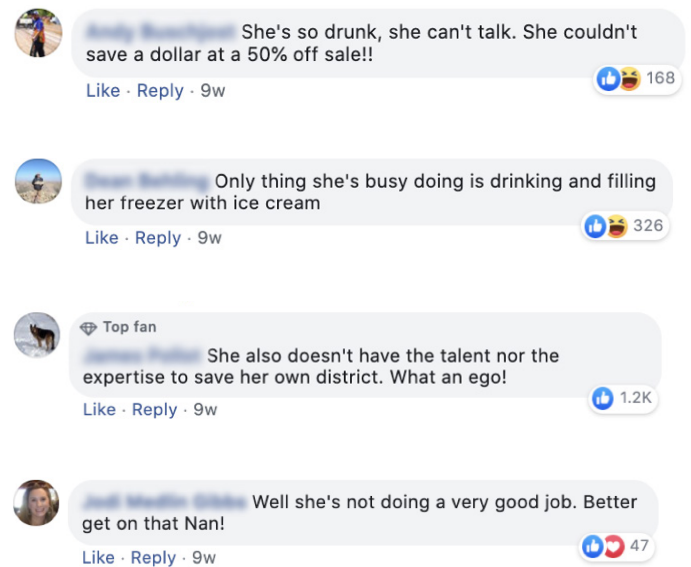


### Nancy Pelosi

The top Facebook posts mentioning Nancy Pelosi were also the most commented on overall, with the top three receiving 58,341, 35,936 and 20,954 comments, respectively. These high counts can be attributed to the number of subscribers to the posts’ authors: Donald Trump Jr, Fox News and conservative pundit Ben Shapiro. Some 16% of the comments – not all of which were abusive – called for Pelosi to be removed from office, to retire, or to be ‘voted out’. Around one-fifth (22%) of the comments about her were classified as abusive, the second highest percentage of abusive comments received by any of the candidates. Figure 9 gives some examples of comments directed at Pelosi.

Despite the posts on her varying in subject matter, the nature and language of abuse was similar and attacked her over her age and professional capability, or alleged that she was drunk. Interestingly, while abusive content related to Pelosi’s age also called into question her competence, abusive content related to McConnell’s age did not, demonstrating that women are more likely to have their competence questioned than men.

Figure 9 Examples of abusive comments directed at Nancy Pelosi



**Ilhan Omar**

In our analysis 10% of comments about Ilhan Omar were classified as abusive. Among all the candidates, this marked the greatest variance with our Twitter findings as the proportion of abuse received by Omar on Twitter during the same period was an estimated 39.1%. This difference can predominantly be attributed to the pages which produced the posts with the highest comment count: Omar’s official Facebook page, actor George Takei’s Facebook page and the progressive news organisation NowThis (Figure 10).

Figure 10 Comments about Ilhan Omar



Similar to what we observed on Twitter, abusive comments directed at Omar mentioned her Somali background and urged her to 'go back'. Content also included false allegations that Omar was facing a 40-year prison sentence.<sup>31</sup> Figure 11 gives some examples.

Figure 11 Examples of abusive comments directed at Ilhan Omar





### Alexandria Ocasio-Cortez

The top posts about Alexandria Ocasio-Cortez had a cumulative comment count of 15,730. This was the fourth highest comment count of all the candidates, behind McConnell (25,943), Scott (42,060) and Pelosi (146,140). Ocasio-Cortez received the highest level of abuse among all the candidates, with 28% of comments classified as abusive. Figure 12 gives some examples of these comments.

Four of the five posts were published by public figures and pages affiliated with the Republican party, specifically Donald Trump Jr, Senator Ted Cruz, 'Donald Trump Is My President' and 'Donald Trump 2020 Voters'.

The posts varied in subject matter, ranging from Ocasio-Cortez winning her New York primary to a statement she made about budget cuts to the New York Police Department. The language and nature of the abuse found in the top comments were almost entirely unrelated to the subject matter, and instead echoed the abusive language we found on Twitter. Commenters questioned her intelligence ('low IQ', 'simple minded'), criticised her appearance ('horse') and belittled her for her previous job as a bartender.

Figure 12 Examples of abusive comments directed at Alexandria Ocasio-Cortez

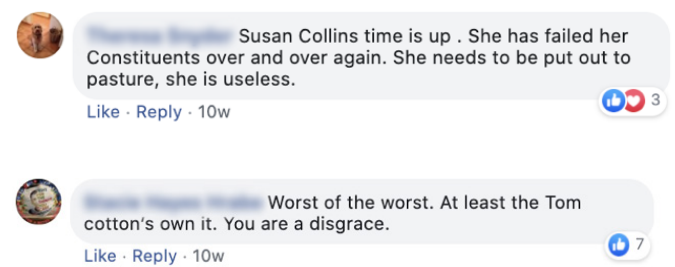


### Susan Collins

Posts mentioning Susan Collins which received the highest comment count related to her vote confirming Justice Brett Kavanaugh to the Supreme Court, and Kavanaugh's ensuing vote against abortion rights in Louisiana on 29 June – similar to what we observed on Twitter. Figure 13 shows some examples of comments directed at Collins.

Some 16% of the comments were abusive, and all related to her support for Justice Kavanaugh. They predominantly attacked Collins' intelligence and character, using terms such as 'fool', 'disgrace', 'puppet', 'useless'. While most comments made towards Collins could not be classified as abusive, 78% were highly critical, with several relating to her re-election.

Figure 13 Examples of critical comments directed at Susan Collins





## Conclusion and recommendations

### Conclusion

Despite commitments made by Facebook<sup>32</sup> and Twitter<sup>33</sup> to protect users from abuse on their platforms, this research shows that abuse targeting US political candidates continues to thrive online. In the context of continuing lockdowns ahead of the presidential election in 2020, online communication has come to play an even more prominent role in political discourse than before. The high level of abusive content directed at public figures who are in the political limelight demonstrates not only the fraught, polarised political context in which the election is taking place, but also the platform that social media companies provide to those seeking to attack political candidates and often to threaten their safety. The ability to target and amplify hate and abuse towards political candidates has been hyper-charged by the business models of social media companies that grow quickly, promote outrage or sensationalism, and are unable to deal with the subsequent harms rife across their platforms.

Our findings highlight a marked difference in the nature of the abuse directed towards female and male candidates: while men mostly received generalised attacks (about their political stances), abuse directed at women tended to be gendered and highly personal. Their skills, competency and place in the political arena were more likely to be called into question. Female candidates are disproportionately targeted by online abuse on Facebook and Twitter; they receive 12% more abuse in Facebook comments than men. On Twitter 16.5–39.1% of all comments about them are abusive, compared with just 5.8–27.8% of comments about male candidates being abusive. Female candidates from

ethnic minority backgrounds are particularly vulnerable to high levels of consistent online abuse.

In highlighting the gendered and intersectional dimension of online abuse in the US presidential campaign, this report holds wider implications for the realm of modern political communication environments. Online abuse poses a challenge for our democracies and the right to open debate, subjecting candidates to often consistent violent and aggressive language. Previous research has highlighted the extent to which online abuse can lead candidates to step down from or restrict their political activities,<sup>34</sup> and the scale of online abuse threatens diversity and representation in politics. This is a prominent issue in its own right: despite the progress made in the 2018 midterm elections, where a record number of women were elected to the House of Representatives, in 2020 women hold only 23.7% of the seats in the US Congress.<sup>35</sup> This gender disparity has long-term implications, perpetuating a gender bias in decision-making processes and subsequent policies implemented by Congress.

While Facebook's Bullying & Harassment Policy<sup>36</sup> and Twitter's Abusive Behaviour<sup>37</sup> and Hateful Conduct<sup>38</sup> policies aim to encompass a wide range of abusive and violent behaviours, they are enforced inconsistently. The definition of abuse adopted in this research takes a broader understanding than platforms' terms of service. Consequently, some of the content we classified as abusive did not violate the platform's policies. While most abusive content we analysed in this research is not illegal and does not violate platforms' terms of service, its widespread presence on social media creates a climate of intimidation and incivility towards political figures, threatening the democratic debate.

We also identified Facebook comments and tweets that did violate policy guidelines yet remained live on the platform. The reasons for this are often unclear, but there are a number of factors to consider. This research involved careful, often time-consuming manual detection of abusive content, but technology companies are increasingly scaling back their use of manual analysis of their content in favour of automated systems<sup>39</sup> and outsourcing content moderation to third-party vendors.<sup>40</sup> Moreover, Twitter and Facebook rely in part on reactive user flagging systems that require abusive content to be reported by users or victims.

“ We identified Facebook comments and tweets that did violate policy guidelines yet remained live on the platform

”

Finding an effective balance between automated and manual review systems for abusive content will be critical to the companies' efforts to begin to undermine the prominence of abuse on their sites successfully. The recommendations below suggest steps for companies and democratic governments to take that could chart a path in this direction.

## Recommendations

Online abuse is just one of the many faces of harm prevalent on sites like Facebook and Twitter. For a number of years, these platforms have attempted to make voluntary changes in policy and practice to address the enormous scale of illegal and potentially harmful content and behaviour on their platforms, from disinformation to hate speech and terrorist content. Under pressure from media, policymakers and advertisers, platforms have taken steps to broaden and detail their terms of service, improve their transparency reporting and build content moderation teams. Yet this has failed to stem the tide of hate and abuse across social media sites.

The findings of this report, combined with those of ISD's consistent reporting on the prevalence of hate, extremism and disinformation on social media platforms, underscore the limits of self-regulation and self-reporting by social media companies in dealing with harmful content online. Democratic governments must act swiftly and proportionately to enforce responsibilities on companies to mitigate the risks of their products and platforms on users and societies. Instead of focusing on determining the legitimacy or not of individual pieces of content, they should concentrate instead on the systems, design choices and decisions that are in place to govern the online information environment, which is currently so hospitable to hatred and harassment.

The recommendations below propose two approaches. One is a set of voluntary steps for platforms to take to enable more effective and sustainable protections for all users affected by harassment and abuse on their sites; the other is a regulatory approach for governments that seek to use their democratic mandate to curb the worst excesses of companies that host and curate potentially harmful content online.

Other stakeholders have a role to play in protecting democracies from the damage of systemic harassment of public figures. Civil society organisations can continue to highlight and expose the scale of online abuse taking place on social media platforms and the extent to which it affects high-profile figures in politics, including women and minorities.

While this research report provides some insights into the nature and scale of abuse online, we were not able to cover the full spectrum of manifestations of online abuse and the communities that are often heavily targeted by it. There is much more to be done to understand and therefore respond effectively to this broad set of issues in order to protect a robust democratic discourse and debate on- and offline.

“ Democratic governments must act swiftly and proportionately to enforce responsibilities on companies ”

## Recommendations for social media platforms

- **Social media platforms need to provide greater transparency about their content moderation policies, processes and enforcement outcomes relating to harassment and abuse**, specifically:
  - The type of content that falls in and outside their relevant policies
  - The financial and human resources allocated to content moderation, including the linguistic and cultural contextual expertise of those teams
  - The appeals and redress processes in place for wrongful removals

- o Regular reporting on the type of content removed for harassment, bullying and abuse, including the proportion of content that is removed under these policies that is image-based or video-based.
  - As part of transparency efforts, **social media platforms should archive and preserve all data relating to content removals undertaken during the COVID-19 pandemic through AI**, in order to enable future study on the successes, failures and risks of reliance on these processes for decision-making about content moderation. ISD signed an open letter in April 2020, along with 45 other digital rights and research organisations, to call for social media platforms to enable the preservation of and access to content removed by AI decision-making during the pandemic.<sup>41</sup> The letter also called for additional transparency reporting to provide further information to users and researchers on the nature and potential impacts of those currently opaque systems. These systems play an important role in enabling the identification of content violating company terms of service. It is therefore imperative that there is greater understanding of their potential inaccuracies or biases, as well as their strengths. AI moderation systems can disproportionately mislabel or wrongly categorise content from minorities, yet companies have a growing reliance on AI to keep costs low and scale high. These systems should undergo periodic, independent review to stress-test these flaws, and ensure there is an appropriate balance between technical solutions and human oversight (see recommendations for democratic governments, below).
  - **Human moderators need to receive appropriate training and guidance on detecting and assessing abuse targeting high-profile individuals, including gender-based and intersectional abuse.** To understand the scale of and adequately address gender-based and intersectional abuse on their platforms, social media companies should formalise training from expert organisations on issues of harassment and identity-based abuse for their workforces, at the policy design and enforcement levels. Company managers should regularly update policies to reflect new patterns of illegal and harmful behaviour and evolution in abusive language informed by the latest research.
- They should have precise policies on the use of abusive or false images and videos.
- **Social media platforms should put in place measures to address and minimise the abuse and harassment of politicians and high-profile figures on their platforms.** They should provide guidance and support to political candidates on steps to remain safe online while campaigning. Facebook’s introduction of new platform features that protect prominent figures could be an example of good practice in addressing this type of abuse, if enforced appropriately and comprehensively on the platform, which has yet to be seen. Such initiatives should be replicated across platforms that currently lack sufficient policies to address systematised abuse against public figures. The scale of abuse highlighted in this research shows the problem is widespread on social media and requires companies to take adequate measures to address it. Measures put in place by platforms should emphasise greater control for users over their settings and online experiences. The development of new features and options on platforms should be informed by existing and new research into the impact of abuse on victims, and through consultation with those victims or their representatives. One example of such a change to platform features is the introduction of a new Twitter setting to limit replies only to a user’s followers, which provides a user with greater control over the content they receive on the platform.
- ## Recommendations for democratic governments
- **Democratic governments should pursue regulation that moves away from siloed, content-based rules attempting to address each type of illegal activity or breach of rights separately and should instead pursue regulation requiring transparency and accountability for the processes and systems that order, curate, promote, target, amplify or, in many cases, profit from user-generated content.** Transparency is a prerequisite for users to understand how technology platforms and products are involved in promoting illegal or harmful activity such as abuse and harassment online, and a necessary first step

in designing effective and proportionate mitigation for these negative effects. Democratic governments must move away from attempting to dictate the rights and wrongs of specific content removals and towards a joined-up approach to dealing with the means of distribution and decision-making of these content-hosting sites. The UK's Online Harms White Paper has done this by charting how platforms have a 'duty of care' to deal proactively and systemically with possible risks users might encounter when using them. The EU has the opportunity to design responsible and sustainable digital regulation through the Digital Services Act.

- **Transparency and mandatory access to algorithmic systems must be the core of any regulatory set-up.**<sup>42</sup> Transparency from technology companies has improved over the last five years, with the emergence of self-reported statistics on content moderation efforts and associated policy decisions. However, this leaves companies to mark their own homework, while ignoring the bigger picture questions over how algorithmic systems and design decisions are themselves affecting outcomes relating to abuse or harassment, among other harms. It is of central importance that governments, civil society and the public are able to understand better the impact the internet has on society and democracy in order to encourage its positive effects and curb negative ones. There are four areas in which democratic governments should seek to institute mandatory, regulated transparency from content-hosting platforms: content moderation, advertising, complaints and redress, and algorithmic decision-making.<sup>43</sup>
-

## Endnotes

- 01 The term 'female' is used to reference the gender to which the candidates identify. Only cisgender politicians have been selected for this study. We acknowledge the absence of transgender candidates presents a gap in understanding the full spectrum of online abuse. Including transgender candidates was methodologically challenging owing to the low number of transgender candidates with a comparable social media following compared with cisgender candidates. More research is needed into online abuse targeting transgender candidates.
- 02 The term 'male' is used to reference the gender to which the candidates identify. As noted above, only cisgender politicians have been selected for this study.
- 03 Emma Llansó, 'COVID-19 content moderation research letter', Center for Democracy & Technology, 22 April 2020, <https://cdt.org/insights/covid-19-content-moderation-research-letter/>
- 04 HM Government, 'Online Harms White Paper, CP 57, 2019', [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/793360/Online\\_Harms\\_White\\_Paper.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf)
- 05 Demos, doteveryone, Global Partners Digital, ISD and Open Rights Group, 'Algorithm inspection and regulatory access', April 2020, <https://www.isdglobal.org/wp-content/uploads/2020/04/Algo-inspection-briefing.pdf>
- 06 ISD, 'Extracts from ISD's submitted response to the UK Government Online Harms White Paper', July 2019, <https://www.isdglobal.org/wp-content/uploads/2019/12/Online-Harms-White-Paper-ISD-Consultation-Response.pdf>
- 07 Richard Vaughan, 'General election 2019: online abuse of candidates is four times worse than politicians experienced in 2017', inews, 12 December 2019, <https://inews.co.uk/news/politics/general-election-2019-online-abuse-candidates-four-times-worse-2017-mps-373983>
- 08 'Uncivil' was defined as: 'a tweet containing at least one of the following elements: (1) swear words; (2) vulgarities; (3) insults; (4) threats; (5) personal attacks on someone's private life; or (6) attacks targeted at groups (hate speech)', <https://journals.sagepub.com/doi/pdf/10.1177/2053168018816228>
- 09 Ludovic Rheault, Erica Rayment and Andreea Musulan, 'Politicians in the line of fire: incivility and the treatment of women on social media', *Research and Politics*, January–March 2019, 1–7, <https://journals.sagepub.com/doi/pdf/10.1177/2053168018816228>
- 10 Alexis Grenell, 'Should we have a right to troll politicians on Twitter?', *The Nation*, 18 July 2019, <https://www.thenation.com/article/archive/troll-twitter-aoc-saladino/>
- 11 Michael Brice-Saddler, 'He easily found hundreds of death threats against Rep. Ilhan Omar. He wants Twitter to stop them', *Washington Post*, 16 April 2019, <https://www.washingtonpost.com/technology/2019/04/16/he-easily-found-hundreds-death-threats-against-rep-ilhan-omar-he-wants-twitter-stop-them/>
- 12 Amy K. Lehr and Efthimia Maria Bechrakis, 'Against the odds: overcoming online harassment of women in politics', Center for Strategic & International Studies, 8 November 2018, <https://www.csis.org/analysis/against-odds-overcoming-online-harassment-women-politics>
- 13 Rheault, Rayment and Musulan, 'Politicians in the line of fire'.
- 14 Elle Hunt, Nick Evershed and Ri Liu, 'From Julia Gillard to Hillary Clinton: online abuse of politicians around the world', *Guardian*, 27 June 2016, <https://www.theguardian.com/technology/datablog/ng-interactive/2016/jun/27/from-julia-gillard-to-hillary-clinton-online-abuse-of-politicians-around-the-world>
- 15 Inter-parliamentary Union, 'Sexism, harassment and violence against women parliamentarians', 2016 <http://archive.ipu.org/pdf/publications/issuesbrief-e.pdf>
- 16 Maggie Astor, 'For female candidates, harassment and threats come every day', *New York Times*, 2018, <https://www.nytimes.com/2018/08/24/us/politics/women-harassment-elections.html>; Frances Perraudin and Simon Murphy, 'Alarm over number of female MPs stepping down after abuse', *Guardian*, 31 October 2019,

- <https://www.theguardian.com/politics/2019/oct/31/alarm-over-number-female-mps-stepping-down-after-abuse>; Caroline Spelman, 'After seeing what I've been through, my children wouldn't follow me into politics', *The Times*, 31 October 2019, <https://www.thetimes.co.uk/article/after-seeing-what-ive-been-through-my-children-wouldnt-follow-me-into-politics-s78fhwx5k>; Jacob Jarvis, 'Heidi Allen received abusive email accusing her of 'killing a baby' before she resigned', *Evening Standard*, 30 October 2019, <https://www.standard.co.uk/news/politics/heidi-allen-received-abusive-email-accusing-her-of-killing-a-baby-after-she-had-an-abortion-a4274096.html>; Jess Aloe and April McCullum, 'Kiah Morris: the Vermont incidents that led to a black lawmaker's resignation', *Burlington Free Press*, 2018, <https://eu.burlingtonfreepress.com/story/news/2019/01/02/kiah-morris-vermont-incidents-led-black-lawmaker-resign/1990565002/>
- 17 Mariana Spring and Lucy Webster, 'A web of abuse: how the far-right disproportionately targets women', *BBC News*, 15 July 2019, <https://www.bbc.co.uk/news/blogs-trending-48871400>
  - 18 BBC Staff, 'Anna Soubry urges police action after "Nazi taunts" outside Parliament', 7 January 2019 <https://www.bbc.co.uk/news/uk-politics-46785357>
  - 19 BBC News, 'Woman jailed for life for attack on MP Stephen Timms', 3 November 2010, <https://www.bbc.co.uk/news/uk-england-london-11682732>
  - 20 Sentencing remarks of Mr Justice Wilkie, Judiciary of England and Wales, 23 November 2016, <https://www.judiciary.uk/wp-content/uploads/2016/11/sentencing-remarks-r-v-thomas-mair.pdf>
  - 21 Katie Rogers, 'Trump tells congresswomen to 'go back' to the countries they came from', *The New York Times*, 14 July 2019, <https://www.nytimes.com/2019/07/14/us/politics/trump-twitter-squad-congress.html>
  - 22 Women's Media Center, 'Online Abuse 101', <https://www.womensmediacenter.com/speech-project/online-abuse-101/>
  - 23 <https://www.statista.com/statistics/268136/top-15-countries-based-on-number-of-facebook-users/>
  - 24 'Most popular mobile social networking apps in the United States as of September 2019, by monthly users', Statista, <https://www.statista.com/statistics/248074/most-popular-us-social-networking-apps-ranked-by-audience/>
  - 25 Keywords selected for the automated system were: bitch; fuck off; fuck you; ugly; dirty; stupid; idiot; cow; witch; asshole; bastard; dick; dickhead; nasty; whore; slut; skank; cunt; son of a bitch; motherfucker; go to hell; douchebag; wimp; coward; moron; coon; traitor; un-American; negro; nigger; go\_back\_to; muslim; immigrant; terrorist; get out; fraud; scum; disgrace.
  - 26 Jason Wilson, 'Dripping with poison of antisemitism: the demonization of George Soros', *Guardian*, 25 October 2018, <https://www.theguardian.com/us-news/2018/oct/24/george-soros-antisemitism-bomb-attacks>
  - 27 For this research we used social media analytics tool CrowdTangle (<https://www.crowdtangle.com/>). The platform queries all posts from Facebook that are 'made by a public page, group, or (possibly) verified public person, who has ever (since 2014) had more than 110,00 likes, OR has ever been added to a CrowdTangle list by anyone, AND are posted without the poster aiming at a particular audience using Facebook targeting and gating tools (e.g. age-gating for alcohol pages, geo-gating if the content has country rights restrictions, targeting to women, etc.). There are no explicit restrictions on language or country.' More information is available at <https://help.crowdtangle.com/en/articles/3213537-crowdtangle-codebook>.
  - 28 As when we investigated levels of abuse of candidates on Twitter, we were not able to collect enough mentions of Jennifer Gonzalez on Facebook to be significant.
  - 29 Total count after the application of a relevancy classifier.
  - 30 For example, the most commented-on posts mentioning Marsha Blackburn were all uploaded between 1 and 5 July 2020 and referred to her criticism of the National Basketball Association (NBA) for its stance on China. Of the comments qualitatively analysed, only 40% referred to Blackburn and tended to agree with her remarks. Similarly, four of the five top posts collected for Tim Scott related to Scott's comment that President Trump's retweet of video showing a Trump supporter shouting 'white power' was 'indefensible'. While the cumulative comment count for Tim Scott was the third highest with 42,060 comments, no comments in our sample were classified as abusive.



- 
- 31 Reuters Staff, 'False headline claim: Ilhan Omar faces 40 years in prison and deportation', Reuters, 19 May 2020, <https://www.reuters.com/article/uk-factcheck-omar-prison-deportation/false-headline-claim-ilhan-omar-faces-40-years-in-prison-and-deportation-idUSKBN22V291>
  - 32 Facebook, 'Bullying and harassment', in Community Standards, 2020, <https://www.facebook.com/communitystandards/bullying>
  - 33 Twitter, 'Hateful conduct policy', 2020, <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy>
  - 34 Astor, 'For female candidates, harassment and threats come every day'; Perraudin and Murphy, 'Alarm over number of female MPs stepping down after abuse'; Spelman, 'After seeing what I've been through, my children wouldn't follow me into politics'; Jarvis, 'Heidi Allen received abusive email accusing her of 'killing a baby' before she resigned'; Aloe and McCullum, 'Kiah Morris: the Vermont incidents that led to a black lawmaker's resignation'.
  - 35 'Women in the US Congress 2020', Center for American Women and Politics, <https://cawp.rutgers.edu/women-us-congress-2020>
  - 36 Facebook, 'Bullying and harassment'.
  - 37 Twitter, 'Abusive behavior', 2020, <https://help.twitter.com/en/rules-and-policies/abusive-behavior>
  - 38 Twitter, 'Hateful conduct policy'.
  - 39 Shannon Bond, 'Facebook, YouTube warn of more mistakes as machines replace moderators', NPR, 31 March 2020, <https://www.npr.org/2020/03/31/820174744/facebook-youtube-warn-of-more-mistakes-as-machines-replace-moderators?t=1600436207439>
  - 40 Charlotte Jee, 'Facebook needs 30,000 of its own content moderators, says a new report', MIT Technology Review, 8 June 2020, <https://www.technologyreview.com/2020/06/08/1002894/facebook-needs-30000-of-its-own-content-moderators-says-a-new-report/>, Elizabeth Dwoskin, Jeanne Whalen and Regine Cabato, 'Content moderators at YouTube, Facebook and Twitter see the worst of the web — and suffer silently', Washington Post, 25 July 2019, <https://www.washingtonpost.com/technology/2019/07/25/social-media-companies-are-outsourcing-their-dirty-work-philippines-generation-workers-is-paying-price/>
  - 41 Llansó, COVID-19 content moderation research letter'.
  - 42 Demos et al., Algorithm Inspection and Regulatory Access.
  - 43 ISD, 'Extracts from ISD's submitted response to the UK Government Online Harms White Paper'.
-



Beirut | Berlin | London | Paris | Washington DC

Copyright © Institute for Strategic Dialogue (2020). Institute for Strategic Dialogue (ISD) is a company limited by guarantee, registered office address PO Box 75769, London, SW1P 9ER. ISD is registered in England with company registration number 06581421 and registered charity number 1141069. All Rights Reserved.

[www.isdglobal.org](http://www.isdglobal.org)