



∞ Meta

Human rights report

Insights and actions
2023

Table of Contents

About this report	3
Executive summary	5
Foreword	8
AI in the context of human rights	9
Taking an open approach	9
Addressing potentially harmful generative AI outputs	11
Human rights risk management	13
Day-to-day due diligence	13
Content policy	16
Product development	17
How we prepare for and respond to critical events and global crises	19
How Meta prepares for elections	19
Brazil	20
Pakistan	22
Response to global crises	24
Sudan	24
Nagorno-Karabakh	26
Migration in Central and South America	28
Human Rights Impact Assessments	29
Israel and Palestine: Update on our actions	29
Temporary policy measures	30
Temporary product measures	31
End-to-end encryption: Update on our actions	32
Issue spotlights	33
Combating security threats	33
Child and youth safety	34
Child exploitation	34
Youth safety and well-being	36
Stakeholder engagement	37
Marginalized groups and human rights defenders	37
Trusted Partners	39
Case study: Surfacing novel harms in Haiti	40
Case study: Ramping up reporting in Ethiopia	40
Case study: Mitigating intra-religious conflict in Bangladesh	40
International and multi-stakeholder organizations	41
Transparency and remedy	44
Responding to government requests	45
Oversight Board	50
Looking forward	53
Annex	55
How human rights are governed and managed at Meta	55
Training Meta employees on human rights	55
Links to referenced reports	56



About this report

This is our third annual Human Rights Report. It builds on Meta’s work and reflects progress made on our commitments to the [United Nations Guiding Principles on Business and Human Rights](#) and our [Corporate Human Rights Policy](#). It covers our actions from January 1, 2023, through December 31, 2023. We report on Meta services and products including Facebook, Messenger, Instagram, WhatsApp, Threads and Reality Labs.

Human rights is a topic of critical importance to our stakeholders and to our company. The content in this report reflects the feedback received from our stakeholders following the release of our [2022 Human Rights Report](#) and is grounded in our [Comprehensive Human Rights Salient Risk Assessment](#), which is summarized in the 2022 report. The purpose of the assessment was to identify and prioritize our most significant potential adverse human rights impacts¹ on people, including users and others who may be affected by our actions across the enterprise globally. This report highlights how we at Meta managed, across global operations, the following eight salient human rights:

Top 8 potential salient human rights from our Comprehensive Human Rights Salient Risk Assessment

Salient human right

Report section

Freedom of opinion and expression

Content policy, How Meta prepares for elections, Response to global crises, Israel and Palestine: Update on our actions, Child and youth safety, Responding to government requests, Oversight Board

Privacy

Content policy, End-to-end encryption: Update on our actions, Combating security threats, Child and youth safety, Responding to government requests

Equality and non-discrimination

AI in the context of human rights, Product development, Israel and Palestine: Update on our actions

Life, liberty and security of person

How Meta prepares for elections, Response to global crises

Best interests of the child

Content policy, Child and youth safety, Stakeholder engagement

Public participation, to vote, and to be elected

How Meta prepares for elections: Brazil, Pakistan

Freedom of association and assembly

End-to-end encryption: Update on our actions, Response to global crises: Sudan

Right to health

Response to global crises: Sudan

¹ The term “adverse human rights impact” is in line with the UN Guiding Principles on Business and Human Rights and means an impact that occurs when an action removes or reduces the ability of an individual to enjoy his or her human rights.



Our salient human rights risks are found at the top of sections in this report.

We engaged with many internal and external stakeholders and strive to provide a representative picture of our work across multiple teams around the globe. The report is intended to provide an overview of the human rights issues we deal with as a company, whilst guiding the reader to explore more in-depth content as needed.

This report complements the most recent [Meta Responsible Business Practices Report](#). We [report](#) separately on our efforts to identify and mitigate the risks of modern slavery and human trafficking in our business operations and supply chains. Links to other Meta disclosures can be found in the [Annex](#) to this report.

Our 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

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 Facebook and Instagram, 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.





Executive summary



With this Human Rights Report, we seek to give insight into the policies, processes and standards Meta uses to manage human rights risks for billions of people, at scale, and in real time. We aim to convey how we strive to respect the rights of users, and other communities, across our work and the systems we use every day, from the way we respond to crises and plan for elections to how we respond to government demands.

This year's report covers our activities of the calendar year 2023 and reflects feedback from stakeholders regarding [last year's Human Rights Report](#). This year's report highlights how we at Meta managed across global operations the following eight salient human rights: freedom

of opinion and expression; privacy; equality and non-discrimination; life, liberty and security of person; best interests of the child; public participation, to vote, and to be elected; freedom of association and assembly; and right to health.

Meta's services and apps continued to be used for good — helping to build community, foster innovation and mobilize resources to support humanitarian relief, among other efforts. As challenges emerged, our [Corporate Human Rights Policy](#) informed our actions to address issues such as the use of our services in countries in conflict, cybersecurity and overreaching government requests for data.



We continue our work to keep human rights in mind in how we understand and develop artificial intelligence (AI). Our generative AI products allow people to advance their human rights in novel ways, but we recognize that these systems are imperfect and present opportunities for continued improvement. To mitigate potential risks, we developed an [Acceptable Use Policy](#), incorporated into the [Community License Agreement](#) for our open Llama 2 model, and provided safety tools to help developers build responsibly. We believe that an open approach can drive innovation through iteration and lead to better, safer products through ongoing feedback, scrutiny, development and mitigations.

Proactive engagement with stakeholders is central to our approach to human rights and is embedded in our Corporate Human Rights Policy. Our services development, content policies and moderation, and Community Standards are [informed](#) by insights from a wide variety of civil society organizations, human rights defenders, marginalized groups, international organizations, investors, advertisers, users and [Trusted Partners](#). These groups range across the political and ideological spectrum to provide diverse inputs. In January 2023, our network of Trusted Partners included over 400 non-governmental organizations, humanitarian agencies, human rights defenders and researchers from 113 countries around the globe.

2023 marked the 10-year anniversary of Meta joining the [Global Network Initiative](#) (GNI). We apply, and are accountable to, the GNI principles for how technology companies should respect the freedom of expression and privacy rights of their users when addressing government requests and restrictions. In the report, we outline how the GNI guided our response to government requests, including restrictions to content and access to user data.

We also worked to integrate human rights principles to prepare for and respond to critical events and issues in 2023. The report covers our readiness for heightened challenges in high-risk countries, how we prepare for planned critical events, such as elections, and our coordinated response to crisis events. In 2023, Meta enabled preparation for over 130 elections worldwide. We report here on our response to conflicts in Sudan and Nagorno-Karabakh, as well as human rights issues related to migration in Central and South America. Our crisis response efforts involved many different experts from across content policy, global operations, human rights and product, and other internal experts who assessed and responded to these situations, leveraging our [Crisis Policy Protocol](#). We also describe how we have followed up on the recommendations from previous human rights due diligence regarding [Israel and Palestine](#) and our actions in the most recent conflict, as well as an update on end-to-to-end encryption.



We reinforced our commitment to transparency and remedy as central tenets to human rights risk management. We continued to engage the [Oversight Board](#), an industry-leading effort created to help Meta answer some of the most difficult questions around freedom of expression and safety online. 2023 was a landmark year for the Oversight Board, as they surpassed their goal of 50 decisions annually, more than tripling the number of decisions made in 2022.

In addition, we improved our user experience on Facebook and Instagram for content restricted by Meta based on legal requests from governments and courts. In a majority of cases, our notifications informed users which state authority sent the request resulting in the restriction and the country in which the content is restricted.

In 2023, we submitted our first set of takedown requests from Austria, Vietnam, Singapore, India and Mexico to the [Lumen](#) database. This independent research project, hosted by Harvard University's Berkman Klein Center for Internet and Society, enables researchers to study online content takedown requests from governments and private actors. This is a further step toward enabling the global community to analyze, report and advocate for the digital rights of internet users.





Foreword

We continue to confront new and complex challenges as the world evolves. Our response to these challenges is guided by our [Corporate Human Rights Policy](#) and our commitment to the [United Nations Guiding Principles on Business and Human Rights](#) and [UN Global Compact](#). And we have checks and balances. The [Oversight Board](#) offers expert and independent oversight to ensure fair and balanced decision-making, particularly concerning issues of freedom of expression.

Our policies and procedures also evolve as technology and society change. We consider input from a wide range of stakeholders with an emphasis on multiple dimensions of diversity, including viewpoint diversity. We seek feedback from employees, experts and human rights defenders; marginalized groups and international organizations, and a broad range of civil society groups when revising these policies and procedures.

We use a variety of assessment tools and protocols across the business to identify, prevent and mitigate human rights risks that may be linked to Meta products, policies and operations globally. We report back on previous assessments to share our progress and open channels for feedback.

In 2023, we applied our risk management tools when responding to crisis situations, including conflicts and wars in Africa and the Middle East, as well as to plan for predictable world events, such as the significant number of elections. As such, our policies are dynamic

to ensure that we take an appropriate, tailored response to each situation while balancing competing interests and human rights considerations.

Meta has been a pioneer in artificial intelligence (AI) development for more than a decade. We know that progress and responsibility must go hand in hand. Generative AI tools offer vast opportunities, and we believe that it is both possible and necessary for these technologies to be developed in a way that respects human rights.

A strong and open digital market is of vital importance to help people, society and countries thrive and build community. Today, more than ever, we need technology to help bring the world closer together. We strive to play our part with human rights firmly fixed in our effort.

Signed by:



Nick Clegg,
President of Global Affairs



Jennifer Newstead,
Chief Legal Officer



AI in the context of human rights

Relevant salient risks:

- freedom of opinion and expression
- equality and non-discrimination
- privacy

Rapid advancements in artificial intelligence (AI) — particularly [generative AI](#) — commanded global attention in 2023. Given the increased attention and potential opportunities and challenges AI presents, we provide an extended discussion of our AI work from a human rights perspective.

In 2023, we released our open [Llama 2](#) large language model (LLM), [Meta AI assistant](#), [Emu](#) image generation model and [in-app creative tools](#).

As sophisticated AI chatbots and image generation tools entered into widespread use, they also attracted public and regulatory oversight. For example, the European Union advanced its landmark [AI Act](#), while the United States launched the [White House Voluntary Commitments on AI](#) with support from Meta and other AI developers.

We are committed to developing and deploying AI responsibly while mitigating potential adverse human rights impacts. AI is included in our [Corporate Human Rights Policy](#), which recognizes the importance of the [OECD Principles on Artificial Intelligence](#), widely adopted and endorsed by the G20.

We believe that responsible AI can be a powerful tool for advancing human rights.

We believe that responsible AI can be a powerful tool for advancing human rights. Our generative AI products allow people to exercise their freedom of expression in novel ways, improve access to information and education, and improve accessibility. For example, AI-powered captioning and image recognition on Ray-Ban Meta smart glasses improve accessibility for people with disabilities.

We leverage AI to try to rapidly detect and respond to potentially policy-violating content and unwanted interactions across a growing number of languages to prevent it from escalating. Many of our [Data for Good efforts](#) leverage AI to support humanitarian responses to crises and help facilitate public health work.

Taking an open approach

Meta is striving to take a responsible approach to developing and deploying generative AI products and models.



We believe that an open approach can drive innovation through iteration and lead to better, safer products through ongoing feedback, scrutiny, development and mitigations from the community. It can also:

- advance fairness and support diverse views by enabling impacted rights-holders to better identify potential biases.
- lower barriers to innovation, resulting in economic benefits.
- more easily allow global communities to tailor AI advancements for specific languages and contexts.

However, we acknowledge that an open approach can also present potential human rights risks, including that developers may ignore Meta's responsible use instructions and deploy models in unsafe ways. To attempt to mitigate these risks, we carry out safety testing and have developed a detailed [Acceptable Use Policy](#), incorporated into the [Community License Agreement](#) for our Llama models. We also shared a detailed [Responsible Use Guide](#) and provided safety tools, through our [Purple Llama](#) project, including [Llama Guard](#), to help developers' efforts to build responsibly with these openly available foundation models and deploy them safely for their own use cases.

We believe that an open approach can drive innovation through iteration and lead to better, safer products through ongoing feedback, scrutiny, development and mitigations from the community.

[Deploying AI safely](#) is a shared, ecosystem-wide responsibility, which is why we've collaborated for many years with organizations that are working to build safe and trustworthy AI. For example:

- We've been working with [MLCommons](#) and a global set of partners to create [responsibility benchmarks](#) in ways that benefit the open source community.
- We participated in multi-stakeholder initiatives like the [Partnership on AI](#) to address issues ranging from responsible deployment of foundation models to the most appropriate ways to identify synthetic content.
- We co-founded the [AI Alliance](#), a coalition of companies, academics, advocates and governments working to develop tools that enable an open and safe AI ecosystem.

We consulted with human rights stakeholders to help them understand and provide feedback on our use of AI. As part of these efforts, we engaged with the United Nations Human Rights Office’s [B-Tech Project on its work on generative AI](#), informing [three foundational papers](#) grounded in the UN Guiding Principles on Business and Human Rights. We also engaged with an advisor to the Office of the UN Special Rapporteur for Freedom of Opinion and Expression and UNESCO to help build knowledge of foundational AI concepts and risks, and facilitated informed discussion to understand concerns. Additionally, we [ran a consultative Community Forum](#), conducted in partnership with Stanford University, that incorporated human rights questions to help Meta and other companies, researchers and governments make decisions about generative AI chatbots. [Findings](#) emphasized the importance of the human rights framework and of incorporating information from human rights sources.

Addressing potentially harmful generative AI outputs

We’re conscious that generative AI technology might produce outputs that may, for example, generate potentially hateful, offensive or discriminatory content; reinforce biases; present inaccurate information; and/or raise privacy considerations. We also know bad actors may try to abuse the models and tools we release, in deliberately harmful ways. We’ve taken a range of steps to mitigate these risks in our foundation models, including Llama and Emu. We believe that taking an open, iterative approach will help us improve in real time if these issues arise.

The UN B-Tech Project [highlights](#) a number of relevant risk areas across several internationally recognized rights. We look to this guidance, and Meta’s [salient human rights risks](#), to inform our work.

For our foundation models [Llama 2](#) and [Emu](#), these mitigations, in part, aim to address significant potential human rights risks, by addressing risks in training, addressing contextually sensitive associations, and conducting safety evaluations and tuning.

As we continue to further develop our foundation models, we’ve shared [detailed information](#) about our responsible approach publicly.

In addition to our proactive efforts to mitigate risks associated with potentially harmful AI-generated outputs, we also work to respond to potentially harmful outputs as we become aware of them.



Image generated by Meta AI



In the days following the terrorist attacks on October 7 in Israel, stakeholders notified us of several problems with our tools, all of which we quickly acknowledged and worked to fix. The main problems were machine translations in which the word “terrorist” was appended to translated Instagram profile bios of some users, and AI-generated stickers that depicted Palestinians as violent.

As soon as we became aware of these issues, our engineering teams began an investigation to determine root causes and implement fixes. Our teams identified that these issues appeared to be related to model hallucinations and the training data used. Both issues are well-documented challenges for AI-powered products (see [here](#) and [here](#)).

We delivered an emergency fix to our machine translation product within 90 minutes of learning that there was an issue. We also worked as quickly as possible to mitigate potentially problematic associations in generated image and sticker outputs related to October 7. For example, we added known surnames to the model to prevent hallucinations. In addition, we fine-tuned our foundation models for image generation to better address a wide range of potentially problematic associations related to conflicts, and deployed them across our products.

See the [Israel and Palestine: Update on our actions](#) section for further details.

We look forward to ongoing engagement with a range of rights-holders with diverse views as we strive to put safety, responsibility and human rights at the center of our work to develop and deploy generative AI products. We will also take appropriate measures to comply with legal obligations in this area.

Human rights risk management

Every day, we actively seek to translate human rights principles into meaningful action. The foundation of our work is the [Meta Corporate Human Rights Policy](#). Our commitment and approach are informed by the [United Nations Guiding Principles on Business and Human Rights](#), plus international and regional human rights standards listed in our Corporate Human Rights Policy.

Our work includes, but is not limited to, our due diligence efforts, [stakeholder engagement](#), [product advisory](#), [human rights training](#) and providing data-driven advice to all aspects of the business. In this report, we focus on the different types of human rights due diligence we do at Meta. To undertake due diligence, we apply a range of assessment tools and protocols to identify, prevent and mitigate risks that may be linked to our products, policies and operations globally.

This work informs how we think about regulation, alongside our [Global Network Initiative \(GNI\)](#) commitments.

The regulatory environment is changing and has implications for



human rights as evidenced by recent global trends. For example, during 2023, we saw the UK Online Safety Act being enacted and the European Union’s Digital Services Act come into force, as did other online safety laws, including in Singapore, Taiwan and Uruguay.

The chart on page 15 captures some of the actions we take for human rights risk management. Further details are also available in our [2021](#) and [2022](#) Human Rights Reports.

Day-to-day due diligence

Through our day-to-day due diligence work, many different teams collaborate to continuously integrate human rights into our activities. For example, human rights and civil rights experts at Meta provide ongoing advice to assess and mitigate human rights risks on content policy, product development, crisis and conflict response, and election preparation. This work often involves matters related to hate speech,



misinformation, graphic content that may document human rights abuses, and measures to help protect human rights defenders, as well as other human rights-related issues.

With 3.19 billion people on average using at least one of our apps daily as of December 2023, the potential human rights impacts of our products vary significantly across time, location, content and affected communities. For this reason, we prioritize our work based on our 2022 [Comprehensive Human Rights Salient Risk Assessment](#), which analyzed impacts across all internationally recognized human rights. The analysis prioritized the eight most salient risks using [UN Guiding Principles on Business and Human Rights criteria](#), outlined in the graphic below.

We often grapple with how to prioritize our actions regarding potential risks to human rights. This tension is inherent to the human rights framework and is brought to the fore when we seek to reconcile many — sometimes competing — human rights. We seek to prioritize the most salient human rights risks, meaning those with the potential of the most severe negative impacts to stakeholders.

We seek to prioritize the most salient human rights risks, meaning those with the potential of the most severe negative impacts to stakeholders.

Our prioritization is based on the guidance of the UN Guiding Principles on Business and Human Rights and our Corporate Human Rights Policy, and enriched by stakeholder engagement organized around multiple dimensions of diversity. For example, prioritization is necessary to enable the use of our products and services by vulnerable categories of individuals — such as youth — within the broad and diverse overall user population. Meta looks to the [UN Convention on the Rights of the Child](#), in particular the principle of the “Best Interests of the Child,” to inform how it serves youth. This at times leads to complex decisions between which potential rights impacts take priority: for example, acknowledging the safety and well-being of children while respecting their right to seek information privately on topics such as politics, health or gender identity. The [Best Interests of the Child framework](#) informed the adoption and design of features like [Family Center](#) and [Parental Supervision](#).



Our approach to human rights risk management

Human rights commitments

We implement our commitment to human rights by approaches set out in the UNGPs, including:

Applying human rights policies

Conducting human rights due diligence and disclosure

Providing access to remedy

Maintaining oversight, governance and accountability

Protecting human rights defenders

Prioritization of rights

Comprehensive Human Rights Salient Risk Assessment identified eight human rights as priorities:

Freedom of opinion and expression

Privacy

Equality and non-discrimination

Life, liberty and security of person

Best interests of the child

Public participation, to vote, and to be elected

Freedom of association and assembly

Right to health

Key tools and examples

We deploy the following tools and processes to mitigate risks, with some illustrative examples covered in this report:

Content policy development → Dangerous Organizations and Individuals update

Product advisory → Generative AI; Threads

Human rights embedded into election preparation → Pakistan elections

Human rights embedded into coordinated response efforts → Prioritization for Active Response Countries status

Crisis response → Nagorno-Karabakh and Sudan conflicts

GNI framework for government requests → Content restrictions based on local law and access to user data

Oversight Board → Input on referral cases, e.g., hostages in international law

Stakeholder engagement → Co-design sessions for parents and teens

Protection of human rights defenders → Human Rights Defender Fund



Relevant salient risks:

- freedom of opinion and expression
- privacy
- equality and non-discrimination
- life, liberty and security of person

Content policy

As part of our foundational due diligence work, our human rights experts support the content policy development process. This means that they review proposals to update our content policies through the lens of human rights law to take account of freedom of expression and protection from discrimination, as well as other human rights. In 2023, this included refining our [Dangerous Organizations and Individuals policy](#). This work was conducted in different phases in 2023, and we published an [update](#) on our work in January 2024.

Under the Dangerous Organizations and Individuals policy, we do not allow organizations or individuals that proclaim a violent mission or are engaged in violence to have a presence on Meta’s services. Our global approach is based on [comprehensive definitions](#) of dangerous organizations’ types and tiers, as published in our Transparency Center. We have an independent process to identify dangerous organizations and individuals, with each case reviewed by relevant experts against corroborated evidence. Our definitions and thresholds are agnostic to region or ideology and have been developed in consultation with experts and academics from around the world. In addition, we enforce on U.S.-designated Foreign Terrorist Organizations, Specially Designated Narcotics Trafficking Kingpins, and Specially Designated Global Terrorists.

In line with [Oversight Board](#) and [Business for Social Responsibility](#)’s Israel and Palestine human rights due diligence recommendations, we [updated](#) our Dangerous Organizations and Individuals policy in the following ways:

- We updated our policy to allow more social and political discourse, including about elections, conflict resolution, and disaster and humanitarian relief, while continuing to meet our legal obligations.
- We heard feedback that our previous definition of “praise of a dangerous organization or individual” was too broad, so we updated our policy to make it more nuanced and proportionate. We now prohibit “glorification” of the violence and hate of dangerous organizations and individuals, which is more narrowly defined than “praise.”
- We simplified the way we assess and categorize dangerous actors depending on their ties to offline harm and violence.
- We updated our delisting process to provide more detailed and comprehensive criteria that must be satisfied for a dangerous organization or individual to be considered for delisting. This update enables us to ensure our designations can reflect changing circumstances.



In September 2023, we also made changes to our [Violence and Incitement policy](#) to refine language on high-severity violence. In March 2023, we made changes to our [Bullying and Harassment policy](#) to include a definition of “public figures” as recommended by the Oversight Board. In June 2023, we held a Policy Forum on Violent and Graphic Content in Human Rights Abuse Contexts in response to an Oversight Board recommendation in the [Sudan graphic video case](#). We considered whether our policies on violent and graphic content struck an appropriate balance between respect for the privacy and dignity of victims, freedom of expression and the well-being of our communities. Our analysis indicated that the balance was appropriate, and we did not change our policies.

To support transparency, policy changes are reflected in our [Community Standards](#) and can be viewed in the relevant change log and the [public minutes for our Policy Forums](#).

In addition, we updated our [penalty system](#) in 2023. The new system reduced the number of restriction periods and aims to help ensure a proportionate response to policy violations. We will still apply account restrictions to persistent violators, typically beginning at the seventh violation, after we’ve given sufficient warnings and explanations to help the person understand why we removed their content. For more serious violations, such as posting content that includes terrorism, child exploitation, human trafficking, suicide promotion, sexual exploitation, the sale of non-medical drugs or the promotion of

dangerous organizations and individuals, we will continue to apply immediate consequences.

Product development

Our engineers strive to creatively solve real-world problems. Human rights guide our work developing responsible innovation practices, and we are committed to due diligence for new products.

In 2023, we launched a number of exciting new products, such as [Threads](#) in July, [Ray-Ban Meta smart glasses](#) and [WhatsApp channels](#) in September, and [Quest 3](#), our mixed reality headset, in October. For each of these products, human rights considerations were embedded into the development process.

Additionally, we continued to build for the metaverse, including to help enhance education and counter bias and hate. For example, we launched multiple virtual reality experiences to build community, including [MLK: Now is the Time](#), [Inside the Mosque](#) (a [two-part](#) experience), and [ABLE](#).

In a fast-paced and iterative environment, teams build, test, refine and deploy products and services — including those enabled by AI. We build with [privacy in mind](#), and we’ve designed product risk mitigation approaches that match the particular needs of our dynamic product development process. Early in the development process, product teams apply guidance on responsible innovation, consistent with human rights principles. This helps the teams anticipate

Relevant salient risks:

- freedom of opinion and expression
- privacy
- equality and non-discrimination



and mitigate potential harms to individuals, communities and society. Our civil rights and human rights experts may provide advice or conduct rapid sprints, for example, to identify highest-risk countries for product launches.

For example, during the Threads product development process, our human rights experts leveraged existing product and policy review processes to assess potential human rights impact. They worked with the Threads team to implement mitigations, including in relation to potential government requests to censor content.

We are also continuing to develop Project Height, a framework for product teams to assess civil rights concerns presented in product launches. This enables product development teams to build with civil rights in mind during the product development process. It complements other risk processes we have in place to assess the safety of our products.



MLK: Now is the Time by Meta Quest



Relevant salient risks:

- life, liberty and security of person
- public participation, to vote, and to be elected
- freedom of opinion and expression
- equality and non-discrimination
- privacy
- best interests of the child

How we prepare for and respond to critical events and global crises

We work across the company to integrate human rights principles into Meta’s integrity efforts.³ This includes how we prepare for heightened challenges in high-risk countries, as well as planned critical events, such as elections, and our coordinated response to crisis events. This coordinated work across specialized teams helps us to anticipate and respond efficiently to escalated tensions wherever they occur.

How Meta prepares for elections

2023 saw many countries going to the polls—and even more preparing for a busy 2024 electoral cycle. In 2023, Meta enabled preparation for over 130 elections worldwide, including in Pakistan, Argentina, Türkiye and Nigeria. We share insights and our actions from two relevant elections, in Brazil and Pakistan, below.

Our specialized teams drive Meta-wide preparations to put protections in place ahead of global elections.

³“Integrity” is a term of art inside Meta. Trust and safety, computer and account security, the reduction of bad experiences, related privacy issues and more are handled by a network of teams within Meta that often take “Integrity” as part of their name. Among many things, these teams build tools to prevent harm, moderate our platforms and enforce our policies.

These may include advanced security operations to tackle adversarial threats, updated policies and processes to remove potentially harmful content from our platforms, our industry-leading global fact-checking network, and transparency around political and social issue ads.

We also continually assess the risks of imminent harm during critical events, such as elections, so we can respond with targeted, time-limited policy and product actions that will help keep people safe.

While most of our work is proactive and undertaken well in advance of election dates, we also prepare to react to high-risk events. In these cases, Meta has multiple approaches we can take to address a range of scenarios—for example, changing or limiting product features, introducing messaging rate limits and limiting distribution of content.

Some of the ways we identify potentially harmful trends are by gathering information from public reports, reviewing recommendations from our Trusted Partners, making



ongoing observations of content trends, human rights due diligence and examining assessments from our intelligence experts. This information, along with existing Community Standards and enforcement systems, helps to inform the type of product and policy mitigations we may use to prevent abuse during high-risk elections. This includes our [Crisis Policy Protocol](#) and separate measures we may take to adapt our systems in response to [heightened risk](#). We also provide transparency tools, such as our [Ad Library](#).

Brazil

We started preparing for Brazil’s [2022 presidential election](#) a year before the vote. We continued our efforts during the post-election period, which enabled us to quickly respond to the [January 8, 2023, attacks](#) on the Brazilian Congress, Supreme Court and other public buildings. Meta also designated the post-election unrest as a crisis under the Crisis Policy Protocol to help the company assess how to mitigate content risks.

Our preparations for the 2022 election included mapping risk scenarios for during and after the election. We considered international standards on freedom of expression and other human rights. We also made

changes to products and adapted policies to protect the integrity of the Brazilian election. This included partnering with the Superior Electoral Court to provide people with access to reliable information about voting.

We deployed a range of tools and methods to combat potential hate speech, incitement to violence and misinformation before, during and after the attacks on January 8, 2023. We designated Brazil as a temporary high-risk location shortly before the election, so that we removed calls to bring armaments or to enter government buildings forcibly, as noted in the [Oversight Board’s “Brazilian General’s Speech” case](#).

From the start of the election campaign on August 16, 2022, until the end of January 2023, [we removed](#) more than 1 million pieces of content on Facebook and more than 960,000 pieces of content on Instagram for violations of our violence and incitement policies in Brazil. This content included posts calling for military intervention.

During this five-month period in Brazil, we removed pieces of content that violated our hate speech policies:

 **570k+**

 **520k+**

We also removed pieces of content in Brazil for violating our bullying and harassment policies:

 **380k+**

 **630k+**

Gender-based violence

The threat of gender-based harassment and violence against women candidates, journalists and human rights defenders often escalates during elections.

To mitigate these risks and as part of our broader approach to increase protection of women on platforms and connect them to public services, our actions in Brazil included:

- partnering with the Ministry of Women to launch [an official channel on WhatsApp in Brazil](#) enabling women to pass on complaints or request information from the Ombudsman about violence against women, as an additional means of communication with the Ombudsman. Users are able to obtain information about the relevant laws, find addresses of services dedicated to assisting women and speak directly to attendants.
- publishing [an online guide](#) in Portuguese to protect women against online violence.
- engaging with industry groups, civil society organizations and regulators, including the Ministry of Women, on joint initiatives to address gender-based violence and harassment.



Pakistan

Although the Pakistan elections took place in 2024, our electoral preparations began in 2022. They included company-wide efforts across product, policy and operations to avoid and mitigate risks to human rights related to the use of our platforms.

Among other actions, we developed signals to identify content with rights-related concerns and high-priority escalations. In line with the [GNI principles](#), we published a real-time [case study](#) in our Transparency Center about takedown requests received in December 2023 from the Pakistan government that did not violate our Community Standards or local law.

Stakeholder engagement was critical. Prior to the election, we held briefings with regional human rights stakeholders on election integrity efforts. We also engaged with the Election Commission to explain our approach to government action requests and human rights. The internet was throttled during the run up to the election, and Meta shared [information](#) with stakeholders on how to set up a proxy server to connect to WhatsApp.

As a member of multi-stakeholder groups, including the [GNI](#) and [Asia Internet Coalition](#), Meta contributed to statements of concern about the impacts on rights of internet disruptions and rights-infringing legislation prior to the elections.

Several months ahead of the elections, we also took measures to provide human rights defenders and other vulnerable individuals with enhanced platform protections. This included providing protection against content removals based on false accusations of Community Standards violations. Our approach was inclusive, in accordance with our [Corporate Human Rights Policy](#), which adopts the broad definition of human rights defenders in the [UN Declaration on Human Rights Defenders](#).





Readiness for challenges in Active Response Countries

Active Response Country classification is determined by an evidence-based process, steered by specialized teams and our human rights experts, and is informed by our [human rights due diligence process](#) for reviewing and prioritizing countries with the highest risk of offline harm and violence. This triggers an in-depth risk-mitigation strategy. This may include intensified risk monitoring and mitigation efforts and additional investment, such as in relation to content review across conflict-relevant languages to remove policy-violating content and tailored product support. Due to the complexity of the risks, we also draw upon Trusted Partners and third-party fact-checkers to support our efforts in Active Response Countries.



Response to global crises

We respond to unexpected crises around the world, including interstate conflicts, terrorist attacks and non-violent crises such as environmental disasters. Our crisis response efforts involve many experts from across content policy, global operations, human rights and product, and other internal experts who assess and respond to the situation, employing our [Crisis Policy Protocol](#). As described in our [2022 Human Rights Report](#), the protocol guides our expedited use of targeted or exceptional policy levers to mitigate potential harm consistent with observed risks and informed by past crisis interventions and human rights principles.

Our Comprehensive Human Rights Salient Risk Assessment helps us prioritize between different potential human rights risks during conflict situations—for example, between the right to liberty and security of the person, the right to freedom of expression, and the safety and well-being of our communities. As part of our toolbox, we may adopt a number of temporary measures in order to help keep people safe and mitigate the risks that our platforms could be used to further exacerbate tensions both online and offline. We may consider relevant public international law developments, including international humanitarian law—the law of armed conflict—in line with UN Guiding Principles on Business and Human Rights guidance on heightened human rights due diligence

during times of conflict. We set out below illustrative examples of our work. We provide an [update](#) on our actions on Israel and Palestine human rights due diligence and related conflict in the next subsection.



Sudan

In April 2023, fighting erupted between the Sudanese Armed Forces (SAF) and the Rapid Support Forces (RSF), causing mass displacement, food insecurity, lack of access to healthcare and other challenges requiring humanitarian aid. Using the Meta Crisis Policy Protocol, Sudan was assessed to be at the highest level of risk.

Guided by our Crisis Policy Protocol, we took actions to respect human rights and minimize risks of harm, as well as to support the provision of humanitarian aid.

Relevant salient risks:

- life, liberty and security of person
- freedom of opinion and expression
- freedom of association and assembly
- privacy
- right to health



In August 2023, we designated the RSF under our [Dangerous Organizations and Individuals policy](#), which does not allow organizations or individuals that proclaim a violent mission or are engaged in violence to have a presence on our platforms. This helped curb the spread of potentially harmful content. The decision was taken after engagement with UN agencies and on the basis of human rights due diligence research and reports by major non-governmental organizations covering grave human rights violations. It was reviewed and endorsed by the Oversight Board. We also designated Sudan as a temporary high-risk location, which helped remove content that contained calls for civilians to take up arms or bring them to a specific location in the country. We worked with third-party fact-checkers to debunk, label and reduce the visibility of conflict-related misinformation on-platform. We also removed misinformation that Trusted Partners assessed could contribute to the risk of imminent physical harm or violence.

Due to the mass displacement from the capital, Khartoum, many of our stakeholders, including government officials, civil society, media and human rights defenders, were difficult or impossible to reach. This required identifying and engaging with new groups to support humanitarian efforts, as well as to help identify and stop the spread of potentially harmful content and misinformation. We engaged with the most active civil society groups, including doctors, media, citizen journalists and women's groups, both within Sudan and primarily within the diaspora. We briefed them on content policy and provided them with digital safety training and Meta crisis tools. This included running workshops

on how to use WhatsApp Communities, which were later used to coordinate evacuations and humanitarian support.

Despite the deteriorating security situation, our Trusted Partners continued to report content related to violence and incitement, misinformation, hate speech, and bullying and harassment. Trusted Partner assessments helped enforce our Misinformation and Harm policy on content using out-of-context images to stoke tensions, as well as on claims made against humanitarian agencies operating in Sudan.

Similar to our approach to [other conflicts](#), we removed some violating graphic content without applying [strikes](#)—the penalties for violations that result in escalating [account restrictions](#) as they accumulate—to avoid overly penalizing or restricting users who were trying to raise awareness of the conflict's impacts. This approach was informed by an Oversight Board recommendation in the [Sudan graphic video case](#) in 2022. This case prompted a review in 2023 of our policies on sharing violent and graphic content for raising awareness of human rights abuses and a more detailed explanation of how we apply our [newsworthiness allowance](#).

There was an acute shortage of medicines available in the country due to a lack of healthcare infrastructure, so we made temporary policy changes to enable users to request or donate medications, normally prohibited by our restricted goods policies. In addition, we worked with the Sudan Medical Specialization Board to launch the [SMSB Sudan Clinic](#), a telemedicine app using the



Building resilience: digital security and content policy training

2023:

6

training sessions conducted

300

beneficiaries impacted

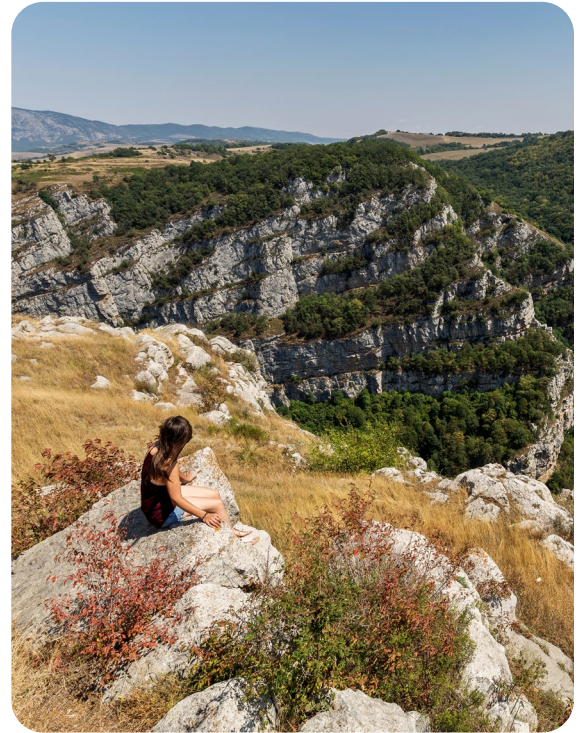
After the conflict:

2

training sessions conducted

70

beneficiaries impacted



WhatsApp API. This was offered through our WhatsApp social impact fee waiver program — a program for government and nonprofit partners whereby we waive our fees for the cost of messages.

We also needed to balance mobility and safety messaging — the communication of information that helps individuals travel while helping ensure their safety — without supporting the irregular migration to neighboring countries of the Central African Republic, Chad, Egypt, Ethiopia, Libya and South Sudan. This is because the risk of human exploitation increases as the needs of the displaced population rise, leading to the emergence of bad actors in these areas. We made efforts to identify bad actors and to remove content that could put people at risk, including that of human smuggling.

Nagorno-Karabakh

On September 19–20, 2023, Azerbaijan initiated a military offensive in the self-declared Republic of Artsakh, part of the disputed Nagorno-Karabakh region. The offensive prompted a flight of ethnic Armenians from the area, triggering a refugee crisis.

A cross-functional team, including human rights experts, had been monitoring the situation and swiftly mobilized resources and applied our [Crisis Policy Protocol](#). Internal experts with language skills and contextual knowledge handled the increased prevalence of hate speech, bullying, graphic content and references to dangerous organizations and individuals. We also engaged with UN agencies on the ground to better understand the dynamics

Relevant salient risks:

- freedom of opinion and expression
- life, liberty and security of person
- privacy



of the crisis. Meta provided additional protections to civic actors on both sides to reduce risks of impersonation, harassment and hacks.

As in other conflicts, we had to decide how to balance preserving the privacy and dignity of prisoners of war with the benefits of documenting human rights abuses. Meta Community Standards prohibit the posting of information that could reveal the identities or locations of prisoners of war. This is consistent with the company’s human rights responsibilities under the UN Guiding Principles on Business and Human Rights. These responsibilities are heightened during armed conflict and must be informed by the rules of international humanitarian law. Revealing identities and locations of prisoners of war can put at risk their safety, privacy and dignity, and the safety of their families.

At the same time, such exposure can inform public debate and raise awareness of potential mistreatment, including violations of international human rights and international humanitarian law. It can also build momentum for action that protects rights and ensures accountability. The scale and speed at which imagery of prisoners of war can be shared via social media complicates the task of resolving these competing interests.

In 2023, the [Oversight Board](#) upheld our decision to leave up a Facebook post that included a video depicting identifiable prisoners of war with the added “mark as disturbing” warning screen. The Board agreed with Meta, saying that in this situation, the posts could inform public debate and raise awareness of potential mistreatment, including violations of international human rights and international humanitarian law, which “outweighed the risk to the safety and dignity of the prisoners of war.”

Evidence retention

We support justice for all international crimes. We’ve worked for several years to explore rights-respecting initiatives for evidence retention and disclosure, consulting extensively with civil society, academia and international prosecutorial experts and bodies.

As [stated](#) in our response to recommendations from the Oversight Board, we worked to develop an approach to allow international courts and UN-authorized accountability mechanisms — such as fact-finding missions and Commissions of Inquiry — to make requests to us for extended retention of data that is relevant to their ongoing investigations. This work is largely completed and is distinct from our [longstanding policies for responding to preservation requests from law enforcement](#).

We briefed a range of UN-authorized mechanisms and special rapporteurs on our approach and outlined the process for making extended retention requests to Meta. We will carefully review all requests received for consistency with our policies and applicable law.

This is a novel area without established or tested best practices, and there remain significant legal, privacy and policy considerations inherent to this work. We expect to share further updates on our work in this area in our [biannual updates on the Oversight Board](#) and our annual Human Rights Report.



Relevant salient risks:

- life, liberty and security of person
- privacy

Migration in Central and South America

Human smuggling may result in the exploitation of vulnerable individuals who are trying to leave their country of origin, often in pursuit of a better life. We made efforts to remove content that facilitates or coordinates the exploitation of humans, including human trafficking and smuggling.

In 2023, we identified potential human rights risks regarding smuggling of migrants across North, Central and South America, including along the Darién Gap, a stretch of dangerous and densely forested jungle across northern Colombia and southern Panama. These are areas of increased risk of illegal commercial activity, online presence of dangerous organizations, and misinformation for migrants and refugees, especially as some attempt to make their way up to the U.S.-Mexico border.

In response to these risks, we applied our [Community Standards](#) and business policies on [Commerce](#) that prohibit human exploitation, including human smuggling. We collaborated with civil society and authorities to tackle misinformation, and we continued to monitor and enforce our policies on violating organic, paid and commercial content. We also made it easier for third-party fact-checkers to find and rate content related to the U.S.-Mexico border, because we recognize that speed is especially important during critical events. We used keyword detection to group related content in one place for our fact-checking partners, who rated content on our platform and published articles in English and Spanish addressing misinformation.

We also support migrant and refugee communities and non-governmental



organizations through integrity and educational programs such as We Think Digital, which focuses on digital skills to identify misinformation, frauds and scams.

Human Rights Impact Assessments

Methodologies for human rights due diligence are evolving. We consider potential risks and learn from past efforts to determine how we might improve our approach and strengthen our efforts that worked well.

We undertake Human Rights Impact Assessments for specific countries, apps, services, hardware or strategic initiatives as guided by our Comprehensive Human Rights Salient Risk Assessment. We perform these assessments to help us anticipate potential impacts, particularly in preparation for new products and features, such as [end-to-end encryption](#). Human Rights Impact Assessments involve [stakeholder engagement](#), a fundamental component of human rights due diligence. When we undertake an assessment, each recommendation often requires multiple teams and workstreams to implement the recommendations across the company.

We've shared summaries and recommendations from Human Rights Impact Assessments in the past, and they are listed in the [Annex](#). We also provide updates to actions taken in response to the assessments conducted on the

potential human rights risks related to Meta platforms in [the Philippines](#), [India](#), [end-to-end encryption](#), and [Israel and Palestine](#).

Israel and Palestine: Update on our actions

Shortly after we published our [September 2023 Israel and Palestine Human Rights Due Diligence Update](#), the October 7 terrorist attacks by Hamas against Israel took place, followed by Israel's military response in Gaza and further involvement and escalation by other regional actors. We know that people in the region and around the world have felt deeply impacted by our response to the ongoing conflict situation in the region.

In the immediate aftermath of the attack, Meta designated the violence at the highest level of our Crisis Policy Protocol and implemented immediate crisis response measures, including a dedicated 24/7 cross-functional team. Core human rights principles guided our actions. We looked to the [UN Guiding Principles on Business and Human Rights](#), as foundational to our [Corporate Human Rights Policy](#), to prioritize and mitigate the most salient human rights risks. We also used international humanitarian law as an important reference.

We initially shared details of our response in a blog post in [English](#), [Arabic](#) and [Hebrew](#) on October 13, 2023, and provided further updates on October 18, December 5 and December 8. We've continued to refine our

Relevant salient risks:

- life, liberty and security of person
- freedom of opinion and expression
- equality and non-discrimination
- privacy



approach to reflect changing dynamics, including the ongoing humanitarian crisis in Gaza and the hostages still being held in captivity by Hamas.

Balancing the safety and well-being of our communities and voice on our platforms is especially difficult in conflict situations — and even more challenging when it involves U.S.-sanctioned entities, such as Hamas, Hezbollah and Palestinian Islamic Jihad, which are also designated as dangerous organizations under our policies.

The work we already did in response to the [Israel and Palestine human rights due diligence](#) helped inform our approach. For example, we were able to route Arabic content more appropriately across our systems to enable greater precision. We also increased our content moderation resources in the Hebrew market. For a more detailed update, see [Meta Update 2024: Israel and Palestine Human Rights Due Diligence](#).

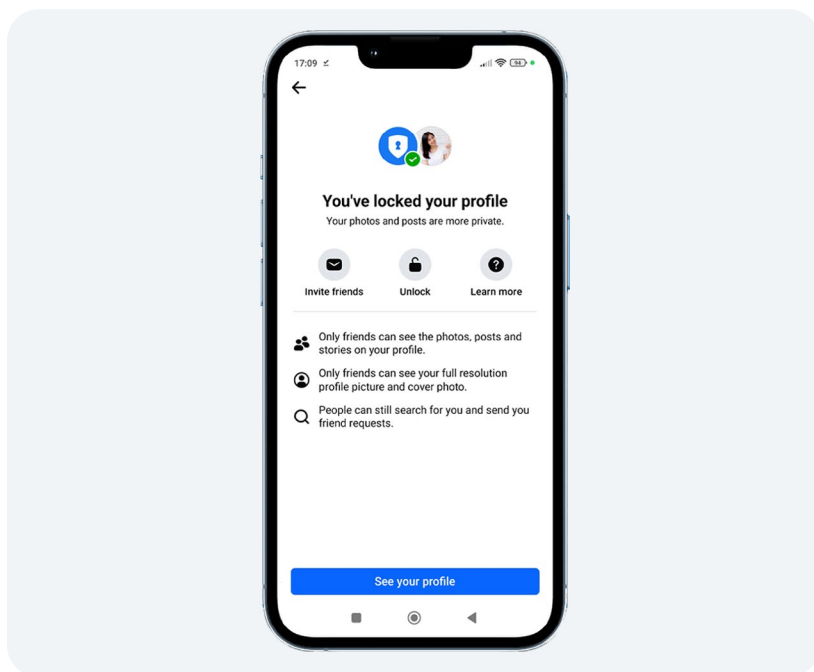
Our response to the conflict went beyond the measures outlined in the [due diligence](#) and were also guided by the use of our [Crisis Policy Protocol](#). We adopted a number of [time-bound temporary product and policy measures](#) to mitigate salient human rights risks. We know that in some cases, these measures — for example, lowering thresholds for automated enforcement — can inadvertently limit discussion of critical world events, impacting voice. That’s why we seek to take steps that are time-bound and proportionate to the risks as we are aware of them.

Temporary policy measures

We initially relied on our Dangerous Organizations and Individuals (DOI) policy and expanded our Violence and Incitement policy to remove content that depicted identifiable hostages being kidnapped or held in captivity, even when it was done to condemn or raise awareness of their situation. We did so to protect the privacy and dignity of the victims in line with international humanitarian law standards and prevent Hamas propaganda from appearing on our platforms, consistent with our DOI policy.

As the war continued, we saw that people began sharing hostage content to rebut emerging narratives denying that the October 7 terrorist attacks happened. We adopted a more nuanced approach accordingly and allowed people to share content that depicted hostages being kidnapped to raise awareness or condemn the attacks. We continued to remove content and err on the side of safety if the user’s intent in sharing the content was unclear.

We made other changes aimed at protecting voice. For some policy areas, like certain types of [violent and graphic content](#), we removed violating content without applying [strikes](#) — the penalties for violations that result in escalating [account restrictions](#) as they accumulate — to avoid overly penalizing or restricting users who were trying to raise awareness of the conflict’s impacts.



Temporary product measures

We adopted a number of temporary product measures to help keep people safe and mitigate salient human rights risks as outlined in our [blog post describing our response to the conflict](#). This included adjustments to confidence thresholds for automatically actioning content, blocking certain hashtags from search and product changes to address unwanted and problematic comments, as well as safety measures such as the [lock your profile](#) tool.

We also temporarily reduced the threshold at which borderline or potentially violating content — like images or videos depicting graphic violence — may be made ineligible for recommendation. This measure applied to unconnected content — that is, content from people that someone hasn't already chosen to follow that may appear on surfaces like Feed, Search, Explore and Reels.

As with our temporary policy measures, we made other changes aimed at protecting voice. In response to a large spike in the use of our products, we temporarily adjusted some automated rate limits designed to prevent behaviors defined as [spam under our policies](#) — such as high-frequency posting — to make them more permissive, reducing the risk of restrictions on legitimate users.

Separately, our broader policies on [recommendability](#) and [demotion](#), which are *not* a part of our crisis response actions, may impact the visibility of pieces of content related to the conflict, such as graphic and violent content.

Meta conducted frequent investigations to assess the performance of our content moderation systems and whether they were over-enforcing or under-enforcing on content. Moving forward, we identified areas for continued improvement, including prioritizing review of Arabic content by dialect and enhanced monitoring and oversight of our image and video banks to mitigate potential risks of over-enforcement and related impacts on freedom of expression. This work is ongoing.

For more detailed information on our response, refer to our letters in response to queries we received from civil society organizations [here](#) and [here](#). For more details about our 2024 Israel and Palestine Human Rights Due Diligence Update, see [here](#).



Relevant salient risks:

- privacy
- freedom of opinion and expression
- freedom of association and assembly
- equality and non-discrimination

End-to-end encryption:

Update on our actions

In December 2023, we began rolling out default end-to-end encryption for all personal chats and calls on Facebook and Messenger. This marked a major milestone in our journey, launched in 2019, to expand default end-to-end encryption to all our messaging apps. We also continue to work on end-to-end encryption for Instagram Direct. WhatsApp has provided default end-to-end encryption since 2016.

End-to-end encryption enhances security and privacy by ensuring no one sees your message except you and the person you are messaging. This is an extra layer of security that keeps your messages and calls with family and friends protected and means that nobody, including Meta, can see or listen to what’s being sent or said. When you report an end-to-end encrypted

conversation, recent messages from that conversation will be sent securely from your device to our Help Teams.

The human rights framework has been key to guiding our work to expand end-to-end encryption. End-to-end encryption directly enables the right to privacy, which in turn enables the realization of a diverse range of other human rights, including freedom of expression, association, opinion, movement and bodily security. At the same time, there is always a risk of individuals abusing any messaging service, including encrypted features, to harm others.

In 2022, we shared in full the results of an independent Human Rights Impact Assessment, conducted by BSR, that we commissioned to examine the potential human rights benefits and risks of our plans to expand default end-to-end encryption. The assessment made 45 recommendations to Meta to maximize potential benefits and mitigate potential adverse impacts. Since our initial response to these recommendations, we have continued work to implement our stated commitments and provided extensive public updates, including in January and August 2023, as well as a detailed safety white paper.



Issue spotlights

Relevant salient risks:

- privacy
- life, liberty and security of person

Combating security threats

For the past six years, Meta has released public reports of our efforts to [detect and counter security threats](#) on our platforms. Our post-disruption analysis and investigations were part of the basis for [the Citizen Lab](#) and 45 civil society groups to bring a case to the [Summit for Democracy](#) to ban the sale of spyware until safeguards are in place to prevent human rights abuses. Subsequently, U.S. President Joe Biden issued an [Executive Order](#) to prohibit U.S. federal entities from using spyware to enable human rights abuses.

Some examples of threats we mitigated in 2023 relate to the surveillance-for-hire industry. Surveillance-for-hire companies include businesses that offer services to monitor and track individuals or organizations for various purposes. There are concerns around privacy and misuse of surveillance-for-hire by authoritarian governments, criminal organizations and other bad actors who attempt to spy on regime critics, opposition figures, journalists and human rights defenders to suppress freedom of expression.

In 2023, we [identified and took down](#) six separate networks of accounts on Facebook and Instagram linked to eight firms within the surveillance-for-hire industry from Italy, Spain and the United Arab Emirates. Some of their tactics included social engineering and phishing across the internet and using personas posing as protestors, journalists and young women to trick people into sharing their emails and phone numbers and clicking on malicious links.

“The rise of the private spyware industry democratizes access to sophisticated surveillance capabilities while obscuring the clients who buy and deploy those services. This indiscriminate, opaque targeting poses a significant threat to human rights defenders. That’s why we are committed to working with civil society partners and across the industry to hold abusive spyware companies accountable.”

David Agranovich, Director, Threat Disruption, Meta



Relevant salient risks:

- best interests of the child
- privacy
- freedom of opinion and expression

Child and youth safety

Online safety for children and youth⁴ is an internet-wide challenge that requires effort and collaboration across the industry, government and civil society as we seek to help keep users safe while providing space to exercise a range of human rights—freedom of expression and access to information, among others.

We’ve spent a decade working on this issue and continue to use a rights-based approach for the design of our services and content policy. Our [Best Interests of the Child framework](#) is informed by the [UN Convention on the Rights of the Child](#), as well as regulation and guidance such as the UK Age Appropriate Design Code, the Irish Data Protection Commission’s Children’s Fundamentals and the French *Commission Nationale de l’Informatique et des Libertés* (CNIL) Recommendation on Minors.

Child exploitation

Child safety is always a priority for us. We seek to help prevent harm from happening by enforcing our policies and developing preventative tools. We’ve worked diligently to make reporting tools for child exploitation easier to find, which led to a 75% increase in reports sent to us by teens

⁴There is no universal definition of “youth.” The [UN Convention on the Rights of the Child](#) defines a “child” as every human being below the age of 18 (Article 1), which also encompasses teens above 13 years of age. In this report, we use “youth” and “teen” interchangeably.

in Q1 2023 over Q1 2022 on Messenger and Instagram. We also introduced additional innovative ways to proactively find, remove or restrict accounts that may violate our child safety-related policies. For example, we launched proactive technology to automatically detect and disable accounts (and any accounts linked to them) if they engage in on-platform activity and interactions that demonstrate sexual interest in children. More than 90,000 accounts were identified and automatically [removed](#) from August 1, 2023, to December 31, 2023, as a result of this method. We strive to proactively detect and restrict accounts belonging to adults at the earliest signs of unwanted interactions with or about teens, or attempting to engage in potentially unsafe contact.





As a founding member of [Take It Down](#), a tool designed to proactively prevent young people’s intimate images from spreading online, we worked with the [National Center for Missing and Exploited Children](#) to expand it to more languages and encouraged more industry players to join, resulting in access for millions more teens.

We are also a founding member of [Lantern](#), a new program from the Tech Coalition launched in November 2023 that enables tech companies to share signals about predatory accounts and behaviors across online platforms. Business for Social Responsibility conducted a [Human Rights Impact Assessment](#) of this program.

“Preventing child exploitation is one of the most important challenges facing our industry today. Online predators are determined criminals who use multiple apps and websites to target young people. They also test each platform’s defenses, and they learn to quickly adapt. That’s why now, as much as ever, we’re working hard to stay ahead. In addition to developing technology that roots out predators, we hire specialists dedicated to online child safety and we share information with our industry peers and law enforcement.”

Antigone Davis, Global Head of Safety, Meta



Youth safety and well-being

We want young people to have safe, age-appropriate experiences online while respecting their digital rights. Our approach is focused on preventing harm in the first place, giving people ways to control their experience, and responding to violations of our policies quickly. We have developed numerous [tools, features and resources](#) to support teens and their parents. This work is ongoing.

In 2023, we continued expanding our work around our [parental supervision tools](#) across our services. We conducted [co-design sessions](#) with parents, teens, guardians and experts through [Trust, Transparency and Control Labs](#) (TTC Labs) — a cross-industry effort to put people in control of their privacy — and worked with youth advisors to understand teen behaviors to inform product design. We also launched [parent-managed accounts](#) for preteens aged 10–12 on Meta Quest 2 and 3.

We regularly consult with experts in adolescent development, psychology and mental health to help keep our platforms safe and age-appropriate for young people, including continually developing our understanding of which types of content may be better-suited for teens. Consultations took place with [safety partners](#), youth groups and advisors from around the world, as well as UNICEF to brief and engage them with our work.

Taking into account expert viewpoints, we developed [new protections](#) to give more age-appropriate experiences on our apps, for example, including hiding more search results, prompting teens to update their privacy settings on Instagram, and automatically placing teens into the most restrictive content control setting on Facebook and Instagram.

Stakeholder engagement

We conducted [co-design sessions](#) through [TTC Labs](#) to engage users, our community and experts to help create positive online experiences for young people by listening to their needs and concerns to inform our product design.

87 teens

71 guardians

221 experts

13 countries

Stakeholder engagement



Proactive engagement with a wide spectrum of stakeholders is central to our approach to human rights at Meta. It brings external knowledge and diverse views and feedback into our product and content policy development processes and helps provide accountability and transparency. The groups we engage with range from international organizations, varied civil society groups and academics to marginalized and underrepresented communities, as well as users such as teens, parents and caregivers. We gather perspectives from all sides of an issue and from all regions of the world to ensure balance.

In addition, we continued to engage our Trusted Partners to identify trends, better understand challenges to our escalation channels and explore how we can make such mechanisms more effective for civil society partners.

Proactive engagement with a wide spectrum of stakeholders is central to our approach to human rights at Meta.

Marginalized groups and human rights defenders

We recognize the importance of meaningful engagement with stakeholders from marginalized communities, taking account of variations within region and even country, and we seek to inform ourselves about these differences.

We strive to listen to and seek advice from human rights experts and activists, academics and others on a wide range of issues and brief them on relevant developments at Meta. Their insights inform areas such as development



of our content policies and enforcement to advance voice and safety.

Around the world, human rights defenders who use digital platforms as instruments for their activism frequently encounter intimidation, oppression and legal challenges as a consequence of their online endeavors aimed at advancing and protecting human rights. This is particularly noticeable among marginalized groups, such as women and the LGBTQIA+ community.

Guided by our [inclusivity framework](#), we use roundtables, workshops and one-on-one meetings to understand relevant content policy issues, identify policy gaps and pursue opportunities for engagement on policy development. For example, we worked with the [Minority Rights Group](#) to conduct eight global convenings in countries deemed at risk of armed conflict or societal violence.

We also engaged with over 250 stakeholders from diverse backgrounds, including women, members of the LGBTQIA+ community, religious minorities, ethnic minorities and indigenous groups, to gather valuable insights on the policy challenges faced by these communities. Based on feedback from over 19 consultations and workshops with LGBTQIA+ organizations and human rights defenders in the Middle East and North Africa, we collaborated with the Jordan Open Source Association (JOSA) to release the [Digital Security Toolkit for Human Rights Defenders](#). This guide, available in both [Arabic](#) and [English](#), identifies leading practices in digital security, offering security features, tips and reactive measures to enhance online safety for activists and human rights defenders in the Middle East and North Africa. We also launched an LGBTQIA+ influencer campaign in Sub-Saharan Africa to raise awareness about safety resources for LGBTQIA+ activists and highlight our commitment to protecting high-risk users in the region.

Following a successful two-year Asia Pacific-focused pilot, in 2023, we expanded our [Human Rights Defender Fund](#), operated in partnership with [Civil Rights Defenders](#), to support human rights defenders globally and renewed the funding for an additional USD \$500,000 to be applied in 2024.

AfricanDefenders

Building on the learnings and principles of our [Asia Pacific Human Rights Defender Fund](#) and in partnership with [AfricanDefenders](#), we launched the [African Human Rights Defender Fund](#), designed to support human rights defenders who are targets of harassment, persecution and/or prosecution because of their online activities in support of human rights. That includes members of marginalized groups advocating for their rights, such as activists and citizen journalists, non-violent political activists, women human rights defenders and the LGBTQIA+ community. The micro-grants will be used for support in emergency situations and to pay for new devices and security technologies, temporary relocation, and emergency legal and safety assistance.

Trusted Partners

In January 2023, our network of [Trusted Partners](#) included over 400 non-governmental organizations, humanitarian agencies, human rights defenders and researchers from 113 countries around the globe. In addition to our ongoing, wide-ranging and diversified stakeholder engagement, these partners bring a wealth of knowledge and experience to help inform our content moderation and the development of effective and transparent policies. They also have access to an escalation channel to report potentially harmful content and account security issues, which we act upon as appropriate to help keep users safe.

In 2023, reports from our Trusted Partners helped us remove 49,600 pieces of content that violated our policies. Trusted Partners also helped us uncover and take down six deceptive networks engaged in [coordinated inauthentic behavior](#) in the U.S., Georgia, Burkina Faso and Togo. In addition, we took action against three separate [cyber espionage operations in South Asia](#). These networks were designed to manipulate public debate for a strategic goal, which can pose a threat to human rights.

In 2023, we undertook a comprehensive review of our channels for receiving emergency content and account security-related escalations from civil society partners. A key part of this work was the organization of a global summit that was attended by policy, operations and legal experts from across Meta as well as by 15 civil society partners from 13 countries across Asia Pacific, the Middle East, Europe, Latin America and Africa. Through this engagement, we were able to holistically assess our civil society escalation channels and formulate a shared vision for their future—improving the prioritization of reports and strengthening processes for emergency response and insight-gathering.





Case study: Surfacing novel harms in Haiti

July 7, 2023, marked the two-year anniversary of the assassination of Haitian President Jovenel Moïse. Tensions remain high in the country, exacerbated by the resurgence of gang violence. Our Trusted Partners reported content related to multiple violation areas including Violence and Incitement, Dangerous Organizations and Individuals, Bullying and Harassment, and Graphic Violence. They also alerted us to the presence of a new vigilante movement, Bwa Kale, on our platforms. These insights enabled our operations and investigative teams to leverage keywords and proactively identify potentially violating content, effectively mitigating offline harm.

Case study: Ramping up reporting in Ethiopia

Given heightened sociopolitical tensions and ongoing violence in Ethiopia, we met with our Trusted Partners monthly to discuss content

trends and reinforce practices for reporting high-severity content through the Trusted Partner Channel. We also ran a bi-monthly Ethiopia Dialogue Series with civil society from Ethiopia and the diaspora. Our partners in Ethiopia continued to report content and provided harm assessments for misinformation and veiled threat reports. By working with these partners, we were able to better identify key issues, including out-of-context imagery, fake accounts spreading mis/disinformation, verified accounts leveraging influence to spread misinformation, hate speech, coordinated harassment and doxxing, politically charged terms initially coined by public figures, and dangerous organizations and individuals.

Case study: Mitigating intra-religious conflict in Bangladesh

In March 2023, intra-religious riots broke out between Sunni Muslims and Ahmadi Muslims in Bangladesh, a Muslim minority community. Our Trusted Partners in Bangladesh reported content

including harmful misinformation, hate speech, and violence and incitement targeted at the Ahmadi community. By leveraging previously designated, pre-reviewed harmful claims, we were able to swiftly enforce on misinformation reports. Trusted Partners also provided critical local signals that enabled us to remove content under escalation-only policies, including veiled threats, as well as insights into local hate speech terms.

International and multi-stakeholder organizations

As technology advances and its impact on human rights increases, we continue to build our engagement with a wide range of international government bodies, including:

- [The Office of the High Commissioner for Human Rights](#)
- [UNESCO](#)
- [The Office on Genocide Prevention and the Responsibility to Protect](#)
- [The Office of the Secretary-General's Envoy on Technology](#)
- [UNICEF](#)
- [UNHCR](#)
- [The Internet Governance Forum](#)
- [The United Nations Foundation](#)
- [OECD](#)
- [The World Economic Forum](#)
- Various regional organizations



We are a member of various multi-stakeholder initiatives, such as [Global Internet Forum to Counter Terrorism](#), the [Global Network Initiative](#) and the [Freedom Online Coalition Advisory Network](#). We also proactively engage in industry-wide collaborations to provide input and support government policy.

We participated in the second [Summit for Democracy](#) and publicly shared commitments in response to the summit's call to the private sector to advance democracy. We were recognized for our co-leadership on developing industry principles to "Counter Rising Threats Posed by Cyber Mercenaries" and provided input into the call for the private sector to support [Democracy-Affirming Anti-Censorship Technologies to Combat Authoritarian Regimes](#). Additionally, we were recognized for our work to enable secure and private connections, and countering government abuse of digital technology featuring WhatsApp by proxy and [surveillance-for-hire](#).



We participated in consultations on the development of the [Global Digital Compact](#) with the [Office of the Secretary-General's Envoy on Technology](#) to support the need for a globally harmonized approach to internet governance and protecting human rights online, including the rights to privacy and freedom of expression. We also participated in consultations on the [UN Code of Conduct for Information Integrity](#) and provided a detailed submission.

We continued to meet regularly with officials of the Office of the UN High Commissioner for Human Rights, independent experts and special rapporteurs to discuss global and country-specific issues and participate actively in the [B-Tech Project](#). We made a pledge of commitment to human rights on the occasion of the 75th anniversary of the Universal Declaration of Human Rights. This included renewed commitments to human rights due diligence and disclosure and engagement with civil society and affected communities, among others.

Meta engaged in UNESCO's [Internet for Trust collaborative effort](#), participating in the February Summit in Paris and in the multi-stakeholder process to develop human rights-centric guidance on digital platform regulations.

In collaboration with the [World Economic Forum](#), we participated in an extensive project to build consensus for emerging tech policy. We helped develop and endorsed the [Global Principles on Digital Safety: Translating International Human Rights for the Digital Context](#), which establishes a framework designed to help governments and online service providers to advance digital safety through a multi-stakeholder approach. We are also part of the [AI Governance Alliance](#) and the [Defining and Building the Metaverse](#)

“We can’t do this on our own. We need to have these strategic engagements and strategic partnerships for this project to have any impact.”

Lene Wendland, Chief, Business and Human Rights Unit, UN Human Rights

initiative, which seeks to guide the development of an ethical and responsible metaverse. Last year, we joined the [AI Governance Alliance](#), which aims to establish global standards for AI governance, positively contributing to technological advancements to societal well-being, guided by commitments to ethics and diversity.

We began work with the [UN Commission on Narcotic Drugs](#) to set up The Prevent Alliance, a public-private initiative between Meta, Snap, the U.S. government and the UN Office on Drugs and Crime aimed at preventing the abuse of digital platforms for illicit and harmful activities related to the non-medical use of synthetic drugs.

We made a [pledge](#) to the [Global Refugee Forum](#) in support of the rights of refugees, asylum seekers and internally displaced persons.



Data for Good at Meta

Our Data for Good program continues to work with humanitarian agencies, nonprofits, researchers and governments around the world to support marginalized populations. For example, this program continues to support the work of nonprofits such as IMPACT Initiatives as they help resettle those displaced by the war in Ukraine. It also supports international agencies, including the International Organization for Migration on improved [global metrics](#) to support migrant populations globally and those displaced

by natural disasters. Additionally, Data for Good at Meta contributes significantly to the work of development institutions, including the World Bank, through large-scale collaborations like the Development Data Partnership, which enables large multilateral institutions to better target programs that serve communities in need. For example, AI-driven tools built by Data for Good, such as the Relative Wealth Index, have helped the World Bank better understand how air pollution affects the poorest populations in low- and middle-income countries.





Transparency and remedy



We are committed to transparency and remedy as central tenets to human rights risk management. Both are core elements of the [United Nations Guiding Principles on Business and Human Rights](#). Our [Transparency Reports](#), published regularly, and our annual Human Rights Report give visibility into how we enforce our [Community Standards](#) and [Community Guidelines](#), respond to government requests and protect intellectual property. We also work hard to continuously improve accessibility to our Community Standards, which were available in 90 languages at the end of 2023. We believe the Oversight Board continues to provide access to remedy that is unique in the social media industry.



Responding to government requests

Relevant salient risks:

- freedom of opinion and expression
- privacy
- right to a fair trial
- right to an effective remedy

We do not take action on content or disclose user data simply because a government entity requests it. As members of the [Global Network Initiative](#) (GNI), we are committed to the GNI Principles on Freedom of Expression and Privacy (GNI Principles), which are based on internationally recognized human rights standards. This is important because the principles give us guidance on how to respond to government requests and minimize negative impacts on the freedom of expression and privacy of our users.

Over the years, governments’ requests have evolved. Government actors ask us to take a wide range of actions beyond restricting

10 years with the Global Network Initiative

“The GNI principles are our north star for managing government takedown requests. Requests to restrict content based on local law often throw up very real and far-reaching human rights dilemmas. When faced with conflicts between local laws and our human rights commitments, we look to the GNI, internationally recognized human rights standards and our Corporate Human Rights Policy to guide us.”

Siobhán Cummiskey, Content Policy Director, Meta

The [GNI](#) is a multi-stakeholder initiative whose members commit to upholding rights to freedom of expression and privacy as defined in a key human rights treaty, the International Covenant on Civil and Political Rights (ICCPR). The GNI sets a global standard for responsible company decision-making to promote and advance freedom of expression and privacy rights across the technology ecosystem. In the context of the UN Guiding Principles on Business and Human Rights, the GNI Principles serve as a link between human rights obligations of governments to protect and the corporate responsibility to respect those rights.

2023 marked the 10-year anniversary of Meta joining the GNI and our commitment to human rights. We apply, and are accountable to, the GNI principles for how technology companies should respect the freedom of expression and privacy rights of their users when addressing government requests and restrictions. We’re regularly independently assessed on our implementation.



access to content in one country. Since 2020, the number of government requests has increased, and so has the volume of content restricted by Meta based on local law.

Government requests include both restrictions to content and access to user data. These requests engage the rights to freedom of expression and privacy, respectively. When we receive a report to restrict content or an access to user data request from a government entity or a court, we review it following a consistent global process for both situations ([here](#) and [here](#)).

- **Government officials sometimes make requests for data about people who use our platforms as part of official investigations.** We continue to apply [GNI Principles](#) to our response to all forms of requests from governments. The vast majority of these requests relate to criminal cases. In many of these cases, these government requests seek basic subscriber information, such as name, registration date and length of service. Other requests may also seek IP address logs or account content. We have strict [guidelines](#) in place to deal with all government data requests. We report the number of requests received, the number of users/accounts requested, and the percentage of requests where we produced some data. We have publicly [reported](#) this data along with our policies for responding to requests since 2016.
- **Government authorities sometimes ask for content to be removed for allegedly violating local law.** We also may get an order from a court requiring content removal. If we determine that the content does not go against our policies, we may restrict access to the content in the country where it's said to be unlawful after a careful

legal and human rights assessment conducted in line with our [Corporate Human Rights Policy](#) and our commitments as a member of the GNI. We also inform the affected user, if not legally prohibited from doing so. We consider whether any mitigation measures should be deployed, such as taking steps to limit the impact of the request, including complying under protest (which in some cases has led to a revocation of such requests), or a legal appeal by Meta of the takedown order, real-time transparency via a case study in our [Transparency Center](#) and/or submitting a copy of the request to the [Lumen database](#). In cases where we believe that requests from the government or orders from courts are not legally valid, are overly broad or are inconsistent with international human rights standards, we may request clarification, appeal or take no action.

- **We occasionally receive legal requests or orders from governments and courts to take novel actions based on local law requirements,** including applying a correction notice on content (e.g., stating that the content is considered by local authorities to be misinformation), limiting a user's access to specific features, restoring content previously removed for violating our policies, obligation to automatically restrict content at scale and in country, and restricting content globally although it violates only one country's local laws. While we respect the law in countries where we operate, when faced with conflicts between domestic legal obligations and our human rights and transparency commitments, we will seek to honor the principles of internationally recognized human rights to the greatest extent possible. We may request clarification, challenge the request or take no action.



- In some instances, we may be legally prohibited from publishing takedown requests or certain information contained within them. In such cases, we will aim to publish as much information as possible about the order and its existence without violating our legal obligations.

How to read content restriction reports

We receive reports from governments, regulators and courts, as well from non-government entities and members of the public, for content alleged to violate local laws. The [Content Restrictions Report](#), published bi-annually since 2013, details instances where we restricted access to content on Facebook and Instagram based on local law.

- The section on [country-specific data](#) provides details on the volume, nature and type of content we restrict every six months based on local laws. Since 2023, our reports now also include information where in limited countries we are obligated based on local law requirements to automatically restrict content, at scale and in country, or apply a correction notice on content in country.
- The section on [global restrictions](#) details the occasional instances when we were legally compelled to restrict the availability of content globally due to

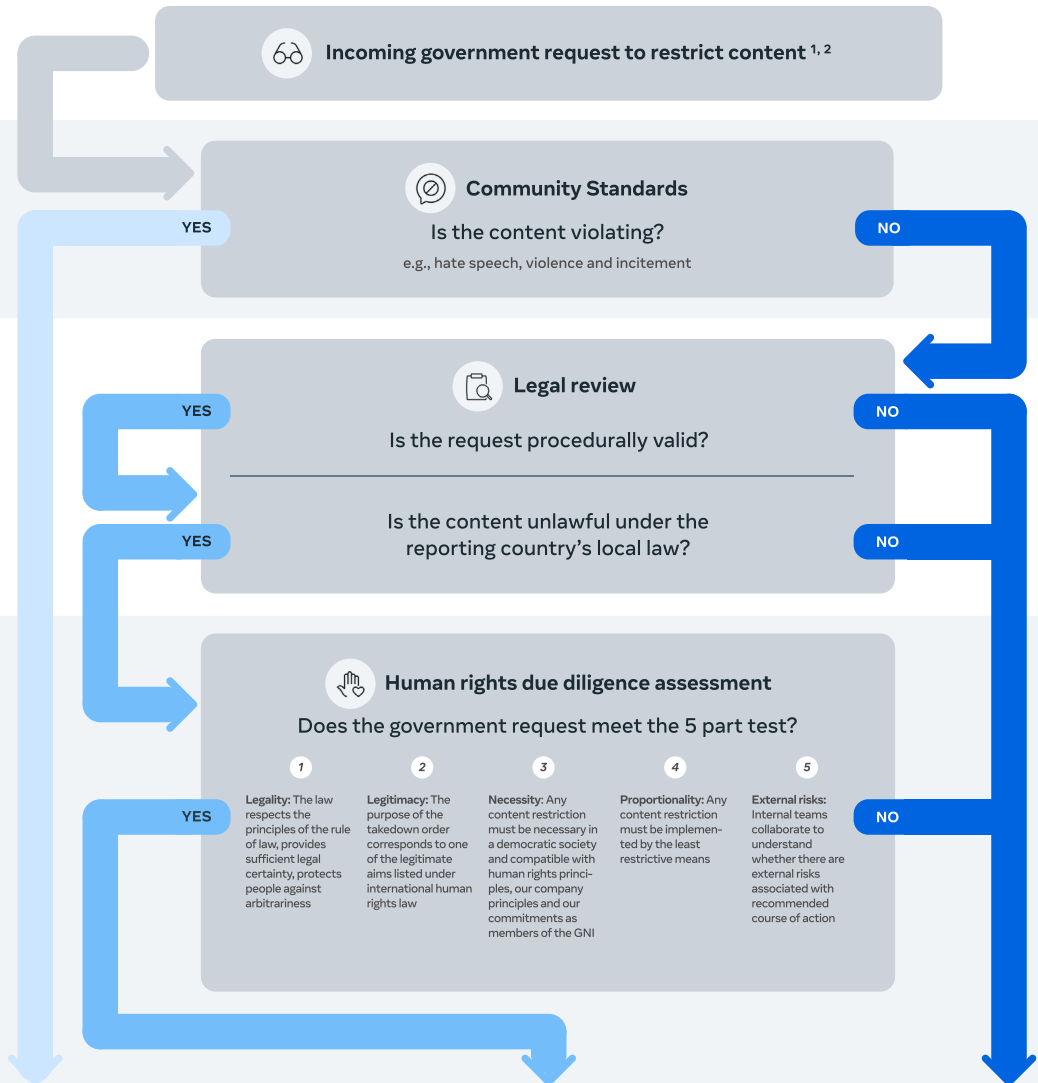
extra-territorial legal demands, although the content at issue went against only the local law of a specific country.

- The section on [case studies](#) provides details about some of the content takedown requests we received from governments and courts. Since 2023, we are focused on publishing only case studies of high public interest.
- The section on [how we assess reports of content violating local law](#) contains information about our review process for government takedown requests and mitigation measures we undertake to limit the impact of the restrictions in line with international human rights standards.

We published data on content restrictions based on local law for H2 2023 in May 2024 (see [here](#)).



What's the life cycle of a government request?



Outcome

We may deploy mitigation measures such as appeals or the sharing information about the government request to restrict content externally to limit the impact of the request.

Action under our policies

When we identify content that goes against our Facebook Community Standards or Instagram Community Guidelines, the appropriate action will be taken (e.g., the content will be removed or age-gated, etc.).

Restrict access in jurisdiction where it is unlawful

We tell people when we restrict something they posted based on a report that the content goes against local law, and we also tell people when they try to view something that is restricted in their country on the basis of a government request to restrict content. We also notify the relevant government authority that action has been taken in response to their request.

No action OR seek further information

We may engage with the authority to seek further information about the request.

¹ This illustration reflects our process for reviewing and responding to government requests to restrict content on Facebook and Instagram, as described in our Transparency Center. For a fulsome description, please refer to the Transparency Center (link here). Some differences may apply for ads and other Meta products.

² We may be compelled to deviate from the lifecycle outlined above during emergency situations. While uncommon, occasionally a country's law may obligate us to automatically restrict access to content, at scale in specific countries, based on local law requirements. In such cases, we will continue to look to our Global Network Initiative commitments and Corporate Human Rights Policy to guide our approach.



Update on our participation in Lumen

In March 2022, we committed to [participation in Lumen](#), an independent research project hosted by Harvard University’s Berkman Klein Center for Internet and Society. The project enables researchers to study content takedown requests from governments and private actors concerning online content. In November 2023, we submitted our first set of takedown requests from Austria, Vietnam, Singapore, India and Mexico to [Lumen](#), which are accessible on their database. This will further enable the global community’s efforts to analyze, report and advocate for digital rights of internet users. The principles that guide our approach for sharing information publicly about government takedown requests can be found in our [Content Restrictions Report](#).

Improved notice for users

To further enhance our transparency and in line with the right to an effective remedy and right to a fair trial standards, we improved our notice for users in 2023 on Facebook and Instagram for content restricted by Meta based on legal requests from governments and courts. In a majority of cases, our notifications inform users which state authority sent the request resulting in the restriction and the country in which the content is restricted. It does not apply to content that we removed for violating our policies.





Relevant salient risks:

- right to an effective remedy
- right to a fair trial
- freedom of opinion and expression
- life, liberty and security of person
- public participation, to vote, and to be elected
- equality and non-discrimination
- best interests of the child
- privacy

Oversight Board

The Oversight Board is an industry-leading effort created to help Meta answer some of the most difficult questions around freedom of expression and safety online, what to take down, what to leave up, and why. It is an independent body that can review cases from either Meta directly or people on Facebook, Instagram or Threads who disagree with our content moderation decisions. In addition to issuing binding decisions on whether to leave up or take down content, the Board also issues recommendations to help improve our content moderation, as well as Policy Advisory Opinions delivered at our request. The Oversight Board applies global human rights standards to provide Meta with critical expert insights and recommendations to help ensure decisions about our policies and products are in the best interests of users.

2023 was a landmark year for the Oversight Board, as they surpassed their goal of 50 decisions annually, more than tripling the number of decisions made in 2022.

The Oversight Board made significant changes to their Bylaws to provide for [expedited reviews](#) in exceptional circumstances, including when content could result in urgent, real-world consequences. This process enables accelerated content decisions to be issued within 30 days. The Board used the expedited review process for the first time for two cases related to the [Israel-Hamas conflict](#).

The Board decisions covered many different issues across a wide range of countries, including the Iran protest slogan “Death to Khamenei,” the Cambodia Prime Minister Hun Sen case, gender and nudity, Armenian prisoners of war, the Brazil elections and women’s protests in Cuba among others. In addition to its decisions, the Board made 60 recommendations to Meta in 2023. It also published a Policy Advisory Opinion on the Removal of COVID-19 Misinformation. For more details, see the Oversight Board’s transparency reporting, [here](#), [here](#) and [here](#).



Newsworthiness allowance

In response to Oversight Board recommendations, we provided additional transparency around how and when we apply our [newsworthiness allowance](#) for content that may violate our [Facebook Community Standards](#) or [Instagram Community Guidelines](#) but is in the public interest to keep visible.

- From June 1, 2022, through June 1, 2023, we documented 69 newsworthiness allowances.
- Nine (~13%) of those documented allowances were issued for posts by politicians.
- Of the 69 allowances, we documented a total of 17 scaled allowances, meaning that each allowance applies to more than one piece of content.



Our cross-check program

2023 also marked a significant turning point in our cross-check program, as it kicked off with our initial response to the Oversight Board's 33 recommendations targeting the program's technical sophistication, operational rigor, fairness, governance and transparency. We took steps to increase the involvement of Trusted Partners and civil society in the development of cross-check governance through feedback sessions with partners from nine different regions.



In 2023, we completed work on a total of 61 recommendations out of 122 recommendations and continue our efforts on many others (for more details, see our [Oversight Board biannual updates](#)). The recommendations we undertook spanned our operations, policies and products, contributing to broad and meaningful improvements across the company and our global community.

2023 Oversight Board recommendations

Oversight Board
recommendations

66

(91 in 2022)

Meta assessment
and/or implementation
in progress*

69

(75 in 2022)

Recommendations
fully implemented*

61

(14 in 2022)

* Some assessments and/or implementations in progress or recommendations fully implemented include recommendations from prior years (see our [2022 Human Rights Report](#) for further details).



Looking forward

In 2023, we witnessed many fast-paced developments, as described in this report, and many opportunities to include human rights in decision-making. We sought to do so. Yet broader sociopolitical contexts are unstable, and many data sources indicate an erosion of freedom of expression and democratic norms. Artificial intelligence (AI) technologies have existed for decades, but public awareness of the opportunities and risks they can create exploded in 2023.

The fast-evolving regulatory landscape holds great promise as well as challenges for human rights: it may drive significant innovation in how, and how broadly, companies apply human rights risk management aligned with key human rights principles, including the UN Guiding Principles on Business and Human Rights. But the demands of compliance could also crowd out the desire and room for innovation. It's becoming increasingly clear that sustained human rights risk management at scale involves ongoing use of UN Guiding Principles on Business and Human Rights-aligned data and tools that are

“We are surrounded by opportunities to drive human rights considerations into everyday decisions by business and governments.”

John Ruggie, author of the United Nations Guiding Principles on Business and Human Rights

embedded within broader company processes.

Meta seeks to innovate in our human rights risk management methodologies, including in our engagement with stakeholders. Experiential learning, simulation exercises and case studies appear to hold the promise of great practice, but have yet to become mainstream — we strive to strengthen our work in these areas. We also recognize the importance of concerted efforts and innovation to address the potential human rights risks linked to generative AI, as we've discussed in this report.

We know that we still have a long way to go. We want to build on our work and lessons learned to contribute to the evolution of knowledge and good practice for the tech



sector and the business and human rights communities.

We are poised at a challenging but promising moment for human rights in the tech world. All of us, including companies, civil society, governments, the UN system, investors and regulators will need to work to develop new skills and knowledge — and new ways of collaborating — to shape, share and learn what human rights risk management can look like at scale.

Future waves of immense creativity are gathering: we'll need multiple brave people, experiments and entrepreneurial interventions to seize them. Fortunately, the human rights movement has never lacked the courage to innovate. Let's get out there.



Annex

How human rights are governed and managed at Meta

Clear management and governance structures enable us to advance our work toward respecting human rights across our programs, services and policies. Our human rights experts guide the implementation of our [Corporate Human Rights Policy](#), which is overseen by the President of Global Affairs and Chief Legal Officer.

The human rights experts' 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 gives guidance 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 and 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 experts.

Training Meta employees on human rights

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.

Our *Bigger than Meta: Human Rights* training launched in 2022, and continued throughout 2023. The training highlights the potential and actual real-life impacts of our 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 our [civil rights](#) training, which centers on the principles of non-discrimination, justice and fairness.

Our human rights training objectives are also supported by our mandatory annual privacy training. This focuses on developing our collective ability to protect individuals—including, in particular, marginalized categories of individuals—against harms emanating out of the processing of people's data. This training supports the protection of people's rights to privacy and data protection.



Links to referenced reports

- [2024 Sustainability Report](#)
- [2024 Responsible Business Practices Report](#)
- [2022 Human Rights Report, 2021 Human Rights Report](#)
- [2023 Anti-Slavery and Human Trafficking Report](#)
- [2024 Anti-Slavery and Human Trafficking Report](#)
- [2023 Conflict Minerals Report](#)
- [2023 Update on Civil Rights Progress](#)
- [Meta Transparency Reports](#)
- [Regulatory and Other Transparency Reports](#)
- Human Rights Impact Assessments: [End-to-end Encryption](#), [Philippines](#), [Myanmar](#), [Indonesia](#), [Cambodia](#), [India](#), [Sri Lanka](#), and [Israel and Palestine](#)

