Revealed: the Facebook loophole that lets world leaders deceive and harass their citizens

Monday 3 May 2021, by WONG Julia Carrie (Date first published: 12 April 2021).

Facebook has repeatedly allowed world leaders and politicians to use its platform to deceive the public or harass opponents despite being alerted to evidence of the wrongdoing.

The Guardian has seen extensive internal documentation showing how Facebook handled more than 30 cases across 25 countries of politically manipulative behavior that was proactively detected by company staff.

The investigation shows how Facebook has allowed major abuses of its platform in poor, small and non-western countries in order to prioritize addressing abuses that attract media attention or affect the US and other wealthy countries. The company acted quickly to address political manipulation affecting countries such as the US, Taiwan, South Korea and Poland, while moving slowly or not at all on cases in Afghanistan, Iraq, Mongolia, Mexico and much of Latin America.

“There is a lot of harm being done on Facebook that is not being responded to because it is not considered enough of a PR risk to Facebook,” said Sophie Zhang, a former data scientist at Facebook who worked within the company’s “integrity” organization to combat inauthentic behavior. “The cost isn’t borne by Facebook. It’s borne by the broader world as a whole.”

Facebook pledged to combat state-backed political manipulation of its platform after the historic fiasco of the 2016 US election, when Russian agents used inauthentic Facebook accounts to deceive and divide American voters.

But the company has repeatedly failed to take timely action when presented with evidence of rampant manipulation and abuse of its tools by political leaders around the world.

Facebook fired Zhang for poor performance in September 2020. On her final day, she published a 7,800-word farewell memo describing how she had “found multiple blatant attempts by foreign national governments to abuse our platform on vast scales to mislead their own citizenry” and lambasting the company for its failure to address the abuses. “I know that I have blood on my hands by now,” she wrote. News of the memo was first reported in September by BuzzFeed News.

Zhang is coming forward now in the hopes that her disclosures will force Facebook to reckon with its impact on the rest of the world.

“Facebook doesn’t have a strong incentive to deal with this, except the fear that someone might leak it and make a big fuss, which is what I’m doing,” she told the Guardian. “The whole point of inauthentic activity is not to be found. You can’t fix something unless you know that it exists.”

Liz Bourgeois, a Facebook spokesperson, said: “We fundamentally disagree with Ms Zhang’s characterization of our priorities and efforts to root out abuse on our platform.

“We aggressively go after abuse around the world and have specialized teams focused on this work. As a result, we’ve taken down more than 100 networks of coordinated inauthentic behavior. Around half of them were domestic networks that operated in countries around the world, including those in Latin America, the Middle East and North Africa, and in the Asia Pacific region. Combatting coordinated inauthentic behavior is our priority. We’re also addressing the problems of spam and fake engagement. We investigate each issue before taking action or making public claims about them.”

Facebook did not dispute Zhang’s factual assertions about her time at the company.

With 2.8 billion users, Facebook plays a dominant role in the political discourse of nearly every country in the world. But the platform’s algorithms and features can be manipulated to distort political debate.

One way to do this is by creating fake “engagement” – likes, comments, shares and reactions – using inauthentic or compromised Facebook accounts. In addition to shaping public perception of a political leader’s popularity, fake engagement can affect Facebook’s all-important news feed algorithm. Successfully gaming the algorithm can make the difference between reaching an audience of millions – or shouting into the wind.

Zhang was hired by Facebook in January 2018 to work on the team dedicated to rooting out fake engagement. She found that the vast majority of fake engagement appeared on posts by individuals, businesses or brands, but that it was also being used on what Facebook called “civic” – ie political – targets.

The most blatant example was Juan Orlando Hernández, the president of Honduras, who in August 2018 was receiving 90% of all the known civic fake engagement in the small Central American country. In August 2018, Zhang uncovered evidence that Hernández’s staff was directly involved in the campaign to boost content on his page with hundreds of thousands of fake likes.

One of the administrators of Hernández’s official Facebook Page was also administering hundreds of other Pages that had been set up to resemble user profiles. The staffer used the dummy Pages to deliver fake likes to Hernández’s posts, the digital equivalent of bussing in a fake crowd for a speech.

This method of acquiring fake engagement, which Zhang calls “Page abuse”, was made possible by a loophole in Facebook’s policies. The company requires user accounts to be authentic and bars users from having more than one, but it has no comparable rules for Pages, which can perform many of the same engagements that accounts can, including liking, sharing and commenting.

The loophole has remained open due to a lack of enforcement, and it appears that it is currently being used by the ruling party of Azerbaijan to leave millions of harassing comments on the Facebook Pages of independent news outlets and Azerbaijani opposition politicians.

Page abuse is related to what Russia’s Internet Research Agency did during the 2016 US election, when it set up Facebook accounts purporting to represent Americans and used them to manipulate individuals and influence political debates. Facebook called this “coordinated inauthentic behavior” (CIB) and tasked an elite team of investigators, known as threat intelligence, with uncovering and removing it. Facebook now discloses the CIB campaigns it uncovers in monthly reports, while removing the fake accounts and Pages.

But threat intelligence – and numerous Facebook managers and executives – resisted investigating both the Honduras and Azerbaijan Page abuse cases, despite evidence in both cases linking the

abuse to the national government. Among the company leaders Zhang briefed about her findings were Guy Rosen, the vice-president of integrity; Katie Harbath, the former public policy director for global elections ; Samidh Chakrabarti, the then head of civic integrity; and David Agranovich, the global threat disruption lead.

The cases were particularly concerning because of the nature of the political leaders involved. Hernández was re-elected in 2017 in a contest that is widely viewed as fraudulent. His administration has been marked by allegations of rampant corruption and human rights violations. Azerbaijan is an authoritarian country without freedom of the press or free elections.

Hernández did not respond to queries sent to his press officer, attorney and minister of transparency. After publication of this article, the YAP, Azerbaijan's ruling party, denied any connection to the Pages leaving harassing comments on the 6 March Azad Soz post.

It took Facebook nearly a year to take down the Honduras network, and 14 months to remove the Azerbaijan campaign. In both cases, Facebook subsequently allowed the abuse to return. Facebook says that it uses manual and automated detection methods to monitor previous CIB enforcement cases, and that it "continuously" removes accounts and Pages connected to previously removed networks.

The lengthy delays were in large part the result of Facebook's priority system for protecting political discourse and elections.

"We have literally hundreds or thousands of types of abuse (job security on integrity eh!)," Rosen told Zhang in an April 2019 chat after she had complained about the lack of action on Honduras. "That's why we should start from the end (top countries, top priority areas, things driving prevalence, etc) and try to somewhat work our way down."

Zhang told Rosen in December 2019 that she had been informed that threat intelligence would only prioritize investigating suspected CIB networks in "the US/western Europe and foreign adversaries such as Russia/Iran/etc".

Rosen endorsed the framework, saying: "I think that's the right prioritization."

Zhang filed dozens of escalations within Facebook's task management system to alert the threat intelligence team to networks of fake accounts or Pages that were distorting political discourse, including in Albania, Mexico, Argentina, Italy, the Philippines, Afghanistan, South Korea, Bolivia, Ecuador, Iraq, Tunisia, Turkey, Taiwan, Paraguay, El Salvador, India, the Dominican Republic, Indonesia, Ukraine, Poland and Mongolia.

The networks often failed to meet Facebook's shifting criteria to be prioritized for CIB takedowns, but they nevertheless violated the company's policies and should have been removed.

In some of the cases that Zhang uncovered, including those in South Korea, Taiwan, Ukraine, Italy and Poland, Facebook took quick action, resulting in investigations by staff from threat intelligence and, in most cases, takedowns of the inauthentic accounts.

In other cases, Facebook delayed taking action for months. When Zhang uncovered a network of fake accounts creating low-quality, scripted fake engagement on politicians in the Philippines in October 2019, Facebook left it to languish. But when a tiny subset of that network began creating an insignificant amount of fake engagement on Donald Trump's Page in February 2020, the company moved quickly to remove it.

In several cases, Facebook did not take any action.

A threat intelligence investigator found evidence that the Albanian network, which was mass-producing inauthentic comments, was linked to individuals in government, then dropped the case.

A Bolivian network of fake accounts supporting a presidential candidate in the run-up to the nation's disputed October 2019 general election was wholly ignored; as of Zhang's last day of work in September 2020, hundreds of inauthentic accounts supporting the politician continued to operate.

Networks in Tunisia and Mongolia were similarly left uninvestigated, despite elections in Tunisia and a constitutional crisis in Mongolia.

Amid mass protests and a political crisis in Iraq in 2019, Facebook's market specialist for Iraq asked that two networks Zhang found be prioritized. An investigator agreed that the accounts should be removed, but no one ever carried out the enforcement action, and on Zhang's final day, she found approximately 1,700 fake accounts continuing to act in support of a political figure in the country.

Ultimately, Zhang argues that Facebook is too reluctant to punish powerful politicians, and that when it does act, the consequences are too lenient.

"Suppose that the punishment when you have successfully robbed a bank is that your bank robbery tools are confiscated and there is a public notice in a newspaper that says, 'We caught this person robbing a bank. They shouldn't do that,'" Zhang says. "That's essentially what's going on at Facebook. And so what's happened is that multiple national presidents have made the decision that this risk is enough for them to engage in it.

"In this analogy, the money has already been spent. It can't be taken back."

Julia Carrie Wong

Click here to subscribe to ESSF newsletters in English and/or French.

P.S.

The Guardian

https://www.theguardian.com/technology/2021/apr/12/facebook-loophole-state-backed-manipulation