∞ Meta

Meta Update:
Philippines Human Rights
Impact Assessment

Table of Contents

Introduction

In December 2021, Meta[1] published an [independent Human Rights Impact Assessment](#) (HRIA) on potential human rights risks in the Philippines related to its platforms, along with a detailed, recommendation-by-recommendation [response](#). Meta provided a further update to those responses, with attention to measures put in place prior to the 2022 presidential election, in its July 2022 [Annual Human Rights Report](#).

The project was undertaken by [Article One Advisors](#), independently of Meta, in 2020 and 2021. Article One's methodology was based on the [United Nations Guiding Principles on Business and Human Rights](#) and included qualitative interviews with 32 stakeholders, including journalists, child protection specialists, and civil society activists, as well as a survey of 2,000 Meta users in the Philippines. This implementation update is part of our [broader commitment](#) to meaningful transparency about our human rights due diligence and about our integrity work.[2]

---

[1] On October 28 2021, Facebook, Inc. changed its name to Meta Platforms, Inc. For consistency, this report uses "Meta" to refer to the company both before and after October 28 2021. References to "Facebook" apply only to the social media platform, not the company as a whole. Further, this response references steps taken, or plans to take steps, by Meta as a company regarding a specific entity. Such a statement is not intended to imply that Meta would, or will, take steps regarding all entities. No statement in this response is intended to create — or should be construed as creating — new obligations (legal or otherwise) regarding the application of a policy or procedure to other products or entities. For example (and in contrast to other Meta technologies), WhatsApp is an end-to-end encrypted messaging and calling application with unique human rights touchpoints. This response's discussion of content moderation and related issues on Facebook and Instagram do not apply to WhatsApp. Unless a policy or commitment is specified as applying to WhatsApp, it does not apply to WhatsApp.

[2] Meta's publication of this update should not be construed as an admission, agreement with, or acceptance of any of the findings, conclusions, opinions, or viewpoints identified by Article One, or the methodology employed to reach such findings, conclusions, opinions, or viewpoints. Likewise, while Meta references steps it has taken, or plans to take, that may correlate to points Article One raised or recommendations it made, these also should not be construed as an admission, agreement with, or acceptance of any findings, conclusions, opinions, or viewpoints.

Implementation Update

Overview

The HRIA made 40 recommendations covering incitement of violence, surveillance, online sexual exploitation, trafficking, extremist activity, and corporate accountability. Meta committed to implementing 24 recommendations, partly implementing seven recommendations, and to assess the feasibility of nine recommendations.

For ease of reading, we have categorized this update to implementation of the human rights due diligence (HRDD) recommendations to match the order in which they appear in the HRDD executive summary. We are categorizing our responses as follows:

- **Complete:** We have completed full or partial implementation in line with our response to the recommendation and will have no further updates in the future.
- **In progress:** We are continuing to make progress on our response to the recommendation and may have further updates on the recommendation in the future.
- **No further updates:** We will not implement the recommendation and will have no further updates in the future.

Updates to our responses to recommendations

1. Misinformation and disinformation

a) Develop a risk-mitigation plan for the 2022 presidential election, focusing especially on the organized use of disinformation campaigns.

Original commitment: Implementing
Status: Complete

In line with our approach to other major global elections, Meta made it a high priority to invest in risk-mitigation and preparation efforts for the 2022 presidential election in the Philippines. These efforts were informed and guided by recommendations of the HRIA, among other things.

We built new products and developed stronger policies in collaboration with the Commission on Elections, election watchdogs, independent fact-checkers, and civil society organizations. Details on these actions can be found in our Newsroom post on how Meta is preparing for the Philippines' 2022 general election, in our response to the HRIA, and in the Preparation for Philippines Presidential Election 2022 section of our 2022 Annual Human Rights Report.

We also took a proactive action on malicious actors, which combines expert investigations and scaled detection and enforcement. Details on this action can be found in our Q2 2022 Quarterly Adversarial Threat Report.

b) Expand efforts to track and remove coordinated disinformation campaigns that rely on real profiles.

Original commitment: Implementing
Status: Complete

We launched a policy against certain mass harassment/brigading in August 2021 — strongly inspired by the experiences documented in the HRIA — after consultation with key human rights stakeholders in the Philippines and around the world.

We have strengthened our ability to take action against coordinated social harms by authentic users. Ahead of the 2022 election, we manually disrupted and took down about 10,000 accounts for violating Inauthentic Behavior (IB) policies, where malicious actors

used tactics to inflate the distribution of content that included election-related posts. We also combined automated and scaled detection of these IB patterns (alongside expert review), resulting in enforcement against over 15,000 additional accounts.

c) Implement the recommendations of Facebook's Civil Rights Audit related to voter interference in the Philippines.

Original commitment: Implementing
Status: Complete

In response to feedback from our global stakeholders and the important work of the Civil Rights Audit in the United States, we have significantly expanded our global policies addressing voter interference. Our Community Standards now [prohibit](#) offers to buy or sell votes; content advocating or providing instructions for voter fraud; misrepresentation of dates, locations, times, and methods for voting or voter registration; misrepresentation of who can vote and requirements for voting; and content implying that people will be infected with COVID-19 (or other diseases) if they vote. These policies apply globally, including in the Philippines.

We've also [introduced](#) measures to reduce the risk that potentially harmful Groups are recommended to people, and we let people know when they're about to join a Group that has Community Standards violations. We limit invitation notifications for these Groups, so people are less likely to join. For full details on Meta's follow-up to the Civil Rights Audit, see [here](#).

d) Scale Facebook's commitment to the EU Code of Practice on Disinformation to include the Philippines.

Original commitment: Assessing feasibility
Status: No further updates

Meta's approach to addressing misinformation and disinformation is global. It includes our partnerships with third-party fact-checkers, as well as our work to support and elevate sources of quality information, to address misrepresentation and the coordinated use of fake accounts and to prohibit misinformation in advertisements in accordance with our Advertising Policies and Community Standards.

We support multi-stakeholder, collaborative, and iterative approaches that self-regulatory initiatives — like the EU Code of Practice on Disinformation and the Australian Code of

Practice on Disinformation and Misinformation — facilitate. Such approaches strike a balance between freedom of expression, safety, and security and provide a forum for ongoing dialogue between stakeholders to find solutions as the digital ecosystem evolves.

e) Expand Ads Library to the Philippines to ensure transparency on political advertisements.

Original commitment: Implementing
Status: Complete

Advertisers in the Philippines are required to complete our ad authorizations process and include "Paid for by" disclaimers on ads about elections, politics, and certain categories of social issues. In early 2022, we began requiring anyone running ads about certain categories of social issues in the Philippines to get authorized and show the organization or person who is running the ad with disclaimers.

Ads about social issues, elections, or politics that run in the Philippines will also appear in the Ads Library so that everyone can see what ads are running, who saw them, and how much was spent. This fully searchable archive stores these ads for seven years. Ahead of the 2022 election, Meta, together with COMELEC (the Philippines Commission on Elections), conducted capacity-building training for political parties and candidates on ads transparency and with non-government organizations for ads about social issues, elections, or politics.

f) Update policies to forbid political ads that include disinformation on Meta's platforms. This could include having political ads fact-checked by International Fact-Checking Network-certified fact-checkers before they are approved for dissemination on Meta's platform.

Original commitment: Implementing in part
Status: Complete

We require advertisers to follow our Advertising Policies, in addition to our Community Standards. We prohibit ads containing claims that have been debunked by third-party fact-checkers, unless they consist of direct and unedited speech by politicians.

We do not allow third-party fact-checking on the direct and unedited speech of politicians because we believe that by censoring the speech of politicians, we'd leave people less informed about what their candidates or elected officials say, and leave politicians less accountable for their words.

Given the volume of advertising and the difficulty of assessing often opinion-based claims included in political and social-issues advertising, it is not a viable or a desirable use of inherently limited fact-checking resources to have each political ad systematically assessed by a third-party fact-checker for accuracy prior to approval.

Such an approach may also limit the journalistic independence of our fact-checking partners, who are currently free to choose the claims they wish to assess. In practice, it may also create an unintended bias against advertisers speaking about certain types of issues or based in certain areas where fact-checking capacity is relatively limited.

g) Include state actors in the existing Dangerous Organizations and Individuals (DOI) Policy.

Original commitment: Assessing feasibility
Status: No further updates

Meta does not currently include state actors in its DOI Policy. However, we have numerous mechanisms in place to prevent state actors from abusing our platform. These include our policies on labeling state-controlled media, as well as our disrupting and taking down disinformation networks, including in the Philippines prior to the 2022 election.

h) Assess opportunities to expand fact-checking of political leaders beyond misinformation that may result in imminent harm.

Original commitment: Implementing in part
Status: Complete

We assess such opportunities on a continuous basis. For example, in 2020–2021, we expanded prohibitions against certain COVID-19, vaccine, and electoral content, as noted in the paragraphs above. We also pioneered new methods of enforcing misinformation policies and claim-checking at scale in the U.S., Myanmar, and other elections.

We are not considering additional Community Standards changes in this area at the time of writing. Our approach to fact-checking is grounded in Meta's fundamental belief in free expression, respect for the democratic process, and the belief that political speech is the most scrutinized speech there is. Just as critically, by limiting political speech we would leave people less informed about what their elected officials are saying — and leave politicians less accountable for their words.

i) Provide financial support to cover the cost of security and mental health support for fact-checkers at partner organizations.

Original commitment: Assessing feasibility
Status: In progress

Meta has [contributed more than $84 million](#) to programs supporting our fact-checking efforts since 2016, including [support for debunking COVID-19 content](#), a [mentorship program](#), a [Climate Misinformation Grant](#), and an [Accelerator](#) to support fact-checkers' businesses. We are always looking for additional ways to support fact-checking partners based on their needs and will consider options for mental health support as we plan for the future. Fact-checking partners who meet the relevant criteria also will be eligible to apply for our human rights defender and journalism safety initiatives.

j) Expand existing misinformation and harm policy to include further categories of health misinformation.

Original commitment: Implementing
Status: Complete

Our Community Standards are a living document, meaning that we are constantly evolving our policies to keep pace with changing online behaviors. Our teams will actively explore additional policy changes related to health misinformation, taking into account important freedom of expression, access to information, and proportionality considerations and [recent guidance from the Oversight Board](#).

k) Expand investment in digital literacy training and provide funding for journalism schools in the Philippines.

Original commitment: Implementing
Status: Complete

Meta has made a major investment in digital literacy in the Philippines, including launching an expanded version of [We Think Digital](#)/Digital Tayo. This includes interactive content on safety and privacy and modules on news verification, community building, and civic engagement. It continues to grow its partner roster of government agencies, youth organizations, non-government organizations, and community-based groups.

In 2023, we introduced new learning modules around building [Digital Foundations](#), promoting [Digital Wellness](#), identifying and making the most of [Digital Opportunities](#), and

Avoiding Scams. These are additions to currently available resources on Digital Parenting (here and here), Digital Engagement, Digital Empowerment, and Digital Citizenship available in Maingat, Magalang, and Mapanuri.

We also updated the library of resources for various audiences to include materials on digital literacy and citizenship, online safety, civic engagement, and avoiding scams, some of which are available in Filipino.

In January 2023, we held a digital literacy focus group discussion with academic experts and Trusted Partners to discuss misinformation and disinformation in the Philippines and how to improve our digital literacy and citizenship efforts.

There is ongoing work on integration of digital literacy into the learning curriculum of pre-service tertiary and basic education teachers. We have developed a Facebook and Messenger School Guide for students, teachers, and home-based learning facilitators that provides key information on our Terms of Service, Community Standards, and use of Messenger Kids, coupled with best practices and safety and privacy tips to aid online learning, in partnership with the Philippine Department of Education and Stairway Foundation. We also supported a project with FYT, a community-oriented independent media outfit, to train local/campus journalists on digital literacy, visual storytelling, and fact-checking in preparation for the 2022 election.

The Meta Journalism Project has also partnered with Reuters to launch a free online digital journalism foundations training program. This includes modules on digital news gathering, news verification and reporting, publishing on social media, and wellness and resilience training while reporting. In June 2021, we published the Community Standards in Filipino, enabling Filipino users to read them in their native language.

2. Online harassment

a) 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.

Original commitment: Implementing
Status: Complete

We have an extensive suite of product tools designed to empower people to address the potentially negative impacts of unwanted contact. These vary across different apps but

include the ability to unfriend, unlike, ignore, block, restrict, and report. See our [Abuse Resources](#) for more information. More detail on recent safety features can also be found [here](#) and [here](#).

For Safer Internet Day 2022, we introduced the [Instagram Parent's Guide](#) in the Philippines to educate Filipino parents on Instagram's safety features and empower them to initiate conversations with their children on online safety and mental well-being. For Safer Internet Day 2023, Meta developed and launched a [child-friendly webtoon series and choose-your-own adventure games](#) to teach teens about privacy, data protection, and responsible online interaction. The webtoon referred to [tools and resources](#) that help teens keep their accounts safe and secure. The press launch included discussions of the new safety tools, which resulted in several pieces of media coverage.

We also launched the [Family Center](#) for children's month in November 2022, serving as a hub where parents and guardians can access materials and resources to help young people build positive online habits.

b) 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.

Original commitment: Implementing
Status: Complete

We agree that providing tailored guidance within our apps is one of the best ways to help people understand our policies and settings. In line with this recommendation, we provide contextually aware information on privacy, safety, and security features throughout the Facebook app. For example, when reporting a post for harassment, we also share one-tap options to block or unfollow the user or Page that made the post, as well as details on the relevant portions of our Community Standards. Other tools available in the Philippines include comment controls (to control who can comment on your public posts) and a profanity filter.

When we consider a user may be particularly vulnerable to security risks, we provide notifications encouraging them to turn on two-factor authentication and other advanced protections. We develop technology to fight behavior that threatens to keep women and vulnerable users offline. We also offer a number of tools to help people control their

experience on Facebook, such as ignoring unwanted messages and blocking other people without them being notified. Victims and their allies can also report violating behavior.

In consultation with stakeholders, we'll continue to explore additional opportunities to provide such information, including to human rights defenders in the Philippines.

c) Strengthen policy measures against harassment, particularly the organization, incitement, or coordination of harassment, against users.

Original commitment: Implementing
Status: Complete

We updated our Community Standards to prevent calls for coordinating harassment in 2020. We also launched a Mass Harassment/Anti-Brigading Policy in August 2021, a significant step in addressing this complex challenge. To reduce the reach of those looking to use our platform for harmful purposes, we now also make it harder to find and join Groups that violate our policies. In addition, when Group members have repeated violations, we now block them from being able to post or comment for a period of time in any Group and prevent them from inviting others to any Groups and creating new Groups. Details on enforcement of our Bullying and Harassment Policy can be found in the most recent [Enforcement Report](.).

d) 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 human rights defenders (HRDs) and journalists.

Original commitment: Implementing
Status: Complete

We launched multiple changes to our Bullying and Harassment Policy including: (i) clarifying that most HRDs and journalists are "limited public figures" with stricter policies prohibiting bullying and harassment than public figures; (ii) stronger protections against gender-based harassment and sexualization, including of public figures; and (iii) a policy prohibiting certain mass harassment ("brigading"), including against HRDs. We also launched highly relevant product protections, such as [Facebook Protect](.) and the [News Page Index](.).

e) Consider partnering with UN agencies and CSOs to develop an HRD awareness-raising program.

Original commitment: Implementing
Status: Complete

Meta has conducted briefings for HRDs on policies around data sharing and Facebook's Community Standards, as well as training on using Facebook for digital advocacy.

In August 2021, we launched the Online Security and Resiliency Training for HRDs with The Asia Foundation. The series aims to equip HRDs with knowledge on how to stay safe and be more resilient online. It includes digital security, mental wellness, and community engagement. As of April 2023, we have trained 80 organizations. The series will continue in the Philippines and potentially in other countries in the region. We also financially support several global response/advocacy mechanisms for HRDs and will explore global campaign options.

3. Incitement to violence

a) Formally adopt the Rabat Plan "six-part test" into Facebook's Community Standards implementation, specifically in relation to Violence and Incitement.

Original commitment: Implementing in part
Status: Complete

Meta is exploring ways to incorporate learnings from the Rabat threshold test into our policies. While the nuance of the test makes it operationally infeasible to directly apply at scale, we have adapted it into two content policy tools, including an escalation-based framework on hate speech attacking concepts, not people (adopted 2021), as well as an updated framework to evaluate discussions of state use of force by governments.

Meta has developed a prototype test for application of the Rabat Plan of Action for use when assessing potential hate speech. We have also adapted the principles under the Rabat Plan into active content policy tools, including escalation-based frameworks to evaluate speech attacking concepts (vs. people) and content involving state threats to use force. In 2020, Meta worked with the UN Office of the High Commissioner for Human Rights to have the test translated and made available in 26 additional languages.

b) Explore opportunities to share platform insights (not specific user or content data) with accountability or preventive bodies via Data for Good or other mechanisms.

Original commitment: Implementing
Status: Complete

In addition to this disclosure, our Human Rights, Stakeholder Engagement, and Public Policy teams routinely discuss insights on broad content trends — including those related to social media and conflict and incitement to violence — with partners in the human rights community, academia, and international organizations. In doing so, we seek to inform policy development and enforcement, support rapid response, develop best practices, and support accountability for human rights violations.

Meta has proactively partnered with Philippines-based organizations on our Data for Good programs since 2017 in areas such as disaster response and health. Partners include the Philippines Red Cross, Philippine Disaster Resilience Foundation, UNICEF Philippines, and the University of Philippines Resilience Institute. We have financially supported, and participated in, human rights multi-stakeholder conversations related to documentation and disclosure and will actively explore additional ways to share additional insights and data with our partners, consistent with our privacy and other legal obligations.

Our Transparency Center, launched in 2021, is a significant resource, as is our quarterly Community Standards Enforcement Report, which includes information on the prevalence of various types of harmful content over time and volumes of such content removed.

4. Surveillance

a) Collaborate with civil society to advocate for the passing of the Human Rights Defenders Bill.

Original commitment: Assessing feasibility
Status: Complete

We are collaborating with civil society to encourage protections for HRDs, but passage of this bill is stalled.

b) Develop a standalone HRD policy and implement recommendations developed as part of an HRD assessment commissioned by Meta.

Original commitment: Implementing
Status: Complete

Our Corporate Human Rights Policy (launched March 2021) specifically references the needs and risks of HRDs. We have also created an HRD-specific cross-functional working group; significantly revised our Bullying and Harassment Policy to address the vulnerabilities of HRDs; launched new restrictions on calling for or engaging in certain mass harassment; and improved our ability to identify and launch relevant product solutions, such as locked profiles.

Together with The Asia Foundation and AHA! Behavioral Design, Meta also started the Online Security and Resiliency (OSR) program. The program featured modules for HRDs that bolster personal and digital capacity. OSR Team Coaching sessions covered Digital Security, Community Engagement, and Mental Health. The program trained over 80 organizations from June 2021 to April 2023.

In addition, Meta launched the Human Rights Defenders' and Journalists' Fund to provide material support to some HRDs, as well as Online Security and Resiliency Training for Human Rights Defenders to equip HRDs with knowledge on how to stay safe and be more resilient online.

c) Co-develop technical tools with HRDs that work to protect their rights while on Meta platforms.

Original commitment: Assessing feasibility
Status: In progress

Meta's Human Rights Team is actively seeking to develop a more consistent set of co-design opportunities on products to support HRDs on our apps.

5. Online sexual exploitation

a) Use leverage with the government of the Philippines to expand protections against online child sexual exploitation.

Original commitment: Implementing
Status: Complete

In line with the UN Guiding Principles on Business and Human Rights, we have actively sought to support outreach, understanding, and consultation around the drafting of

various online child safety bills that aim to address the issue of online sexual exploitation of children and further define the duties of platforms and tech companies in addressing related issues. This work has entailed, among other engagements, attendance at various Technical Working Group meetings.

We also established active, ongoing coordination mechanisms with the [Inter-Agency Council Against Trafficking](), Department of Justice, Department of ICT, and the Department of Education as well as NGO partners [UNICEF,]() [Plan International](), [Stairway Foundation](), and [Child Rights Network](). This allows us to quickly respond to reports and continuously gather local insights and trends that will help us improve our policies, tools, and resources to better protect children online.

Meta ensured that mandated government agencies (Department of Justice, Philippine National Police, National Bureau of Investigation, and Cybercrime Investigation and Coordinating Center) have access to dedicated reporting channels and provided them with a Community Standards refresher and a briefing on how Meta assesses reports of content violating local law.

Further, Meta onboarded the Department of Migrant Workers (DMW) to a dedicated reporting channel in line with its mandate under the DMW Act. Meta provided a similar onboarding for the Office of the Department of Justice (DOJ) Undersecretary-in-Charge of the Inter-Agency Council Against Trafficking (IACAT), the National Coordinating Council Against Online Sexual Abuse or Exploitation of Children-Child Sexual Abuse or Exploitation Materials (NCC-OSAEC-CSAEM) Executive Director, and the DOJ Office of Cybercrime, in line with their mandates under the Anti-OSAEC Law.

We train and work extensively with law enforcement, including the Philippine National Police's (PNP) Anti-Cybercrime Group, the Women and Children's Protection Center, and the National Bureau of Investigation, including to provide training on how to efficiently submit legal requests. We combat child online exploitation including working with the PNP's [Anti-Cybercrime Group](), the [Women and Children Protection Center](), and the National Bureau of Investigation, along with their foreign law enforcement and NGO partners such as Homeland Security Investigations, the [International Justice Mission](), and others.

We report apparent instances of child sexual exploitation appearing on our site from anywhere in the world to the [National Center for Missing and Exploited Children (NCMEC)](), which then refers cases out to global law enforcement agencies. More details can be found [here](). We have also provided multiple training sessions to law enforcement

agencies over the years in the Philippines organized by the [International Centre for Missing and Exploited Children](#) and NCMEC.

In 2022, Meta helped support the efforts of the government to enact two safety bills — The Expanded Anti-Trafficking in Persons Act of 2022 (EATA) and The Anti-Online Sexual Abuse or Exploitation of Children and Anti-Child Sexual Abuse or Exploitation Materials Act (Anti-OSAEC). Meta continued to actively participate in consultations and made recommendations to help ensure that internet intermediaries are able to cooperate and assist law enforcement in investigating and prosecuting violations of the laws.

b) Commission a child rights impact assessment of Messenger Kids.

Original commitment: Assessing feasibility
Status: No further updates

Messenger Kids was developed specifically to meet child safety needs. Product development included extensive consultation with a team of global advisors and independent experts in the fields of online safety, child development, and media from organizations such as the [Family Online Safety Institute](#), [Center on Media and Child Health](#), [MediaSmarts](#), the [Yale Center for Emotional Intelligence](#), and the [Fred Rogers Center](#). In the Philippines, we consulted the Stairway Foundation, the Child Rights Network, UNICEF, and Plan International. We will actively consider this recommendation while also focusing on future child-related product development due diligence.

c) Evolve default privacy setting for minors (anyone under the age of 18):
  1) Friend requests: from "everyone" to "friends of friends";
  2) Friend list: "public" to "only me";
  3) Email search: "everyone" to "only me";
  4) Telephone search: "everyone" to "only me";
  5) Do not include minors in "Friends to add" recommendations;
  6) Do not link search engine results to Facebook profiles for those under the age of 18;
  7) Ensure the location of minors is never made available.

Original commitment: Implementing in part
Status: Complete

To protect minors, we have numerous special safeguards in place, including: requiring everyone to be at least 13 years old before they can create an account; setting their

accounts to share with "friends only" by default; taking steps to prevent default audience options for posts from including "public"; [restricting](#) the visibility of minors' profiles in public search; prevent public sharing of content (including location, contact information, school or birthday, and friend lists); blocking adult-only features (e.g., Facebook Dating); reminding minors that they should only accept friend requests from people they know, and we do not allow unconnected adults to message minors; allowing anyone — whether they use our services or not — to report suspected underage people on Instagram and Facebook. We have dedicated channels where we review these accounts based on these or other reports.

In addition, minors and adults can also turn on Timeline Review to review posts they're tagged in before the posts appear on their timeline. If a minor or anyone else is tagged in something they don't like, they can remove the tag or ask the person to take the post down. On Facebook, a parent or legal guardian can also request that we remove an image of their child under the age of 13 by [filling out this form](#). A young person between the ages of 13 and 18 can request we remove their image using this "Report a Privacy Violation" feature on our Help Center. We restrict adults from messaging teens they aren't connected to or from seeing teens in their People You May Know recommendations, and we won't show them in teens' People You May Know recommendations. As an extra layer of protection, we're also testing removing the message button on teens' Instagram accounts when they're viewed by suspicious adults altogether. Additional information on how Meta products and tools support age-appropriate experiences can be found [here](#).

In February 2023, [Meta announced](#) that Facebook and Instagram are founding members of [Take It Down](#), a new platform designed to proactively prevent young people's intimate images from spreading online. Take It Down lets young people take back control of their intimate images. We also announced new features to make it even more difficult for suspicious adults to interact with teens on Instagram.

We supported the National Center for Missing and Exploited Children (NCMEC) in the development of Take It Down, building on the success of StopNCII, a platform we developed that helps adults stop the spread of their intimate images online. Take It Down was designed with Meta's financial support. Built in a way that respects young people's privacy and data security, Take It Down allows people to only submit a hash — rather than the intimate image or video itself — to NCMEC. Hashing turns images or videos into a coded form that can no longer be viewed, producing hashes that are secure digital fingerprints. We are working with NCMEC to promote Take It Down across our platforms,

in addition to integrating it into Facebook and Instagram so people can easily access it when reporting potentially violating content.

We also use photo-matching technologies that help us detect, remove, and report the sharing of images and videos that exploit children. These photo-matching technologies create a unique digital signature of an image or video (known as a "hash") that is then compared against a database containing signatures (hashes) of previously identified illegal images to find copies of the same image.

We also run these technologies on links from other internet sites shared on our apps and their associated content to detect known child exploitation elsewhere on the internet. All violating content found is reported to the NCMEC, which works with law enforcement around the world.

Finally, we use artificial intelligence and machine learning to proactively detect child nudity and previously unknown and new child-exploitative content, as well as inappropriate interactions with children, sometimes referred to as "grooming." We provide minors with specific education about what it means to post publicly. We also protect sensitive information, such as minors' contact info, school, and birthday, from appearing in searches to a public audience.

d) 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.

Original commitment: Implementing in part
Status: Complete

Meta collaborated with the Child Rights Network, National Privacy Commission, and creators and youth advocates to launch a [child-friendly webtoon series and choose-your-own adventure games](#) about privacy, data protection, and responsible online interaction.

In addition, we are working with the Department of Education and the Stairway Foundation to develop a Facebook Safety Guide for Schools to support school personnel, home-based learning facilitators, and learners in safely navigating Facebook as a communication platform to share important school updates, information, and announcements.

We have already developed child-friendly safeguarding tools and will continue to do so. For example, we introduced safety prompts in Messenger, which are surfaced to minors before interacting with an adult who has been exhibiting potentially suspicious behavior. For example, if an adult is sending a large amount of friend or message requests to people under 18, this tool alerts the recipients to be cautious when interacting with an adult they may not know, educates them on harassment, and empowers them to ignore, block, or report someone when something doesn't seem right. We also [launched a Youth Portal in our Safety Center in 2018](#), a central place for teens that includes education on staying safe and first-person accounts from teens around the world about how they are using technology in new and creative ways. It features tips on things like security, reporting content, and deciding who can see what they share and some guidelines for how teens can get the most out of the internet.

e) Consider launching a child online safety month or comparable effort to raise awareness of risks and tools through Meta's platforms.

    Original commitment: Implementing
    Status: Complete

    We participate regularly in Safer Internet Day activities. As part of 2021 Safer Internet Day, we ran [an online safety education campaign](#) on Facebook and Instagram from February to March 2021 that has been viewed over 18 million times. For Safer Internet Day in February 2023, Meta collaborated with the Child Rights Network, National Privacy Commission, and creators and youth advocates to launch a [child-friendly webtoon series and choose-your-own adventure games](#) about privacy, data protection, and responsible online interaction.

    To help ensure that materials were aligned with child safeguarding standards and the voices of teens included in the process, Meta worked with child safety practitioners and consulted children with disabilities, LGBT children, and teens from diverse socio-economic backgrounds from across the Philippines. The resulting media coverage helped amplify the resources and new safeguarding tools further.

    We have also worked in partnership with the Child Rights Network, Stairway Foundation, Department of ICT, and [Called To Rescue](#) to educate users on Facebook privacy and security tips. Meta annually participates in Safer Internet Day activities.

    Building on our current participation in other national child safety campaigns, at the time of writing, we were running the #ReportItDontShareIt campaign in partnership with the

Inter-Agency Council Against Trafficking (IACAT), Child Rights Network, and the Stairway Foundation. The campaign raises awareness of the risks of sharing child sexual exploitation material, promotes a shift in local attitudes, and encourages people to report such content.

6. Trafficking

a) Take efforts to exclude from Facebook search results content related to the sale of human organs.

Original commitment: Implementing
Status: Complete

We have updated proactive detection tools with the goal to reduce discoverability of exploitative content and incorporate emerging trends in consultation with the Philippines Department of Health and the National Kidney and Transplant Institute.

b) Update Human Exploitation Policy to prohibit all sales of children, even in cases where the sale may be legally permissible.

Original commitment: Implementing
Status: Complete

We have extensive language prohibiting multiple forms of human exploitation, including child selling for illegal adoption, in our existing policies. This has been updated to unambiguously prohibit all sales of children.

c) Conduct on-platform awareness-raising training related to the risks of illegal recruitment and how to detect potentially exploitative content.

Original commitment: Implementing
Status: Complete

In May 2023, we launched Search Interstitials on Facebook to prevent users from responding to fraudulent job posts. We worked with regional organizations to collect and translate the top keywords associated with this type of labor exploitation. The Search Interstitial surfaces a warning message when users type keywords associated with high-risk fraudulent job advertisements leading to labor exploitation. It blocks users from seeing the search results until they opt into seeing them and connects potential victims to organizations that can provide support and resources.

In partnership with [Stop The Traffik](), we also launched a [Labor Trafficking/Domestic Servitude Awareness Campaign]() in 2021 to provide information to potential domestic workers and low-skilled laborers about their legal rights, local recruitment guidelines for overseas job applications, and available helplines. We worked with NGO partners such as Stop the Traffik to deliver targeted prevention and support ad campaigns. More recently, among the new Digital Tayo modules we introduced in 2023 is the [Avoiding Scams]() module, which covers the common types of scams and offers tips and actions on what to do if scammed.

In January 2022, we launched victim support for people who post content on Facebook looking to sell their own organs (self-identified potential victims of organ trafficking). When self-identified potential victims have content taken down for violating our Community Standards, they will also be directed to support resources that we co-designed with external experts. The intervention has been translated into 60 languages. We developed these interventions based on consultations with expert organizations in the Philippines, including the Inter-Agency Council Against Trafficking, the Department of Health, the National Kidney and Transplant Institute, the [Overseas Workers Welfare Administration](), the [Blas Ople Policy Center](), the Department of Foreign Affairs, and the [Philippine Overseas Employment Administration]().

In April 2023, we hosted a forced criminality and human trafficking conference in Bangkok. We brought together law enforcement and NGOs/IOs/civil society stakeholders for meaningful signal sharing and strategizing and to plan the way forward. Law Enforcement agencies in attendance included the Philippines National Police, National Bureau of Investigation, Department of Justice, Bureau of Immigration, Inter-Agency Council Against Trafficking, FBI Manila, and HSI Manila.

7. Extremist activity

a) 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.

   Original commitment: Assessing feasibility
   Status: No further updates

   At the time of writing, Meta has content moderation capacity in Cebuano and Filipino. We'll evaluate how to prioritize additional capacity based on our criteria for global language prioritization and the UN Guiding Principles on Business and Human Rights.

In addition, we have launched [The Resiliency Initiative](#), which aims to promote tolerance, strengthen interfaith and inter-ethnic understanding, and counter violent extremism by helping to build resilient communities across APAC. The initiative delivers a range of capacity-building programs for civil society organizations to help develop localized positive narrative campaigns online that aim to counter intolerance and discrimination — including in the Philippines.

We have also implemented the Boost Your Business training program in the Bangsamoro Autonomous Region of Muslim Mindanao for digital readiness and opening online economic opportunities for the local community, with [Bayan Academy](#).

8. Corporate accountability

a) Develop a standalone Human Rights Policy.

Original commitment: Implementing
Status: Complete

Meta's [Corporate Human Rights Policy](#) was launched March 2021. Meta's dedicated [human rights website](#) was also launched in May 2022 and contains detailed information on Meta's human rights policies and practices, as well as links to its July 2022 [Annual Human Rights Report](#) (which contains an update on implementation of the Philippines HRIA and detail on preparations for the 2022 election). The website will be regularly updated.

b) Consider adding an independent board member responsible for human rights and/or establishing a Human Rights Advisory Board, similar to the Safety Advisory Board.

Original commitment: Assessing feasibility
Status: In progress

Under our Corporate Human Rights Policy, salient human rights issues are periodically reported to the company Board of Directors. Meta has benchmarked business practice-related human rights stakeholder advisory and consultation mechanisms and will seek to implement such a mechanism. The Meta Human Rights team is also in the process of developing a global stakeholder engagement and outreach program, which may include an expert body or advisory group.

c) Build on transitional justice work to contribute to broader recognition of human rights harms in countries with legacies of human rights abuses.

> **Original commitment:** Implementing
> **Status:** Complete
>
> *Meta is actively incorporating transitional justice and principles into its policy, communications, and human rights work. Transitional justice experiences are highly relevant to Meta and other social media companies because they often confront systems-level challenges and struggle with issues of scale, speed, and remedy.*

d) Continue efforts to align with Santa Clara Principles.

> **Original commitment:** Implementing in part
> **Status:** In progress
>
> *The principles give excellent guidance. We will continue to work to align with the principles, but note content volumes are so large we cannot commit to human review for all cases. Instead, we seek to prioritize based on severity of potential harm in a framework aligned to the UN Guiding Principles on Business and Human Rights.*

e) Formally integrate human rights expectations into the Responsible Innovation process, requiring that all "high-risk products" undergo a human rights review.

> **Original commitment:** Implementing in part
> **Status:** No further updates
>
> *We have aligned Responsible Innovation (RI) principles with human rights guidance, but the process mentioned here has been deprecated. We will take further steps to integrate human rights guidance into RI product risk management processes.*

f) 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.

> **Original commitment:** Assessing feasibility
> **Status:** Complete
>
> *Our Corporate Human Rights Policy requires us to know our salient human rights risks. We have conducted a salient risk assessment across the company, which is an important precursor to this work.*

9. Updated HRIA recommendations (from September 2021)

a) 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.

Original commitment: Implementing
Status: Complete

We are convening two dedicated workshop series with leading misinformation, safety, and expression experts to advise us on best practice research and thinking on effectiveness of misinformation enforcement, with one of them focused on the Asia-Pacific region. We will consider opportunities to share related knowledge.

b) Increase efforts to combat disinformation and misinformation in the Philippines, including: onboarding new third-party fact-checkers to ensure sufficient resourcing to address disinformation in advance of the upcoming election and applying Meta's heightened penalties for public figures who violate Facebook's Community Standards.

Original commitment: Implementing
Status: Complete

As part of our preparation for the Philippines election, we expanded our support and training for the local fact-checking community, in line with the IFCN Principles. In 2021, we launched the Philippine Fact-Checker incubator (PFCI), with five participating organizations. As of 2023, two of the five organizations participating in the PFCI, Press One and MindaNews, have been accredited by the IFCN. The applications of the three other organizations are currently in process.

We also created additional means to strengthen enforcement of misinformation policies (including, among others, voter interference and misinformation that leads to real-world harm). Our policy of heightened penalties against public figures who violate the Community Standards during periods of civil unrest is global, and thus will be applied when policy criteria are met.

We also supported a project with FYT, a community-oriented independent media outfit, to train local and student journalists on digital literacy, visual storytelling, and fact-checking in preparation for the 2022 general election.

Details on these and other actions taken to combat misinformation and disinformation

prior to the 2022 election can be found in the "[Preparation for Philippines Presidential Election 2022](#)" section of our 2022 Annual Human Rights Report.

c) 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.

Original commitment: Implementing
Status: Complete

We seek to continue our strong work prohibiting content that outs specific at-risk groups, including red-tagging. Since 2019, Meta has reviewed and enforced upon content threatening and accusing individuals of being communist ("red-tagging") as a violation of our Coordinating Harm & Promoting Crime Policy.

Meta also addresses red-tagging through policies designed to counter Information Operations, which include policies and enforcement against coordinated inauthentic behavior, inauthentic behavior, and foreign inauthentic behavior. Since 2019 (including a significant action in [2022](#)), we have been removing networks of accounts, Pages, and Groups in the Philippines that worked together to systematically violate our Community Standards and evade enforcement.

We also continue to engage with the [Philippine Commission on Human Rights](#), as well as other local rights organizations. We also completed an Online Security and Resilience Training for Human Rights Defenders, involving 60 organizations in the Philippines, and have plans to expand the same program to Thailand.

In 2022, we briefed members of the Armed Forces of the Philippines about Meta's Community Standards and also provided a digital literacy session focused on critical thinking online.

d) Increase the capacity to mitigate the prevalence of online sexual exploitation and human trafficking on Meta platforms, including:
- investing greater resources in the prevention of child exploitation and sharing of child sexual abuse material (CSAM) in unencrypted spaces
- 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

- expanding approaches to identify and limit dangerous engagements that could result in exploitation.

Original commitment: <u>Implementing</u>
Status: <u>Complete</u>

We lead the industry in our work to prevent and disrupt the sharing of CSAM on both our encrypted and unencrypted services. We will continue to invest deeply in this space, including by expanding our work on identifying and limiting potentially exploitative interactions with children. We are already deploying industry-leading methods and are seeking to pioneer new ones to help address the potential spread of CSAM in encrypted spaces.