A REPORT FROM THE
GLOBAL PROJECT AGAINST HATE AND EXTREMISM

Democracies Under Threat

HOW LOOPHOLES FOR TRUMP'S SOCIAL MEDIA ENABLED THE GLOBAL RISE OF FAR-RIGHT EXTREMISM

Heidi Beirich, PhD | Wendy Via

GPAHE

GPAHE

The Global Project Against Hate and Extremism is a nonprofit organization devoted to building a diverse global community by exposing and countering racism, bigotry, and prejudice and promoting human rights that are central to flourishing, multicultural societies and democracies. We believe that white supremacy, hate, and far-right extremist movements are existential threats to societies and democracies around the globe. Extremists' hateful propaganda and actions don't stop at a country's borders and neither can those who work to stop it.

GLOBALEXTREMISM.ORG

@GLOBALEXTREMISM

CONTACT@GLOBALEXTREMISM.ORG

CONTENTS

ON THE COVER
Prominent far-right populists have built their constituencies and spread hate using social media, in particular on Facebook CEO Mark Zuckerberg's (top right) platform. Depicted clockwise from Zuckerberg are Indian Prime Minister Narendra Modi, Brazilian President Jair Bolsonaro, Dutch anti-Muslim politician Geert Wilders, former U.S. President Donald Trump and Filipino President Rodrigo Duterte.

photos/wikicommnons

This report contains offensive and potentially triggering language.

DEMOCRACIES UNDER THREAT

Executive Summary

THE DECISION BY MULTIPLE SOCIAL MEDIA PLATFORMS to suspend or remove ex-American President Donald Trump after he incited a violent mob to invade the U.S. Capitol on January 6, 2021, was too little, too late. Even so, the deplatforming was important and it should become the standard for other political leaders and political parties around the world that have engaged in hate speech, disinformation, conspiracy-mongering and generally spreading extremist material that results in real world damage to democracies.

For years, Trump violated the community standards of several platforms with relative impunity. Tech leaders had made the affirmative decision to allow exceptions for the politically powerful, usually with the excuse of "newsworthiness" or under the guise of "political commentary" that the public supposedly needed to see. For example, last year Facebook decided to allow a Trump tweet targeting social justice protesters that read "when the looting starts, the shooting starts." The tweet was cross-posted to Facebook and remained on the platform (Twitter took it down). Within days, the post had been shared over 71,000 times and reacted to over 253,000 times. The message was also overlaid onto a photo shared on Trump's Instagram account, which quickly received over half a million likes.

Why did Trump's clearly violative post stay up? Facebook CEO Mark Zuckerberg made the decision, though it was one criticized by many of his employees. "I disagree strongly with how the President spoke about this, but I believe people should be able to see this for themselves, because ultimately accountability for those in positions of power can only happen when their speech is scrutinized out in the open," was Zuckerberg's explanation.

Facebook in particular gives considerable latitude to public figures, codifying in its policies an exception that allows speech by political figures that violate its rules to stay up and prevents political ads from being fact-checked. The policy was created during the 2016 campaign specifically to allow hate and violative material posted by Trump to stay up. In the last year, Twitter, which has long allowed unfettered discourse, rethought its position stating that Trump will not be allowed back on the platform and began to sanction other political figures and political parties in the same way it deals with ordinary citizens. However, this change has not been wholesale and much violative content remains.

By effectively making a special deal for Trump, Zuckerberg and other social media leaders created a cascade of ever-changing policies that allow savvy politicians across the world to harm billions of people with polarizing messages that undermine democracies. It is, of course, the powerful whose hateful or false words have the greatest impact on public safety. Community standards are supposed to exist to protect users from online harm and the public from offline harms driven from the platforms. So creating exceptions (or really excuses) for those with the greatest ability to impact and mobilize people, sometimes into violence, is the exact opposite of what the social media companies say they are committed to doing.

And here's the thing, these exceptions aren't necessary for the many politicians across the globe who are advocating for inclusive policies and strong democracies. That's because the incendiary, divisive rhetoric that comes out of extreme politicians' mouths works well for online engagement, and therefore in ad buys. It's what sells. A recent New York University study found that far-right news services on Facebook consistently received the highest engagement of partisan groups and that frequent readers of far-right content engaged at a 65 percent higher rate. Further evidence that posting far-right content and misinformation pays off.

NOT JUST AN AMERICAN PROBLEM

This isn't just about Trump. It should be understood that authoritarian and far-right extremist politicians and political parties cannot build their constituencies without demonizing individuals and communities with vitriolic rhetoric. Social media is the most effective way to disseminate disparaging messages and disinformation that would never be hosted on other media. Anti-Black racism and the threat of immigrants, Muslims and others divide electorates while mobilizing the far right. Targets also include LGBTQ communities, Jews, the Roma, women, or any community that can be made to be seen as other and dangerous. A case in point is Brazilian President Jair Bolsonaro saying on a Facebook Live broadcast that Indigenous citizens were still "evolving and becoming" human beings. Facebook rejected a plea to remove the content as dehumanizing speech.

AUTHORITARIAN AND FAR-RIGHT EXTREMIST POLITICIANS CANNOT BUILD THEIR CONSTITUENCIES WITHOUT DEMONIZING INDIVIDUALS AND COMMUNITIES WITH VITRIOLIC RHETORIC. SOCIAL MEDIA IS THE MOST EFFECTIVE WAY TO DISSEMINATE DISPARAGING MESSAGES AND DISINFORMATION THAT WOULD LIKELY NOT THRIVE OTHERWISE.

As New Zealand Prime Minister Jacinda Arden emphasized after a white supremacist killed 51 people at Christchurch mosques, "there is no question that ideas and language of division and hate have existed for decades, but their form of distribution, the tools of organization—they are new." Trumpian and hard-line authoritative figures in several countries, as well as racist and xenophobic political parties, are using social media to sow hate and misinformation—and grow their base, radicalizing untold numbers into hateful and extremist ideas. Far-right parties are often early adopters of technology, and social media platforms have been essential to the rise of parties like the xenophobic Alternative for Germany (AfD).

Violence linked to far-right political figures' use of social media is not a matter of opinion. Experts in genocidal processes routinely highlight the power of political figures to incite violence against their perceived enemies. The Dangerous Speech Project has developed a model of the type of language, speakers and audiences that are most closely linked to mass violence and genocide. The model describes the particular danger of influential speakers with audiences susceptible to inflammatory messages because they are fearful or resentful. Also problematic are dehumanizing speech, coded language, attacks on women and the impression that members of a target group might damage the purity or cleanliness of the audience group. That is a good description of Trump, his online activities and those who made up the bulk of his audience. And it is true of other far-right leaders.

There is no doubt that Trump's and others' social media posts targeting marginalized communities have resulted in offline harm. For example, academic research has established the link between Trump's online speech and offline violence. One study directly tied Trump's anti-Muslim tweets with rising anti-Muslim sentiment and hate crimes. Since Trump's use of social media is similar to that of other authoritarians such as Bolsonaro or Philippine President Rodrigo Duterte, it is reasonable to assume that it has the same devastating impact in other countries.

It's not just individuals who are dangerous. Bigoted political parties, which also have mass reach online, have the same impact. The researchers who studied Trump's anti-Muslim posts found a direct correlation between social media posts by the far-right AfD party and hate crime in Germany. The data revealed that AfD's Facebook and Twitter posts against refugees, mostly Muslim, "show that right-wing anti-refugee sentiment ... predicts violent crimes against refugees in otherwise similar municipalities with higher social media usage."

HARM TO DEMOCRATIC SYSTEMS

Violative content by public figures is also harming inclusive democratic systems. Multiple studies have shown the retreat of democracy in dozens of countries, including the U.S. As citizens have moved away from strong support for democracy, far-right parties and authoritarian figures—many of whom made their way from the fringe by harnessing social media—have seen their ranks grow and their anti-democratic and extremist ideas mainstreamed.

In February 2020, U.N. Special Rapporteur Fernand de Varennes said, "The last decade has seen minorities around the world facing new and growing threats, fueled by hate and bigotry being spewed through social media platforms." He denounced the "banalization of bigotry," and the increasing "otherization and dehumanization of minorities through social media." The U.N.'s Rabat Plan of Action requires political leaders to refrain from any

incitement, to speak out firmly and promptly against hate speech and to never justify violence by prior provocation. This includes in the social media context. And the reasons are clear: leaders' words speak volumes compared to ordinary citizens. Unfortunately, many far-right political figures reject this advice.

The Carnegie Endowment for International Peace made this same argument while supporting Trump's permanent ban on Facebook, writing, "The argument that Facebook should permit political figures to post content even if it violates its community standards is a loophole that illiberal leaders are all too willing to exploit. In truth, it should be the opposite: public figures with a megaphone must have a greater responsibility to refrain from harmful speech. A speaker with elevated political status and a wide reach has a much higher capacity to incite violence and harm through their words."

This report documents how exceptions from social media platform policies given to far-right political figures and political parties is causing online and offline harm, mainstreaming hate and affecting government policies and spreading disinformation and conspiracy theories in eight countries and one region. In these cases, there is at least some documentation, largely by civil society actors and journalists; unfortunately, for most parts of the world, little to no research exists on how social media is impacting political systems. But the experiences examined here suggest that the harms will be found in other contexts as well, especially since adequate content moderation is even less likely to be comprehensive in areas outside those investigated here. This report shows that if platform policies are not applied equally, globally, and to everyone, democracies will continue to suffer and violence, including hate crimes and terrorism, will increase.

Key Findings

1. Social media companies have made exceptions for politicians and do not enforce their hate speech and fact-checking rules for political figures globally, allowing them to sow division and hate and build their audiences.
2. Social media exceptions for political figures and A.I. systems that target engagement have increased polarization and sown division in multiple societies.
3. Social media has been fundamental to the rise of far-right and authoritarian politicians and governments.
4. Social media companies lack cultural and language competencies to globally enforce their rules so that societies and democracies are not harmed.
5. There is little research in many parts of the world into online hate speech, disinformation, and abuses by far-right, biogted politicians, political parties and governments. As a result, little is known about the scope of these problems.

Recommendations

1. End newsworthiness and political commentary exceptions and apply all policies globally.
2. Apply fact-checking standards to political advertising.
3. Design and implement preventative genocide protocols.
4. Fashion A.I. systems so that the pursuit of engagement does not favor hate content, conspiracies, polarization and disinformation. Never monetize any content of this kind.
5. Conduct human and civil rights audits everywhere a platform is available. Particular attention should be paid before platforms are allowed to be used for political campaigns.

UNITED STATES

Trump Dictates the Terms

IN JANUARY 2021, THEN-PRESIDENT DONALD TRUMP was indefinitely suspended from Facebook, Instagram and YouTube and expelled from Twitter and Snapchat, with other social media platforms also taking steps to address his abuse of their services. The mass actions against Trump content came after it became clear that he had used his online megaphone to incite the white supremacists, neo-Nazis, militia members, QAnon adherents and others who stormed the U.S. Capitol on January 6, in an effort to disrupt the certification of the election.

As the riot unfolded, Trump defended them in a tweet, "These are the things and events that happen when a sacred landslide election victory is so unceremoniously & viciously stripped away." Hours earlier, Trump had posted a tweet attacking Vice President Mike Pence even as rioters, some of them chanting "Hang Mike Pence," came within striking distance of him as he was evacuated from the senate chamber.

That finally appeared to be a step too far for social media. But the fact that it took an actual insurrection, planned and encouraged on the companies' own services, to get Facebook, Twitter, et al., to move is unbelievably discouraging. And even then, Facebook COO Sheryl Sandberg tried to deflect responsibility, telling Reuters, "I think these events were largely organized on platforms that don't have our abilities to stop hate and don't have our standards and don't have our transparency."

LIKE MOST SOCIAL MEDIA PLATFORMS, FACEBOOK IS AN AMERICAN COMPANY, AND IT CHOSE TO ABANDON ITS OWN POLICIES TO GIVE AN AMERICAN, TRUMP, UNFETTERED ACCESS TO ITS PLATFORM.

The damage Trump did with his online activity became much clearer when he was gone. According to Zignal Labs, online misinformation about election fraud plunged 73 percent in the weeks following Twitter's decision to ban Trump on January 8. Other forms of online misinformation also plummeted. Zignal found, "Mentions of the hashtag #FightforTrump, which was widely deployed across Facebook, Instagram, Twitter and other social media services in the week before the rally, dropped 95 percent. #HoldTheLine and the term 'March for Trump' also fell more than 95 percent." It helped that some of Trump's enablers were also deplatformed.

HOW DID WE GET HERE?

For years, Trump used social media platforms to spread hate, disinformation and conspiracy theories. Facebook is a particularly bad actor in this mess. About a quarter of Trump's 6,081 Facebook posts in 2020 contained extremist rhetoric, misinformation about the coronavirus, the election or his critics, according to an analysis by Media Matters for America. The research determined that Facebook basically failed to limit the reach of, or block, Trump's propaganda, which was shared and liked more than 927 million times.

Like most social media platforms, Facebook is an American company, and it chose to abandon its own policies to give an American, Trump, unfettered access to its platform. Its "newsworthiness" exception for political figures was specially created in the lead up to the 2016 election to allow violative hate material posted by Trump to stay up.

An early case was candidate Trump's announcement of the Muslim Ban in a video posted in 2015. Many Facebook employees found the video to violate the company's community standards, but its executives made a contorted decision to allow the video to stay up. Monika

After George Floyd's tragic murder at the hands of police on May 25, 2020, Trump threatened social justice protesters from his social media accounts. Facebook did not take the post down.

JOSH MUN / ALAMY STOCK PHOTO

Bickert, Facebook's vice president for policy, said the company kept the video up because executives interpreted Trump's comment to mean that he was not speaking about all Muslims, but rather advocating for a policy position on immigration as part of a newsworthy political debate. Over time, this loophole for Trump became big enough to drive a truck through.

Facing a barrage of complaints about Trump's violations, Facebook stuck to its guns. In 2019, Nick Clegg, then the newly hired head of global affairs and communications and a former British deputy prime minister, repeated Facebook's position that politicians would not be held to account on the platform. Claiming that, aside from speech that causes violence or real-world harm, which seemingly no longer included hate speech, Facebook would allow politicians to express themselves virtually unchecked on social media. Facebook's network of independent fact-checkers, which had been established as a key part of the company's response to disinformation, would not evaluate their claims and the community guidelines would largely not apply to politicians. Facebook did not want to be an "arbiter of truth."

One former executive, Yael Eisenstat, who worked to improve the political ads process, wrote in 2019 that the controversy over allowing lies in political advertising was "the biggest test of whether [Facebook] will ever truly put society and democracy ahead of profit and ideology." Trump's ads were notable for disparaging comments about his opponents, calling Senator Elizabeth Warren "Pocahontas" and House Speaker Nancy Pelosi "a liar and a fraud." Ads about immigration used especially dark rhetoric and imagery, stoking fears of "caravan after caravan" of migrants coming to the U.S. or urging voters to vote yes or no on whether to "deport illegals." His video ads featured Trump warning that democrats were "openly encouraging millions of illegal aliens" to "destroy our nation."

Facebook's cover for Trump has had multiple, negative effects. It has stymied the company's efforts against disinformation and misleading news and allowed conspiracy theories to proliferate. The company even altered its news feed algorithm to neutralize false, but insistent, claims that it was biased against conservatives. That latter decision warped the platform fundamentally, pushing Facebook into more deferential behavior toward its growing number of right-leaning and extreme users, tilting the balance of news people see on the network.

It got worse in 2020 as Trump ramped up his election rhetoric. In late April, he tweeted a series of posts against the COVID lockdowns, reading "LIBERATE MINNESOTA," as well as other states. This began a right-wing backlash that led extremists into the streets to protest the pandemic measures. For extremists in militias and white supremacist groups, Trump's tweets were a license to riot. Just weeks after protests erupted in Minneapolis in the wake of George Floyd's murder at the hands of police on May 25, 2020, Trump used his social media megaphone to post, "Any difficulty and we will assume control but, when the looting starts, the shooting starts." This phrase was used by a racist Miami police chief in the 1960s and has been widely interpreted as a violent threat against protesters. Twitter quickly hid the post for glorifying violence, as it had done for countless of Trump's lies related to the election at this point.

ANY DIFFICUTLY AND WE WILL ASSUME CONTROL BUT, WHEN THE LOOTING STARTS, THE SHOOTING STARTS. THANK YOU!

@REALDONALDTRUMP, MAY 29, 2020
BANNED FROM TWITTER;
KEPT BY FACEBOOK

Facebook chose a different path, ignoring its rules that bar speech that inspires or incites violence. The company decided to allow Trump's tweet, which was cross-posted to Facebook, to remain on the platform. Within days, it had been shared over 71,000 times and reacted to over 253,000 times. The message was also overlaid onto a photo shared on Trump's Instagram account, which quickly received over half a million likes.

Why did Trump's clearly violative post stay up? Facebook CEO Mark Zuckerberg, in a decision criticized

by more than 5,000 of his employees, made the call against advice of staff. And it reportedly came after a personal phone call from Trump. "I disagree strongly with how the President spoke about this, but I believe people should be able to see this for themselves, because ultimately accountability for those in positions of power can only happen when their speech is scrutinized out in the open," was Zuckerberg's explanation. Facebook also decided to leave up Trump posts that spread misinformation about voting by mail.

By the time Facebook's own civil rights auditors issued their final report in July 2020, the damage of the Trump loopholes was clear. The auditors found that Facebook's moderation of Trump's use of the platform has undermined its broader civil rights efforts, singling outposts that lied about mail-in ballots or incited violence, all of which Facebook allowed to stand.

"These decisions exposed a major hole in Facebook's understanding and application of civil rights," the auditors wrote. "While these decisions were made ultimately at the highest level, we believe civil rights expertise was not sought and applied to the degree it should have been and the resulting decisions were devastating. Our fear was (and continues to be) that these decisions establish terrible precedent for others to emulate." It was later disclosed that Zuckerberg was involved in the decision to leave the posts up.

In 2020, Trump's abuse of the platforms exploded. He used Facebook and other platforms to tout misleading information about coronavirus cures, election fraud and the motives of protesters, frequently and falsely targeting antifa as a cause of violence (in fact, most violence, up to and including murder, during the social justice protests, was conducted by far-right extremists). His claims that the election was stolen, which spread like wildfire across Facebook Stop the Steal groups, came to an explosive conclusion in the January 6 insurrection.

Trump has lost much of his online presence, but YouTube has said that his channel will be restored when the "risk of violence passes," and he may regain his Facebook account. On January 21, the company forwarded its decision to its new Oversight Board, which is expected to rule within 90 days.

TRUMP AND TWITTER

Twitter was just as lax as Facebook in terms of letting Trump post whatever he wanted for many years. The list of lies, conspiracies, threats and hate Trump put up on his Twitter account is long, but there is a difference. As Twitter began to enforce its policies against everyone starting in mid-2020, the company did not create loopholes for Trump, repeatedly labeling his posts

Social media companies finally took action against Trump after he incited the Jan. 6 Capitol insurrection from his online accounts.

DIEGO MONTOYA / ALAMY STOCK PHOTO

about voting and the election as untrue. And when it suspended his account on January 8, the company made clear it would be permanent.

But there was still considerable damage in terms of Trump spreading hate and misinformation starting right from when he was a candidate. In 2015, he tweeted out a false chart that claimed that 81 percent of white murder victims are killed by black people, a white supremacist talking point. The fake statistics were first posted by a neo-Nazi Twitter account. In November 2017, Trump retweeted three inflammatory and unverified anti-Muslim videos from Britain First, a racist group that was banned by the U.K. government. One of the videos purported to show an assault by a Muslim immigrant, but the assailant was neither Muslim nor an immigrant. Trump's promoting inflammatory content from an extremist group was without precedent among modern American presidents. Trump's sharing of the tweets was praised across far-right circles, increased anti-Muslim content on social media and elevated the profile of Britain First.

On July 2, 2017, Trump tweeted a video of himself attacking Vince McMahon during a WrestleMania event, but altered the video to place the CNN logo over McMahon's face. News reporters rightly took Trump to task, including CNN's Brian Stelter, who said Trump was "encouraging violence against reporters" and "involved in juvenile behavior far below the dignity of his office." Trump subsequently said that CNN took the post too seriously, adding that CNN has "hurt themselves very badly."

In August 2018, Trump tweeted that he had asked his secretary of state to "closely study the South African land and farm seizures and expropriations and the large scale killing of farmers," another white supremacist talking

point. South Africa's Minister for International Relations and Cooperation rebuked Trump, saying he was expressing "right-wing ideology" and added that the South African government had requested an explanation from the U.S. embassy, which did not defend Trump's tweet. There are no reliable figures that suggest that white farmers are at greater risk of being killed than the average South African.

In July and August 2019, Trump retweeted anti-Muslim British bigot Katie Hopkins. Among other things, Trump retweeted Hopkins' attack on London mayor Sadiq Khan in which she blamed him for the city's violent crime rate. Twitter permanently deleted Hopkins' account in June 2020 for violating its "Hateful Conduct" policy. In 2020, violence became more obvious in Trump's tweets. That May, Trump retweeted a video in which one of his supporters, Couy Griffin, a New Mexico county commissioner and founder of "Cowboys for Trump," said, "The only good Democrat is a dead Democrat." A day later, Trump tweeted the infamous, "When the looting starts, the shooting starts," which importantly was flagged by Twitter as "glorifying violence."

Twitter has now rid itself of the Trump problem, and most companies are rethinking political exceptions. But not Facebook. The company insists the use of incendiary populist language predates social media, so its spread is unrelated to Facebook. This position completely ignores how Facebook has manipulated the online space in favor of extremism and how political abuse of social media has altered the American political landscape.

BRAZIL

Facebook's WhatsApp Gets a Violent, Bigoted President Elected

IN 2018, JAIR BOLSONARO, RIGHT-WING APOLOGIST for Brazil's military dictatorship and former Army officer, was elected as Brazil's president after a coordinated, deceptive campaign run largely from Facebook-owned WhatsApp. Bolsonaro previously served in various elected positions representing the state of Rio de Janeiro, while advancing a far-right, bigoted agenda including vociferous opposition to the LGBTQ community, women's equality, affirmative action and secularism.

Bolsonaro has made his bigoted views plain, at times with violent rhetoric. In a 2002 interview, Bolsonaro told a Brazilian newspaper, "If I see two men kissing in the street, I will beat them." He publicly defended beating gay children by saying, "If your child starts to become like that, a little gay, you take a whip and you change their behavior." In a 2013 BBC documentary, Bolsonaro said "we Brazilians do not like homosexuals."

His views on women are equally disturbing. In a 2017 speech, Bolsonaro said he had five children, that the first four were male, and that for the fifth, he produced a daughter out of "a moment of weakness." In a 2014 Congressional debate, Bolsonaro said that minors should be treated as adults if they commit heinous crimes such as murder or rape, to which Human Rights Minister Maria do Rosário responded by calling him a "rapist." Bolsonaro then stated that Rosário was "not worth raping; she is very ugly."

Often referred to as the "Trump of the Tropics," Bolsonaro is an open admirer of Trump and employed similar campaign tactics, including a far-right agenda, hardline attacks on opponents and use of incendiary rhetoric on social media. And like Trump, this incendiary rhetoric is correlated with increasing hate crimes. A survey conducted by Gênero e Número tracked violence against LGBTQ people during and after Brazil's 2018 presidential campaign. It found that over 50 percent of respondents suffered from some form of violence due to their sexual orientation. At least 92 percent claimed that such violence increased following Bolsonaro's election.

BOLSONARO'S RISE TO THE PRESIDENCY WAS PROPELLED BY SOPHISTICATED AND COORDINATED MISINFORMATION CAMPAIGNS RUN FROM SOCIAL MEDIA.

WHATSAPP CENTRAL TO BOLSONARO'S ELECTION

Bolsonaro's rise to the presidency was propelled by sophisticated and coordinated misinformation campaigns run from social media. An investigation by the Brazilian newspaper, *Folha,* revealed that days before the late October runoff between Bolsonaro and his opponent, the Worker Party's Fernando Haddad, a conservative Brazilian business lobby bankrolled a multi-million-dollar smear campaign in which Bolsonaro supporters delivered daily misinformation through WhatsApp to millions of Brazilians' phones. This material included doctored photos, audio clips manipulated to misrepresent Haddad's policies and fake "fact-checks" discrediting authentic news stories.

Aos Fatos, a Brazilian fact-checking organization, analyzed WhatsApp misinformation from the election and found more than 700 false or misleading posts being shared. These rumors distorted at least four types of information: statements by political candidates, news on electronic voting and legislation, the nature of protests and the outcomes of opinion polls.

The messages were largely intended to reach

right-leaning political groups, Catholic and evangelical churches, trade and business associations and military groups.

Targeting WhatsApp users was strategic because it is an essential communication tool in Brazil, where it is used by about 120 million of Brazil's 210 million citizens. It has become endemic because mobile phone companies offer free access in prepaid mobile-internet plans to specific applications, usually Facebook, WhatsApp and Twitter. This means that most Brazilians have unlimited social media access, but they have to pay to use other aspects of the internet. As a result, 95 percent of all Brazilian internet users say they mostly use messaging apps and social media when online.

The damage to the election was considerable given the wide reach of these free apps. According to polls conducted a few days before the first round of the presidential elections, a staggering 87 percent of users claimed they had received fake news via WhatsApp. A Brazilian university study determined that out of the top 50 images circulating in political WhatsApp groups during the first round, only four were real, the rest being hoaxes or mass distortions.

The fallout of *Folha's* report on these machinations was significant and WhatsApp issued an apology. "Every day, millions of Brazilians trust WhatsApp with their most private conversations," wrote its vice president in *Folha*. "Because both good and bad information can go viral on WhatsApp, we have a responsibility to amplify the good and mitigate the harm." The company announced it would purge thousands of Brazilian spam accounts, label messages to show that they had been forwarded, tighten rules on group messaging and partner with Brazilian fact-checking organizations to identify false news.

Of course, by that time, the damage was done. Bolsonaro won and built a constituency for his extreme views. Today, social media in Brazil is very much a battleground in which politicians leverage their followers for political advantage, and the vast majority of the offensive content in circulation is propagated by individuals with extreme right sympathies.

A Joint Parliamentary Inquiry Committee eventually determined that there had been a "hate cabinet" coordinating these operations run by Carlos Bolsonaro, one of the president's sons. This hate cabinet oversaw a sprawling network of blogs and social media profiles actively spreading disinformation and threatening opponents using YouTube, Facebook, WhatsApp and Instagram. It took until July 2020 for Facebook to finally act against the coordinated pro-Bolsonaro activity on the platform that had distorted and undermined the 2018 election. That month Facebook removed dozens of accounts for "coordinated inauthentic behavior."

BOLSONARO'S PRESIDENCY AND SOCIAL MEDIA

Bolsonaro leverages social media to press his extremist agenda and bigoted beliefs. He tweets aggressively, streams weekly Facebook Live videos and posts content on his YouTube channel which has more than 3 million subscribers. By creating his own online media empire, he has been able to present his abhorrent views unfiltered to the public owing to the platforms' loopholes for politicians, particularly Facebook and YouTube.

Bolsonaro has also been a font of coronavirus lies on social media. He accused the W.H.O. of being a "partisan political organization" and "one of the least scientific" organizations in the world. He mentioned hydroxychloroquine as a coronavirus cure in 13 of 14 live broadcasts on YouTube and Facebook monitored by Vanessa Barbara from June to September 2020. In late 2020, he told Brazilians not to deal with COVID-19 like "a country of fags" (Brazil has the second-highest death toll worldwide after the U.S.).

Bolsonaro also lies about the environment and the Amazon. In 2019, he claimed that the fires in the Amazon were a fake news story created by Brazilian newspapers and propagated by foreign media. If he admits to a fire outbreak, Bolsonaro blames Indigenous people, calling them "caboclos" (people of mixed Indian and white origin) and riverside dwellers. "It's their culture," he says.

Until 2020, Bolsonaro was allowed to post at will across platforms, routinely engaging in hate speech and disinformation. In March 2020, after egregious amounts of disinformation about the pandemic had been spread by Bolsonaro, Facebook finally acted. This required Facebook to abandon its policy of not fact-checking political figures in order to prevent the spread of potentially harmful coronavirus misinformation. Facebook removed a video shared by Bolsonaro where he claimed, without evidence, that "hydroxychloroquine is working in all places." Twitter also removed two posts, and YouTube pulled two videos from Bolsonaro's official account for violating its policies.

FACEBOOK FACES SUPREME COURT ORDER

In August 2020, Facebook finally complied with an order by Brazil's Supreme Court to block access worldwide to the accounts of a dozen of Bolsonaro's top allies. The group is accused of spreading fake news against judges. This move came after a May decision by the Supreme Federal Tribunal ordering the block of 16 Twitter accounts and 12 Facebook accounts connected to Bolsonaro supporters who allegedly violated laws on hate speech. The accounts were also linked to

the spreading of fake news during Brazil's 2018 election. Facebook claimed the measure was a threat to freedom of speech and said it would appeal the order, saying in a statement that the order was extreme, "conflicting with laws and jurisdictions worldwide."

Once Trump was banned from most social media after the January 6 Capitol insurgency, Bolsonaro was quick to realize he might be next. He urged his followers to move with him to Telegram, an app infested with neo-Nazis, where he set up his own channel.

GERMANY

Facebook Fuels an Anti-Muslim Party's Rise

THE ALTERNATIVE FUR DEUTSCHLAND (AfD), the most far-right political party to enter the German parliament since the Nazi era, has social media to thank for its rise. As in many other countries, this openly racist and xenophobic political party was able to harness the online space, and Facebook in particular, to grow its ranks and push its dangerous messaging right into the heart of German politics.

Launched in 2013, the AfD is rabidly anti-Muslim, anti-refugee and anti-LGBTQ. When German Chancellor Angela Merkel opened Germany's borders in 2015 to more than a million asylum seekers fleeing the Syrian war, AfD harnessed anti-Muslim and anti-refugee sentiment to propel its growth. In 2016, the party adopted a specific anti-Muslim position that, "Islam is not a part of Germany," even though the country has nearly 2 million Muslim citizens. The party is closely linked to other extremists including neo-Nazis, anti-Muslim movements and white supremacist Identitarians.

AFD'S RISE FROM A FRINGE PARTY IN 2013 TO AN INCREASINGLY EXTREMIST FORCE TO BE RECKONED WITH IS DEEPLY TIED TO THE PARTY'S HARNESSING OF SOCIAL MEDIA.

Predominantly using Facebook, the AfD has spread its bigoted messaging while avoiding questions from mainstream press about its policies and beliefs. One year after its 2014 founding, the party won seven seats in the European parliament elections. In 2017, the AfD gained seats in 14 of 16 German state-level parliaments. In October 2017, it became the first far-right party to be elected to the Bundestag in over half a century, becoming the third-largest party with 94 seats.

Since 2017, the AfD has been increasingly open to working with far-right extremist groups, in particular the anti-Muslim Pegida movement. The extremism of one of the party's factions, Der Flugel, or The Wing, led the Federal Office for the Protection of the Constitution to place The Wing under surveillance in March 2020 as "a right-wing extremist endeavor against the free democratic basic order" that is "not compatible with the Basic Law." One year later, the entire party was put under surveillance, a decision that is currently under review by the courts. The head of the office called its leaders

On Aug. 30, 2020, neo-Nazis, QAnon supporters, and other far-right extremists tried to storm Germany's parliament building. Elected officials blamed the Alternative for Germany for mobilizing the mob.

COLIN MCPHERSON / ALAMY STOCK PHOTO

"right-wing extremists."

The demonization of Muslims and refugees driven by AfD messaging has had real world harms in addition to making the party electorally successful. In February 2020, 11 people were killed and five others wounded in a shooting spree by a far-right extremist targeting two shisha bars in Hanau. Though the attack was conducted by a lone actor, several commentators quickly pointed out that the AfD had helped poison the discourse around immigrants. "One person carried out the shooting in Hanau, that's what it looks like," the head of the Social Democratic Party, Lars Klingbeil, told broadcaster ARD, "but there were many who provided him with the ammunition, and the AfD is definitively among them."

Besides the Hanau attacks and additional mosque plots, Muslims have faced substantial hate violence. An analysis by the Left Party concluded that, every other day in 2019, a mosque, a Muslim institution or a religious representative was targeted by an anti-Muslim attack. At least 15 mosques were attacked between April and June of 2020 and dozens of Muslims were physically assaulted or verbally harassed.

FACEBOOK DRIVES AFD GROWTH

In Germany, Facebook has a 65 percent market share compared to Twitter's 21 percent and is the primary driver of toxic and divisive content. Facebook also drives online news coverage in Germany. In 2019, more than a quarter of German adults reported getting their daily news on Facebook.

AfD's rise from a fringe party in 2013 to an increasingly extremist force to be reckoned with is deeply tied to the party's harnessing of social media. Starting in 2016, the AfD built up a large following on both Facebook and Twitter by sharing a high volume of sensationalist tweets and posts. For example, shortly after the August 2017 Islamic State terrorist vehicle attack in Barcelona, the AfD posted a picture of bloody tire marks with the headline: "Mrs. Merkel, the victims of your political rampage are not forgotten! But how many have to die before you understand?" An analysis of the AfD's material found that the party's inflammatory postings were far more popular online than those of other political parties. To hone its digital media operation, the AfD hired Harris Media, an Austin, Texas-based firm that works with far-right candidates including Trump, Marine Le Pen's National Front party in France and Israeli Prime Minister Benjamin Netanyahu.

Soon after landing in Berlin in early September 2017, Harris' vice president for content production, Joshua Canter, went to a meeting at Facebook's Berlin offices.

Canter's assignment was to use digital ads to micro-target Germans whose backgrounds made them likely AfD converts. The meeting included the company's Berlin advertising staff and Sean Evins, the head of politics and government for Europe, the Middle East, and Africa, who pitched Canter on using Facebook Live. Canter explained to Bloomberg that he used the AfD's 300,000 Facebook likes to target millions of other Germans who might be receptive to the party's message using Facebook's lookalike audiences tool. That process generated a new group of 310,000 people who were most similar to AfD fans. A key to Canter's strategy was introducing negative campaign themes about Merkel linked to a website featuring a flickering image of Merkel's face and a counter displaying the number of people killed or injured by terrorists in Germany. The AfD tried to buy Google ads for "Angela Merkel" to drive traffic to the site, but Google demurred.

Research showed that half of the retweeted messages during the 2017 campaign were about the AfD, and its Facebook posts were shared five times more than those of any other party. Research by Bavarian academics made clear that the AfD's social media tactics were central to its rise. The evidence also indicated that automated accounts contributed to AfD's online superiority. When the election results came in, the AfD won 12.6 percent of the vote—more than double the five percent needed to claim seats in the Bundestag—making it the third-most-popular party in the country. Merkel won reelection, but it was her party's worst result since 1949. The German government had thought Russians would help the AfD, but the AfD's foreign assistance was American.

In advance of the European Union parliamentary elections in 2019, the AfD's messaging was enhanced by bots

and a large, dense network of suspect accounts promoting AfD Facebook posts. AfD maintained 1,663 Facebook pages, more active pages than all the other German political parties combined. Its content was shared between five and seven times more than all the other parties together and received four times the comments. Research showed that the AfD shared content from several sources accused of misrepresentation or outright misinformation such as tichyseinblick.de, epochtimes.de and jungefreiheit.de. In addition, a network of roughly 200,000 accounts liked or promoted AfD pages and content. The densely networked accounts engaged in what appeared to be coordinated behavior.

HUNGARY, POLAND AND THE BALKANS

Social Media Benefits Illiberal Democracies

POLAND AND HUNGARY HAVE A LOT IN COMMON, and not in a good way. They're both former Eastern bloc countries with illiberal democracies whose leaders curtail a free press, interfere with the judiciary, claim that Christianity is under attack, demonize migrants and LGBTQ communities, rewrite their Nazi-era history, and push antisemitic conspiracy theories and tropes. In both countries, social media has been key to the rise of illiberal leaders. And now, leaders in both countries have had a visceral reaction to the deplatforming of Trump, as they've followed his online playbook.

HUNGARY // Ruling Party Weaponizes Facebook Against Opponents

Prime Minister Viktor Orbán, congratulated by Trump for doing a "tremendous job" for his anti-immigrant policies and "putting a block up" to protect Christian communities, is now—in effect—Hungary's elected dictator.

Since his first election in 2010, Orbán has ruthlessly taken apart Hungary's democracy, bit by bit. He's undermined elections, stacked the courts with his allies and taken control of more than 90 percent of the country's media. Amidst the coronavirus pandemic, he had Parliament pass a new law that allows him to rule by decree, with no end date, seized and redistributed pubic funding meant for opposition political campaigns and passed a law that makes "spreading a falsehood" a crime, punishable by up to five years in prison. This law was almost immediately abused when two men were arrested for Facebook posts deemed critical of Orbán. In May 2020, Freedom House downgraded Hungary from a "semi-consolidated democracy" to a "hybrid regime."

Orbán and his Fidesz Party have spread hateful rhetoric and implemented oppressive policies and laws regarding immigrants, Muslims, women, Jews and LGBTQ people. During the 2015 Syrian refugee crisis, he shut the borders completely in the name of protecting

Hungary's Christian democracy. Orbán has a blatantly racist "zero tolerance" policy toward immigration, saying these measures are necessary to "ensure the survival of the Hungarian nation," and has denounced the EU for what he says is its desire to fill up Europe with Muslims. "Those who decide in favor of immigration and migrants, no matter why they do so, are in fact creating a country with a mixed population" and the left-wing is "the gravedigger of nations, the family and the Christian way of life."

In December 2020, Orbán's government voted to limit adoption to married heterosexual couples. Exceptions can be made for single parents but only with the approval of the family affairs minister, effectively halting adoptions by LGBTQ parents. The constitution was also amended to make it clear that only "traditional" households are families. "The mother is a woman, the father is a man. ... Hungary protects the institution of marriage ... between a man and a woman, as well as the family as the basis for the survival of the nation." Hungary allows only civil unions for same-sex couples. Parliament had already "banned legal gender recognition" which prevents transgender and intersex people from changing their gender or assigned birth sex.

This September, despite the near collapse of Hungary's educational system during the pandemic, Orbán will be introducing a new national curriculum which will make antisemitic authors mandatory reading and see history books rewritten to downplay any Hungarian involvement in the Holocaust, pushing instead a narrative of pride in the nation. This follows the forced relocation to outside the country of most of the operations of Budapest's Central European University, founded by philanthropist George Soros, so that the government could exert more political influence on the Hungarian Academy of Sciences.

FACEBOOK IS THE PREFERRED OUTLET

All of this, all the destruction of democracy and a free and open society, has been manipulated through coordinated social media campaigns, specifically on Facebook. Orbán leads Hungarian politicians with over one million Facebook followers. Other Fidesz leaders, including the Justice Minister Judit Varga and Foreign Minister Peter Szijjarto, are active there as well. Hungarian politicians are generally prolific on Facebook, and to a lesser degree, Twitter, using social media to campaign and make major policy announcements.

Facebook was absolutely crucial to the Fidesz win in the 2018 elections when Orbán was re-elected and Fidesz gained a two-thirds majority in Parliament. According to the *Budapest Beacon*, itself a casualty of Orbán's takeover of independent media, Fidesz operated the "mother of all activist networks," reportedly developed after carefully studying Trump's 2016 campaign, requiring all Fidesz MPs and candidates to submit names of colleagues who could be turned into "social media soldiers." These people were then summoned to Fidesz headquarters where they were met by Fidesz vice-president and campaign manager Gábor Kubatov and subjected to a day-long training on Facebook fan-pages and Fidesz's internal online network. This online network is a messaging system by which all candidates, political associates and social media volunteers managing the Facebook pages of Fidsz candidates can receive and assign tasks.

FIDESZ OPERATED THE "MOTHER OF ALL ACTIVIST NETWORKS" DEVELOPED AFTER CAREFULLY STUDYING TRUMP'S 2016 CAMPAIGN.

Every day, Fidesz headquarters sent out directives ranging from the sharing of posts to "occupying" the comment section of a post by an opposition candidate and spreadsheets of recommended daily messages. This national network of volunteer social-media activists was required by party headquarters to follow every command word for word, and warnings were issued to activists who did not follow orders. Once the directives were issued, the social media soldiers then passed them to local activists, pro-Fidesz NGOs and other allies reportedly through private Facebook groups which were secretly monitored by a Fidesz staffer. All of the party's Facebook statistical information was carefully reviewed on a monthly basis.

Now with the 2022 elections looming, Hungary's failing economic performance during the pandemic and its ailing healthcare sector are "Fidesz' weaknesses," says Andras Biro-Nagy, who heads the Budapest Policy Solutions research institute. Current polls show Fidesz neck and neck with the opposition alliance, and the banning of Trump on social media is cause for real concern for Orbán and the Fidesz party. After the ban, Fidesz Justice Minister Judit Varga lashed out against Facebook, accusing it of having "secretly and for political reasons" partially blocked access to her profile page. After claiming "tech giants can decide elections," Varga accused social media platforms of "reducing the visibility of conservative, right-wing views."

Thus the rush to introduce legislation that would prevent Facebook and others from deplatforming Fidesz

politicians, although there is no evidence that Facebook is contemplating such a move. Varga plans to introduce the regulatory legislation in the spring of 2021.

Orbán and Fidesz have a tight control on the media in Hungary, but social media has been a wild west. Facebook is Hungary's most popular site with more than 5.4 million users out of a 9.8 million population. Agostan Mraz of the Nezopont think-tank has said, "Orbán long ago realized how important media regulation is in politics, now it's social media's turn." Fidesz also significantly outspends opposition parties on Facebook advertising. Fidesz's Kubatov has said that social media "has taken the leading role (from television) in political campaigns."

In an effort to maintain its Facebook dominance, Orbán and Fidesz have a new program for the 2022 elections and hope to recruit more young people into the movement, according to reporting by *DW*. In a video produced by Megafon.hu, a young woman tells the viewers that most Hungarians have conservative beliefs and that those beliefs shouldn't just be shared at home, but also on Facebook. The video offers trainings that will turn the viewer into a "professional Facebook warrior."

Megafon's founder, Istvan Kovacs, insists that the company is a privately funded nonprofit, but Kovacs has close ties to Fidesz and learned from the Trump campaign and the U.S. how to use social media. He's also said, "Facebook will determine" the outcome of the 2022 elections and that "we need to outdo the left."

POLAND // Legislating Against Community Standards

In response to Twitter banning Trump and Facebook and YouTube suspending him, Poland's ministry of justice has introduced legislation that would make it illegal for the social media companies to delete content or accounts that do not violate Polish law, even if they violate the companies' community standards. Companies could face fines up to $13.4 million dollars if they do not restore the content upon order of the government.

The Polish government is keenly aware that it has much to lose if it's not able to use social media. "These politicians are able to galvanize more support on social media than mainstream politicians and parties have managed," says Ralph Schroeder of the Oxford Internet Institute. "The reason is that social media gives them a means to express ideas that cannot be expressed in traditional news media or in traditional party affiliations."

MARGINALIZED COMMUNITIES AND MINORITIES IN POLAND HAVE BEEN UNDER FEROCIOUS ATTACK ... WITH AN ESCALATION IN THE MONTHS LEADING UP TO AND SINCE THE JULY 2020 ELECTIONS.

Rafal Pankowski, co-founder of the Polish anti-racism group 'NEVER AGAIN' Association, said this legislation could set "a dangerous precedent internationally. One might expect other nationalist and authoritarian governments ... to act similarly, in order to protect the hate speech against minorities that has so often led to violence."

Marginalized communities and minorities in Poland have been under ferocious attack in recent years, with an escalation in the months leading up to and since the July 2020 elections. They're faced with a polarizing social climate and a government that is backsliding on democratic protections, with displays of open disdain for the LGBTQ community, women's rights and the Jewish community, among others. This is happening in public venues, on television and especially on social media such as Facebook, Twitter and YouTube.

Jaroslaw Kaczynski, leader of the governing Law and Justice party, has repeatedly told supporters that Poles will not be forced "to stand under the rainbow flag," and that homosexuality is a "threat to Polish identity, to our nation, to its existence and thus to the Polish state." Nearly 100 Polish towns and communities have declared themselves to be "LGBT-Free Zones."

It is in this climate that the current president, Andrzej Duda, who was endorsed by Trump, was narrowly reelected in July 2020 after running on a right-wing populist platform rife with antisemitism and anti-LGBTQ policies and rhetoric. Implying that his opponent would be influenced by Jewish special interests, Duda vowed that he would never sign a bill which "treat[s] one ethnic group more favorably than others." This was in reference to discussions about property restitution of the nearly 3 million Polish Jews who were killed in the Holocaust.

The campaign was marked by vicious attacks claiming that the LGBTQ rights movement is "worse than communism" and that "LGBT are not people." He

vehemently opposed marriage equality, officially proposed a ban against adoption by same-sex couples, and signed a declaration to help families by "protecting children from LGBT ideology," with a ban on "propagating LGBT ideology in public institutions."

On August 7, 2020, there were demonstrations in Warsaw and other cities after the arrest of Margot, a renowned LGBTQ rights activist in Poland, along with two others. Police brutally arrested dozens during the demonstrations, including peaceful protesters and bystanders. There are allegations of beatings, people being held without charges, and invasive strip searches including a body search of a trans woman performed by a man.

In September 2020, Przemyslaw Czarnek also of the Law and Justice Party was appointed minister of education and science. Czarnek has 29,000 followers on Twitter and 15,000 on Facebook, and has previously said, "Let us defend the family against this kind of corruption, depravity, absolutely immoral behavior, let us defend us against the LGBT+ ideology and finish listening to this idiocy about human rights or equality. These people are not equal to normal people, let's end this discussion," and "There's no doubt, that LGBT+ ideology grew out of... the same root as Germany's Hitlerian National Socialism, which was responsible for all the evil of World War II." He has also said that a woman's primary function is to have children and start early, fulfilling God's mission for her, instead of delaying children while building a career.

This vitriolic messaging does have an effect. In a 2019 survey, men under 40 said that the biggest threat to Poland was "the LGBT movement and gender ideology." Another recent survey revealed that 55 percent of Poles believe that "Jews have too much influence in the world" and 19 percent believed it was a good thing that World War II resulted in fewer Jews in Poland.

THE BALKANS // An Unmoderated Space

The Balkan countries have struggled since the fall of the Eastern bloc with fragile democracies, wars, genocide, government corruption and economic insecurity. For most Balkan countries, there's also been protracted membership discussions to join the European Union, which many hope will address the growing far-right political momentum that results in anti-democratic societies and bigoted and oppressive policies.

Shockingly little is known about how Facebook and other social media is used in the region, so the question becomes why social media companies would allow exceptions to their community standards for politicians in an area of the world that is particularly fragile and subject to hate violence? According to DataReportal, Facebook and Twitter are the most popular platforms in the Balkans with about 3.7 million social media users in Serbia, 1.1 million in North Macedonia, 390,000 in Montenegro and 1.7 million in Bosnia and Herzegovina. Facebook is more popular than Twitter in all countries. These large numbers make Facebook attractive to politicians and advertisers, but they also make the users vulnerable to hate speech and misinformation.

SOCIAL MEDIA COMPANIES CANNOT ADEQUATELY PROTECT THEIR USERS IN ENGLISH, MUCH LESS THE LESSER-KNOWN LANGUAGES OF THE BALKANS. ALMOST NO RESEARCH ON THE CONTENT AND HATE SPEECH IN THESE LANGUAGES HAS BEEN DONE ... .

Content moderation is severely lacking throughout the region. According to a survey conducted by the Balkan Investigative Reporting Network (BIRN), more than half of hate speech and threats of violence remain on Facebook and Twitter in Bosnian, Serbian, Montenegrin or Macedonian and are available to users even after they've been reported to the companies, and after the companies have confirmed that the content is in violation of the rules. Chloe Berthelemy, a policy advisor at European Digital Rights said that, "because the dominant social media platforms reproduce the social systems of oppression, they are also often unsafe for many groups at the margins. Furthermore, those social media networks are also advertisement companies. They rely on inflammatory content to generate profiling data and thus advertisement profits. There will be no effective, systematic response without addressing the business models of accumulating and trading personal data."

Belgrade-based digital rights NGO, SHARE Foundation,

told *Balkan Insight* that, "When it comes to relatively small language groups in absolute numbers of users, such as languages in the former Yugoslavia or even in the Balkans, there is simply no incentive or sufficient pressure from the public and political leaders to invest in human moderation." Sanjana Hattotuwa, special advisor at ICT4Peace Foundation told *Balkan Insight*, "And in many cases, these markets are out of sight and out of mind, unless the violence, abuse or platform harms are so significant they hit the *New York Times* front-page." These comments are reminiscent of how hateful content was handled in Myanmar in the years leading up to the Rohingya genocide, when the country had no Burmese content moderators and military officials were able to use Facebook at will, ultimately using the platform to wage genocide.

The companies themselves have refused to answer questions about their policies and staffing in the region. Facebook will not disclose the number of human content moderators it has in any given country or language, telling *Balkan Insight* in its reporting on the BIRN survey that "it isn't accurate to only give the number of content reviewers. That alone doesn't reflect the number of people working on a content review for a particular country at any given time" (This statement is how Facebook always replies to questions about its moderators, no matter who asks or what country they're in. They've used these exact words when answering questions from U.S. senators and congressmen). And given the complexity of the Balkan languages, artificial intelligence algorithms meant to clean up content are likely doomed to fail.

These companies cannot adequately protect their users in English, much less other languages in other alphabets, including the languages spoken in the Balkans. Almost no research on the content and hate speech in these languages has been done, and there is a very incomplete picture of the impact of the social media companies on democracies and societies in the region. We know that the platforms, particularly Facebook, are widely used by political parties and elected officials. We know that there has been a backsliding of democracy in the region, specifically Bulgaria, and that anti-Roma and anti-LGBTQ hate speech and discriminatory policies are rampant. What we don't know is if Facebook, Twitter and others will take responsibility and create solutions for the long term.

INDIA

Facebook Makes Hindu Nationalism a Force

IN 2014, NARENDRA MODI LED HIS PARTY, the Bharatiya Janata Party (BJP), to a commanding lead in India's lower house of parliament, landing him the prime ministership, the top Indian perch from which to demonize Muslims and other populations Modi has consistently attacked. Modi and the BJP reached these heights through the effective use of social media, in particular Facebook, which has a long and deep relationship with him.

The impact has been devastating. Freedom House downgraded India from a "free" democracy to "partly free" in its 2021 report citing increasing pressure on human rights groups, intimidation and harassment of journalists and academics and policies that stigmatize and harm religious minorities, particularly Muslims.

SINCE BEING ELECTED, MODI, HIS BJP AND OTHER HINDU NATIONALIST SUPPORTERS HAVE USED FACEBOOK TO RUN AN ONLINE CAMPAIGN OF TERROR AGAINST MUSLIMS AND OTHERS THAT HAS LED TO REAL WORLD VIOLENCE.

Modi was already known to be a dangerous figure when Facebook chose to engage closely with him. Most notably, in February 2002, while head of the Gujarat government, Modi allegedly encouraged massive anti-Muslim riots. As the state was overcome with violence and over a thousand Muslims were murdered, leaders of the BJP and its even more nationalist ally, the Vishwa Hindu Parishad, gave speeches provoking Hindus to teach Muslims a lesson. Modi himself gave an incendiary speech, mocking riot victims and calling relief camps for Muslims "child-producing factories." The intensity and brutality of the violence unleashed against Muslims in 2002 led the Supreme Court of India to describe the Modi's Gujarat government as, "Modern day Neros who looked the other way while young women and children were burnt alive."

That ugly history did not deter Facebook, which instead saw a massive prize in the Indian market. India is Facebook's largest market, with 328 million using the social media platform in 2020. Another 400 million rely on Facebook's messaging service WhatsApp. The BJP, which has more than 16 million followers on its page, was Facebook India's biggest advertising spender in 2020. Ties between the company and the Indian government run even deeper, as the company has multiple commercial ties, including partnerships with the Ministry of Tribal Affairs, the Ministry of Women and the Board of Education. Both CEO Mark Zuckerberg and COO Sheryl Sandberg have met personally with Modi, who is the most popular world leader on Facebook. Before Modi became prime minister, Zuckerberg even introduced his parents to him.

FACEBOOK INDIA BACKS MODI

In 2014, after Modi's win, Facebook's top Indian staffer, Ankhi Das, wrote a celebratory opinion piece about how Modi successfully harnessed Facebook to propel his election to the highest office. She wrote, "From the start, Modi ran the campaign like a U.S. presidential election and took a commanding, front-row seat in building a community and driving engagement." Das cited his 8 million Facebook fans in 2013 that grew in a year to over 11 million. She praised his online tactics: "As the national campaign momentum picked up, Modi's fan

base increased by 28.7% crossing 14 million fans by May 12—the second most 'liked' political on Facebook after Obama." Das spoke of how Facebook's election tracker ranked Modi the "no. 1 leader" throughout the campaign. Revealing how Facebook helped spread news of Modi, she described how the company encouraged voters to share how they voted, messages seen by over 31 million voters. Das was downright giddy in her piece, writing that when the results were called, "Modi's photo with his victory wall message generated more than a million likes and shares."

Since being elected, Modi, his BJP and other Hindu nationalist supporters have used Facebook to run an online campaign of terror against Muslims and others that has led to real world violence. The situation is so dire regarding the anti-Muslim bias of Facebook India that the company's senior executives were summoned before a parliamentary committee for a closed-door hearing on September 2, 2020. The hearing followed allegations that Das, who posted anti-Muslim material on her own Facebook page, had prevented the removal of hate speech and anti-Muslim posts by BJP politicians in order to protect and promote the party and Modi.

Das' inability to carry out her responsibilities to remove hate speech in an objective manner caused immense real-world harm. For example, Facebook appeared to play a pivotal role in the February 2020 New Delhi riots in which more than 50 people died and thousands of homes and several mosques were destroyed. While both Hindus and Muslims were affected in the riots, Muslims were targeted in far greater numbers by mobs of young men, many of whom had traveled into the city to harass Muslims after seeing fake news shared widely on Facebook that Muslim religious leaders were calling for Hindus to be kicked out of Delhi.

One post by a BJP member, who is also a member of the right-wing militant Hindu organization Bajrang Dal, prompted hundreds to comment that they and their Hindu "brothers" would join the fight to defend Delhi from the Muslims. And two days before the anti-Muslim riots began in Delhi, a member of Modi's cabinet said Muslims should have been sent out of India to Pakistan in 1947 during the partition of India. Ultimately, the Delhi State Assembly's Peace and Harmony Committee said it had prima facie found Facebook guilty of aggravating the Delhi riots, and posited that it should be investigated for every riot since 2014. In the wake of the violence, hundreds of Muslim families fled New Delhi.

In May 2020, a BJP member of parliament in West Bengal, Arjun Singh, posted an image on Facebook that he wrongly claimed was a depiction of a Hindu who had been brutalized by Muslim mobs. It was captioned: "How long will the blood of Hindus flow on in Bengal...we will not stay quiet if they [Muslims] attack ordinary people." Four hours later, an angry mob of about 100 Hindus descended on a town in West Bengal and a Muslim shrine was vandalized. Facebook failed to remove the posts until after the company experienced backlash as a result of the violent attacks, which local Muslims alleged had been incited by Singh's post. Dozens of Muslims have been lynched since 2012 by vigilantes, with many of the incidents triggered by fake news shared on WhatsApp.

On Jan. 25, 2020, supporters of human rights march before the Indian High Commission in London to protest Prime Minister Narendra Modi's government's attacks on Muslims and other minorities.

PENELOPE BARRITT / ALAMY STOCK PHOTO

When T. Raja Singh, another member of the BJP, called for the slaughter of Rohingya Muslim refugees, threatened to demolish mosques and labeled Indian Muslim citizens as traitors, Facebook's online security staff determined his account should be banned for not only violating its community standards, but also for falling under the category of "Dangerous Individuals and Organizations." Das stepped in to protect Singh from punitive action, because "punishing violations by politicians from Mr. Modi's party would damage the company's business prospects in the country," according to Facebook employees. Outrage in response to these disclosures forced Facebook to finally ban Singh from the platform in early September 2020.

Das ultimately stepped down in October 2020 after her protection of anti-Muslim hatred on Facebook was exposed in a series of news articles, but by then much damage had been done.

FACEBOOK IGNORES CIVIL SOCIETY PLEAS

Facebook's anti-Muslim actions in India have been repeatedly called out by civil society actors. In October

2019, a report by the nonprofit organization Avaaz accused Facebook of having become a "megaphone for hate" against Muslims in the northeastern Indian state of Assam—where nearly 2 million people, many of them Muslims, have been stripped of their citizenship by the BJP government. Another report, by the South Asian human rights group Equality Labs, found "Islamophobic [anti-Muslim] content was the biggest source of hate speech on Facebook in India, accounting for 37 percent of the content," and that 93 percent of the hate speech they reported to Facebook was not removed. They also reported on how Facebook is being used to spread hate speech and misinformation accusing Muslims of deliberately infecting non-Muslims and Hindus with COVID-19, again contributing to potential violence against Muslims.

In September 2020, a letter signed by 41 civil rights organizations from around the world including Global Project Against Hate and Extremism called on Facebook to put an end to anti-Muslim hate on its platform and, among other requests, immediately suspend Das, who was still on staff at that time, to protect the safety and security of Muslims.

Meanwhile, as the Modi government was stripping Muslims of their rights, Facebook was taking WhatsApp accounts away from Muslims in Kashmir. The government had suspended Internet in the region to prevent communication, and Facebook's policy automatically discontinues WhatsApp participation after 120 days without use. As a result, the government prevented Muslims in the region from organizing, and Facebook contributed by further reducing communication opportunities.

Many, however, question the utility of continuing to urge Facebook to address hate on the platform driven by the BJP and other Hindu nationalist organizations in India. Malay Tewari, a Kolkata-based activist, argued Facebook "rarely" responded to his complaints about BJP-linked posts and "quite strangely, Facebook posts which expose the propaganda or hate campaign of the BJP, which do not violate community standards, are often removed." Indian journalist Rana Ayyub agreed, "For years now, verified Facebook pages of BJP leaders such as Kapil Mishra have routinely published hate speeches against Muslims and dissenting voices. The hate then translates into deadly violence, such as the February anti-Muslim attacks in Delhi that left many people dead in some of the worst communal violence India's capital has seen in decades... It's clear that Facebook has no intention of holding hatemongers accountable and that the safety of users is not a priority."

In late August 2020, after much bad press, Facebook, in a supposed effort to evaluate its role in spreading hate speech and incitement to violence, commissioned an independent report by the U.S. law firm Foley Hoag on the platform's impact on human rights in India. The report findings are pending. Meanwhile, Das' interim replacement, Shivnath Thukral, appears to have his own problems. It was Thukral who had ignored Avaaz's 2019 flag of anti-Muslim hate speech by a BJP leader. And, perhaps no surprise, Thukral worked for the BJP during its 2014 election campaign.

NETHERLANDS

Racist Political Leaders Rise Through Social Media

IN THE LEAD UP TO THE NETHERLANDS' MARCH 17, 2021 snap elections, Facebook ran a full-page ad on February 26, and then a half-page ad on March 3, in several Dutch newspapers implying that it had 35,000 content moderators ensuring that hate speech and disinformation were not swirling around the upcoming elections. Facebook would like the Dutch to believe that they are committed to accuracy and can be trusted to provide reliable information on the elections.

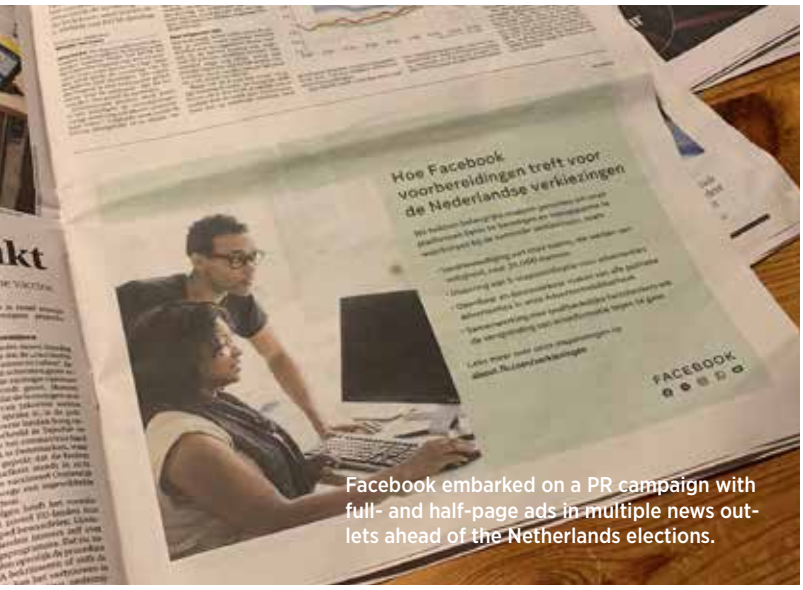
But once you go to the Facebook website link in their ads, it's clear that the 35,000 is a global number and that Facebook is firmly maintaining its position that politicians' content is not subject to fact-checking. In addition to lax enforcement of its hate speech and community standard rules when it comes to politicians, Facebook and other social media's lack of fact checking is a carve out that has had far-reaching consequences around the world, including the Netherlands. Indeed, far-right extremist politicians there have skillfully used the platforms and tools to spread their bigoted views and misinformation, push for discriminatory policies and grow their constituencies.

FACEBOOK RAN FULL- AND HALF-PAGE ADS IN SEVERAL NEWSPAPERS IMPLYING THAT HATE SPEECH AND DISINFORMATION WERE NOT SWIRLING AROUND THE UPCOMING ELECTIONS ... IN REALITY, THE SMALL PRINT REINFORCED THEIR REFUSAL TO FACT-CHECK POLITICIANS.

The lack of fact-checking for political messaging led to a scandal when respected Dutch organizations working with Facebook to fact-check materials were caught unaware of this exemption, and ultimately severed their relationship with the company.

In 2017, Facebook partnered with NU.nl, a Dutch news outlet, and Leiden University to act as fact checkers. False information wouldn't necessarily be removed, but it would be downgraded in the algorithm and flagged, a typical Facebook approach. (In 2018, Facebook launched an ad campaign in the Netherlands and other parts of the globe to teach people how to recognize fake news on Facebook.) But, in 2019, NU.nl resigned from their work with Facebook saying, "What is the point of fighting fake news if you are not allowed to tackle politicians?" (Leiden University had resigned the year before). When asked by NPR why NU.nl left the partnership, editor-in-chief Gert-Jaap Hoekman said, "The direct reason why we quit was that Facebook emphasized that political speech is not a part of this program. And that was for us, well, quite a big problem."

The disagreement between Facebook and NU.nl began in earnest because of an ad by Esther de Lange of the Christian Democratic Appeal party that NU.nl had labeled 'unsubstantiated.' The ad referred to the amount of Romanian farmland owned by non-Europeans. Facebook stepped in and told NU.nl to remove the flag saying that they were never intended to fact-check politicians. According to reporting by NPO3, this was the first that NU.nl learned of this restriction other than Nick Clegg, Facebook's VP of global affairs and communications, saying in September of 2019, "From now on we will treat speech from politicians as newsworthy content that



Facebook embarked on a PR campaign with full- and half-page ads in multiple news outlets ahead of the Netherlands elections.

PHOTO/KOEN DE REGT

should, as a general rule, be seen and heard."

NU.nl had enough when Facebook ordered them to remove flags placed the year before on two posts by far-right, anti-immigrant and anti-Muslim parties, the Party for Freedom (PVV) led by Geert Wilders and the Forum for Democracy (FvD) led by Thierry Baudet. The PVV post claimed that rival parties had voted to support child marriage. The flagged FvD post claimed that the Netherlands will have a predominantly immigrant (non-white) population within two generations. Both claims were patently false and meant to mislead and frighten voters. The flags disappeared from the posts and NU.nl quit Facebook, leaving politicians like Wilders and Baudet to continue to spread their hateful rhetoric. And they have done so aggressively.

Geert Wilders, founder and leader of PVV, is often referred to as the Dutch Donald Trump for his looks, anti-Muslim rhetoric and his strategic use of social media to reach large audiences with his divisive language. He has nearly 1 million followers on Twitter and nearly 400,000 Facebook fans.

Wilders is best known for his vehement opposition to Muslims and Islam and counts among his American allies anti-Muslim hate group leaders David Horowitz and Pamela Geller. He has been endorsed by American white supremacist David Duke for his virulent anti-Muslim speech. Over the course of his career, Wilders has advocated for the closure of all mosques, the prohibition of new mosque construction and a ban on Muslim immigrants. He produced and released a short film, *Fitna*, in 2008 to illustrate a perverted interpretation of the Koran and has compared the Koran to Hitler's *Mein Kampf*.

Playing on Dutch fears about the economy, unemployment, crime and fears that Muslim immigrants have not integrated into Dutch society, Wilders has seen his party steadily rise in power since its founding in 2006, and it is expected to be the second-largest after the March 17 elections. In a recent typical Facebook post, he said, "The figures show that the integration of non-Western immigrants has completely failed. If a terrorist beheads a French teacher, 125,000 Muslims in the Netherlands sign a petition to ban Mohammed cartoons!" He routinely uses hashtags like #stopislam on his social media accounts and cross-posts offensive material, once calling immigrants "scum."

His public comments have gotten him charged and tried by the Netherlands authorities for hate speech – twice. In 2011, he was charged with inciting hatred and discrimination against Muslims but was acquitted, further emboldening him to spread his hateful speech and push for discriminatory policies against immigrants and Muslims. In 2016, he was convicted for offending a group based on their race, Muslims, and inciting discrimination for remarks he made at a 2014 rally. The conviction for inciting discrimination was overturned on appeal in 2020, but the conviction for insulting a group based on race stood. He reinforced his position recently, saying he has no regrets and that, "The immigration of non-western immigrants is an existential problem."

When accused by a rival candidate of engaging in racism and exclusion and asked about diversity in government at a debate in February, Wilders said, "The first person of color I would like to defend is Zwarte Piet," the racialized character in Dutch tradition that is presently represented by wearing black face and now condemned by many Dutch as a racist tradition. He went on to say, "The PVV holds on to national traditions well, for us Zwarte Piet will forever remain black!" On Twitter, he declared that he would make Zwarte Piet Minister of Culture in his cabinet, to preserve diversity and Dutch culture. Even so, social media companies allow him to continue using their platforms. (The Facebook Oversight Board is set to review a take down of a Zwarte Piet image in connection with Facebook's banning of black face and caricatures of Black people.)

For all of Wilders' extreme positions and comparisons to Trump, he has been challenged in recent years by Thierry Baudet who has effectively used social media to grow support for his FvD party, especially among younger voters. Baudet is an outspoken supporter of Trump, even echoing Trump's claims in a radio interview that the U.S. election was stolen. He is perhaps even more like Trump in his virulent anti-immigrant rhetoric and spread of conspiracy theories. The FvD website calls for expulsion of non-whites to their home countries,

questions climate change and rejects any special treatment of "identity" groups based on religion, ethnicity and gender, among others. Shockingly, he has also called for the formation of a "civilian army" on Twitter and YouTube to be active on election day.

This support of a civilian army should, however, come as no surprise. In the midst of the Capitol riots on January 6, Baudet retweeted one of his own 2016 tweets saying that Trump would be not only a great leader for the U.S., but for the West as a whole. Baudet deleted the tweet and then denied having posted it, but Politwoops had already captured it.

Baudet has pushed the dangerous Great Replacement conspiracy theory that has inspired several mass murders. In a March 2020 parliamentary debate, he said the EU was setting up a ferry service "to transfer immigrants from Africa to Europe, to weaken national identities so that there will be no more nation-states." The Great Replacement conspiracy theory is an international movement that believes that white people are being replaced, or genocided, by a group of elites, often Jews. This theory, and reference to it, were specifically banned on Facebook after the Christchurch, N.Z., massacre of Muslims by a shooter who live-streamed his attack on Facebook. The shooter was inspired by the Great Replacement theory and had been radicalized online, mostly on YouTube, according to a New Zealand government report.

Baudet has also been known to tweet videos made by a women's group called 120 Dezibel, which claims they are afraid of migrants and compare the lack of protection from them to the German attitude to the Holocaust after WWII. 120 Dezibel is associated with Generation Identity in Germany which is a large transnational white nationalist group and perhaps the biggest proponent of the Great Replacement conspiracy theory.

In November 2020, it was reported that several members of the FvD youth party had been expelled for using WhatsApp and Instagram to share anti-Semitic and anti-LGBTQ messages. One 23-year-old said, "Jews have international pedo networks and help women en masse into pornography." This after reports in April of the same conduct for which little to no action was taken because the party did not want to be "thought police." In February of this year, Baudet was once again tormented by screen shots of a WhatsApp conversation. In this one, he claims that white people are more intelligent than Latinos and Black people. An FvD staffer says, "a people that never has to plan for winter develops differently." Baudet goes on to say, "Moreover: African Americans have lived in America for 150 years. Still score 40 IQ points lower."

PHILIPPINES

Facebook Uplifts a Serial Human Rights Violator

AS HE WAS RUNNING FOR PRESIDENT IN 2016, RODRIGO DUTERTE participated in a debate hosted by an online publication, The Rappler, at a Manila university. Maria Ressa, co-founder of the site and a world-renowned journalist, hosted the debate. It was broadcast across the country, with the questions coming from users on Facebook, which had co-hosted the forum and where it was live-streamed to millions.

For Ressa, it was an exciting moment. "Duterte's campaign on social media was groundbreaking," she told Bloomberg in 2017. This changed to devastating, as the publication later found itself in Duterte's crosshairs.

In the Philippines, Facebook rules the internet. Smartphones are more numerous than people, and 97 percent of Filipinos have Facebook accounts. Duterte was introduced to the Filipino population through The Rappler event, and its popularity on Facebook led him to quickly understand that the election would be fought online. He hired strategists to build out his online presence and, with the direct help of Facebook, engineered a network of Facebook pages and bloggers worldwide. Duterte grew his support network aggressively. They came to be called the Duterte Die-Hard Supporters (DDS), an obvious reference to another DDS, Duterte's Davao Death squad, thought to have killed hundreds of people while Duterte was mayor of Davao City.

Duterte's track record of human rights violations was already terrifying by the time he ran for president. The DDS allegedly engaged in summary executions in Davao City, including of street children. The group is estimated to be responsible for the killing or disappearance of more than 1,000 people between 1998 and 2008. For years, Duterte had been heavily criticized by numerous organizations for condoning and even inciting executions, including by the U.N. Assembly of the Human Rights Council. In 2009, it said, "The Mayor of Davao City has done nothing to prevent these killings, and his public comments suggest that he is, in fact, supportive." His behavior should have been cause for concern by the time he ran for the presidency with Facebook's aid. Once Duterte was elected in May 2016, he turned Facebook into a weapon against his enemies.

DUTERTE: A FACEBOOK CREATION

In March 2015, Facebook CEO Mark Zuckerberg announced Internet.org, a free service created by Facebook and intended to give the world's billions of unconnected people access to the internet. Facebook wanted to become the internet and it did so in the Philippines, with Duterte one of its most popular users.

The 2016 election was marked with misinformation and threats. Facebook quickly started receiving complaints about inauthentic pages run by Duterte supporters, many of whom were circulating aggressive messages, insults and threats of violence. The campaign also ramped up false information. In March, one of the campaign's Facebook pages posted a fake endorsement from Pope Francis, reading "Even the Pope Admires Duterte" beneath the Pope's image. The Catholic Bishops' Conference of the Philippines posted a statement clarifying, "May we inform the public that this statement from the Pope IS NOT TRUE. ... We beg everyone to please stop spreading this."

Duterte ended up dominating the political conversation so thoroughly that in April 2016, a month before the vote, Facebook glowingly called him the "undisputed king of Facebook conversations." After he was elected president, his communications secretary thanked Facebook, "During the campaign period, Facebook was quite a valuable tool for the President's base of supporters in organizing gatherings and spreading news about campaign activities," adding the platform was

"comparable to having instant-access live radio and television facilities."

Thanks to heavy subsidies that keep Facebook free to users on mobile phones, Facebook has completely saturated the Philippines, and in effect is the internet there. And because using other data, like accessing a news website via a mobile web browser, is expensive just as in other developing countries like Brazil, for most Filipinos the only internet access is through Facebook. The platform is a leading provider of news and information, and it was what enabled Duterte to ride a wave of populist anger to the presidency.

Since taking office, Duterte has shaken Filipino democracy to its foundations. His administration has threatened the justice system by ousting the chief justice of the Supreme Court, jailed opponents on baseless charges, attacked press freedoms and sanctioned the extrajudicial executions of more than 12,000 Filipinos suspected of selling or using drugs. In 2017, Human Rights Watch described Duterte's government as "a human rights calamity."

Filipino president Rodrigo Duterte has a dismal human rights record but a thriving Facebook presence.

PHOTO/WIKICOMMONS

DUTERTE WEAPONIZED FACEBOOK

Once in office, Facebook worked closely with Duterte's administration, offering special services so it could maximize the platform's potential. Duterte often banned news organizations from covering his events, including the inauguration, which instead was streamed on Facebook. Throughout Duterte's term, Facebook has been used as a key amplifier of pro-administration narratives and sentiment. As an example, nearly two dozen pro-Duterte Facebook pages and websites shared the fake news that Chief Justice Maria Lourdes Sereno tried to leave the country to escape the impeachment complaint filed against her. Ellen Tordesillas, president of Vera Files, a Facebook fact-checking partner in the Philippines, told Buzzfeed the "majority" of false posts that her organization checks "definitely" come from pro-administration Facebook pages or were inspired by the president's remarks.

As The Rappler became more critical of the government, Ressa's news site found itself under attack from the Duterte regime. "The [attempted] shuttering of Rappler—an organization whose credibility was undermined as a result of fake news and trolling circulating on Facebook—is a tragic reminder to Facebook of the central role it plays in shaping political discourse," Carly Nyst, a human rights lawyer, told CNBC in 2017. "It is increasingly untenable for Facebook to deny its role in facilitating the Duterte regime's clampdown on critical voices," she added. Ressa herself faces multiple different trumped up charges, including a conviction for cyber-libel in June 2020, as well as verbal attacks by Duterte, multiple investigations, tax fraud charges and the revocation of her publication's license. She has also been relentlessly harassed with online threats, often driven by pro-Duterte allies. Amnesty International described Ressa's conviction as a sham.

"IT IS INCREASINGLY UNTENABLE FOR FACEBOOK TO DENY ITS ROLE IN FACILITATING THE DUTERTE REGIME'S CLAMPDOWN ON CRITICAL VOICES."

CARLY NYST, HUMAN RIGHTS LAWYER, CNBC, 2017

In the last year, Duterte's government has used Facebook as a tool in its campaign against those refusing to agree to COVID lockdown measures, whom the government has threatened to kill. On March 24, 2020, police in San Isidro forced alleged curfew violators to sit in the hot sun, and the local government's Facebook page posted a photo of them, saying, "Everyone violating the curfew will be placed here." Other types of punishments are also highly demeaning. On April 5, three members of the LGBTQ community in Pandacaqui were "ordered to kiss each other and do a sexy dance in front of a minor," as punishment for violating the curfew, and the incident was streamed live on Facebook by the highest elected official in the village. Another Facebook Live post showed

detainees in Pandacaqui forced to sign bail papers with their own sweat, while being threatened with paddling.

FACEBOOK FINALLY TAKES ACTION

In September 2020, Facebook dismantled a network of fake accounts that originated in China and targeted the Philippines, including some that criticized the Communist Party of the Philippines and its armed wing, the New People's Army, longtime opposition to Duterte. The company said it removed 155 accounts, 11 pages, nine groups and six Instagram accounts for violating its policy against foreign or government interference functioning as "coordinated inauthentic behavior." The activity originated in China and focused primarily on the Philippines and Southeast Asia.

The Filipino Facebook activity reached hundreds of thousands of people and included content supporting Duterte and his daughter, Davao City Mayor Sara Duterte-Carpio, who is now running for president. (Duterte can only serve one term; elections will be held in 2022). One network had links to the Filipino military and police and appeared to have accelerated its operations between 2019 and 2020. About 280,000 people were reached with posts in English and Filipino, about domestic politics, the military's anti-terrorism activities, proposed legislation as well as criticism of communism, youth activists and opposition organizations. The earliest example dated back to 2015, and included the "red-tagging," (labeling people as communist) of critics of Duterte, which has in some cases led to the murders of those tagged.

Duterte was enraged by the takedowns, likely because of years of friendly interaction with the company. "Facebook, listen to me," he said in a televised address, "We allow you to operate here hoping that you could help us. Now, if government cannot espouse or advocate something which is for the good of the people, then what is your purpose here in my country?" "What would be the point of allowing you to continue if you can't help us?"

CONCLUSION

Social Media: a Disaster for Human Rights and Democracy

SOCIAL MEDIA, AND FACEBOOK IN PARTICULAR, has had a horrifically damaging effect on democracies, societies and vulnerable populations around the world. Bigoted populist leaders and far-right political parties across the globe have harnessed the power of social media to achieve political heights likely previously unattainable.

Facebook insiders have said the same. In September 2020, a 6,600-word memo leaked to Buzzfeed by a departing Facebook data scientist, Sophie Zhang, made clear that the problem of distorting political systems is even more wide-ranging than documented in this report.

Zhang, who worked on Facebook's site integrity team, made explosive allegations that Facebook had simply failed to take action on fake accounts that were manipulating elections and politics in multiple countries. She described several instances of heads of government and political parties using fake accounts or misrepresenting themselves to sway public opinion. In several countries, including Azerbaijan, Honduras, Ecuador, India, Spain and others, she also found evidence of coordinated campaigns of varying sizes to boost or hinder political candidates or outcomes.

Zhang rejects the idea that Facebook is run by malicious people hoping to achieve a particular outcome, but also said, "It's an open secret within the civic integrity space that Facebook's short-term decisions are largely motivated by PR and the potential for negative attention," having been told directly that anything published in the *New York Times* or *Washington Post* would obtain elevated priority. Outside of the U.S. and Western Europe, where Facebook often faces pushback from activists, media and governments, there appears to be little concern over what Facebook is doing to political systems. She wrote, "A manager on Strategic Response mused to myself that most of the world outside the West was effectively the Wild West with myself as the part-time dictator – he meant the statement as a compliment, but it illustrated the immense pressures upon me."

Which leads one to ask, why would a company whose mission is to "give people the power to build community and bring the world closer together" wait until a big media story about yet another tragedy laid at the feet of Facebook to be proactive in preserving human rights?

This report has documented how social media has altered the political fates of eight countries and one region around the world, with a combined population of more than 2 billion. In these countries, social media has helped raise the profiles of far-right politicians and political parties, allowed them to recruit and indoctrinate growing audiences and spread their anti-human rights agenda, resulting in punishing policies for many vulnerable communities. In many cases, demonized populations are now facing rising levels of violence up to and including genocide as a result.

Even when harms have been exposed, almost always by civil society actors and journalists, an all-consuming profit motive and drive to conquer new internet territories has repeatedly overridden considerations around the protection of users and the moral imperative to support democracy and human rights. Additional investments in user protection usually only come after tragic events, like the Rohingya genocide in Myanmar where Facebook played a key organizing role.

Again and again, we have witnessed a complete refusal by the tech world to acknowledge how their platforms have been co-opted and manipulated to malignant ends. The willful naiveté and arrogance of the social media leaders, with their declarations around uniting the world, and their utter blindness and disregard that the opposite might happen, has cost lives and damaged democracy.

In the U.S., Western Europe and a few other cases such as Brazil and India, major documentation efforts have been undertaken by civil society actors to expose

how these platforms are allowing far-right movements to grow and act as fonts of hate speech, violence and disinformation. But that sort of research is lacking in many countries and especially in dozens and dozens of languages. There is no way to know how badly tech platforms are damaging human rights and democracy in many places—the evidence just doesn't exist.

But you can be sure that the platforms won't step up and take responsibility for their damaging actions, as their track records have shown. Either they won't investigate because it might lead to a need to alter their operations in a way that reduces engagement and growth, and thus profits, or they are simply unconcerned with what happens outside their areas of interest.

And they've constantly reiterated the mantra that rising extremism is not their fault. In June 2020, Clegg told *The Washington Post* that populism wasn't invented in Silicon Valley, pointing to centuries of political history before social media companies' existence. "From the Arab Spring to local candidates challenging political incumbents, social media has also helped to open up politics, not favor one side over the other," Clegg added, claiming research has shown that "polarization has fallen in many countries with high internet use."

Again, Clegg is wrong. Research, including by Facebook itself, has shown that the platforms' A.I. mechanics preference and push polarizing and outrageous content and aid divisive politicians. A former Facebook A.I. researcher told MIT Technology Review that his team conducted multiple studies that confirmed platforms that maximize engagement, like Facebook, "increase polarization." A.I. models honed to increase engagement learn to feed users more extreme content, and over time, "they measurably become more polarized." Given the state of online content moderation, we can be sure to have more polarization wherever Facebook is used, and extremist politicians will benefit, along with hate groups and conspiracy theorists.

This situation is not tenable. The social media companies have more power and money than many countries. But they also have a responsibility to the global citizenry. Their repeated displays of ignorance or their disregard for human rights and democratic societies can no longer be ignored. They must be held to account for their actions and forced to disclose how they will protect human rights and democracy going forward. This requires governments and international bodies to step in. Lives are at stake.

The new Facebook Oversight Board has said that it has the authority to decide "how Facebook treats posts from public figures that may violate community standards," including against hate speech and that it "won't shy away from the tough cases and holding Facebook accountable." We'll see.

Global Project Against Hate and Extremism would like to thank András Bíró-Nagy, Russell Estes, Willemijn Keizer, Rafal Pankowski and Muslim Advocates for lending their support, advice, skills and expertise to this report. Their contributions were invaluable.


GPAHE
Global Project Against Hate and Extremism