∞ Meta

# Human rights report

Insights and actions
2022

# Table of Contents

# 01    Executive summary



This is Meta's second annual Human Rights Report, covering the calendar year 2022. It demonstrates how we are living up to the commitments made in our Corporate Human Rights Policy, which is based on the United Nations Guiding Principles on Business and Human Rights (UNGPs). The report shows how we are building on our work to respect human rights, how we assess potential impacts that could be linked to our services, and what we are doing to mitigate these risks.

Over the years, Meta's services and apps have helped grassroots movements grow rapidly, challenge orthodoxies, advocate for rights, and in doing so, change the world. 2022 was no exception. As new challenges emerge, our human rights work evolves to address such issues as the use of our services in countries in conflict, threats to an open internet, and the rise of artificial intelligence.

We have strengthened our governance systems to advance our work toward respecting human rights across all our products and services. This includes continuing to empower the Oversight Board, which published 12 content moderation decisions and made 91 recommendations in 2022. Our efforts to strengthen governance also include training our employees. During the year, we launched human rights training titled *Bigger than Meta: Human Rights* company-wide.

In 2022, Meta passed its Global Network Initiative assessment, which included an examination of our policies and processes as well as eight specific illustrative case studies involving government takedown requests and data requests.

Human rights due diligence is central to our Corporate Human Rights Policy. In this report, we disclose key findings from our Comprehensive Human Rights Salient Risk Assessment (CSRA) and provide an update on our implementation of the recommendations borne of the due diligence work.

In 2022, we conducted the CSRA in partnership with Business for Social Responsibility (BSR), an independent organization with expertise in human rights. The analysis identified eight potential salient risks as highest priority, based on UNGP criteria. These were 1.) Freedom of Opinion and Expression; 2.) Privacy; 3.) Equality and Non-Discrimination; 4.) Life, Liberty, and Security of Person; 5.) Best Interests of the Child; 6.) Public Participation, to Vote, and to Be Elected; 7.) Freedom of Association and Assembly; and 8.) Right to Health. The assessment found that the potential risks relating to how Meta delivers our services and profit models are varied and nuanced.

# As new challenges emerge, our human rights work evolves to address such issues as the use of our services in countries in conflict, threats to an open internet, and the rise of artificial intelligence.

We also describe how we have followed up on the recommendations from previous human rights due diligence regarding the Philippines, Israel and Palestine, and India.

Further, we continue to address the recommendations outlined in Meta's 2021 Civil Rights Audit. In 2022, an additional 32 of the 117 recommendations and action items were implemented or represented ongoing work, which puts us at a completion percentage of 84%.

We listen and learn from external stakeholders to live up to our human rights responsibilities. Stakeholder engagement is embedded throughout our Corporate Human Rights Policy, and we are focused on deeper engagement with marginalized communities. Our services development, content policies and moderation, and Community Standards are informed by insights from civil society organizations, human rights defenders, marginalized groups, international organizations, Trusted Partners, investors, advertisers, and users.

We prioritize building partnerships where we face the most salient human rights risks. Our network of Trusted Partners comprises over 400 non-governmental,

not-for-profit, national, and international organizations in 113 countries who report content, accounts, and behavior that we review in context. In 2022, we expanded our network of Trusted Partners across 36 new countries.

Local context is vital to understanding potential human rights risks, and we are making progress in supporting more languages to help us do so. We recognize human rights are under particular threat during times of crisis. Meta assesses risks of imminent harm — even before a crisis — and seeks to respond with specific policies, services, and operational actions to respect human rights. In this report, we review the progress of our Crisis Policy Protocol to ensure our policy response to crises is principled and calibrated. We provide insights from our work in Russia, Ukraine, and Iran, as well as around elections in Brazil, Kenya, and Nigeria.

We continue our work to keep human rights central to how we understand and implement the rapidly advancing technology of artificial intelligence (AI). AI models may support people's ability to exercise their rights but can also exhibit problematic biases or discriminatory impacts and generate problematic content. We are committed to addressing these issues in an open and collaborative way and proud of our wide-ranging efforts to solve some of the hardest problems social media and AI present.

Fully aware of our responsibilities, we strive to advance our human rights principles at scale for the benefit of our users and societies at large, giving people the power to build community.

## 02  About this progress report

Meta's[1] second annual Human Rights Report reflects our progress on our commitments to the United Nations Guiding Principles on Business and Human Rights and our Corporate Human Rights Policy. We report on how we are addressing our human rights impacts. This includes a review of our recent salient risk assessment, insights from our human rights due diligence, and the actions we are taking in response.

Our first annual Human Rights Report was published in July 2022 and covered our learnings and progress from January 1 2020 through December 31 2021.

This, our second annual Human Rights Report, covers our learnings and progress from January 1 2022 through December 31 2022. It builds on the foundational first report. We report on Meta services including Facebook, Instagram, WhatsApp, Messenger, and Reality Labs.

Our reporting is complemented by Meta's First Responsible Business Practices Report, which is aligned with the Global Reporting Initiative (GRI) and Sustainable Accounting Standards Board (SASB) standards. As stated in the Meta 2023 Responsible Business Practices Report, human rights scored highly as a priority topic in Meta's updated priority topic assessment.

Meta's Corporate Human Rights Policy applies enterprise-wide. Each Meta service and entity has its own policies and procedures with, at times, varying human rights impacts. This report references actions taken by Meta as a company regarding one or more Meta entities. Statements are not intended to imply that Meta took that same action regarding all entities.[2]

In writing this report, we engaged with many internal and external stakeholders and sought to balance brevity with completeness.

[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 services and apps, not the company as a whole.

[2] For example, WhatsApp, an end-to-end encrypted messaging and calling application, has unique human rights touchpoints. This report's discussion of content moderation and related actions on Facebook and Instagram does not apply to WhatsApp and, unless a policy or action is specified as applying to WhatsApp, it does not apply to WhatsApp. Further, while many actions described in this report apply to Instagram and Facebook, there are intentional distinctions in policies and procedures between the services. If a policy is labeled a "Facebook" policy, it may not apply to Instagram. No statement in this report 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 services or entities.

Meta's mission is to give people the power to build community and
bring the world closer together. Our services and apps help make it
possible for grassroots movements to flourish and challenge established
authority and orthodoxy. A protestor in Iran, an LGBTQIA+ activist
in Uganda, a journalist in Kyiv — they, and everyone who uses our
services, are the reason we work hard to respect human rights.

With billions of people using our apps every day, we seek to meet the
responsibilities of operating at such a scale. With the speed of technological and
political change, understanding and addressing the ways our services impact
human rights is a constantly evolving challenge. Meta has been a leader in AI
research and development for over a decade, and we continuously adapt our
approach to live up to our human rights responsibilities and remain vigilant for
emerging risks as powerful new tools such as generative AI rapidly evolve.

The human rights issues we confront are complex, vary across countries and
regions, and are influenced by the evolution of government regulations around

the world. For example, the European Union's Digital Services Act, enacted in 2022, creates significant due diligence and reporting obligations for Meta but is explicitly linked to the region's fundamental rights framework. This is against the backdrop of the evolving threat to the open internet and the fundamental rights of our users that comes from the spread of an authoritarian internet model, with citizens segregated from the rest of the global internet.

"More than 3.5 billion people use Meta's apps — Facebook, Instagram, WhatsApp, and Messenger — every month. It means that between a third and half of all human beings on Earth use them. And in doing so, they have access to an interconnected world of people, ideas, news, communities, and commerce unconstrained by local or national boundaries. This scale of connectedness is unprecedented in human history."

Nick Clegg, President, Global Affairs

In this report, we share our progress toward respecting human rights across all our programs, services, and policies. We continue to explore new avenues to incorporate human rights into our services and practices. This includes our work advancing democracy, ensuring an open and interoperable metaverse, and finding ways to help ensure AI serves communities fairly.

All our efforts are grounded in our commitment to our Corporate Human Rights Policy and the United Nations Guiding Principles on Business and Human Rights (UNGPs). We don't do this work alone. The Oversight Board plays a critical, industry-leading role as an independent and expert arbiter that reviews content decisions and offers policy advisory opinions and recommendations. As per its charter, the Oversight Board pays particular attention to the human rights norms protecting free expression.

Our work is also informed by the principles and practices of specialized multilateral and UN organizations and multi-stakeholder alliances. We are members of the UN Global Compact and support its principles. And we continue to participate in the UN Office of the High Commissioner for Human Rights B-Tech Project, which works to develop authoritative guidance and resources for implementing the UNGPs in the technology sector. As a member

of the Global Network Initiative (GNI), we are committed to its principles and implementation guidelines. In 2022, we passed the GNI tri-annual assessment.

We also continue to explore new forms of governance. For example, we piloted Community Forums to bring together diverse groups of people from all over the world to discuss tough issues and share their perspectives on a set of recommendations.

This report is intended to give greater insight into the work Meta has done to identify, mitigate, and prevent human rights risks in 2022. As always, there is more work to be done, and we look forward to more reporting on our progress in the years to come.

**SIGNED:**

**Nick Clegg,**
President of Global Affairs

**Jennifer Newstead,**
Chief Legal Officer

## 04    How Meta manages and governs human rights work



Clear management and governance structures enable us to advance our work toward respecting human rights across all our programs, services, and policies. Our Meta Human Rights team guides the implementation of the Corporate Human Rights Policy, which is overseen by the President of Global Affairs and Chief Legal Officer.

As described in our first annual Human Rights Report, the team's tasks include promoting the policy's integration into existing and developing policies, programs, and services; undertaking due diligence; and supporting the training of employees on the policy. The policy guides teams to build rights-respecting products, respond to emerging crises, and work with speed and agility to embed human rights at scale.

Our Corporate Human Rights Policy commits us to periodic reporting to the Board of Directors on key human rights issues. The Audit & Risk Oversight Committee of the Board oversees the company's various risk exposures, including human rights-related risks, and the steps management has taken to monitor or mitigate such exposures. The committee is periodically briefed on existing and developing work conducted by the Human Rights team.

## Training Meta employees

*How* we build at Meta is as important as *what* we build. Our human rights training gives employees a better understanding of their responsibilities and the knowledge and skills needed to fulfill them.

*Bigger than Meta: Human Rights,* our training launched in 2022, highlights the potential and actual real-life impacts of Meta's services, policies, and business decisions on human rights. It seeks to promote a human rights mindset in our day-to-day work, encouraging respect for human rights to the benefit of all who use our services.

The training complements Meta's civil rights training, which was launched in July 2022 and centers on the principles of non-discrimination, justice, and fairness.

## Appeals and access to remedy

At Meta, we strive to provide pathways for stakeholders to report concerns, for Meta to review them, and for Meta to provide remedy and remediation consistent with UNGP 31. Meta maintains multiple grievance pathways, identified in the Code of Conduct and in the Help Center on platforms and apps, including an appeals process to the first-of-its-kind Oversight Board.

When content is removed, we notify users and clearly identify pathways for content removal decisions to be appealed. Data regarding removal decisions and other aspects of policy enforcement are published through the Transparency Center. For more information, see our first annual Human Rights Report.

## Oversight Board

The Oversight Board is a unique, industry-leading effort to respect the right to free expression. It provides the public with a mechanism to make additional appeals to an external body. In 2022, Meta underscored its commitment to the Oversight Board by approving a new three-year, $150 million grant to the Oversight Board Trust. This enables the Board to continue operating as a separate, independent entity.

The Oversight Board makes fully independent content moderation decisions and recommendations about content policy, services, and operations. In 2022, the Oversight Board published 12 decisions; nine of these overturned Meta. These decisions covered content in a wide range of countries, including Colombia, Ethiopia, Afghanistan, Croatia, Sudan, India, and the United States. Issues included the depiction of graphic violence in the content of a military coup, mention of

the Taliban in news reporting, and description of sexual violence against minors. These decisions all provided access to remedy for users of the platform.

In addition to its decisions, the Oversight Board made 91 recommendations in 2022, 89 of which are under consideration for implementation, in progress of implementation, or fully implemented.

**2022 Oversight Board Recommendations**

| Oversight Board recommendations | Meta assessment and/or implementation in progress | Recommendations fully implemented |
|---|---|---|
| 91 | 75 | 14 |

# Global Network Initiative

Meta has been a member of the Global Network Initiative (GNI) since 2013. GNI is a growing alliance of internet and telecommunications companies and diverse civil society groups, including human rights and press freedom groups from around the world. We are committed to GNI's Principles on Freedom of Expression and Privacy (GNI Principles), which are based on internationally recognized laws and standards. These commitments were reaffirmed in our Corporate Human Rights Policy in 2021.

All GNI companies undergo an independent assessment of their implementation of the principles every two to three years.

In 2022, Meta passed its GNI assessment, which included examining our policies and processes as well as eight specific illustrative case studies involving government takedown requests and data requests. The GNI Board unanimously determined that Meta had made "good faith efforts to implement the GNI Principles, with improvement over time," the standard against which member companies are assessed. GNI will release a separate public assessment report covering all company assessments for the 2021–2022 cycle.

# 05

## How we identify and address salient human rights risks



In 2022, we undertook a Comprehensive Human Rights Salient Risk Assessment (CSRA), a foundational program tool to help guide future risk management. The purpose of the CSRA was to identify and prioritize Meta's most salient inherent human rights risks across the enterprise globally, defined as Meta's most significant potential adverse human rights impacts[3] on people, including users and others who may be affected by Meta's actions. "Inherent" risks refer to risks that exist before processes or mitigations are put in place and are contrasted with "residual" risks, which refers to risks that exist after processes have been implemented to address inherent risks.

**How it was conducted:** Business for Social Responsibility (BSR) conducted desktop research and extensive engagement with internal and external stakeholders, involving more than a dozen meetings with civil society groups, UN representatives, and investors across multiple regions and dimensions of diversity. The assessment also drew upon insights gained from previously conducted human rights due diligence and stakeholder engagement. The assessment prioritized risks based on the definitions of "severity" (scope, scale, remediability) and "likelihood" as defined under the UNGPs and the UNGPs Reporting Framework:

[3]   The term "adverse human rights impact" is in line with the UNGPs and means an impact that occurs when an action removes or reduces the ability of an individual to enjoy his or her human rights. This report should not be construed to suggest that any particular individual or community experienced an adverse impact, unless otherwise specifically stated.

1. **Scope:** How many people could be affected by the potential adverse human rights impact?
2. **Scale:** How serious would the potential adverse human rights impact be for those affected?
3. **Remediability / Irreversibility:** Will a remedy restore those potentially affected to the same or equivalent position?
4. **Likelihood of occurrence:** What is the likelihood of the potential human rights impact occurring?

In line with the UNGPs, the CSRA paid particular attention to the rights, needs, and challenges of individuals who may be at heightened risk of becoming vulnerable or marginalized.

**What we learned:** The analysis identified as highest priority across global operations the following eight salient human rights:

1. **Freedom of Opinion and Expression**
2. **Privacy**
3. **Equality and Non-Discrimination**
4. **Life, Liberty, and Security of Person**
5. **Best Interests of the Child**
6. **Public Participation, to Vote, and to be Elected**
7. **Freedom of Association and Assembly**
8. **Right to Health**

## Our salient risks

The CSRA analyzed inherent salient risks across all internationally recognized human rights across the global enterprise and prioritized the eight most salient using UNGPs criteria and the UNGPs Reporting Framework. These risks are examined below, with illustrative examples of some potential risk factors and a selection of Meta's prevention and mitigation efforts.

# Freedom of Opinion and Expression

The Right to Freedom of Opinion and Expression is central to what we at Meta believe and work to protect.

| Examples of potential inherent salient human rights risks identified in CSRA | Examples of how Meta addresses the potential risk |
|---|---|
| Internet disruptions and blocks to social media prevent people from exercising their freedom of expression and cut them off from receiving and sending vital news and information | We recently launched a feature that allows people to connect to WhatsApp through a proxy server when their internet connection is disrupted or access to WhatsApp is blocked. |
| Overbroad government limits on content | We assess valid legal government takedown requests (TDRs) according to GNI Principles and are independently assessed in our implementation regularly. We pursue extensive operational controls to scrutinize validity of TDRs and provide information on government TDRs in the Transparency center. We conduct regular revisions of content policies, seeking diverse stakeholder input, against international human rights norms. |
| Meta's content moderation enforcement may limit freedom of expression | AI models predict whether a piece of content is hate speech or violent and graphic content. Separate systems — our enforcement technologies — determine whether to take an action, such as deleting, demoting, or sending the content to a human review team for further review. By enforcing our policies, we seek to mitigate risks of hate speech, violence and incitement, and other human rights harms while upholding freedom of expression. |

## Privacy

"Getting privacy right is a continual, collective investment across our company and is the responsibility of everyone at Meta to advance our mission."

Michel Protti, Chief Privacy Officer for Product

The Right to Privacy is core to Meta's mission and a necessary condition for the realization of other human rights, such as freedom of expression, freedom of assembly and association, and freedom of belief and religion.

We are committed to improve and adapt our practices and policies to anticipate and address challenges to people's privacy on our services and apps.

| Examples of potential inherent salient human rights risks identified in CSRA | Examples of how Meta addresses the potential risk |
|---|---|
| Overbroad or unnecessary government demands for user data | We protect users from unlawful or overbroad government data requests through our dedicated Law Enforcement Response team.<br><br>We push back on government requests that do not meet internationally recognized standards, including those set out in our Corporate Human Rights Policy, and challenge them in court if necessary.<br><br>We respond to valid legal requests where we have a good-faith belief that we're required by law to do so. We also respond to certain legal requests where not compelled by law but where we have a good-faith belief that a response:<br>• is required by law in another jurisdiction in which a user may be located,<br>• affects users in that jurisdiction, and<br>• is consistent with internationally recognized standards, including our Corporate Human Rights Policy. |

(cont.)

(cont.)

When we comply with government data requests, we strive to notify people when their data has been requested, and we publish a transparency report on government requests we receive.

Content or behavior on Meta apps may adversely impact privacy and data protection rights (e.g., doxxing, scraping)

Meta has robust Community Standards and Community Guidelines that govern how users can use our apps. We do not allow content that shares, offers, or solicits personally identifiable information or other private information that could lead to physical or financial harm. We remove content that goes against our policies as soon as we become aware of it.

Using automation to access or collect data from Meta's services and apps without our permission ("scraping") is a violation of our terms of service. We have a dedicated team, technical measures, and processes that focus on detecting, investigating, and blocking unauthorized scraping attempts.

Meta or third parties may use sensitive or highly personal user data, including those related to protected characteristics, for targeted advertising in ways that have adverse impacts on privacy and data protection rights

Any information we collect and process depends on the service we are offering. To help us in this task, we are developing privacy-enhancing technologies (PETs) that help minimize data collection and use. Our team is dedicated to building PETs for use by teams across Meta, focused on key areas like de-identifying data at collection and enabling teams to implement end-to-end encryption (E2EE) in our apps and services.

In consultation with civil rights experts, policymakers, and other stakeholders, we rolled out changes to improve user control over ads, and we removed ad targeting options that relate to topics that people may perceive as sensitive.

(cont.)

(cont.)

We collaborated with the U.S. Department of Justice to develop and launch the Variance Reduction System (VRS), a new technology that will help distribute certain ads on our apps in a more equitable way. We launched VRS in the U.S. for housing ads and will expand it to employment and credit ads.

The CSRA identified several risk pathways relating to targeted advertising while emphasizing the nuanced nature of these risks. For example, while ad-based business models can create risks to the right to privacy, it is possible to establish mitigations to address these risks while also preserving the broad ability for people around the world to express their ideas and convene. Likewise, advertising is a critical driver of opportunities for small- and medium-sized enterprises, which otherwise would lack viable access to potential customers.

# Equality and Non-Discrimination

The Right to Equality and Non-Discrimination provides for equal protection against any discrimination. While the democratization of technology continues to make the internet more accessible to a wider range of people, not all communities share the same experiences online, especially marginalized individuals and communities.

We strive to create a more inclusive and equitable online environment for all users through our Community Standards and Community Guidelines, which prohibit hate speech, bullying, and harassment. Our Human Rights and Civil Rights teams work together to strive to respect the equal rights of individuals and enhance equity on all our technologies.

| Examples of potential inherent salient human rights risks identified in CSRA | Examples of how Meta addresses the potential risk |
|---|---|
| Content adversely impacting equality and non-discrimination (e.g., hate speech) | Our Facebook Community Standards, including policies regarding hate speech, which relate to protected characteristics, outline what is and isn't allowed on Facebook. The Instagram Community Guidelines outline what is and isn't allowed on Instagram. We take action against content that violates our policies. Our Advertising Standards provide policy detail and guidance on the types of ad content we allow and the types of ad content and targeting we prohibit.

We continued our work to build responsible AI systems to reduce bias and improve fairness, including in areas like ad delivery and reducing the prevalence of harmful language in generative AI applications. We took a major step forward to help researchers measure fairness in our AI models. A lack of diverse data — or data that represents a wide range of people and experiences — can lead to AI-powered outcomes that reflect problematic stereotypes or fail to work equally for everyone. That's why we worked to prepare Casual Conversations v2, a |

(cont.)

(cont.)

consent-driven, publicly available dataset that enables researchers to better evaluate how fair and robust certain types of AI models are, with the goal of making them more inclusive.

Some languages and dialects may be more challenging to moderate than others

We use a combination of AI and human content moderators to help remove policy-violating content on our apps.

We use human and machine-based translation to enhance content moderation. We are constantly evolving our reporting flows and integrity classifiers in the languages we support.

### Hate speech

We removed approximately 50 million pieces of hate speech content from Facebook in 2022, 91% of which we proactively found and took action on. We removed 16 million pieces of hate speech content from Instagram, 92% of which we proactively found and took action on. Hate speech made up less than 0.02% of viewed content on Facebook and Instagram.

For more on how Meta's Civil Rights team addresses equality and non-discrimination, see this progress report and this update.

## Accessibility

As technology advances, it's important that people with disabilities are not left behind. We strive to continually improve our efforts to ensure accessibility across our services and apps. Making these accessible to people with disabilities is a key commitment in our Corporate Human Rights Policy, where we cite the Convention on the Rights of Persons with Disabilities as a core international standard.

**Our actions: Expanding cross-industry efforts to improve accessibility**
Virtual, augmented, and mixed reality technologies offer particular promise for people with disabilities, whether through assistive technologies, rehabilitation

## 91%

of the 50 million pieces of hate speech content removed from Facebook in 2022 was proactively found and acted on

or training solutions, or experiences driving awareness and insight. Meta continues to partner with the XR Association on projects to improve accessibility knowledge and cross-industry efforts such as the launch of a new accessibility-focused site, xraccessibility.github.io, in collaboration with XR Access. Recent innovations include Ray-Ban Stories smart glasses powered by Meta, which feature hands-free messaging and calling on WhatsApp and Messenger, as well as image and video capture using voice commands and touch controls.

In addition to embedding features in our services, Meta has developed and shared technical recommendations for developers to create accessible virtual reality apps. We also worked with the disability community and industry organizations to develop and launch XR accessibility guidelines for developers in partnership with the XR Association.

# Life, Liberty, and Security of Person

The Right to Life, Liberty, and Security of Person concerns freedom from physical confinement as well as protection from injury to the body and mind. For Meta, respecting this human right includes the risk that content may provoke harm, including human trafficking, state-sponsored online threats, and non-state groups engaged in or advocating for violence or hate.

| Examples of potential inherent salient human rights risks identified in CSRA | Examples of how Meta addresses the potential risk |
|---|---|
| Bad actors:<br><br>• Exploiting Meta services and apps to coordinate online or offline harm<br>• Misusing services and apps for cyber-attacks or phishing<br>• Threatening and harassing human rights defenders (HRDs), activists, and other vulnerable groups | Our security policies are designed to identify and defend against adversarial threats. Meta has tools and resources for HRDs and journalists, which were built with direct feedback from HRDs. |
| | We expanded the availability of Facebook Protect to more people, particularly HRDs, allowing them to have more control over their accounts, including who they allow to see their posts.<br><br>The Safety Guide in our Privacy Center centralizes a range of resources to help people manage their privacy settings to protect themselves from online surveillance or unauthorized or compelled access. It also provides help to find support if they're experiencing bullying and harassment. We collaborated with experts and HRDs across a range of issues to create policies, tools, and resources including digital literacy, child protection, and tips for parents. |
| | Meta invests in countering the misuse of our services by authoritarian governments, terrorist groups, or other bad actors who attempt to surveil regime critics, opposition figures, and HRDs. We take action when we see such activity, block their domain infrastructure from being shared on our services, and, where feasible and as appropriate, notify people who we believe were targeted by these malicious operations. |

# More ways we respect the Right to Life, Liberty, and Security of Person

The following initiatives highlight our approaches to help groups and individuals at risk.

### Protecting HRDs and journalists

For HRDs, journalists, and other vulnerable people, we developed resource tools in the Meta Safety Center. Our Journalist Safety Guide helps ensure journalists are equipped to manage their online safety with tools and resources to safeguard their sources, contacts, logins, and personal information. We also support the international news community by sponsoring programs via our Meta Journalism Project.

In 2022, we took these actions:

- Gave activists and journalists more protections from threats of violence, including removing content that "outs" people as activists in situations that could endanger them.
- Secured the accounts of HRDs who have been arrested or detained in authoritarian states or conflict settings against unauthorized access by local law enforcement, security services, and conflict actors.
- Advanced the Meta Human Rights Defender Fund and Journalist Safety Initiative for Asia Pacific, through which the international human rights group Civil Rights Defenders provided urgent evacuation, temporary relocation, digital security measures, security trainings, and legal aid to more than 500 HRDs, mainly in South and Southeast Asia. Civil Rights Defenders have also conducted security assessments, facilitated implementation of security measures, and provided other security-related supports to HRDs and organizations in the regions.
- Intensified engagement with HRDs to give continuous feedback on the impact of regional and global defender threats and trained HRDs on the Safety Guide in the Privacy Center.
- Started a free, online digital security and safety training course for journalists and HRDs, created and managed by the International Center for Journalists in partnership with the Border Center for Journalists and Bloggers.

In 2022, we presented the Safety Guide in the Privacy Center to the Office of the UN Special Rapporteur on HRDs.

### Human trafficking and exploitation

In an effort to disrupt and prevent harm, we remove content that facilitates or coordinates the exploitation of people, including human trafficking.[4]

In 2022, we updated our Human Exploitation Policy to prohibit content that solicits smuggling services. We also provide informational resources on Facebook about the risks of engaging with smugglers, the signs of potential exploitation, and ways to seek legal migration, including asylum. We developed these resources in consultation with experts, including the International Organization on Migration.

---

[4] For more on human trafficking and exploitation, see page 32 of our first annual Human Rights report.

We continue to allow content that solicits or shares information about personal safety and border crossing and how to leave a country or seek asylum. We allow content condemning or raising awareness of human trafficking.

**Helplines in Bihar State, India, for trafficked persons**
Human trafficking is particularly pervasive in the Indian state of Bihar. To assist victims, Meta supported the development of a helpline on WhatsApp, which assists the Bihar State Labor Department tracking system. Users can report incidents of child labor to a specific WhatsApp number, which provides services in English, Hindi, and Magahi. The helpline is also used to monitor progress made by rehabilitated child laborers in the state.

**Surveillance-for-hire**
With help from the surveillance-for-hire industry, some governments and bad actors have access to the capability to silence voices by targeting and intimidating activists, dissidents, journalists, and researchers through abusive spyware. This spyware targets people without their knowledge, with no regard to human rights, civil liberties, and the rule of law.

Surveillance-for-hire practices abuse our services as a vector to plant malware on our users' mobile devices. We have disrupted and reported on many of these operations. In 2022, we collaborated with industry partners to propose principles and practices on preventing and mitigating this threat. We released these principles at the Summit for Democracy under the aegis of the Cybersecurity Tech Accord. In our Policy Recommendations from December 2022, we offered a roadmap for other stakeholders, including governments, the surveillance-for-hire industry itself, and civil society to join forces in promoting rights and freedoms online.

**Non-consensual intimate image sharing (NCII)**
Sharing intimate images or videos of someone, either on or offline, without their consent is a violation of that person's rights to privacy and security. Images are often spread across multiple platforms and apps. NCII is often called "revenge porn," as the sharing of intimate content may be an act of perceived "revenge" to a person after a relationship breakdown.

Meta has collaborated with nonprofits and other companies to produce a free tool, StopNCII.org, that enables would-be victims to flag images so that participating companies can detect and remove the images to prevent them from being shared online. With one action, individuals can request removal of NCII across multiple platforms and apps. StopNCII.org, which builds on technology developed by Facebook and Instagram, is available in 22 languages and works across Facebook, Instagram, and the platforms of other participating companies, including TikTok and Bumble.

## Best Interests of the Child

The Best Interests of the Child gives children the right to have their best interests considered in all decisions and actions that affect them, including protection from human trafficking. Meta's Best Interests of the Child framework aligns with the fundamental values of the UN Convention on the Rights of the Child.

We want young people to have safe, positive experiences online. Online child protection is a top priority at Meta, and we offer tools for teens to spend more meaningful time online, including more ways to focus and set boundaries, such as break reminders, expert tips, quiet mode, nudges to switch topics, and the option to hide "like" counts on posts in Feed. Our approach to building these technologies is grounded in research and direct feedback from parents, teens, and experts in mental health and child psychology.

| Examples of potential inherent salient human rights risks identified in CSRA | Examples of how Meta addresses the potential risk |
|---|---|
| Children exposed to unwanted or inappropriate content | Our policies require that everyone be at least 13 years old to create an account on Facebook and Instagram. We write our policies to help support a safe experience for those 13 and above, and we consult child safety experts in the development of these policies. We age gate certain content (e.g., certain kinds of graphic violence); we work to avoid recommending certain non-violating content to minors (e.g., alcohol, tobacco); and we have a sensitive content control that we default to the strictest setting for minors. We prohibit content that sexualizes minors, and when we become aware of apparent child sexual exploitation, we report it to the U.S. National Center for Missing and Exploited Children (NCMEC). |
| | We also have more restrictive ads policies for minors (e.g., in addition to prohibiting ads for restricted goods or services, like alcohol or tobacco, we also do not allow ads for weight-loss products for minors). |
| | Our policies prohibit, and Facebook and Instagram seek to prevent, content that |

| | |
|---|---|
| (cont.) | (cont.)<br>encourages suicide, self-injury, eating disorders, or bullying and harassment. We remove such content when we find it and continue to improve the technology used to detect and remove it. |
| Sexual exploitation or abuse of children | The child protection section in our Safety Center highlights our comprehensive approach to child safety, which includes policies prohibiting child exploitation and cutting-edge technology to prevent, detect, remove, and report policy violations.<br><br>We've built over 30 tools to support children, teens, and their parents, including, on Instagram, supervision tools in our Family Center that help parents and their children navigate their time online together.<br><br>We use sophisticated detection technology across Facebook and Instagram aimed at detecting, removing, and preventing the sharing of images and videos that exploit children. We report apparent instances of child exploitation appearing on our sites to the NCMEC and agencies around the world. In 2022, we supported NCMEC on the development of the Take It Down Platform for minors who are worried about their intimate images being shared on online platforms and apps. |

## Public Participation, to Vote, and to Be Elected

The Right to Public Participation, to Vote, and to Be Elected in free and fair elections is a cornerstone of democracy. It is also crucial to the rule of law, social inclusion, economic development, and the advancement of all human rights.

Protecting the integrity of elections on our services and apps is one of our highest priorities. Meta invests significant effort and resources to protect elections online before, during, and after election periods. We aim to give people a voice and help them participate in the civic process while combating hate speech, interference, and foreign influence.

| Examples of potential inherent salient human rights risks identified in CSRA | Examples of how Meta addresses the potential risk |
|---|---|
| Content adversely impacting public participation, voting, or running for office | Attempts to interfere with or suppress voting are not allowed. We work to identify and remove this content, no matter who it comes from.<br><br>We provide an industry-leading level of transparency around political advertising and pages so users can see who is trying to influence votes. In some markets, we prohibit ads that discourage voting, prematurely claim victory in an election, attempt to delegitimize the election, or are inconsistent with health authorities on voting safely.<br><br>Hate speech can impact voters. Our Community Standards cover policies on harassment, incitement, violence, and hate speech and ban attacks on people based on characteristics like ethnicity or religion. We remove content that violates these rules as soon as we find it. |
| Efforts to prevent people from voting, increases in spam, foreign interference, or reports of content that violates our policies | We actively search content for any organizations that Meta has banned from attempting to violate our policies or cause offline harm.<br><br>In 2022, we updated Facebook Protect, which offers security tools and additional protections to candidates, their campaigns, and elected officials. |
| Coordinated bad actors interfering with elections | Our advanced security operations support election security by taking down manipulative campaigns, identifying emerging threats, and deploying our industry-leading global fact-checking network.<br><br>We use AI to identify fake accounts and remove the vast majority of these accounts before they are reported. We have removed over 150 networks of coordinated inauthentic behavior since 2017. |

# More about our work on national elections

For elections where there is greater risk of violating content, we bring together a specialist, cross-functional team with expertise in election integrity, misinformation, safety, human rights, and cybersecurity, with relevant language expertise in that market. The team monitors and responds to emerging risks in real time. It also monitors news coverage and election-related activity across other social networks and traditional media. These efforts give us a comprehensive view and help track what type of content may go viral.

In 2022, we built and strengthened teams to remain ready for elections around the world, including in Brazil, Kenya, and Nigeria, as well as midterm elections in the U.S. We made investments, engaged stakeholders and experts, established partnerships, and improved technologies.

### 2022 U.S. Midterm Elections

In the run-up to these elections, we worked with state and local elections officials to send in-Feed voting alerts, resulting in more than 80 million election-related notifications. We fought both foreign and domestic operations and exposed and disrupted dozens of networks that attempted to interfere with the elections. We banned more than a thousand militarized social movements and took down tens of thousands of QAnon pages, groups, and accounts from our apps. We also removed the original Stop the Steal group and banned over 270 white supremacist organizations.

**Brazil**

General elections occurred in Brazil in early October 2022, with a second round of the presidential election later that month. The elections were held in the midst of an economic crisis and concerns about election delegitimization claims. Our work on election integrity is continuous, and we intensified our efforts for the 2022 Brazilian elections a year in advance. We focused on four key areas in preparation for the election: preventing interference, collaboration with electoral authorities, fighting misinformation and disinformation, and increasing transparency in advertisements.

We activated our Elections Operations Center to allow integrity specialists in Brazil and around the world to identify potential threats on our services and apps in real time. We used human review and AI to help us, from August to October, proactively

detect and remove over 310,000 pieces of content from Facebook and Instagram from users in Brazil for violating our  Violence and Incitement Policy and over 290,000 pieces for violating our Hate Speech Policy. In addition, we reduced the distribution of content that our technology identified as likely to be violating those policies, to prevent it from spreading quickly. We posted election day reminders on Facebook and Instagram as well as reminders to update voter identification cards.

From January 1 2022 to October 2 2022, over 30 million people clicked on electoral information labels added to posts about elections on Facebook in Brazil and were directed to official information in the Tribunal Superior Electoral (TSE) Authority's website. Over 4.7 million people subscribed to TSE's chatbot on WhatsApp to receive authoritative information. To combat online violence against women in politics, we launched a guide with the support of the TSE Court and the Women's Democracy Network in Brazil.

To fight misinformation, we worked (and continue to work) with independent fact-checking agencies to verify the veracity of posts, reels, stories, videos, pictures, and comments on Facebook and Instagram in Portuguese. In 2022, we expanded from four to six the number of partners in our fact-checking program in Brazil.

To fight disinformation, our internal teams found and removed networks that violated our Coordinated Inauthentic Behavior Policy.

To increase transparency in ads, we expanded transparency policies beyond political ads to advertising on Facebook and Instagram in Brazil on social issues such as economics, safety, and education. We rejected 135,000 ad submissions from domestic and foreign sources that included Brazil in their targeting.

More information on our efforts in the Brazilian elections can be found here and here.

**Kenya**
Kenya held its general elections in August 2022. Previous electoral processes had been marred by human rights abuses. In advance of the 2022 elections, we engaged with key human rights defenders (HRDs), UN agencies, journalists, and civil society members, taking a number of steps to bolster our safety and integrity measures, including expanding our Trusted Partner program.

As with similar elections, we created a cross-functional team with language expertise to lead our work around the election. Our efforts focused on vulnerable groups. We enrolled women leaders in Facebook Protect to help safeguard high-target accounts, adopted stronger security protections (including monitoring for potential hacking threats), and conducted harassment sweeps of Facebook and Instagram. We also developed detection and action systems to help protect journalists and HRDs.

We hosted a series of training sessions for HRDs and independent women journalists. We also worked on combating the spread of misinformation, both on- and off-platform, by removing content intended to suppress the vote. This included removing out-of-context or false claims of acts of violence. Our actions were informed, in part, by the advice of local partners.

We increased investment to our Kenyan third-party fact-checkers such as AFP, Africa Check, and PesaCheck. We also made grants to local newsrooms across Kenya to boost local language fact-checking. To drive education on misinformation and hate speech, we developed on-platform and local radio campaigns in multiple languages and launched two supporting services — Trending Event and CrowdTangle Live Display — to help third-party checkers identify and act on possible misinformation.

# Freedom of Association and Assembly

The Right to Freedom of Association and Assembly is essential to democracy and foundational to many other rights guaranteed under international law, including the right to freedom of expression and to take part in the conduct of public affairs.

| Examples of potential inherent salient human rights risks identified in CSRA | Examples of how Meta addresses the potential risk |
|---|---|
| Content on Meta platforms may lead some users to feel unable to freely gather on Meta apps | Hate speech, bullying, and other forms of harassment can impact a person's right to assembly and association. We remove this type of content when it is reported by the victim because it prevents people from feeling safe and respected on Meta apps and offline. Meta has extensive policies against bullying and harassment and violence and incitement, including content that targets private individuals through unwanted Pages, Groups, and Events.<br><br>We seek to remove mass-directed harassment because it cuts across many rights, including the right to assembly. It includes targeting individuals at heightened risk of offline harm, including HRDs, minors, victims of violent events or tragedies, and opposition figures in at-risk countries during election periods. |
| Overbroad government-imposed limits to content that restrict the right to assembly and association | WhatsApp's end-to-end encryption feature protects the right to privacy and thus users' ability to exercise their rights to assembly and association.<br><br>We assess valid legal government takedown requests (TDRs) according to Global Network Initiative Principles and are independently assessed in our implementation regularly.<br><br>We pursue extensive operational controls to scrutinize validity of TDRs and provide information on government TDRs in the Transparency center. |

(cont.)

(cont.)

We conduct regular revisions of content policies, seeking diverse stakeholder input, against international human rights norms.

Meta respects the right of our vendors' employees to organize unions. Union organization by vendor employees does not alter our decision to work with or engage that vendor. We believe it is important that the companies we work with do not oppose or inhibit the right of their employees to unionize.

# Right to Health

According to the World Health Organization, "the enjoyment of the highest attainable standard of health is one of the fundamental rights of every human being without distinction of race, religion, political belief, economic, or social condition."

Meta respects the Right to Health by increasing access to credible health information, enabling people with similar health issues to connect with one another, and empowering them to make informed decisions about their health and well-being.

| Examples of potential inherent salient human rights risks identified in CSRA | Examples of how Meta addresses the potential risk |
|---|---|
| Medical misinformation or disinformation on Meta platforms | In 2020, Meta adopted a policy to remove health misinformation if:<br>• there is a public health emergency,<br>• leading global or local health authorities tell us a particular claim is false, and<br>• those organizations or authorities tell us the claim can directly contribute to the risk of imminent physical harm.<br><br>Meta continues to use AI tools to scale fact-checkers' work to detect false and misleading health information, including by classifying an image according to its visual content. This allows us to enforce bans on violating ads and commerce listings for certain medical products. |
| Policy-violating content that incites or is intended to encourage offline harm | Meta's Safety Center includes our Suicide Prevention resources and other information to help people who may be experiencing suicidal thoughts.<br><br>Our policies prohibit, and Facebook and Instagram seek to prevent, content that encourages suicide, self-injury, eating disorders, or bullying and harassment. We remove such content as soon as we find it and continue to improve the technology used to detect and remove it. |

Updates on due diligence



We endeavor to assess human rights impacts through human rights due diligence as laid out in our Corporate Human Rights Policy and in alignment with UNGPs 17 and 21. This includes taking appropriate action and implementing findings, monitoring implementation, and reporting annually on relevant insights and actions.

In 2022, we published findings from Human Rights Impact Assessments (HRIAs) and human rights due diligence exercises on Israel and Palestine, India, and End-to-End Encryption. We continued our implementation of the recommendations from these and previous assessments. We also advanced due diligence efforts on the potential human rights impacts of the metaverse.

## Philippines

We published the findings of and our response to the independent HRIA we commissioned in the Philippines in December 2021. The HRIA made

recommendations covering incitement of violence, surveillance, online sexual exploitation, trafficking, extremist activity, and corporate accountability.

Some of our subsequent actions include the following:

- We built new services and developed stronger policies in collaboration with the Commission on Elections, election watchdogs, independent fact-checkers, and civil society organizations during the 2022 Philippines election. These efforts expanded our ability to remove violating content and networks, connect more people to credible election information, promote digital literacy and civic engagement, and provide more transparency into political ads.

- We made the Facebook Ad Library available in the Philippines and required advertisers to complete our ad authorizations process and include "Paid for by" disclaimers on ads about elections, politics, and certain categories of social issues.

- We made several updates to our Community Standards, including expanding protections for public figures such as journalists and human rights defenders (HRDs).

- We removed more types of violating content, and our policies now provide stronger protections against gender-based harassment. We launched new policies against mass harassment and brigading and removed coordinated efforts of mass harassment that target individuals at heightened risk of offline harm. This includes attacks against dissidents.

- Partnering with the International Center for Journalists and the Border Center for Journalists and Bloggers, we rolled out a free digital security and safety program to help journalists and HRDs protect their digital assets and counter online harassment. Meta's Journalist Safety Hub centralizes all resources and tools available on our apps.

More details can be found in the Human Rights Implementation Update 2023.

## Israel and Palestine

We published the report from Business for Social Responsibility (BSR) on its independent Human Rights Due Diligence (HRDD) exercise into the impact of our policies and processes in Israel and Palestine during May 2021. This was in response to the Oversight Board's request for an independent review to determine if Meta's content moderation in Arabic and Hebrew, including its use of automation, had been applied without bias.

The report surfaced industry-wide, long-standing challenges around content moderation in conflict-affected regions and the need to protect freedom of expression while reducing the risk of online services being used to spread hate or incite violence. The report also highlighted the complexities of

managing issues surrounding the conflict due to social and historical dynamics, various fast-moving violent events, and the actions and activities of terrorist organizations. BSR found no evidence of intentional Meta bias on these issues.

During the reporting period, some subsequent actions included:

**Policy**

- Revising our definition of the term "praise" in our Dangerous Organizations and Individuals (DOI) Policy following extensive stakeholder engagement with academics, civil society actors, and other experts in counterterrorism, human rights, and freedom of expression from around the world.

**Enforcement**

- Following an extensive internal assessment, we determined that creating specialized routing for a variety of Arabic dialects across our systems would contribute to greater precision in our Arabic content moderation for high-severity content. We are now reviewing the development of mechanisms to efficiently route content by Arabic dialect to improve the accuracy of Arabic content review and better prevent under- and over-enforcement issues.
- We have conducted analysis on building a dialect-specific Arabic classifier for detection of any content in that language. Based on these results and input from linguists and language models experts, we will add expanded language identification functionality to our systems that can recognize content in different Arabic dialects.
- We are committed to updating our classifiers to regularly improve accuracy and performance.
- We launched a Hebrew classifier that proactively detects and actions violating hostile speech content in Hebrew.

For a more detailed update on progress made on the implementation of recommendations, see Meta Update: Israel Palestine Due Diligence Update.

# India

In Meta's first annual Human Rights Report, we provided a detailed summary of the India Human Rights Impact Assessment, which was developed after consulting The Danish Institute for Human Rights Guidance on HRIA of Digital Activities Reporting and Evaluation.

Based on the guidance of this group and other human rights experts, we believe the form of disclosure mitigates security risks as per UNGP 21(c). In this report, we detail actions we have already taken or intend to take. We will continue to study its findings and consider its

recommendations as a baseline to identify and guide related actions. We will track our implementation regularly through 2023 and beyond.

Since then, we have made the following progress:

- **Stakeholder Engagement:** In response to insights related to Meta's potential to deepen our stakeholder engagement in India comprehensively, Meta is refining its approach to civil society engagement and outreach, including a specific program of engagement with India civil society and community organizations.

- **Transparency Reporting:** Our reporting for India has become more detailed. In addition to our reporting on government requests for user data, we continue to provide a detailed India report every month, which includes information on actions taken against violating content on Facebook and Instagram, grievances received from users in India, and orders received from the recently formed Grievance Appellate Committees constituted under the IT Rules. We continue to refine our systems for gathering metrics on content removed as a result of government requests and public reporting, reflected in our reporting to the Oversight Board in response to its recommendations (Q4 2022 and Q1 2023).

- **Expanded Partnerships:** We have made good progress on our commitment to expand partnerships with India civil society including on digital literacy, women and child safety, and countering extremism. The Women's Safety Hub was relaunched as the Meta Safety Center in 2023 and translated into Assamese, Bengali, Gujarati, Hindi, Marathi, Punjabi, Tamil, Telugu, Kannada, and Malayalam. We have expanded participation by Indian organizations in our Resiliency Initiative and are preparing to launch our Search Redirect Program, which redirects users using extremism and violence-related search terms toward resources, education, and outreach groups.

- **Content Moderation:** In response to insights related to policies to further prevent discriminatory and hateful content against vulnerable groups, Meta has developed a prototype test for application of the Rabat Plan of Action for use when assessing potential hate speech. We have 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. We have increased the number of groups designated across all categories of Designated Organizations and Individuals in India to help ensure that our platform is not abused by bad actors.

In addition, Meta's experience in ensuring election integrity informs its preparations for upcoming state and general elections. These efforts include activating our elections operation center, ensuring content reviewers provide support in 20 Indian languages, expanding our independent fact-checking partners from seven to 11 (now covering 15 Indian languages), enforcing political ads transparency, and working closely with both electoral authorities and civil society.

## 07  Stakeholder engagement: How external parties strengthen our work



Engagement with external stakeholders around the world helps us live up to our human rights responsibilities and creates an important lever for accountability and transparency. In particular, we recognize the importance of meaningful engagement with stakeholders from underrepresented and marginalized communities. Stakeholder engagement is embedded throughout our Corporate Human Rights Policy and related work. We strive to listen to and seek advice from human rights experts and activists, academics, and others, as well as brief them on ongoing developments at Meta. Their insights inform areas such as development of our content policies and enforcement and our approach to algorithmic ranking and recommendations.

In 2022, Meta strengthened and expanded our ongoing external stakeholder engagement with the following groups.

| Stakeholders | Consultation and collaboration efforts |
|---|---|
| Civil society organizations | Support for Community Standards review and integrity efforts in key areas such as elections |
| Human rights defenders | Measures to protect account security and minimize spread of misinformation and violating content |
| Vulnerable or marginalized groups | Feedback mechanisms for content policy development and building safety services |
| International organizations | Sharing of perspectives on digital human rights policies and improving communication in high-priority countries |
| Trusted Partners | For this global program, improvements in identifying and flagging violating content, contributing local knowledge to our policy enforcement, and building safety services |
| Investors and advertisers | Consultations and briefings on Meta's human rights work including human rights due diligence |
| Users | Input on contentious issues and decentralized decision-making |

Some examples of our work include:

## Helping create secure online spaces and support election integrity in Nigeria

In the lead-up to the 2023 General Election in Nigeria, we engaged with a broad range of stakeholders. The engagements focused on our election policies and efforts to uphold the integrity and accuracy of information on our apps. We gathered insights about violating content trends, helping us to improve its detection and to maintain information integrity.

**Our actions:**

- Carried out digital security and safety training for vulnerable groups, including female politicians and female public figures, human rights defenders (HRDs), and activists
- Conducted a Dangerous Organizations and Individuals (DOI) Policy roundtable in Nigeria as well as two roundtables in London with the Nigerian diaspora
- Collaborated with Trusted Partners through workshops and consultations to help them identify and escalate content including hate speech, misinformation, violence and incitement, voter interference, and bullying and harassment

## Engaging marginalized and underrepresented communities in content policy

We made progress toward our goal to bring the voices of marginalized and underrepresented communities into content policy development. We developed an Inclusivity Framework to ensure our diverse stakeholders are considered in the development of policy as well as to inform Community Standards.

**Our actions:**

- **Solicited feedback on Community Standards:** Leaders from faith, indigenous, and ethnic communities, who are underrepresented on our apps, took part in eight roundtables across Europe, the Middle East, and Africa to discuss how effective our Community Standards are at creating an inclusive environment across Meta services and apps. These included roundtables in post-election Kenya and analysis of the migrant crisis in Poland.
- **Strengthened relationships with the LGBTQIA+ community:** We increased engagements with groups and representatives in Europe, the Middle East, and Africa on the impact of Meta's content policies on users, particularly regarding hate speech, bullying, and harassment. These included LGBTQIA+ HRDs, civil society organizations, academic scholars, and activists from the United Kingdom, Spain, Germany, Kenya, Ghana, Namibia, Uganda, Egypt, Sudan, Zimbabwe, South Africa, Gambia, and Tanzania.
- **Expanded outreach to Caribbean civil society:** We listened to concerns around the spread of potentially violating content and led discussions on digital rights, online safety and security, women's safety, and LGBTQIA+ rights. Experts and civil society groups from Haiti, Trinidad and Tobago, Jamaica, Panama, the Dominican Republic, and the Cayman Islands participated in the first roundtable of its kind.
- **Broadened engagement with faith communities:** We expanded our policy engagements to include Muslim women in the UK, Muslim communities in the Sahel, and the Sikh community in the UK, as well as the Bahá'í community with representatives from Canada, the U.S., Chile, Brazil, Germany, Spain, Switzerland, South Africa, United Arab Emirates, Egypt, and Tunisia.

- **Produced a guide on engaging persons with disabilities:** In collaboration with disability rights organizations working at the intersection of disability and tech policy, we produced a guide for our content policy teams.

**Perspectives from Indigenous Communities on non-medical drugs**
We engaged with stakeholders in North America, Gabon, Cameroon, and Zimbabwe to bring their voices into policy development on non-medical drugs in religious or traditional contexts. We explored how to balance freedom of expression about traditional and religious substances against the safety risks involved with the potential promotion of non-medical drugs, such as ayahusca. Among those consulted were a broad range of academics and civil society organizations, including scholars focused on health, medicine, anthropology, clinical psychology, drug policy, ethnography, ethno-ornithology, religion, and psychedelic drug research, as well as legal experts, regulators, freedom of expression advocates, health professionals, traditional/religious healer associations, community leaders, and indigenous groups. The input we received was used to craft a revision of our approach to positive discussion of non-medical drugs in traditional or religious contexts.

## Engagement with international organizations

The rapid growth and impact of technology has prompted many international organizations to create policies that address human rights in the digital world. We regularly collaborate with a wide range of international government bodies, including the Office of the High Commissioner for Human Rights, UNESCO, the Office on Genocide Protection and the Responsibility to Protect, UNICEF, the World Economic Forum, the Internet Governance Forum, UNHCR, OECD, the UN Foundation, and various regional organizations. We provide insight into our Community Standards, how they are enforced, and how we seek to protect HRDs. In 2022, we intensified our engagements with the UN at the country level, including in Ethiopia, Kenya, Ukraine, Haiti, and Afghanistan.

**Some examples of our actions:**
- UNICEF and Meta worked closely to ensure that critical information was available for Ukrainians, along with ad credits to support fundraising and advocacy messages in top global markets, Ukraine, and neighboring countries such as Poland and Romania.
- We met regularly with officials of the Office of the UN High Commissioner for Human Rights to discuss global and country-specific issues and participate actively in the B-Tech Project.
- We provided training on Community Standards and reporting

mechanisms for UN country teams in Ethiopia, Nigeria, and Haiti.

- We engaged with the <u>Office of the Secretary General's Envoy on Technology</u> on the Global Digital Compact and the need for it to support a globally harmonized approach to internet governance that promotes the free flow of data. This is essential for protecting human rights online, including the rights to privacy and freedom of expression.

- We held meetings with senior UN officials and diplomats around the High Level Week of the UN General Assembly to discuss human rights and related issues.

- We engaged with a number of UN Special Rapporteurs of the <u>UN Human Rights Council</u>, including those focused on HRDs, freedom of expression, Afghanistan, and Myanmar, amongst others, to share information and better understand their concerns as they relate to the role of Meta's services.

- We continued to respond to requests from the UN's <u>Independent Investigative Mechanism for Myanmar</u> for data to assist accountability efforts.

## Scaling identification and escalation of issues with Trusted Partners

We prioritize building partnerships in regions of the world where we face the most salient human rights challenges.

Our network of <u>Trusted Partners</u> comprises over 400 non-governmental, not-for-profit, national, and international organizations in 113 countries who report content, accounts, and behavior that we review in context. These Trusted Partners are important allies in our efforts to understand the local impact of our services. They help us stay abreast of emerging trends, can flag violating content that we may miss, and improve policies and enforcement in areas such as misinformation and harm. Their reports provide valuable insights to help inform policy development and help keep our users safe.

**Our actions:**
- In 2022, we expanded our network of Trusted Partners across 36 countries, including Yemen, Ukraine, Ethiopia, Mali, Kenya, Nepal, Haiti, and Cambodia.
- Our partners have alerted us to developments and risks that we might otherwise have missed. For example:
    - Trusted Partners reported a spike in sextortion taking place on our apps in Yemen, which led to an investigation and ultimately the removal of a large number of Facebook users, business accounts, and Instagram accounts.
    - In Cambodia, Trusted Partners flagged Facebook posts recruiting people for call centers in Southeast Asia that contained key indicators of human trafficking and labor exploitation. Over 30 posts were taken down.
    - Trusted Partners in Pakistan flagged misinformation around the Transgender Persons Act and targeted content that put individuals at risk. The content was removed, and we took steps to ensure we would catch further instances of similar policy-violating content.

# Innovating through Community Forums

Meta continues to explore ways to integrate more voices beyond Silicon Valley into decision-making. We piloted Community Forums based on deliberative democracy, a style of decision-making used by governments and organizations globally. It brings together representative groups of people to have structured conversations and make decisions about particularly complex or contentious issues.

**Our actions:**

In 2022, we held a model of group deliberation based on Deliberative Polling®, a method created at the Stanford Deliberative Democracy Lab. We collaborated with the lab as well as the Behavioural Insights Team, a data analytics group based in the UK, to ensure the process was fair and independent.

This forum was the largest deliberative polling forum to date, with over 6,300 people from 32 countries. It focused on addressing bullying and harassment in social virtual spaces.

Participants received independently vetted educational materials on the topic, deliberated among themselves in multiple small-group sessions, and had the opportunity to pose questions to independent experts from around the globe. These included cybersecurity specialists, human rights lawyers, bullying/harassment activists, experts from immersive virtual reality, and former brand safety leads for Twitter, TikTok, and Second Life, as well as UN and World Economic Forum advisory committee members and international Press Freedom Award winners.

Following these sessions, participants were polled about their support for the proposals they'd been discussing to help inform services and app development for virtual experiences. This event was part of our broader effort to create programs that foster more inclusive decision-making and enable more people to have a voice in the development of our apps and technologies.

Overall, participants strongly supported the notion that both creators and platform owners are responsible for monitoring the worlds they create, whether public or members-only spaces.

Participants supported the use of tools like automatic speech detection, video capture, and visible and invisible moderators. Support was stronger for these tools in public spaces versus members-only spaces.

For more detailed results, see Stanford's results report.

## 08    Language



Language is one of the most important factors influencing Meta's human rights impacts. The importance of language was underscored in stakeholder consultations in the Comprehensive Human Rights Salient Risk Assessment. In 2022, we increased our capacity to support more languages and enabled more people to exercise their rights online, including access to remedies.

Understanding local context is important for understanding potential human rights risks. That's why language capacity at Meta includes both human and machine-based translations. Language support includes language detection and language-based classifiers, content moderation, and translation of user interfaces, as well as other content like the Help Center or Community Standards.

We are continuously working on adding support for more languages. Between 2021 and the end of 2022, our team extended professional support for Facebook's user interface in 30 additional languages. The list includes languages that are underrepresented on the internet and for which linguistic sourcing and translation are a challenge, such as Javanese, Shona, Kirundi, Sinhala, Nepali, and Assamese.

Additionally, by the end of 2022, our Community Standards were available in 77 languages.

**No Language Left Behind**

Meta's No Language Left Behind (NLLB) is a first-of-its-kind AI breakthrough project that uses open-source models capable of delivering evaluated, high-quality translations directly between 200 languages — including lesser-spoken languages like Asturian and Luganda. It aims to give people the opportunity to access and share web content in their native languages, thus communicating with anyone, anywhere, regardless of their language preferences.

The technology behind the NLLB-200 model, now available through the Wikimedia Foundation's Content Translation tool, is supporting Wikipedia editors as they translate information into their native and preferred languages. This helps make more knowledge available in more languages for Wikipedia readers around the world.

During times of crisis, human rights are often under particular threat. At such times, Meta assesses on- and off-platform risks of imminent harm and responds with specific policies, services, and operational actions to respect rights.

In 2022, Meta strengthened our organizational response to support people during periods of escalated tension in any country, including during conflict, elections, or civil unrest. This response brings together a cross-functional team that focuses on prioritization of support and resources to efficiently anticipate and respond to moments of escalated tension anywhere in the world. The work is organized around global elections and active response countries as well as global readiness to anticipate and prepare for future critical events. A set of integrity standards is developed and refined to enable users to report violating content and is represented through nuanced country-language expertise.

## Crisis Policy Protocol

In August 2022, based on a recommendation from the Oversight Board, we launched our internal Crisis Policy Protocol (CPP) to help ensure our policy response to crises is principled and calibrated.

The CPP is a dynamic framework that allows us to identify emergent crisis situations and assess their relative severity. It guides our expedited use of targeted or exceptional policy

levers to mitigate potential harm consistent with observed risks and informed by past crisis interventions. As a result, the CPP helps us balance a consistent global response to crises with the ability to adapt to quickly changing local conditions. To develop the CPP, we consulted with over 50 global external experts in human rights, conflict prevention, hate speech, humanitarian response, and national security. We also undertook original research.

This protocol was designed to complement and bolster existing, company-wide crisis response efforts. The CPP helps ensure Meta uses a holistic and interdisciplinary policy approach to drive continuous improvement in our crisis response across services, policy, and operations teams.

Notable examples of our crisis response activities included:

## Russia and Ukraine

The Russian invasion of Ukraine in February 2022 posed significant human rights challenges for Meta. Propaganda and disinformation to influence public opinion proliferated online, as did surveillance, disruptions to internet services, and promotion of violating content. We needed to accelerate our content moderation efforts while balancing freedom of expression to help Ukrainian citizens coordinate humanitarian relief efforts and express their resistance to the invading forces.

In Russia, both Facebook and Instagram were blocked, and Meta Platforms, Inc., was added to extremist organization lists by the authorities. However, WhatsApp continued to be available.

**Due diligence:**

Since the onset of the war, our human rights due diligence has included interpreting relevant international humanitarian law and human rights standards to help our teams navigate the challenges. These efforts have included reviewing videos identifying prisoners of war as well as those promoting hate speech, misinformation, and Russian state-controlled propaganda denying that Russia had invaded Ukraine.

We engaged with human rights defenders (HRDs), activists, journalists, and civil society and human rights groups in both Russia and Ukraine to understand the issues they faced. We also engaged with the UN and other international institutions to deepen our understanding of the issues at stake and gain further insights into content moderation best practices in times of war.

**Our actions:**

Meta took a number of steps to address the challenges, including:

- **Setting up a special operations center** staffed 24/7 by experts from across the company, including Russian and Ukrainian language capacity, to monitor and respond to the rapidly evolving war in real time.

- **Intensifying our efforts to tackle violating content** such as violence and incitement, hate speech, coordinated inauthentic behavior, misinformation, and disinformation. This included the expansion of our third-party fact-checking capacity in Russian and Ukrainian languages across the region and the provision of additional financial support to Ukrainian fact-checking partners. We also onboarded five new Trusted Partners, convening listening sessions with them to improve our understanding

of content-related risks and develop appropriate reporting mechanisms.

- **Globally demoting content from Facebook Pages and Instagram accounts from Russian state-controlled media outlets** and making them harder to find across our apps. We also started <u>labeling and demoting posts and stories</u> that contain links to Russian state-controlled media websites on Facebook and Instagram. In the EU and Ukraine, we blocked all Russian state media at the insistence of their respective governments.

- **Rolling out safety features** focused on HRDs and other vulnerable groups, including the ability for people to lock their Facebook profiles, removing the ability to view and search friends' lists, and adding tools on Messenger to help protect people from being targeted.

- **Helping ensure HRDs'** and independent media's accounts received added protection against harassment or impersonation and that their voices were not unduly restricted through our cross-check program. This work is ongoing.

- **Helping predict refugee flows** by our <u>Data for Good</u> team sharing privacy-preserving datasets with Trusted Partners.

- **Helping Ukrainians** and others in the region find reliable information from local UN agencies and Red Cross societies by temporarily running Community Help as a central resource on Facebook.

## Iran

In September 2022, 22-year-old Mahsa Amini was arrested by Iran's morality police for "improperly" wearing her hijab. She died while in police custody, sparking widespread protests within Iran and around the world that continue. In response to the protests, Iranian authorities clamped down aggressively on freedom of speech and freedom of assembly, as well as limiting the use of the internet and apps like Instagram and WhatsApp, which are widely used by Iranians.

Meta moved swiftly and took a number of steps to help keep people connected and safe on our apps. This included setting up a dedicated Farsi-speaking team focused on addressing issues stemming from the protests. This team ensured we were applying our policies correctly and keeping our users safe. We continue to monitor the situation closely.

**Our actions: Helping protestors shine a light**

Instagram has been widely used by Iranians to shed light on the protests and the reality of events on the ground. Hashtags related to the protests in Iran were used on Instagram more than 160 million times in the first five months of the protests. Protestors shared their Instagram footage with international media outlets, many of which couldn't report from Iran.

- **Helping people connect:** We launched Instagram Lite in Iran in 2022 to help people access Instagram when bandwidth is reduced. In addition, WhatsApp <u>launched</u> a proxy-server capability that allows people to connect when their internet connection is disrupted or blocked, which was particularly important for communities in Iran.

- **Protecting activists and HRDs:** Social media accounts of activists and journalists are often targeted during protests with bullying, harassment, impersonation, and hacking attempts. We worked with Iranian activists to protect their accounts and help them raise awareness of human rights abuses.

Our approach to human rights and responsible product development



## Metaverse

In 2021, Meta created the <u>XR Programs and Research Fund</u>, a two-year, $50 million investment in programs and external research with industry partners, civil and human rights groups, governments, nonprofits, and academic institutions. These investments focus on areas that Meta must get right to realize the potential benefits of the metaverse. All proposals must align with Meta's work in one or more of the four policy priority areas:

- Economic opportunity: how we can give people more choice and maintain a thriving digital economy
- Privacy: how we can build meaningful transparency and controls into our products
- Safety and integrity: how we can help keep people safe on our platforms and give them tools to take action or get help if they see or experience something they're not comfortable with
- Equity and inclusion: how we can make sure these technologies are designed inclusively and in a way that's accessible

Partners in this initiative include universities and research institutions, international organizations, civil rights groups, and NGOs drawn from or working with underrepresented communities from every region in the world.

## Artificial intelligence

We believe that artificial intelligence (AI), developed and deployed responsibly, can be a powerful tool for advancing human rights. AI plays an important role in helping us address many of the potential adverse human rights impacts discussed in this report. For example, AI powers have scaled detection and review of problematic content across a growing number of languages, advanced identification of potentially unwanted contact and interactions, improved accessibility through automated caption generation, and enhanced a range of our Data for Good efforts.

At the same time, the development and use of AI poses novel human rights risks. Among other potential risks, AI models may exhibit problematic biases or discriminatory impacts and generate problematic content or engage in over-enforcement that adversely impacts freedom of expression.

We are committed to addressing these issues in an open and collaborative way. For example, demographic data is essential in assessing whether a service, app, or process  treats all groups fairly. Gathering this data, however, is sensitive and raises important concerns about how to protect people's privacy.

We put responsibility and human rights at the center of our approach. AI is embedded in our Corporate Human Rights Policy. We also recognize the importance of the OECD Principles on Artificial Intelligence, which explicitly refer to human rights and are widely adopted and endorsed by the G20. We hope to foster more collaborative and transparent dialogue across disciplines and audiences about the path ahead for these critical issues.

We created a dedicated, cross-disciplinary Responsible AI (RAI) team that seeks to ensure that AI at Meta benefits people and society. The team includes ethicists, social and political scientists, policy and rights experts, researchers, and engineers. The RAI team works closely with the Human Rights team to help ensure that our machine-learning systems are designed and used responsibly.

We also pioneered an approach to AI system cards to provide standardized transparency into how AI systems — not just individual models — function, beginning with a pilot focused on Instagram Feed ranking.

**Five Pillars of Responsible AI**

Responsible AI at Meta is built on five pillars:

- Privacy and security
- Fairness and inclusion
- Robustness and safety
- Transparency and control
- Accountability and governance

We've made progress in addressing many of these issues, and there remains much work to be done. In 2022, we worked to:

- **Expand language support**
  - Our open-source No Language Left Behind effort harnesses AI to deliver high-quality translations directly between 200 languages.
  - Our Massively Multilingual Speech (MMS) recognition model supports speech-to-text and text-to-speech for 1,107 languages and supports language identification for over 4,000 languages. Some of the supported languages, such as Tatuyo, have only a few hundred speakers, and for most of these languages, no prior speech technology exists. We publicly share the models and code.
  - Wikipedia teamed up with Meta AI to fact-check citations using Meta's open-sourced AI program, Sphere.
- **Address bias**
  - In our drive toward ensuring that AI systems treat everyone fairly, we introduced and open-sourced several datasets and models responsible for addressing this gap in privacy-protected ways.
  - We trained an AI model for reducing demographic biases in text — a demographic text perturber — which can help break stereotypical associations present in natural language programing (NLP) data.
  - We introduced and open-sourced two new datasets to help measure fairness and mitigate potential bias in NLP models. These datasets include a more comprehensive representation of different demographic dimensions to measure fairness in these models, including terms for gender identity, age, race, and people with disabilities.
- **Better understand problematic content associations**
  - We assembled a cross-disciplinary team, including people from our Civil Rights, Human Rights, Engineering, RAI, AI Research, Policy, and Product teams, to better understand problematic content associations in several of our end-to-end systems. We worked to

implement technical mitigations to reduce the chance of the problematic content associations occurring on our apps that use AI models.

- **Assess fairness**
  - One way we address AI fairness is through the creation and distribution of more diverse datasets. We also, through our Casual Conversations v2 project, designed a large, consent-driven database to measure algorithmic bias and robustness.
  - Meta AI Research began exploring new public fairness indicators to quantitatively assess three well-documented types of potential harms and biases in computer vision models. These fairness indicators complement existing approaches to responsible AI development, such as data and model documentation. The indicators are specifically designed to adapt and evolve as research advances and new approaches emerge.

We also made advances in SEER (SElf-SupERvised), Meta AI Research's self-supervised computer vision model that is focused on improving results for diverse image sets — without the need for the careful data curation and labeling that goes into conventional computer vision training.

# Looking forward: Future considerations



Human rights standards and principles remain humanity's best articulation of our universal values and best foundation on which to build legitimate institutions of governance. Internet governance models have made a great deal of progress over the last few years. As this report demonstrates, we've worked throughout 2022 to continue to advance our governance models and make good on our Corporate Human Rights Policy and our commitment to the United Nations Guiding Principles on Business and Human Rights.

This work continues, as does our commitment to progress. We look forward to implementing many of the advanced accountability mechanisms discussed in this report. For example, we continue our work to enable people to participate in governance measures through community forums and assemblies.

We will also continue our work keeping human rights central to how we understand and implement the rapidly advancing science of artificial intelligence (AI). While there are many challenges ahead and much yet to be achieved, we are proud of our wide-ranging efforts to solve some of the hardest problems of social media and AI technologies.

We will continue to rely on human rights due diligence to better ascertain the most important human rights risks faced by people who use our services and to craft policy that addresses those risks.

We operate in a complex and fractured regulatory environment. The borderless and largely free internet that many of us use every day is under threat. It is vital that we work with governments, the United Nations, and civil society to ensure that evolving regulatory frameworks are built on human rights principles and values. In full recognition of our human rights responsibilities, we will strive to advance these principles in our daily work for the benefit of our users, and of societies at large.