

2020

ASSESSING THE
HUMAN RIGHTS IMPACT
OF META'S PLATFORMS
IN THE PHILIPPINES

ARTICLE ONE

EXECUTIVE SUMMARY FOR META

PHOTO: STOCK IMAGE

EXECUTIVE SUMMARY

Meta's platforms have broken down barriers across the world. They have made neighbors of strangers and have, as the company's mission states, "brought the world closer together." Meta has fundamentally altered the ability for people to express themselves freely by democratizing access to powerful channels of communication.

At the same time, in some cases, the company's platforms have been misused by bad actors to infringe on human rights. Bad actors may use the platforms to amplify hate speech, spread misinformation and disinformation, target and harass vulnerable groups, and infringe on users' right to privacy.

The Philippines presents clear challenges when it comes to potential human rights abuses on Facebook, Instagram, and WhatsApp. These challenges include misinformation and disinformation, harassment and red-tagging of vulnerable groups, and violence incited by political leaders and others.

To determine the degree to which Meta's platforms may or may not have contributed to adverse human rights impacts and to mitigate the risk of further adverse impacts, Meta partnered with Article One from February to July 2020 to conduct a country-level human rights impact assessment (HRIA) of its platforms in the Philippines.

The HRIA was designed to pursue the following objectives:

1. Assess the role Meta's platforms play in the Philippines through a human rights lens¹
2. Proactively address risks to avoid harm to rights holders, including the platform's users
3. Determine how best to evolve and apply Meta's global policies and standards in high-risk contexts
4. Enable Meta to maximize opportunities for positive human rights impacts

This executive summary outlines the findings of the HRIA and presents recommendations for Meta to mitigate actual and potential impacts in line with the expectations of the United Nations Guiding Principles on Business and Human Rights (UNGPs).

¹ Informed by the United Nations Guiding Principles on Business and Human Rights (UNGPs) and the Global Network Initiative (GNI) Principles.

I. METHODOLOGY

The HRIA methodology was informed by guidance from the UNGPs and the Global Network Initiative (GNI), as well as by Article One's award-winning methodology for and experience in conducting HRIAs around the world.

The HRIA emphasized the voices of rightsholders potentially affected by Meta's platforms and their legitimate representatives. It paid special attention to ensuring direct engagement with Philippines-based civil society organizations, human rights defenders (HRDs), users of Meta's platforms, and international experts.² Given the risks posed by the COVID-19 pandemic, in-person engagements were deemed too risky to both participants and researchers. Therefore, Article One conducted all engagement via video conferences and telephone surveys. Engagements included interviews with 14 Meta representatives; 22 Philippines-based stakeholders, including civil society organizations, journalists, academics, and human rights defenders; and 10 international experts. In addition, Article One partnered with Nielsen to conduct a survey of 2,000 Facebook users in the Philippines, including 1,500 urban users, 500 rural users, and 75 journalists. These engagements began the first week of March and concluded the first week of June 2020 and this report was finalized in August 2021.

Article One applied the methodological approach outlined in the UNGPs to determine the saliency of and Meta's responsibility for the actual and potential impacts surfaced during the assessment. To determine relative priority of salient human rights impacts, Article One

assessed their scale, scope, and remediability. Using guidance from UN Human Rights, Article One evaluated Meta's relationship to adverse impacts by examining whether Meta may have incentivized harm, facilitated harm, failed to conduct adequate human rights due diligence, or failed to recognize the extent to which the company knew or should have known about adverse impacts.³

Article One then developed a series of recommendations to support Meta in maximizing positive human rights impacts and mitigating adverse ones in the Philippines.

II. CONTEXTUAL ANALYSIS

Despite declining poverty rates in the Philippines, the country continues to face human rights challenges, in particular, related to President Rodrigo Duterte's ongoing anti-drug campaign. While official policy calls for the arrest of alleged drug users and dealers, human rights experts report that, in practice, extrajudicial killings of drug dealers and users regularly occur. Recent estimates suggest there have been over 22,000 extrajudicial killings of this nature.⁴ Amnesty International views extrajudicial killings in the drug war as tantamount to crimes against humanity. Furthermore, Amnesty International has reported that no meaningful investigations or prosecutions for any of the drug war killings have taken place, that police in the Philippines operate with near total impunity, and that Duterte has pledged to pardon police officers implicated in these killings.⁵

As part of the drug war, the Duterte administration is accused of persecuting critics, political foes, journalists, and HRDs. By labelling



Step 1: Country Engagement



Step 2: Desk Review



Step 3: Stakeholder Mapping



Step 4: Stakeholder Engagement



Step 5: Meta & International Expert Engagement



Step 6: Analysis & Internal Report

² Human rights defenders are defined as "all persons, who individually or in association with others, act to promote or protect human rights peacefully." Journalists, who report on human rights issues, may be considered within this definition. ³ OHCHR: "B-Tech: Taking Action to Address Human Rights Risks Related to End Use" (2020). ⁴ Human Rights Watch: "World Report" (2019). ⁵ Amnesty International, "Philippines: UN investigation urgently needed into Duterte administration's murderous 'war on drugs'" (2019).

these groups as communists—i.e., by “red tagging” them—the Philippines Department of Justice places them at risk of extrajudicial execution.⁶ Between 2016 and 2019, there were 128 documented attacks and threats against the press, including physical attacks, death and bomb threats, red-tagging, and smear campaigns.⁷

These restrictions have only worsened during the current pandemic. In March 2020, Duterte signed into law the Bayanihan to Heal as One Act, which authorizes him to exercise special powers to implement the policies pursuant to the declaration of a state of national emergency. The act criminalizes “creating, perpetrating, or spreading false information regarding the COVID-19 crisis on social media and other platforms,” and has, according to human rights organizations, been used to chill dissent.⁸

III. META'S MARKET PRESENCE

Facebook is widely used in the Philippines. According to 2019 data, 71 percent of Filipinos have internet access. Of that cohort, 97 percent connect to the web through Facebook.⁹ Three of the top five social media platforms in the country are owned by Meta: Facebook, Facebook Messenger, and Instagram. Not only is a large portion of the country on Facebook, but studies suggest that most people spend more than 10 hours a day on social media—twice the global average—and that those with access to the internet trust social media more than they trust traditional news sources.¹⁰

One reason for Meta's market dominance was the introduction of its Free Basics program in 2013. Free Basics provided free access to Facebook and a “*tightly curated set of web content*.”¹¹ By subsidizing access to Facebook, multiple stakeholders reported that “*Facebook has become synonymous with access. Filipinos don't see Facebook as a distinct app, they just see it as the internet*.”¹² The fact that leaving the platform costs money creates a “*closed information environment*” within the country.¹³ In May 2020, Free Basics was reinvented as Discover, an app that treats all websites on the internet equally.

IV. POSITIVE HUMAN RIGHTS IMPACTS

Facebook is not only widely used in the Philippines, it is widely popular. Article One repeatedly heard statements such as the following: “*Filipinos love Facebook. We go to Facebook for anything and everything: Job opportunities, education, personal connections, and work-related matters are on Facebook*.”¹⁴

Such a statement is supported by data from Article One's survey of Facebook users in the Philippines. According to the survey:¹⁵

- 85 percent of respondents found Facebook very useful to connect with friends and family in the Philippines, while 80 percent found it very useful to connect with friends and family abroad.

- 61 percent of respondents found Facebook very useful for accessing entertainment.
- 52 percent of respondents found Facebook very useful for accessing news and information.

In some cases, these positive impacts promote the realization of human rights. For example, Facebook has provided opportunities for small businesses to access new markets, both domestically and internationally. It has provided opportunities for groups who have traditionally been excluded from formal employment to develop and market small businesses from their homes. For historically disadvantaged groups, these opportunities are essential to realizing their human rights to work and to an adequate standard of living. For example, 84 percent of female respondents in Article One's survey reported that Facebook was “somewhat useful” or “very useful” when it came to work or business, as opposed to 78 percent of men.

Civil society groups, women, LGBTQ+ individuals, HRDs, and other groups have historically found limited platforms for their activism. Facebook changed that by allowing these groups to communicate and connect with a broader network in a fast and cost-effective manner. As one stakeholder reported: “*Facebook has definitely allowed critics of [the Duterte] government to have a platform. You don't have to belong to a newspaper anymore to have a voice*.”¹⁶ This positively affects not only the right to free expression, but also the right to participate in government.

⁶ Human Rights Watch: “World Report” (2019). According to the Supreme Court of the Philippines, red-tagging refers to: “*the act of labeling, branding, naming and accusing individuals and/or organizations of being left-leaning, subversives, communists or terrorists (used as) a strategy...by State agents, particularly law enforcement agencies and the military, against those perceived to be 'threats' or 'enemies of the State*.” ⁷ Freedom House: “Freedom in the World Report: The Philippines” (2020). ⁸ Baker McKenzie: “[Philippines: President Duterte Signs Law to Respond to COVID-19 Pandemic](#)” (2020). ⁹ Hootsuite: “[Digital Philippines](#)” (2019); and NYTimes: “[Soldiers in Facebook's War on Fake News Are Feeling Overrun](#)” (2018). ¹⁰ Inquirer.net: “[Online Filipinos trust social media more than traditional media — poll](#)” (2017). ¹¹ Free Basics provides users with free access to the Philippine Star, the Inquirer, the Sun Star and Rappler. Axios: “[The high price of free Facebook in the Philippines](#)” (2018); and Global Voices: “[Free Basics in Real Life](#)” (2017). ¹² Article One interview with Philippines-based stakeholder in April 2020. ¹³ Meta stakeholder interview in November 2018 and Article One interview with Philippines-based stakeholder in April 2020. ¹⁴ Article One interview with Philippines-based stakeholder in April 2020. ¹⁵ Article One commissioned survey of Facebook users in the Philippines conducted in June 2020. ¹⁶ Article One interview with Philippines-based stakeholder in April 2020.

Stakeholders also reported that Meta's suite of products has become even more essential for the protection of human rights during COVID-19. In the words of one stakeholder: "*Facebook is essential and necessary for NGOs and activists to put messages out on COVID-19 and how governments have been handling it.*"¹⁷ Several human rights organizations engaged by Article One reported that they are alerted through social media to potential human rights violations related to government responses to the virus. In addition, as the pandemic has forced the closure of schools and the migration of classes online, stakeholders reported that Facebook Live was being used by the Philippines Department of Education to train teachers on how best to engage students there.¹⁸

V. SALIENT HUMAN RIGHTS RISKS

Meta faces multiple salient human rights risks in the Philippines. These human right risks largely relate to the ways in which the platform exacerbates existing tensions and risks, including misinformation and disinformation, online harassment, incitement to violence, surveillance of vulnerable groups, sexual exploitation of minors, human and organ trafficking, and terrorist organizing.

For most of these human rights risks, there is clear guidance from human rights standards and domestic legislation to inform a rights-compatible response from Meta. However, in the case of misinformation and disinformation and the online harassment of vulnerable groups, the balance between free expression and the

protection of other rights—including the right to participate in government (UDHR 21), the right to non-discrimination (UDHR 2), and the right to security of persons (UDHR 3)—is less clear.

As the UN has outlined:

*All human rights are universal, indivisible and interdependent and interrelated. Nowhere is this interdependence more obvious than in the discussion of freedom of expression in relation to other human rights. The realization of the right to freedom of expression enables vibrant, multi-faceted public interest debate giving voice to different perspectives and viewpoints. Respect for freedom of expression has a crucial role to play in ensuring democracy and sustainable human development, as well as in promoting international peace and security.*¹⁹

Where questions of rights-compatibility arise, we assess existing international guidance and base our recommendations on that guidance and on insights from leading experts on free expression and human rights.

A. DISINFORMATION AND MISINFORMATION²⁰

Online and offline political disinformation and health misinformation are common in the Philippines. Facebook was used as a key platform to promote political disinformation during the 2016 and 2019 elections. It is widely believed that disinformation campaigns were organized and funded by Duterte and his political supporters. However, the widespread nature of online political operations in the Philippines extends well beyond the Duterte administration to candidates across the political spectrum.²¹ Political disinformation was not only used to influence voter perceptions, it was also used to target and attack political opponents, including attacks on Senator Leila de Lima in 2016.

Health misinformation was also found to be a concern in the Philippines. Misinformation in 2017 regarding the dengue vaccine

87% of population-level respondents and **92%** of journalists reported personally seeing political misinformation on Facebook

74% of population-level respondents and **83%** of journalists reported personally seeing health misinformation.

¹⁷ Article One interview with Philippines-based stakeholder in April 2020. ¹⁸ Article One interview with Philippines-based stakeholder in April 2020. ¹⁹ UN Human Rights Council: "[Report of the United Nations High Commissioner for Human Rights on the expert workshops on the prohibition of incitement to national, racial or religious hatred](#)" (2013). ²⁰ Disinformation includes all "forms of false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit." Misinformation on the other hand "is false information, but the person who is disseminating it believes it to be true." European Commission: "[A multi-dimensional approach to disinformation](#)" (2018) and Access Now: "[Fighting Misinformation And Defending Free Expression During Covid-19: Recommendations For States](#)" (2020). ²¹ Ong, Jonathan and Vincent Cabañes: "Architects of Networked Disinformation: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines" and Ong, Jonathan, et al. "Tracking Digital Disinformation in the 2019 Philippine Midterm Election" (2019).

Dengvaxia promoted both on Facebook and on other channels may have contributed to a decline in confidence in vaccines among Filipino parents. Data from before and after the Dengvaxia misinformation campaigns shows that public confidence fell from 82 percent in 2015 to only 21 percent in 2018.²² In 2020, a different public health crisis emerged: the global COVID-19 pandemic. Whereas the Dengvaxia misinformation campaign was largely propagated by the government, Filipino authorities and Meta have clamped down on the spread of misinformation related to the virus.

Disinformation and misinformation on Facebook may contribute to adverse impacts on:

- Life and security of person (UDHR 3) if misinformation contributes to offline harm.
- Access to information (UDHR 19).
- Access to political participation (UDHR 21).
- Health (UDHR 25).
- Freedom from unlawful attacks on one's honor and reputation (ICCPR 17).
- Truth (Resolution 2005/66).

B. ONLINE HARASSMENT

Online harassment has been widely reported on Meta's platforms in the Philippines. Targeted groups include:

- **Journalists:** Female journalists and those focused on government accountability face significant online harassment.²³ According to Rappler founder Maria Ressa, the attacks are “*very personal, criminal, actually: ‘I will kill you. I will rape you.’ I mean, I think I’ve been called every animal you can think of and every threat.*”²⁴ In 2016, at the height of her attacks, Ressa received approximately 90 hate messages per hour. “*Facebook has promoted self-censorship for journalists. No one should have to live through what we get,*” she said.
- **Political Dissidents:** While both defenders and critics of the Duterte administration are attacked online, many stakeholders reported that Duterte defenders are organized and, in many cases, deployed to silence dissent. One stakeholder reported, “*If you criticize the government for being incompetent or you accuse them of not doing their job, all the trolls will target you. It’s a pattern. Many of my friends have been victim to that.*”²⁵
- **LGBTQ+ Users:** LGBTQ+ survey respondents reported the highest incidence of attacks on Facebook, including doxxing²⁶ and threats of arrest, death, and rape.²⁷ Of LGBTQ+ individuals surveyed, 14 percent reported experiencing offline harm due to engagements on Facebook.²⁸

- **Ethnic Chinese Users:** Stakeholders reported high rates of harassment of ethnic Chinese Facebook users in the Philippines. This relates to ongoing geopolitical challenges between China and the Philippines and has been heightened during the COVID-19 epidemic.²⁹

35% of surveyed journalists reported receiving death threats

Article One found that speech on Meta's platforms in the Philippines may have infringed on the right to:

- Dignity (UDHR 1).
- Non-discrimination (UDHR 2).
- Security of person (UDHR 3), if online harassment results in offline harm, a result reported by 21 percent of survey respondents.
- Privacy (UDHR 12).
- Freedom of expression and assembly (UDHR 19 and 20).
- Political participation.
- Freedom from unlawful attacks on one's honor and reputation (ICCPR 17).

²² London School of Hygiene & Tropical Medicine: “[Dramatic drop in public confidence after Philippines dengue vaccine controversy](#)” (2018). ²³ Article One interview with international stakeholder in May 2020. ²⁴ Recode Decode: “[Kara Swisher interview with Maria Ressa](#)” (2018). ²⁵ Article One interview with Philippines-based stakeholder in May 2020. ²⁶ According to DHS, doxxing refers to gathering an individual's Personally Identifiable Information and disclosing or posting it publicly, usually for malicious purposes such as public humiliation, stalking, identity theft, or targeting an individual for harassment. ²⁷ Article One commissioned survey of Facebook users in the Philippines conducted in June 2020. ²⁸ Article One commissioned survey of Facebook users in the Philippines conducted in June 2020. ²⁹ In particular, the Philippines has rejected China's declaration of two new districts that cover areas in the West Philippines Sea. CNN: “[Philippines Rejects China's ‘illegal’ declaration of Two New Districts that Include West PH Sea Areas](#)” (2020).

C. INCITEMENT OF VIOLENCE

Of particular importance for Meta in the Philippines is the degree to which its platforms expand the dissemination and reach of speech that incites hatred and violence. During the assessment, stakeholders largely reported the involvement of government officials—whether directly posting on Facebook, making statements through official government Facebook pages, or being quoted in news articles—in inciting violence. According to OHCHR, statements by Duterte and other political leaders amounted to “incitement to violence.”³⁰

Those most vulnerable to incitement include:

- **Human Rights Defenders:** Many HRDs engaged by Article One reported being attacked on Facebook with threats of death and rape. One powerful tool is the tactic of “red-tagging” HRDs who have been critical of the administration’s human rights abuses.³¹ According to KARAPATAN, 167 red-tagged individuals have been killed in the Philippines since 2016.³² As one stakeholder reported, “*For people in provinces, if you are [red] tagged you are a target. First it is online and then it is offline.*”³³
- **Alleged Drug Users and Dealers:** Social media platforms, including Facebook, have been used to harass alleged users and dealers. Stakeholders did not report the use of Facebook to plan and organize extrajudicial killings against alleged drug users or dealers. Despite this, stakeholders reported that Facebook was used to doxx alleged users and dealers, threaten and harass

them online, and boast about offline attacks. According to multiple stakeholders, online discussion around drug users is largely focused on, as one stakeholder put it, “*promoting the idea that you have to kill drug addicts to make the Philippines a safer country.*” Given this, stakeholders raised concerns about the dehumanizing tone of the conversations about drug use on Meta’s platforms.

Article One found that speech inciting violence on Meta’s platforms in the Philippines may have infringed on the right to:

- Dignity (UDHR 1).
- Non-discrimination (UDHR 2).
- Security of person (UDHR 3), if online harassment results in offline harm, a result reported by 21 percent of survey respondents.

D. SURVEILLANCE

Human rights organizations assert that HRDs across the Philippines face surveillance of their activities.³⁴ This includes:

- **Online Surveillance:** HRDs engaged by Article One reported that their online activity was being monitored by government agents who used the information to raid offices and bring staff in for questioning.³⁵ According to one HRD, “*Each day, we see more and more cases of people being picked up by police because they have posted their opinion on Facebook.*”

- **Requests for User Data:** In the first half of 2020, Meta received 35 law enforcement requests.³⁶ Each request Meta receives is reviewed for legal sufficiency and whether the request is in accordance with Meta’s global policies, including the GNI Principles. In the first half of 2020, Meta’s compliance rate was 23 percent in response to law enforcement requests.³⁷

In addition, according to OHCHR, the special powers the government granted itself to respond to COVID-19 have resulted in increased criminal charges against “*social media users posting content critical of government policies and actions.*”³⁸ This suggests that government surveillance extends beyond HRDs to include everyday users of the internet.

Government surveillance of users’ Facebook activity may infringe on the right to:

- Security of person (UDHR 3).
- Privacy (UDHR 12).
- Freedom of expression and assembly (UDHR 19 and 20).

³⁰ OHCHR: “[Report of the United Nations High Commissioner for Human Rights on the situation of human rights in the Philippines](#)” (2020). ³¹ Rappler: “[Lives in danger as red-tagging campaign intensifies](#)” (2020). ³² Rappler: “[Lives in danger as red-tagging campaign intensifies](#)” (2020). ³³ Article One interview with Philippines-based stakeholder in May 2020. ³⁴ US Department of State: “[Country Report: Philippines](#)” (2019); and Article One interview with Philippines-based stakeholders in April and May 2020. ³⁵ Article One interview with Philippines-based stakeholder in May 2020. ³⁶ Meta [Transparency Report](#) (2020). ³⁷ Article One email conversation with Meta representative in June 2020. ³⁸ OHCHR: “[Report of the United Nations High Commissioner for Human Rights on the situation of human rights in the Philippines](#)” (2020).

E. ONLINE SEXUAL EXPLOITATION

Online sexual exploitation takes multiple forms in the Philippines and affects both children and adults. In 2018, the Office of Cybercrime at the Philippines Department of Justice received at least 600,000 cyber tips (across all web platforms) of sexual images of Filipino children.³⁹ Given these numbers, an estimated one in five Filipino children are vulnerable to online sexual exploitation of children (OSEC).⁴⁰

According to Plan International, an independent development and humanitarian organization, predators use multiple online channels, including chat rooms, instant messaging apps, online forums, emails, personal and group websites, and social networking sites. Meta products, including Facebook, Facebook Dating, Watch Parties, Messenger, and WhatsApp, have been used by predators to various extents in either grooming or abusing children.

In addition, the Philippines continues to experience cases of revenge porn or non-consensual pornography.⁴¹ This act is sometimes paired with “sextortion,” where individuals are extorted with threats to publish compromising images or videos of them.

OSEC and non-consensual pornography on Meta’s platforms infringe on the right to:

- Dignity (UDHR 1).
- Security of person (UDHR 3).
- Privacy (UDHR 12).

38% of population-level respondents and **59%** of journalists reported personally seeing explicit images of children on Facebook.

- Freedom of expression (UDHR 19).
- Right to work (UDHR 23) and adequate standard of living (UHDR 24).
- Protection from sexual exploitation and abuse (CRC 34).
- Freedom from unlawful attacks on one’s reputation (ICCPR 17).

F. HUMAN ORGAN TRAFFICKING

Human and organ trafficking remain challenges across the Philippines and are increasingly facilitated online, including on Meta’s platforms. The types of trafficking on Meta’s platform include commercial adoption, organ sales, and labor trafficking.⁴²

On Facebook, commercial adoption can occur in closed groups with brokers, parents, and buyers often hiding behind fake profiles. Once initial contact is made, parties typically move from the Facebook platform to messaging apps, such as Facebook Messenger.⁴³ Similarly, organ donors advertise or reach out to buyers directly through Facebook groups. Donors share their contact information,

blood type, and price online before negotiating details through apps like Messenger.⁴⁴ Finally, when it comes to labor trafficking, online posts offering overseas positions often seem legitimate, but may ultimately force applicants into situations of debt bondage.

Human and organ trafficking can impact the right to:

- Dignity (UDHR 1).
- Security of person (UDHR 3).
- Freedom from slavery (UDHR 4) .
- Favorable conditions of work (UHDR 23).
- Freedom from sale (CRC Protocol 1).
- Freedom from organ sale (CRC Protocol 2).

³⁹ Rappler: “FAST FACTS: Why online sexual exploitation of children happens in the Philippines” (2020). ⁴⁰ Rappler: “Philippines top global source of child pornography—UNICEF” (2017). ⁴¹ Fourteenth Congress “REPUBLIC ACT No. 9995” (2009) and Article One email conversation with Meta representative in June 2020. ⁴² ILO data finds that one million Filipinos leave the country every year seeking work abroad. Thirty-five percent of population level respondents to Article One’s survey reported seeing posts from illegal recruiting agencies for jobs overseas. ILO: “Labor migration in the Philippines.” ⁴³ CNA: “Facebook: An emerging black market for Philippines’ baby trade” (2020). ⁴⁴ CNA: “Kidney for sale: How organs can be bought via social media in the Philippines” (2019).

G. EXTREMIST ACTIVITY

Militant groups continue to operate in the Philippines, with a strong presence in Mindanao and the Sulu archipelago. Extremist organizing occurs on Facebook in the Philippines despite strong policies against it.

Research conducted by the Asia Foundation and Rappler in 2018 looked at extremist activity on Facebook and found that extremist messaging is highly localized, referring to local grievances at the municipal and provincial levels, and that content is often spread in Moro languages, particularly Maranao, Maguindanaoan, and Tausug.⁴⁵ These trends, among others, pose specific challenges to the detection of extremist content on the platform.

Terrorist activity on Meta's platforms can result in infringements on the right to:

- Dignity (UDHR 1).
- Security of person (UDHR 3).

27% of population-level respondents personally seeing terrorist content on Facebook

⁴⁵ The Asia Foundation and Rappler: "[Understanding Violent Extremism Messaging and Recruitment on Social Media in the Philippines](#)" (2019).

VI. META'S RESPONSE & RECOMMENDATIONS

Under the UNGPs, Meta should exercise due diligence to mitigate adverse human rights impacts of its business. However, prior to 2018, Article One found that Meta's response was slow and, at times, did not sufficiently meet the UNGPs, which, as a result, may have potentially exacerbated impacts. A company representative acknowledged that prior to the Duterte election, it was *"too slow to respond"* and failed to staff its operations appropriately to respond to the unique challenges presented in the Philippines.⁴⁶

Since 2018, however, the company has taken a more proactive approach. It has established a local presence in Manila and has formed partnerships and undertaken open engagement with many civil society organizations and government agencies, including the Philippines Department of Education. In addition, while not framed explicitly as human rights due diligence internally, many of the steps that Meta has taken have been designed to identify and mitigate human rights risks. These steps included:

- Combatting integrity challenges, including removing coordinated inauthentic behavior.
- Establishing classifiers for problematic content, including in Tagalog.
- Investing in content reviewers and limiting, if not eliminating, the backlog for priority issues.
- Establishing a Trusted Partner network and onboarding Third Party Fact Checkers.

- Investing in public-private partnerships in promoting digital literacy, digital parenting, and digital advocacy training nationwide.

Meta's country-level efforts, however, have drawn online attacks against the company's representatives both in the Philippines and globally. Harassment of Meta representatives typically occurs after the company takes measures that are unfavorable to government messaging on the platform.⁴⁷

These country-level investments are supported by greater human rights accountability at Meta's headquarters, including building out strategic response teams focused on offline harm and imminent risk issues; the creation of election integrity teams; the establishment of a civic response team to represent at-risk countries; the development of Meta's Oversight Board; and the hiring of the company's inaugural Human Rights Director. These steps were recognized by stakeholders who largely praised Meta's efforts to combat terrorist activity and online sexual exploitation.

However, some stakeholders still find Meta's efforts to address disinformation, online harassment, and incitement to violence, in particular, to be insufficient. This includes both policy concerns, such as a lack of protections for certain public figures, and operational concerns, such as the limited ability to address platform-coordinated disinformation campaigns that rely on real profiles. As one stakeholder argued: *"Facebook is doing some things, but what they are doing is not enough. They need to do more. We appreciate that*

*there are attempts to address all these issues, even this conversation is an attempt, but we need to do more because harmful impacts are being spread."*⁴⁸

Many stakeholders felt that Meta's efforts were largely a "band-aid solution" and that real change needed to occur at the level of the business model.⁴⁹ Concerns by Philippines-based stakeholders reflect global efforts by human rights organizations, including Amnesty International and OHCHR's B-Tech project, to understand and better mitigate business model challenges to human rights.⁵⁰ These efforts suggest an ongoing need for Meta to look holistically at its human rights impacts and integrate human rights considerations into key business decisions in an ongoing fashion, as outlined in the Corporate Accountability section of the recommendations below.

Based on the findings of the assessment, Article One encouraged Meta to implement key recommendations, including those summarized in the following table:

⁴⁶ Article One interview with Meta representative in March 2020. ⁴⁷ These threats combined with concerns about the safety of stakeholders who participated in the assessment contributed to Meta's decision to issue this executive summary in lieu of the full report. ⁴⁸ Article One interview with Philippines-based stakeholder in April 2020. ⁴⁹ Article One interview with Philippines-based stakeholder in April 2020. ⁵⁰ Amnesty International: *"Surveillance Giants: How the business model of Google and Facebook threatens human rights"* (2019).

RECOMMENDATIONS

MISINFORMATION & DISINFORMATION

- Develop a risk-mitigation plan for the 2022 Presidential elections, focusing especially on the organized use of disinformation campaigns.
- Expand efforts to track and remove coordinated disinformation campaigns that rely on real profiles.
- Implement the recommendations of Meta's [Civil Rights Audit](#) related to voter interference globally, including in the Philippines.⁵¹
- Scale Meta's commitment to the [EU Code of Practice on Disinformation](#) to include the Philippines.⁵²
- Expand Ad Library to the Philippines to ensure transparency on political advertisements.
- Update policies to forbid political ads that include disinformation on Meta's platforms. This could include having political ads fact-checked by IFCN-certified fact checkers before they are approved for dissemination on Meta's platform.
- Include state actors in the existing *Dangerous Individuals and Organizations* policy.
- Assess opportunities to expand fact-checking of political leaders beyond misinformation that may result in imminent harm.
- Provide financial support to cover the cost of security and mental health support for fact checkers at partner organizations.
- Expand existing misinformation and real-world harm policy to include further categories of health misinformation.
- Expand investment in digital literacy training and provide funding for journalism schools in the Philippines.

ONLINE HARASSMENT

- Enable and empower users to take control of their security and privacy by expanding and increasing awareness of platform-based reporting functions, including through more visible product tools.
- In line with guidance from [Amnesty International](#), "offer personalized information and advice based on personal activity on the platform" (e.g., share guidance on privacy and security settings when users make a report of violence and abuse). This guidance should not contain private information, but rather direct users to resources and tools that can help mitigate future risks.
- Strengthen policy measures against harassment, particularly the organization, incitement, or coordination of harassment, against users.
- Adopt a "public interest" consideration, in place of a "public figures" consideration, and provide policy exceptions and product protections for at-risk public figures, including HRDs and journalists.
- Consider partnering with UN agencies and CSOs to develop a program to raise awareness of HRDs.

⁵¹ As this assessment focused on the Philippines the recommendations shouldn't be taken to suggest observations as they relate to other markets. ⁵² As this assessment focused on the Philippines the recommendations shouldn't be taken to suggest observations as they relate to other markets.

RECOMMENDATIONS

-
- INCITEMENT OF VIOLENCE**
- Formally adopt the Rabat Plan “six-part test” into Meta’s Community Standards implementation, specifically in relation to violence and incitement.
 - Explore opportunities to share platform insights (not specific user or content data) with accountability or preventive bodies via Data for Good and/or other mechanisms.
-

- SURVEILLANCE**
- Collaborate with civil society to advocate for the passing of the Human Rights Defenders Bill.
 - Develop a standalone HRD policy and implement recommendations developed as part of an HRD assessment commissioned by Meta.
 - Co-develop technical tools with HRDs that work to protect their rights while on Meta platforms.
-

- ONLINE SEXUAL EXPLOITATION**
- Use leverage with the Philippines government to expand protections against OSEC
 - Commission a child rights impact assessment of Messenger Kids.
 - Evolve default privacy setting for minors (anyone under the age of 18):
 - Friend requests: “everyone” to “friends of friends”
 - Friend list: “public” to “only me”
 - Email search: “everyone” to “only me”
 - Telephone search: “everyone” to “only me”
 - Do not include minors in “Friends to add” recommendations
 - Do not link search engine results to Facebook profiles for those under the age of 18
 - Ensure the location of minors is never made available⁵
 - Consider requiring children to review a child-friendly version of the Community Standards when signing up for a Facebook account and develop child-friendly safeguarding tools, including guidance on privacy settings, data collection, and reporting of grooming.
 - Consider launching a child online safety month or comparable effort to raise awareness of risks and tools through Meta’s platforms.

⁵³ Recommendation provided by UNICEF Philippines.

RECOMMENDATIONS

TRAFFICKING

- Take efforts to exclude from Facebook search results content related to the sale of human organs.
- Update human exploitation policy to prohibit all sale of children, even in cases where the sale may be legally permissible.
- Conduct on-platform awareness-raising training related to the risks of illegal recruitment and how to detect potentially exploitative content.

EXTREMIST ACTIVITY

- Further invest in Moro language capabilities, particularly highlighting key terms in Maranao, Maguindanaoan, and Tausug, to allow for more accurate, proactive review of potential terrorist content.

CORPORATE ACCOUNTABILITY

- Develop a standalone Human Rights Policy to guide Meta's approach to managing human rights.
- Consider adding an independent board member responsible for human rights and/or establishing a Human Rights Advisory Board, similar to the Safety Advisory Board.
- Build on transitional justice work to contribute to broader recognition of human rights harms in countries with legacies of human rights abuses.
- Continue efforts to align with the [Santa Clara Principles](#).
- Formally integrate human rights expectations into the Responsible Innovation process, requiring that all "high risk products" undergo a human rights review.
- Assess which specific aspects of Meta's business model are more or less likely to impact human rights adversely and develop mitigation measures based on the assessment results.

VII. ADDENDUM: COVID-19 UPDATE (AUGUST 2021)

Human rights organizations assert that the Philippine government's response to the COVID-19 pandemic contributed to a deteriorating human rights environment in 2020 and early 2021, both offline and online, with an erosion of political and civil rights mirrored by a decline in internet freedom.⁵⁴

Given the significant human rights impacts of the pandemic, Meta commissioned Article One to conduct a follow-up assessment to better understand how the worsening human rights context was playing out on Facebook.⁵⁵ Based on stakeholder engagement and desk review, the following key findings emerged:

- COVID-misinformation and government surveillance:** Misinformation regarding COVID-19 remains a pressing problem. According to Rappler, 40 percent of the claims it fact-checked on social media in 2020 related to the COVID-19 pandemic.⁵⁶ Health misinformation can impact on the right to health and the right to access information. Stakeholders also reported that Philippine police monitor Facebook posts to identify and arrest people who are alleged to violate COVID-19 lockdown rules.
- Political disinformation:** While political disinformation was flagged as a concern in the 2020 assessment, stakeholders engaged in August 2021 raised additional concerns given the 2022 elections. These included Meta's fact-checking program omitting misinformation spread via Facebook Messenger, the exemption of politicians' posts from fact checking, and the ongoing presence of fake accounts on the platform that form disinformation networks and harass activists.⁵⁷ Political disinformation can impact users' rights to free expression, to access to information, and to political participation.
- Targeting of HRDs and journalists:** In July of 2020, Duterte signed the Anti-Terrorism Law, which human rights advocates argue includes vague provisions to unjustly target critics.⁵⁸ The law has led to an increase in terrorist tagging—an evolution from red-tagging HRDs as communists to terrorists. Research on 20 red-tagging cases in 2020 and 2021 conducted by Article One and Meta did not find conclusive evidence that red-tagging on Facebook, alone, is correlated with imminent arrest or murder. However, the killings of several individuals that had been red-tagged both online and offline indicates that red-tagging remains an urgent and severe threat to HRDs and one that can infringe on the right to life and security of person.
- Online sexual exploitation and human trafficking:** The pandemic has exacerbated systemic economic and societal inequalities that are among the root causes of human trafficking and online sexual exploitation.⁵⁹ Capitalizing on people's loss of livelihoods and the increasing amount of time spent online, traffickers used social media to recruit new victims. Reporting of online child sexual exploitation in the Philippines increased by 265 percent between March and May 2020, from the same period in 2019.⁶⁰ Online sexual exploitation and human trafficking impacts victims' rights to life and security of person and to freedom from torture and forced labor. The right of children to be free from sexual exploitation is also impacted.

The pandemic's impact on the Philippines is ongoing and will continue to shape how Meta addresses existing and intensified human rights concerns related to its operations and products.

⁵⁴ Human Rights Watch, "World Report 2021: Philippines" (2021) and Freedom House, "Freedom on the Net Report 2021: Philippines" (2021). ⁵⁵ During the follow up assessment, Article One spoke with an additional 6 stakeholders and 4 Meta representatives. ⁵⁶ Falsehoods about the pandemic included cures and preventive measures, conspiracy theories and predictions, policies implemented by governments, and falsely attributed statements, among others. Rappler, "Fact-checking the falsehoods that came out of 2020" (2021). ⁵⁷ Rappler, "Facebook, Twitter, TikTok, YouTube still open for abuse in PH polls" (2021). ⁵⁸ Amnesty International, "Philippines: Dangerous anti-terror law yet another setback for human rights" (2020); Human Rights Watch, "World Report 2021: Philippines" (2021). ⁵⁹ UN Office on Drugs and Crime, "The Effects of the COVID-19 Pandemic on Trafficking in Persons and Responses to the Challenges" (2021). ⁶⁰ DOJ-OOC clarified that the increase was related to cybertips that needed further review and possible investigation of law enforcement units. Cybertips may also include identical materials reported by several platforms, or digital images of children that were generated, in good faith, where they are not engaged in real or simulated explicit sexual activities nor lascivious exhibition of private parts of the body. 2020 research by NCMCEC found that "higher numbers of reports can be indicative of a variety of things including larger numbers of users on a platform or how robust an ESP's efforts are to identify and remove abusive content." "Save the Children, "Online sexual abuse of children rising amid COVID-19 pandemic" (2021) and BusinessMirror, "DOJ unit reports 264.63% spike in online child pornography from March 1 to May 24" (2020).

Since the original assessment, Meta has continued investments in combatting salient human rights risks surfaced in the 2020 assessment. In addition, the company has invested in combating COVID misinformation at both the global and country-level. At the country-level, these investments have included:

- Facebook Ads/Pages training for the Philippines Department of Public Health's information team;
- Development of the #FactualSharingisCaring information awareness campaign, which was shared via Digital Tayo partners;
- Development of the COVID-19 Messenger Chatbot, KIRA, in partnership with the Philippines Department of Health, Google, Viber, and other platforms; and
- Deployment of #CheckTheFAQs, an online campaign by the Philippines Department of Health against vaccine and COVID-19 misinformation.

Based on these findings and Meta's responses, we recommend the following actions be taken:

1. Initiate/expand information sharing on user engagement with COVID misinformation. This data can help independent researchers understand how misinformation is spread and the efficacy of mitigation tactics.
2. Increase efforts to combat disinformation and misinformation in the Philippines, including:
 - a. Onboarding new Third Party Fact Checkers to ensure sufficient resourcing to address disinformation in advance of the upcoming elections; and
 - b. Apply Meta's heightened penalties for public figures who violate Meta's Community Standard.
3. Increase capacity to address red-tagging, including: building awareness among local civil society of Meta's red-tagging policies, expanding the trusted partner program, and taking steps to review red-tagged content in a timely manner, consistent with the threat of life.
4. Increase the capacity to mitigate the prevalence of online sexual exploitation and human trafficking on Meta platforms, including:
 - a. Investing greater resources in the prevention of child exploitation and sharing of CSAM in unencrypted spaces;
 - b. Investigating new methods and technologies to address the spread of CSAM in encrypted spaces, while still respecting the right to privacy and other human rights.
 - c. Expanding approaches to identify and limit dangerous engagements that could result in exploitation.

ABOUT ARTICLE ONE

Article One is a specialized strategy and management consultancy with expertise in human rights, responsible innovation, and social impact. We envision a world in which business succeeds and all people are free and equal in dignity and rights.

In support of that vision, Article One partners with the world's leading companies to drive transformative change that places people at the center of business.

[ARTICLEONEADVISORS.COM](https://articleoneadvisors.com)

