



∞ Meta

Relatório sobre Direitos Humanos

Insights e ações
2023



Índice

Sobre este relatório	3
Resumo executivo	5
Apresentação	8
IA no contexto dos direitos humanos	9
Adotando uma abordagem aberta	9
Lidando com os possíveis resultados prejudiciais da IA generativa	11
Gerenciamento de riscos relacionados a direitos humanos	13
Auditoria diária	13
Política de conteúdo	16
Desenvolvimento de produtos	17
Como nos preparamos e respondemos a eventos críticos e crises globais	19
Como a Meta se prepara para eleições	19
Brasil	20
Paquistão	22
Resposta a crises globais	24
Sudão	24
Nagorno-Karabakh	26
Migração nas Américas Central e do Sul	28
Avaliações de impacto sobre os direitos humanos	29
Israel e Palestina: atualização sobre nossas medidas	29
Medidas temporárias sobre políticas	30
Medidas temporárias sobre produtos	31
Criptografia de ponta a ponta: atualização sobre nossas medidas	32
Destaques de problemas	33
Combate a ameaças de segurança	33
Segurança de crianças e jovens	34
Exploração infantil	34
Segurança de jovens e bem-estar	36
Envolvimento das partes interessadas	37
Grupos marginalizados e defensores dos direitos humanos	37
Parceiros confiáveis	39
Estudo de caso: apurando novos tipos de danos no Haiti	40
Estudo de caso: aumentando a divulgação de informações na Etiópia	40
Estudo de caso: mitigando conflitos intra-religiosos em Bangladesh	40
Organizações internacionais e com várias partes interessadas	41
Transparência e reparação	44
Resposta a solicitações governamentais	45
Comitê de Supervisão	50
Olhar para o futuro	53
Anexo	55
Como os direitos humanos são regidos e gerenciados na Meta	55
Treinamento dos funcionários da Meta sobre direitos humanos	55
Links dos relatórios indicados	56



Sobre este relatório

Este é o nosso terceiro relatório anual sobre direitos humanos. Ele se baseia no trabalho da Meta e reflete o progresso alcançado em nossos compromissos com os [Princípios Orientadores sobre Empresas e Direitos Humanos das Nações Unidas](#) e nossa [Política Corporativa sobre Direitos Humanos](#). Ele abrange nossas ações de 1º de janeiro a 31 de dezembro de 2023. Nós divulgamos informações sobre os serviços e produtos da Meta, incluindo o Facebook, Messenger, Instagram, WhatsApp, Threads e Reality Labs.

Os direitos humanos são um tópico de grande relevância para as partes interessadas em nossas atividades e para nossa empresa. O conteúdo deste relatório reflete o feedback recebido das partes interessadas após o lançamento do nosso [Relatório sobre Direitos Humanos de 2022](#) e está fundamentado em nossa [Avaliação Abrangente de Riscos Relevantes a Direitos Humanos](#), que está resumida no relatório de 2022. O objetivo da avaliação foi identificar e priorizar nossos potenciais impactos adversos mais significativos sobre os direitos humanos¹ das pessoas, incluindo usuários e terceiros que possam ser afetados por nossas ações em toda a empresa globalmente. Este relatório destaca como nós, da Meta, gerenciamos, em nossas operações globais, os seguintes oito direitos humanos relevantes:

As oito principais questões de direitos humanos relevantes identificadas em nossa Avaliação Abrangente de Riscos Relevantes a Direitos Humanos.

Direito humano relevante

Seção do relatório

Liberdade de opinião e expressão

Política de conteúdo, Como a Meta se prepara para eleições, Resposta a crises globais, Israel e Palestina: Atualização sobre nossas medidas, Segurança de crianças e jovens, Resposta a solicitações governamentais, Comitê de Supervisão

Privacidade

Política de conteúdo, Criptografia de ponta a ponta: Atualização sobre nossas medidas, Combate a ameaças de segurança, Segurança de crianças e jovens, Resposta a solicitações governamentais

Igualdade e não discriminação

IA no contexto dos direitos humanos, Desenvolvimento de produtos, Israel e Palestina: atualização sobre nossas medidas

Vida, liberdade e segurança pessoal

Como a Meta se prepara para eleições, Resposta a crises globais

Melhor interesse da criança

Política de conteúdo, Segurança de crianças e jovens, Envolvimento das partes interessadas

Participação pública, direito de votar e de ser eleito

Como a Meta se prepara para eleições: Brasil, Paquistão

Liberdade de associação e reunião

Criptografia de ponta a ponta: atualização sobre nossas medidas, Resposta a crises globais: Sudão

Direito à saúde

Resposta a crises globais: Sudão

¹ O termo "impacto adverso sobre os direitos humanos" está em conformidade com os Princípios Orientadores sobre Empresas e Direitos Humanos das Nações Unidas e significa o impacto registrado quando uma ação remove ou reduz a capacidade de um indivíduo de usufruir de seus direitos humanos.



Nossos principais riscos aos direitos humanos estão destacados no início de cada seção deste relatório.

Mobilizamos muitas partes interessadas internas e externas e concentramos nossos esforços em fornecer uma imagem representativa do nosso trabalho em várias equipes no mundo todo. O objetivo deste relatório é fornecer uma visão geral das questões envolvendo direitos humanos com as quais lidamos como empresa e, ao mesmo tempo, orientar o leitor a explorar um conteúdo mais detalhado, conforme necessário.

O presente relatório complementa o mais recente [Relatório sobre Práticas Comerciais Responsáveis da Meta](#). [Relatamos](#) separadamente nossos esforços para identificar e mitigar os riscos da escravidão moderna e do tráfico de pessoas em nossas operações comerciais e cadeias de suprimentos. Encontre os links para outras divulgações da Meta no [Anexo](#) deste relatório.

Nossa Política Corporativa sobre Direitos Humanos se aplica em toda a empresa. Cada serviço e entidade da Meta possui suas próprias políticas e procedimentos, que podem ter impactos diferentes sobre os direitos humanos em algumas situações. Este relatório menciona as ações realizadas pela Meta como empresa em relação a uma ou mais entidades da Meta. As declarações não têm a intenção de sugerir que a Meta tomou a mesma ação relativamente a todas as entidades.²

² Por exemplo, o WhatsApp, um app de mensagens e ligações criptografadas de ponta a ponta, tem pontos de contato exclusivos com os direitos humanos. A discussão deste relatório sobre moderação de conteúdo e ações relacionadas no Facebook e no Instagram não se aplica ao WhatsApp e, a menos que uma política ou ação seja especificada como aplicável ao WhatsApp, ela não se aplica ao WhatsApp. Além disso, embora muitas ações descritas neste relatório se apliquem ao Facebook e ao Instagram, há distinções intencionais nas políticas e nos procedimentos entre os serviços. Se uma política for identificada como pertencente ao "Facebook", ela não poderá ser aplicada ao Instagram. Nenhuma declaração neste relatório tem a intenção de criar — ou deve ser interpretada como criando — novas obrigações (legais ou de outra natureza) em relação à aplicação de uma política ou procedimento a outros serviços ou entidades.





Resumo executivo



Com este Relatório sobre Direitos Humanos, buscamos oferecer uma visão sobre as políticas, os processos e os padrões que a Meta utiliza para gerenciar os riscos aos direitos humanos para bilhões de pessoas, em grande escala e em tempo real. Nosso objetivo é mostrar como reunimos esforços para respeitar os direitos dos usuários e de outras comunidades em nosso trabalho e nos sistemas usados diariamente, desde a maneira como respondemos a crises e planejamos eleições até como lidamos com as demandas governamentais.

O relatório deste ano abrange nossas atividades do ano civil de 2023 e reflete o feedback das partes interessadas em relação ao [Relatório sobre Direitos Humanos do ano passado](#). O relatório deste ano destaca como nós, da Meta, gerenciamos em nossas operações globais os seguintes

oito direitos humanos relevantes: liberdade de opinião e expressão; privacidade; igualdade e não discriminação; vida, liberdade e segurança pessoal; melhor interesse da criança; participação pública, direito de votar e de ser eleito; liberdade de associação e reunião; e direito à saúde.

Os serviços e apps da Meta continuaram a ser usados para o bem, ajudando a construir comunidades, fomentar inovações e mobilizar recursos para apoiar a ajuda humanitária, entre outros esforços. À medida que surgiam desafios, nossa [Política Corporativa sobre Direitos Humanos](#) guiou nossas ações para lidar com questões como o uso de nossos serviços em países em conflito, segurança digital e solicitações excessivas de dados por entes governamentais.

Mantemos nosso compromisso de considerar os direitos humanos em nossa compreensão e desenvolvimento da inteligência artificial (IA). Nossos produtos de IA generativa permitem que as pessoas exerçam seus direitos humanos de maneiras inovadoras, mas reconhecemos que esses sistemas são imperfeitos e oferecem oportunidades para melhorias contínuas. Para atenuar os possíveis riscos, desenvolvemos uma [Política de Uso Aceitável](#), incorporada ao [Acordo de Licença da Comunidade](#) para o nosso modelo aberto Llama 2, e fornecemos ferramentas de segurança para ajudar os desenvolvedores a criar com responsabilidade. Acreditamos que uma abordagem aberta pode impulsionar a inovação por meio da iteração e resultar em produtos melhores e mais seguros por meio de feedback, análise, desenvolvimento e mitigações contínuos.

O envolvimento proativo com as partes interessadas é fundamental para nossa abordagem em relação aos direitos humanos e está incorporado em nossa Política Corporativa sobre Direitos Humanos. O desenvolvimento de nossos serviços, as políticas de conteúdo e a moderação, assim como os Padrões da Comunidade são [influenciados](#) por contribuições de uma grande variedade de organizações da sociedade civil, defensores dos direitos humanos, grupos marginalizados, organizações internacionais, investidores, anunciantes, usuários e [parceiros confiáveis](#). Esses grupos abrangem diversos espectros políticos e ideológicos, oferecendo contribuições variadas. Em janeiro de 2023, nossa rede de parceiros confiáveis incluía mais de 400 ONGs, agências humanitárias e defensores e pesquisadores dos direitos humanos de 113 países.

Em 2023, a Meta celebrou dez anos de adesão à [Global Network Initiative](#) (GNI). Aplicamos e somos responsáveis pelos Princípios da GNI sobre como as empresas de tecnologia devem respeitar a liberdade de expressão e os direitos de privacidade de seus usuários no atendimento às solicitações e restrições do governo. No relatório, descrevemos como a GNI orientou nossa resposta a solicitações governamentais, incluindo restrições a conteúdo e acesso a dados de usuários.

Também trabalhamos para integrar os princípios de direitos humanos na preparação e resposta a eventos e questões críticas em 2023. O relatório aborda nossa disponibilidade para enfrentar desafios intensificados em países de alto risco, como nos preparamos para eventos críticos planejados, como eleições, e nossa resposta coordenada a eventos de crise. Em 2023, a Meta facilitou a preparação para mais de 130 eleições em todo o mundo. Relatamos aqui nossa resposta aos conflitos no Sudão e em Nagorno-Karabakh, bem como questões de direitos humanos relacionadas à migração nas Américas Central e do Sul. Nossos esforços de resposta a crises envolveram diversos especialistas de diferentes áreas, incluindo política de conteúdo, operações globais, direitos humanos e produtos, além de outros especialistas internos que avaliaram e responderam a essas situações, usando nosso [Protocolo de Política de Crise](#). Também descrevemos como seguimos as recomendações da auditoria de direitos humanos anterior em relação a [Israel e Palestina](#) e nossas ações no conflito mais recente, bem como uma atualização sobre a criptografia de ponta a ponta.

Reforçamos nosso compromisso com a transparência e a reparação como



princípios fundamentais no gerenciamento de riscos relacionados aos direitos humanos. Continuamos a consultar o [Comitê de Supervisão](#), uma iniciativa líder no setor criada para ajudar a Meta a responder a algumas das perguntas mais desafiadoras sobre liberdade de expressão e segurança online. O ano de 2023 foi emblemático para o Comitê de Supervisão, que superou sua meta de 50 decisões anuais, mais do que triplicando o número de decisões tomadas em 2022.

Além disso, aprimoramos a experiência do usuário no Facebook e Instagram em relação a conteúdo restrito pela Meta, em conformidade com solicitações legais de governos e tribunais. Na maioria dos casos, nossas notificações informaram aos usuários qual autoridade estadual enviou a solicitação que resultou na restrição, bem como o país em que o conteúdo foi restringido.

Em 2023, enviamos nosso primeiro conjunto de solicitações de remoção da Áustria, do Vietnã, da Singapura, da Índia e do México para o banco de dados [Lumen](#). Este projeto de pesquisa independente, hospedado pelo Berkman Klein Center for Internet and Society da Universidade de Harvard, permite que pesquisadores estudem solicitações de remoção de conteúdo online feitas por governos e agentes privados. Esse é mais um passo para permitir que a comunidade global analise, denuncie e defenda os direitos digitais dos usuários da internet.





Apresentação

Continuamos a enfrentar novos e complexos desafios em resposta às transformações mundiais. Nossa resposta a esses desafios é orientada por nossa [Política Corporativa sobre Direitos Humanos](#) e nosso compromisso com os [Princípios Orientadores sobre Empresas e Direitos Humanos das Nações Unidas](#) e o [Pacto Global das Nações Unidas](#). E temos mecanismos de freios e contrapesos. O [Comitê de Supervisão](#) oferece auditoria especializada e independente para garantir a tomada de decisões justas e equilibradas, especialmente com relação a questões de liberdade de expressão.

Nossas políticas e procedimentos também evoluem frente às mudanças tecnológicas e sociais. Consideramos as contribuições de diversas partes interessadas, com ênfase em várias dimensões da diversidade, incluindo a de pontos de vista. Ao revisar essas políticas e procedimentos, buscamos feedback de funcionários, especialistas e defensores de direitos humanos, grupos marginalizados e organizações internacionais, além de uma ampla gama de grupos da sociedade civil.

Usamos uma variedade de ferramentas e protocolos de avaliação em toda a empresa para identificar, prevenir e mitigar os riscos aos direitos humanos que possam estar relacionados aos produtos, às políticas e às operações da Meta em todo o mundo. Fazemos relatórios sobre avaliações anteriores para compartilhar nosso progresso e abrir canais para feedback.

Em 2023, aplicamos nossas ferramentas de gerenciamento de risco ao responder a situações de crise, incluindo conflitos e guerras na África e no Oriente Médio, bem como ao planejar eventos mundiais previsíveis, como o número

significativo de eleições. Portanto, nossas políticas são dinâmicas para garantir a adoção de uma resposta adequada e personalizada para cada situação, equilibrando interesses conflitantes e considerações de direitos humanos.

A Meta é pioneira no desenvolvimento de inteligência artificial (IA) há mais de uma década. Sabemos que progresso e responsabilidade devem andar de mãos dadas. As ferramentas de IA generativa oferecem grandes oportunidades, e acreditamos que é possível e necessário que essas tecnologias sejam desenvolvidas de forma a respeitar os direitos humanos.

Um mercado digital forte e aberto é fundamental para ajudar as pessoas, a sociedade e os países a prosperar e construir comunidades. Hoje, mais do que nunca, precisamos da tecnologia para ajudar a aproximar o mundo. Buscamos cumprir nosso papel e mantemos os direitos humanos sempre presentes em nossas iniciativas.

Assinado por:



Nick Clegg,
Presidente de
Assuntos Globais



Jennifer Newstead,
Diretora do
Departamento Jurídico

IA no contexto dos direitos humanos

Riscos salientes relevantes:

- liberdade de opinião e expressão
- igualdade e não discriminação
- privacidade

Os rápidos avanços na inteligência artificial (IA), particularmente na [IA generativa](#), chamaram a atenção global em 2023.

Diante da maior atenção e das possíveis oportunidades e desafios que a IA traz, oferecemos uma análise aprofundada do nosso trabalho com IA sob a perspectiva dos direitos humanos.

Em 2023, lançamos nosso grande modelo de linguagem aberto [Llama 2](#) (LLM), o [Assistente da Meta AI](#), o modelo de geração de imagens [Emu](#) e [ferramentas de criação no app](#).

Como os sofisticados chatbots de IA e as ferramentas de geração de imagens passaram a ser amplamente utilizados, eles também atraíram a supervisão pública e regulatória. Por exemplo, a União Europeia avançou com seu marco [Regulamento sobre Inteligência Artificial](#), enquanto os Estados Unidos lançaram o conjunto de diretrizes [Compromissos Voluntários sobre IA da Casa Branca](#) com o apoio da Meta e de outros desenvolvedores de IA.

Temos o compromisso de desenvolver e implantar a IA de forma responsável e, ao mesmo tempo, mitigar os possíveis impactos adversos sobre os direitos humanos. A IA está incluída em nossa [Política Corporativa sobre Direitos Humanos](#), que reconhece a importância dos [Princípios da OCDE sobre Inteligência Artificial](#), amplamente adotados e endossados pelo G20.

Acreditamos que a IA responsável pode ser uma ferramenta poderosa para promover os direitos humanos.

Acreditamos que a IA responsável pode ser uma ferramenta poderosa para promover os direitos humanos. Nossos produtos de IA generativa permitem que as pessoas exerçam sua liberdade de expressão de maneiras inovadoras, melhorem o acesso à informação e à educação, além de aprimorem a acessibilidade. Por exemplo, legendas geradas por IA e reconhecimento de imagens nos óculos inteligentes Ray-Ban Meta melhoram a acessibilidade para pessoas com deficiências.

Recorremos à IA para tentar detectar e responder rapidamente a conteúdos potencialmente violadores de políticas e interações indesejadas em um número crescente de idiomas, a fim de evitar sua propagação. Muitas das nossas [iniciativas da Data for Good](#) utilizam IA para apoiar respostas humanitárias em situações de crise e facilitar o trabalho na área de saúde pública.

Adotando uma abordagem aberta

A Meta está empenhada em adotar uma abordagem responsável no desenvolvimento e na implantação de

produtos e modelos de IA generativa. Acreditamos que uma abordagem aberta pode impulsionar a inovação por meio da iteração e resultar em produtos melhores e mais seguros por meio de feedback, análise, desenvolvimento e mitigações contínuos da comunidade. Ela também pode:

- promover a justiça e apoiar pontos de vista diversos, permitindo que os detentores de direitos afetados identifiquem melhor os possíveis preconceitos;
- reduzir as barreiras à inovação, resultando em benefícios econômicos;
- permitir mais facilmente que as comunidades globais adaptem os avanços da IA a idiomas e contextos específicos.

No entanto, reconhecemos que uma abordagem aberta também pode apresentar riscos potenciais aos direitos humanos, incluindo o fato de que os desenvolvedores podem ignorar as diretrizes de uso responsável da Meta e implementar modelos de maneiras inseguras. Para tentar reduzir esses riscos, realizamos testes de segurança e desenvolvemos uma [Política de Uso Aceitável](#) detalhada, incorporada ao [Acordo de Licença da Comunidade](#) para nossos modelos Llama. Também compartilhamos um [Guia de Uso Responsável detalhado](#) e fornecemos ferramentas de segurança, por meio do nosso projeto [Purple Llama](#), incluindo [Llama Guard](#), para apoiar os esforços dos desenvolvedores em construir de forma responsável com esses modelos fundacionais disponíveis publicamente e implementá-los de maneira segura para seus próprios casos de uso.

Acreditamos que uma abordagem aberta pode impulsionar a inovação por meio da iteração e resultar em produtos melhores e mais seguros por meio de feedback, análise, desenvolvimento e mitigações contínuos da comunidade.

Implementar a IA de forma segura é uma responsabilidade compartilhada em todo o ecossistema, por isso temos colaborado durante muitos anos com organizações que trabalham para desenvolver uma IA segura e confiável. Por exemplo:

Temos trabalhado com a [MLCommons](#) e um conjunto global de parceiros para criar [marcos de responsabilidade](#) que beneficiem a comunidade de código aberto.

- Participamos de iniciativas envolvendo várias partes interessadas, como a [Parceria sobre IA](#), para tratar de questões que vão desde a implantação responsável de modelos fundacionais até as formas mais adequadas de identificar conteúdo sintético.
- Somos cofundadores da [AI Alliance](#), uma coalizão de empresas, acadêmicos, defensores e governos que trabalham para desenvolver ferramentas que possibilitem um ecossistema de IA aberto e seguro.

Consultamos as partes interessadas em direitos humanos para ajudá-las a entender e fornecer feedback sobre nosso uso da IA. Como parte desses esforços, colaboramos com o [Projeto B-Tech do Escritório de Direitos Humanos da ONU em sua atuação sobre IA generativa](#), contribuindo para [três documentos fundamentais](#) baseados nos Princípios Orientadores da ONU sobre Empresas e Direitos Humanos. Também colaboramos com um assessor do Escritório do Relator Especial da ONU para a Liberdade de Opinião e de Expressão e da UNESCO para ajudar a construir conhecimento sobre conceitos e riscos fundamentais da IA, além de facilitar discussões embasadas para entender as preocupações. Além disso, [realizamos um Fórum da Comunidade com fins consultivos](#), conduzido em parceria com a Universidade de Stanford, que incorporou questões de direitos humanos para ajudar a Meta e outras empresas, pesquisadores e governos a tomar decisões sobre chatbots de IA generativa. [Os resultados](#) enfatizaram a importância da estrutura de direitos humanos e da incorporação de informações de fontes de direitos humanos.

Lidando com os possíveis resultados prejudiciais da IA generativa

Estamos cientes de que a tecnologia de IA generativa pode produzir resultados que, por exemplo, gerem conteúdos potencialmente de ódio, ofensivos ou discriminatórios, reforcem preconceitos, apresentem informações imprecisas e/ou suscitem questões de privacidade. Também sabemos que indivíduos mal-intencionados podem tentar abusar dos modelos e das ferramentas que disponibilizamos de maneiras deliberadamente prejudiciais. Tomamos uma série de medidas para mitigar esses riscos em nossos modelos fundacionais, como Llama e Emu. Acreditamos que adotar uma abordagem aberta e iterativa nos ajudará a melhorar em tempo real caso esses problemas surjam.

O Projeto B-Tech da ONU [destaca](#) diversas áreas de risco relevantes em vários direitos reconhecidos internacionalmente. Observamos essa orientação e os [salientes riscos aos direitos humanos](#) da Meta para fundamentar nosso trabalho.

Para nossos modelos fundacionais [Llama 2](#) e [Emu](#), essas medidas de mitigação visam, em parte, abordar possíveis riscos significativos aos direitos humanos, abordando os riscos durante o treinamento, considerando associações contextualmente sensíveis e realizando avaliações de segurança e ajustes.

À medida que desenvolvemos nossos modelos fundacionais, compartilhamos [informações detalhadas](#) sobre nossa abordagem responsável publicamente.

Além de nossos esforços proativos para mitigar os riscos associados a saídas geradas por IA que possam ser prejudiciais, também trabalhamos para responder a esses resultados potencialmente nocivos assim que tomamos conhecimento deles.



Imagem gerada pela Meta AI

Nos dias seguintes aos ataques terroristas de 7 de outubro em Israel, partes interessadas nos informaram sobre diversos problemas com nossas ferramentas, que foram reconhecidos e corrigidos rapidamente por nós. Os principais problemas foram traduções automáticas nas quais a palavra "terrorista" foi adicionada a biografias de perfis de usuários no Instagram, e figurinhas geradas por IA que retratavam os palestinos de forma violenta.

Assim que tomamos conhecimento desses problemas, nossas equipes de engenharia iniciaram uma investigação para determinar as causas principais e implementar correções. Nossas equipes identificaram que esses problemas pareciam estar relacionados a alucinações do modelo e aos dados de treinamento usados. Ambos os problemas são desafios bem documentados para produtos com tecnologia de IA (veja [aqui](#) e [aqui](#)).

Implementamos uma correção de emergência em nosso produto de tradução automática 90 minutos após a descoberta do problema. Também trabalhamos o mais rápido possível para mitigar associações potencialmente problemáticas em imagens geradas e outputs de figurinhas relacionados ao dia 7 de outubro. Por exemplo, adicionamos sobrenomes conhecidos ao modelo para evitar alucinações. Além disso, aperfeiçoamos nossos modelos fundacionais de geração de imagens para lidar melhor com uma ampla gama de associações potencialmente problemáticas relacionadas a conflitos e fizemos as devidas implementações em nossos produtos.

Veja a seção [Israel e Palestina: atualização sobre nossas medidas](#) para obter mais detalhes.

Esperamos ter um envolvimento contínuo com uma série de titulares de direitos com opiniões diversas, enquanto trabalhamos para priorizar a segurança, a responsabilidade e os direitos humanos em nosso trabalho de desenvolvimento e implementação de produtos de IA generativa. Também tomaremos as medidas apropriadas para cumprir as obrigações legais nessa área.



Gerenciamento de riscos relacionados a direitos humanos

Todos os dias, buscamos transformar ativamente os princípios dos direitos humanos em ações significativas.

A base do nosso trabalho é a [Política Corporativa sobre Direitos Humanos da Meta](#). Nosso compromisso e abordagem são orientados pelos [Princípios Orientadores sobre Empresas e Direitos Humanos das Nações Unidas](#), bem como pelos padrões internacionais e regionais de direitos humanos listados em nossa Política Corporativa sobre Direitos Humanos.

Nosso trabalho inclui, entre outros, nossas iniciativas de avaliação cuidadosa, o [engajamento com as partes interessadas](#), a [consultoria sobre produtos](#), o [treinamento em direitos humanos](#) e o aconselhamento baseado em dados para todas as operações da organização. Neste relatório, priorizamos os diferentes tipos de auditoria de direitos humanos realizados na Meta. Para realizar as auditorias, aplicamos diversas ferramentas e protocolos de avaliação a fim de identificar, prevenir e mitigar os riscos que podem estar relacionados aos nossos produtos, às nossas políticas e às nossas operações em todo o mundo.

Esse trabalho orienta nossa forma de pensar sobre regulamentação, em conjunto com nossos compromissos da [Global Network Initiative \(GNI\)](#).



O ambiente regulatório está em transformação, o que impacta os direitos humanos, como demonstram as recentes tendências globais. Por exemplo, em 2023, observamos a promulgação da Lei de Segurança Online do Reino Unido e a entrada em vigor da Lei de Serviços Digitais da União Europeia, além de outras legislações sobre segurança online, incluindo em Singapura, em Taiwan e no Uruguai.

O gráfico na página 15 retrata algumas das ações que adotamos para o gerenciamento de riscos relacionados a direitos humanos. Outros detalhes

também estão disponíveis em nossos Relatórios sobre Direitos Humanos de [2021](#) e [2022](#).

Auditoria diária

Em nosso trabalho cotidiano de verificação, várias equipes contribuem para garantir que os direitos humanos estejam sempre incorporados às nossas ações. Por exemplo, especialistas em direitos humanos e direitos civis da Meta oferecem orientação contínua para avaliar e mitigar riscos relacionados aos direitos humanos em políticas de conteúdo, desenvolvimento de produtos, resposta

a crises e conflitos, e preparação para eleições. Esse trabalho geralmente envolve questões relacionadas a discurso de ódio, desinformação, conteúdo gráfico que pode documentar abusos de direitos humanos e medidas para ajudar a proteger os defensores de direitos humanos, bem como outros problemas relacionados a direitos humanos.

Com 3,19 bilhões de pessoas, em média, usando pelo menos um de nossos apps diariamente em dezembro de 2023, os possíveis impactos sobre os direitos humanos de nossos produtos variam significativamente de acordo com o tempo, o local, o conteúdo e as comunidades afetadas. Por essa razão, priorizamos nosso trabalho com base em nossa [Avaliação Abrangente de Riscos Relevantes a Direitos Humanos de 2022](#), que analisou os impactos em todos os direitos humanos reconhecidos internacionalmente. A análise priorizou os oito riscos mais relevantes utilizando os [Princípios Orientadores sobre Empresas e Direitos Humanos da ONU](#), conforme detalhado no gráfico abaixo.

Frequentemente, enfrentamos o desafio de como priorizar nossas ações em relação a riscos potenciais aos direitos humanos. Essa tensão é intrínseca à estrutura dos direitos humanos e fica evidente quando tentamos conciliar muitos direitos humanos que, às vezes, podem ser conflitantes. Procuramos priorizar os riscos mais importantes a direitos humanos, ou seja, aqueles com potencial para causar os impactos negativos mais graves às partes interessadas.

Procuramos priorizar os riscos mais importantes a direitos humanos, ou seja, aqueles com potencial para causar os impactos negativos mais graves às partes interessadas.

Nossa priorização é fundamentada nas diretrizes dos Princípios Orientadores sobre Empresas e Direitos Humanos da ONU e em nossa Política Corporativa sobre Direitos Humanos, além de ser ampliada pelo engajamento com partes interessadas, organizado com base em diferentes dimensões de diversidade. Por exemplo, a priorização é essencial para garantir que grupos vulneráveis, como os jovens, possam utilizar nossos produtos e serviços dentro da ampla e diversa população de usuários. A Meta se baseia na [Convenção da ONU sobre os Direitos da Criança](#), especialmente no princípio do "Melhor interesse da criança", para orientar a forma como atende os jovens. Isso, em algumas situações, resulta em decisões complexas sobre quais impactos potenciais nos direitos devem ser priorizados: por exemplo, reconhecer a segurança e o bem-estar das crianças e, ao mesmo tempo, respeitar seu direito de buscar informações de forma privada sobre temas como política, saúde ou identidade de gênero. O [princípio do Melhor Interesse da Criança](#) orientou a adoção e a criação de recursos como a [Central de Família](#) e a [Supervisão dos pais](#).



Nossa abordagem em relação ao gerenciamento de riscos a direitos humanos

Compromisso com os direitos humanos

Colocamos em prática nosso compromisso com os direitos humanos por meio de abordagens estabelecidas nos Princípios Orientadores da ONU, incluindo:

Aplicação das políticas de direitos humanos

Condução de auditoria de direitos humanos e divulgação

Garantia do acesso à reparação

Manutenção da supervisão, da governança e da responsabilidade

Proteção dos defensores dos direitos humanos

Priorização de direitos

A Avaliação Abrangente de Riscos Relevantes a Direitos Humanos identificou oito direitos humanos como prioritários:

Liberdade de opinião e expressão

Privacidade

Igualdade e não discriminação

Vida, liberdade e segurança pessoal

Melhor interesse da criança

Participação pública, direito de votar e de ser eleito

Liberdade de associação e reunião

Direito à saúde

Principais ferramentas e exemplos

Utilizamos as ferramentas e os processos a seguir para mitigar os riscos. Veja alguns exemplos ilustrativos neste relatório:

Desenvolvimento da política de conteúdo → Atualização sobre organizações e indivíduos perigosos

Consultoria de produtos → IA generativa; Threads

Incorporação dos direitos humanos à preparação das eleições → Eleições no Paquistão

Incorporação dos direitos humanos aos esforços de resposta coordenada → Priorização para o status de Países de Resposta Ativa

Resposta a crises → Conflitos em Nagorno-Karabakh e no Sudão

Estrutura da GNI para solicitações governamentais → Restrições de conteúdo baseadas na legislação local e acesso a dados de usuários

Comitê de Supervisão → Contribuições sobre casos de referência, por exemplo, reféns no direito internacional

Envolvimento das partes interessadas → Sessões de cocriação para pais, mães ou responsáveis e adolescentes

Proteção dos defensores dos direitos humanos → Fundo para defensores de direitos humanos



Riscos salientes relevantes:

- liberdade de opinião e expressão
- privacidade
- igualdade e não discriminação
- vida, liberdade e segurança pessoal

Política de conteúdo

Como parte do nosso trabalho essencial de auditoria, nossos especialistas em direitos humanos auxiliam no desenvolvimento de políticas de conteúdo. Isso significa que eles analisam as propostas de atualização de nossas políticas de conteúdo sob a ótica da legislação de direitos humanos para levar em conta a liberdade de expressão e a proteção contra a discriminação, bem como outros direitos humanos. Em 2023, isso incluiu o aprimoramento da nossa [Política sobre Organizações e Indivíduos Perigosos](#). Esse trabalho foi realizado em diferentes fases em 2023, e publicamos uma [atualização](#) sobre nossos progressos em janeiro de 2024.

No âmbito da nossa Política sobre Organizações e Indivíduos Perigosos, não permitimos que organizações ou indivíduos que anunciem uma missão violenta ou que estejam envolvidos em violência tenham presença nos serviços da Meta. Nossa abordagem global é fundamentada em [definições abrangentes](#) dos tipos e níveis de organizações perigosas, conforme publicado em nossa Central de Transparência. Temos um processo independente para identificar organizações e indivíduos perigosos, com cada caso analisado por especialistas relevantes com base em evidências corroboradas. Nossas definições e limites são independentes de região ou ideologia e foram desenvolvidos em consulta com especialistas e acadêmicos de todo o mundo. Além disso, aplicamos nossas medidas contra organizações terroristas estrangeiras designadas pelos EUA, chefes do tráfico de narcóticos e terroristas globais especialmente designados.

Em conformidade com as recomendações da auditoria em direitos humanos do [Comitê de Supervisão](#) e da [Business for Social Responsibility](#) sobre Israel e Palestina, [atualizamos](#) nossa Política sobre Organizações e Indivíduos Perigosos das seguintes maneiras:

- Atualizamos nossa política para permitir mais discursos sociais e políticos, inclusive sobre eleições, resolução de conflitos e ajuda humanitária, sem deixar de cumprir nossas obrigações legais.
- Recebemos comentários de que nossa definição anterior de "elogio a uma organização ou indivíduo perigoso" era muito ampla, por isso atualizamos nossa política para torná-la mais sutil e proporcional. Agora, proibimos a "glorificação" da violência e do ódio de organizações e indivíduos perigosos, que é definida de forma mais restrita do que "elogio".
- Simplificamos a maneira como avaliamos e categorizamos agentes perigosos com base em suas conexões com danos e violência no meio físico.
- Atualizamos nosso processo de remoção para incluir critérios mais detalhados e abrangentes que devem ser atendidos para que uma organização ou indivíduo perigoso seja considerado para tal fim. Essa atualização nos permite garantir que nossas designações reflitam as mudanças circunstanciais.

Em setembro de 2023, também fizemos alterações em nossa [Política sobre Violência e Incitação](#) para refinar a linguagem sobre violência de alta gravidade. Em março de 2023, fizemos alterações em nossa [Política sobre Bullying e Assédio](#) para incluir uma definição de "figuras públicas", conforme recomendado pelo Comitê de Supervisão. Em junho de 2023, realizamos um Fórum de Políticas sobre Conteúdo Violento e Explícito em Contextos de Abuso de Direitos Humanos em resposta a uma recomendação do Comitê de Supervisão no [caso Vídeo explícito do Sudão](#). Analisamos se nossas políticas sobre conteúdo violento e explícito estabeleciam um equilíbrio adequado entre o respeito à privacidade e dignidade das vítimas, a liberdade de expressão e o bem-estar das nossas comunidades. Nossa análise indicou que o equilíbrio era adequado, e não alteramos nossas políticas.

Para promover a transparência, as mudanças em políticas são refletidas em nossos [Padrões da Comunidade](#) e podem ser visualizadas no respectivo registro de alterações e nas [atas públicas dos nossos Fóruns de Política](#).

Além disso, atualizamos nosso [sistema de penalidades](#) em 2023. O novo sistema reduziu o número de períodos de restrição e visa ajudar a garantir uma resposta proporcional às violações das políticas. Ainda aplicaremos restrições à conta de violadores recorrentes, normalmente depois da sétima violação, após fornecermos avisos e explicações suficientes para ajudar a pessoa a entender por que removemos seu conteúdo. Para violações mais graves, como publicar conteúdo que contém terrorismo, exploração infantil, tráfico humano, incentivo ao suicídio, exploração sexual, venda de drogas não medicinais ou promoção de organizações e indivíduos perigosos, continuaremos aplicando consequências imediatas.

Desenvolvimento de produtos

Nossos engenheiros concentram esforços para resolver de forma criativa os problemas do mundo real. Os direitos humanos orientam nosso trabalho no desenvolvimento de práticas de inovação responsável, e estamos comprometidos com a realização de uma avaliação cuidadosa de novos produtos.

Em 2023, lançamos vários produtos inovadores, como o [Threads](#) em julho, os [óculos inteligentes Ray-Ban Meta](#) e os [canais do WhatsApp](#) em setembro, além do [Quest 3](#), nosso headset de realidade mista, em outubro. Para cada um desses produtos, as considerações sobre direitos humanos foram incorporadas ao processo de desenvolvimento.

Além disso, continuamos a desenvolver o metaverso, com foco em melhorar a educação e combater o preconceito e o ódio. Por exemplo, lançamos várias experiências de realidade virtual para criar uma comunidade, como [MLK: Now is the Time](#), [Inside the Mosque](#) (uma experiência em [duas etapas](#)) e [ABLE](#).

Em um ambiente dinâmico e iterativo, as equipes desenvolvem, testam, refinam e implementam produtos e serviços, incluindo aqueles habilitados por IA. Construímos com [foco na privacidade](#) e desenvolvemos abordagens de mitigação de riscos de produtos que atendem às necessidades específicas do nosso processo dinâmico de desenvolvimento. Desde o início do processo de desenvolvimento, as equipes de produto aplicam orientações sobre inovação responsável, alinhadas aos princípios de direitos humanos. Isso ajuda as equipes a antecipar e mitigar possíveis danos a indivíduos, comunidades e à sociedade. Nossos especialistas em direitos civis e direitos humanos podem prestar consultoria ou realizar ações rápidas, por exemplo, para

Riscos salientes relevantes:

- liberdade de opinião e expressão
- privacidade
- igualdade e não discriminação



identificar os países de maior risco para o lançamento de produtos.

Por exemplo, durante o processo de desenvolvimento do produto Threads, nossos especialistas em direitos humanos aproveitaram os processos existentes de revisão de produtos e políticas para avaliar o possível impacto sobre os direitos humanos. Eles trabalharam com a equipe do Threads para implementar mitigações, inclusive em relação a possíveis solicitações governamentais para censurar o conteúdo.

Estamos também avançando com o Projeto Height, uma estrutura para as equipes de produto avaliarem as preocupações relacionadas aos direitos civis que surgem durante os lançamentos de produtos. Dessa

forma, as equipes de desenvolvimento de produtos consideram os direitos civis ao longo de todo o processo de criação. Isso complementa outros processos de gerenciamento de riscos implementados para avaliar a segurança dos nossos produtos.



MLK: Now is the Time pelo Meta Quest



Riscos salientes relevantes:

- vida, liberdade e segurança pessoal
- participação pública, direito de votar e de ser eleito
- liberdade de opinião e expressão
- igualdade e não discriminação
- privacidade
- melhor interesse da criança

Como nos preparamos e respondemos a eventos críticos e crises globais

Trabalhamos em toda a empresa para integrar os princípios de direitos humanos nas iniciativas de integridade da Meta.³ Isso abrange nossa preparação para desafios aumentados em países de alto risco, assim como eventos críticos planejados, como eleições, e nossa resposta coordenada a situações de crise. Esse trabalho coordenado entre equipes especializadas nos ajuda a antecipar e responder de forma eficiente a tensões elevadas, onde quer que ocorram.

Como a Meta se prepara para eleições

Em 2023, muitos países foram às urnas, e outros ainda se prepararam para um movimentado ciclo eleitoral de 2024. Em 2023, a Meta permitiu a preparação de mais de 130 eleições em todo o mundo, incluindo no Paquistão, Argentina, Turquia e Nigéria. Compartilhamos a seguir insights e nossas ações em duas eleições relevantes, no Brasil e no Paquistão.

Nossas equipes especializadas lideram os preparativos na Meta como um todo para implementar medidas de proteção antes das

eleições globais. Isso pode incluir operações de segurança avançadas para lidar com ameaças adversárias, políticas e processos atualizados para remover conteúdos potencialmente prejudiciais de nossas plataformas, nossa rede de verificação de fatos global de referência no setor e transparência em relação a anúncios políticos e de questões sociais.

Também avaliamos continuamente os riscos de danos iminentes durante eventos críticos, como eleições, para que possamos responder com ações direcionadas e limitadas no tempo em relação a políticas e produtos que ajudarão a manter as pessoas seguras.

Embora a maior parte do nosso trabalho seja proativa e realizada bem antes das datas das eleições, também nos preparamos para reagir a eventos de alto risco. Nesses casos, a Meta tem várias abordagens que podemos adotar para lidar com uma variedade de cenários, por exemplo, alterando ou limitando os recursos de produtos, introduzindo limites de taxa de mensagens e limitando a distribuição de conteúdo.

Identificamos tendências potencialmente prejudiciais de várias formas, como coletando informações de relatórios públicos, revisando recomendações de nossos parceiros confiáveis, monitorando

³ "Integridade" é um termo técnico na Meta. A confiança e a segurança, a proteção de contas e computadores, a redução de experiências negativas, questões de privacidade relacionadas e outros aspectos são tratados por uma rede de equipes dentro da Meta, que frequentemente incluem "Integridade" em seus nomes. Entre outras funções, essas equipes desenvolvem ferramentas para prevenir danos, moderar nossas plataformas e fazer cumprir nossas políticas.

continuamente tendências de conteúdo, realizando auditorias de direitos humanos e analisando avaliações de nossos especialistas em inteligência. Essas informações, em conjunto com os Padrões da Comunidade e os sistemas de fiscalização existentes, ajudam a definir o tipo de produto e as atenuações de política que podemos usar para evitar abusos durante eleições de alto risco. Isso abrange nosso [Protocolo de Política de Crise](#) e outras medidas que podemos adotar para ajustar nossos sistemas em resposta a [riscos elevados](#). Além disso, fornecemos ferramentas de transparência, como nossa [Biblioteca de Anúncios](#).

Brasil

Começamos a preparação para a [eleição presidencial de 2022](#) um ano antes da votação. Damos continuidade aos nossos esforços durante o período pós-eleitoral, o que nos permitiu reagir rapidamente aos [ataques de 8 de janeiro de 2023](#) ao Congresso Brasileiro, ao Supremo Tribunal Federal (STF) e a outros prédios públicos. A Meta também designou a agitação pós-eleitoral como uma crise, conforme o Protocolo de Política de Crise, para ajudar a empresa a avaliar como mitigar os riscos de conteúdo.

Nossos preparativos para a eleição de 2022 incluíram o mapeamento de situações de risco durante e após a eleição. Consideramos os padrões internacionais sobre liberdade de

expressão e outros direitos humanos. Também fizemos alterações em produtos e adaptamos políticas para proteger a integridade da eleição brasileira. Isso incluiu uma parceria com o Tribunal Superior Eleitoral para fornecer às pessoas acesso a informações confiáveis sobre a votação.

Implementamos uma variedade de ferramentas e métodos para combater possíveis discursos de ódio, incitação à violência e desinformação antes, durante e depois dos ataques de 8 de janeiro de 2023. Designamos o Brasil como um local temporário de alto risco pouco antes da eleição, de modo que removemos os apelos para portar armamentos ou entrar à força em prédios do governo, conforme observado no caso do [Comitê de Supervisão "Discurso de general brasileiro"](#).

Desde o início da campanha eleitoral em 16 de agosto de 2022 até o final de janeiro de 2023, [removemos](#) mais de 1 milhão de itens de conteúdo no Facebook e mais de 960.000 itens de conteúdo no Instagram em virtude de violações das nossas políticas de violência e incitação no Brasil. Esse conteúdo incluía publicações pedindo intervenção militar.

Durante esse período de cinco meses no Brasil, removemos itens de conteúdo que violavam nossas políticas de discurso de ódio:



Mais de 570 mil



Mais de 520 mil

Também removemos itens de conteúdo no Brasil por violação das nossas políticas de bullying e assédio:



Mais de 380 mil



Mais de 630 mil



Violência de gênero

A ameaça de assédio e violência de gênero contra mulheres candidatas, jornalistas e defensoras dos direitos humanos geralmente aumenta durante as eleições. Para mitigar esses riscos e como parte de nossa abordagem mais ampla para aumentar a proteção das mulheres nas plataformas e conectá-las a serviços públicos, nossas iniciativas no Brasil abarcaram:

- parceria com o Ministério da Mulher para lançar um canal oficial no WhatsApp no Brasil, permitindo que mulheres enviem denúncias ou solicitem informações da Ouvidoria sobre violência contra mulheres, como um meio adicional de comunicação com a Ouvidoria. Os usuários podem obter informações sobre as leis relevantes, encontrar endereços de serviços dedicados à assistência às mulheres e falar diretamente com os atendentes.
- publicação de um guia online em português para proteger as mulheres contra a violência online.
- colaboração com grupos do setor, organizações da sociedade civil e órgãos reguladores, incluindo o Ministério da Mulher, em iniciativas conjuntas para lidar com a violência e o assédio de gênero.





Paquistão

Embora as eleições no Paquistão tenham ocorrido em 2024, nossos preparativos para o pleito começaram em 2022 e incluíram esforços em toda a empresa nas áreas de produto, política e operações para evitar e mitigar riscos aos direitos humanos associados ao uso das nossas plataformas.

Entre outras iniciativas, criamos sinais para identificar conteúdos com preocupações relacionadas a direitos e situações que exigem atenção prioritária. Em conformidade com os [Princípios da GNI](#), publicamos um [estudo de caso](#) em tempo real em nossa Central de Transparência sobre os pedidos de remoção recebidos em dezembro de 2023 do governo do Paquistão, que não violavam nossos Padrões da Comunidade ou a legislação local.

O envolvimento das partes interessadas é fundamental. Antes do pleito, realizamos reuniões informativas com as partes interessadas regionais em direitos humanos sobre os esforços de integridade nas eleições. Também fizemos reuniões com a Comissão Eleitoral para explicar nossa abordagem em relação aos pedidos de ação governamental e aos direitos humanos. A Internet foi restringida durante o período que antecedeu a eleição, e a Meta compartilhou [informações](#) com as partes interessadas sobre como configurar um servidor proxy para conexão ao WhatsApp.

Como membro de grupos de várias partes interessadas, incluindo a [GNI](#) e a [Asia Internet Coalition](#), a Meta emitiu declarações de preocupação sobre os impactos das interrupções da internet e da legislação que infringe direitos antes das eleições.

Vários meses antes das eleições, também tomamos medidas para oferecer aos defensores dos direitos humanos e a outros indivíduos vulneráveis proteções aprimoradas na plataforma, o que incluiu proteção contra remoções de conteúdo com base em falsas acusações de violações dos Padrões da Comunidade. Nossa abordagem foi inclusiva, em conformidade com a nossa [Política Corporativa sobre Direitos Humanos](#), que adota a definição ampla de defensores dos direitos humanos contida na [Declaração da ONU sobre Defensores dos Direitos Humanos](#).





Preparação para desafios em Países de Resposta Ativa

A classificação "País de Resposta Ativa" é definida segundo um processo baseado em evidências, conduzido por equipes especializadas e nossos especialistas em direitos humanos, com base em nosso [processo de auditoria de direitos humanos](#) para avaliar e priorizar países com maior risco de danos e violência no meio físico, o que desencadeia uma estratégia abrangente de mitigação de riscos. Isso pode incluir esforços intensificados de monitoramento e mitigação de riscos, além de investimentos adicionais, como a análise de conteúdo em idiomas relevantes para conflitos visando remover conteúdo que viole políticas e suporte personalizado a produtos. Devido à complexidade dos riscos, também recorremos a parceiros confiáveis e verificadores de fatos terceirizados para respaldar nossos esforços nos Países de Resposta Ativa.



Resposta a crises globais

Respondemos a crises inesperadas em todo o mundo, incluindo conflitos interestaduais, ataques terroristas e crises não violentas, como desastres ambientais. Nossos esforços de resposta a crises envolvem diversos especialistas de diferentes áreas, incluindo política de conteúdo, operações globais, direitos humanos e produtos, além de outros especialistas internos que avaliam e respondem a essas situações, aplicando nosso [Protocolo de Política de Crise](#). Conforme descrito em nosso [Relatório sobre Direitos Humanos de 2022](#), o protocolo orienta o uso acelerado de medidas políticas direcionadas ou excepcionais para mitigar possíveis danos, em alinhamento com os riscos observados, e fundamentadas em intervenções em crises passadas e princípios de direitos humanos.

Nossa Avaliação Abrangente de Riscos Relevantes a Direitos Humanos nos ajuda a priorizar entre diferentes riscos em potencial aos direitos humanos em situações de conflito, por exemplo, entre o direito à liberdade e à segurança da pessoa, o direito à liberdade de expressão, e a segurança e bem-estar de nossas comunidades. Como parte do nosso conjunto de ferramentas, podemos adotar diferentes medidas temporárias para ajudar a manter a segurança das pessoas e mitigar os riscos de que nossas plataformas sejam usadas para agravar ainda mais as tensões, online e no meio físico. Podemos levar em conta os desenvolvimentos relevantes do direito internacional público, incluindo o direito humanitário internacional — a lei dos conflitos armados — de acordo com as diretrizes dos Princípios Orientadores sobre

Empresas e Direitos Humanos da ONU, que tratam da intensificação dos processos de auditoria de direitos humanos em períodos de conflito. Apresentamos abaixo exemplos ilustrativos do nosso trabalho. Apresentamos uma [atualização](#) sobre



nossas ações referentes à auditoria de direitos humanos sobre Israel e na Palestina, além do conflito relacionado, na próxima subseção.

Sudão

Em abril de 2023, o conflito entre as Forças Armadas do Sudão (SAF) e as Forças de Apoio Rápido (RSF) resultou em deslocamento em massa, insegurança alimentar, falta de acesso à saúde e outros desafios que demandaram assistência humanitária. Aplicando o Protocolo de Política de Crise da Meta, o Sudão foi avaliado como um local de risco máximo.

Seguindo a orientação do nosso Protocolo de Política de Crise, tomamos medidas para

- Riscos salientes relevantes:
- vida, liberdade e segurança pessoal
 - liberdade de opinião e expressão
 - liberdade de associação e reunião
 - privacidade
 - direito à saúde

respeitar os direitos humanos, minimizar os riscos de danos e apoiar a prestação de ajuda humanitária.

Em agosto de 2023, categorizamos a RSF conforme nossa [Política sobre Organizações e Indivíduos Perigosos](#), que não permite que organizações ou indivíduos que anunciem uma missão violenta ou que estejam envolvidos em violência tenham presença em nossas plataformas. Isso ajudou a reduzir a disseminação de conteúdo potencialmente prejudicial. A decisão foi tomada após a interação com agências da ONU e com base em pesquisas de auditoria de direitos humanos e relatórios de grandes organizações não governamentais que abordam graves violações de direitos humanos. Ela foi analisada e endossada pelo Comitê de Supervisão. Também designamos o Sudão como um local temporário de alto risco, o que ajudou a remover conteúdo que incluía apelos para que civis pegassem em armas ou as levassem para um local específico no país. Trabalhamos com verificadores de fatos independentes para desmentir, rotular e diminuir a visibilidade de desinformação relacionada a conflitos na plataforma. Também removemos desinformação que parceiros confiáveis avaliaram como potencialmente capazes de aumentar o risco de danos físicos iminentes ou violência.

Em razão do deslocamento em massa da capital, Cartum, foi difícil ou impossível contatar muitos de nossas partes interessadas, incluindo representantes do governo, sociedade civil, mídia e defensores dos direitos humanos. Isso exigiu a identificação e o envolvimento com novos grupos para apoiar os esforços humanitários, bem como para ajudar a identificar e interromper a disseminação de conteúdo potencialmente prejudicial e desinformação. Colaboramos com os grupos mais ativos da sociedade civil, incluindo médicos, a mídia, jornalistas cidadãos e grupos de mulheres, tanto dentro do Sudão quanto, principalmente, na diáspora. Nosso papel consistiu em fornecer informações sobre a política de conteúdo e oferecer treinamento em segurança

digital e ferramentas de gerenciamento de crises da Meta, o que incluiu a realização de workshops sobre como usar as comunidades do WhatsApp, que mais tarde foram usadas para coordenar evacuações e apoio humanitário.

Apesar da deterioração da situação de segurança, nossos parceiros confiáveis continuaram denunciando conteúdo relacionado a violência e incitação, desinformação, discurso de ódio, bullying e assédio. As avaliações de parceiros confiáveis ajudaram a aplicar nossa Política sobre Desinformação e Danos em relação a conteúdo incluindo imagens fora de contexto para alimentar tensões, bem como em relação a alegações feitas contra agências humanitárias que operam no Sudão.

De maneira semelhante à nossa abordagem em [outros conflitos](#), removemos alguns conteúdos explícitos violadores sem aplicar [penalidades](#), que poderiam resultar em [restrições de contas](#) crescentes, a fim de evitar penalizar ou restringir excessivamente os usuários que tentavam conscientizar sobre os impactos do conflito. Essa abordagem foi motivada por uma recomendação do Comitê de Supervisão no [caso Vídeo explícito do Sudão](#) em 2022. Esse caso levou a uma análise em 2023 das nossas políticas sobre o compartilhamento de conteúdo violento e explícito para conscientizar sobre abusos de direitos humanos, além de uma explicação mais detalhada sobre como aplicamos nossa [permissão para conteúdo de valor jornalístico](#).

Havia uma escassez aguda de medicamentos no país devido à falta de infraestrutura de saúde, por isso fizemos alterações temporárias nas políticas para permitir que os usuários solicitassem ou doassem medicamentos, algo normalmente proibido por nossas políticas de bens restritos. Além disso, colaboramos com o Conselho de Especialização Médica do Sudão para lançar o [Sudan Clinic](#), um app de telemedicina que usa a API do WhatsApp. Esse recurso foi disponibilizado com o apoio do nosso programa de isenção de taxas de impacto

Construindo resiliência: treinamento em segurança digital e políticas de conteúdo

2023:

6

sessões de
treinamento realizadas

300

beneficiários
impactados

Após o conflito:

2

sessões de
treinamento realizadas

70

beneficiários
impactados



social do WhatsApp , que beneficia parceiros governamentais e sem fins lucrativos ao não cobrar taxas pelo custo das mensagens.

Também precisávamos equilibrar as mensagens sobre mobilidade e segurança — informações que auxiliam as pessoas a viajar e, ao mesmo tempo, a garantir sua proteção — sem apoiar a migração irregular para países vizinhos, como República Centro-Africana, Chade, Egito, Etiópia, Líbia e Sudão do Sul. Isso porque o risco de exploração humana aumenta à medida que as necessidades da população deslocada crescem, resultando na emergência de indivíduos mal-intencionados nessas áreas. Reunimos esforços para identificar agentes maliciosos e remover conteúdos que poderiam colocar pessoas em risco, incluindo publicações relacionadas a contrabando de migrantes.

Nagorno-Karabakh

Em 19 e 20 de setembro de 2023, o Azerbaijão iniciou uma ofensiva militar na autodeclarada República de Artsakh, parte da disputada região de Nagorno-Karabakh. A ofensiva provocou a fuga de armênios étnicos da área, desencadeando uma crise de refugiados.

Uma equipe multidisciplinar, incluindo especialistas em direitos humanos, monitorava a situação e mobilizou recursos rapidamente, aplicando nosso [Protocolo de Política de Crise](#). Especialistas internos com habilidades linguísticas e conhecimento contextual gerenciaram o aumento da prevalência de discurso de ódio, bullying, conteúdo explícito e referências a organizações e indivíduos perigosos. Também colaboramos com agências da ONU no local para compreender melhor a dinâmica da crise. A Meta ofereceu proteções adicionais a atores cívicos de ambos os lados para reduzir os riscos de falsificação de identidade, assédio e invasões.

Riscos salientes relevantes:

- liberdade de opinião e expressão
- vida, liberdade e segurança pessoal
- privacidade

Como em outros conflitos, tivemos que decidir como equilibrar a preservação da privacidade e da dignidade dos prisioneiros de guerra com os benefícios de documentar as violações dos direitos humanos. Os Padrões da Comunidade da Meta proíbem a publicação de informações que possam revelar a identidade ou a localização de prisioneiros de guerra, o que é consistente com as responsabilidades de direitos humanos da empresa de acordo com os Princípios Orientadores sobre Empresas e Direitos Humanos da ONU. Essas responsabilidades aumentam durante conflitos armados e devem ser baseadas nas regras do direito internacional humanitário. Revelar a identidade e a localização de prisioneiros de guerra pode colocar em risco sua segurança, privacidade e dignidade, bem como expor seus familiares.

Ao mesmo tempo, essa exposição pode enriquecer o debate público e aumentar a conscientização sobre possíveis maus-tratos, incluindo violações de direitos humanos internacionais e do direito humanitário internacional. Isso também pode fomentar ações que protegem direitos e garantem a responsabilização. A escala e a rapidez com que imagens de prisioneiros de guerra podem ser compartilhadas nas redes sociais tornam mais difícil a tarefa de resolver esses interesses conflitantes.

Em 2023, o [Comitê de Supervisão](#) manteve nossa decisão de deixar ativa uma publicação no Facebook que incluía um vídeo retratando prisioneiros de guerra identificáveis com a tela de aviso adicional "marcar como perturbador". O Comitê concordou com a Meta, afirmando que, nesta situação, as publicações poderiam enriquecer o debate público e aumentar a conscientização sobre possíveis maus-tratos, incluindo violações de direitos humanos internacionais e do direito humanitário internacional, o que "superava o risco à segurança e dignidade dos prisioneiros de guerra".

Retenção de evidências

Apoiamos a justiça para todos os crimes internacionais. Há vários anos, trabalhamos para explorar iniciativas que respeitem os direitos de retenção e divulgação de evidências, consultando extensivamente a sociedade civil, o meio acadêmico e especialistas e órgãos de justiça criminal internacionais.

Conforme [mencionado](#) em nossa resposta às recomendações do Comitê de Supervisão, trabalhamos para desenvolver uma abordagem que permita que tribunais internacionais e mecanismos de responsabilização autorizados pela ONU (como missões de apuração de fatos e Comissões de Inquérito) façam solicitações a nós para a retenção prolongada de dados relevantes para suas investigações em andamento. Esse trabalho está quase concluído e é distinto de nossas [políticas consolidadas para responder a solicitações de preservação de dados de órgãos de aplicação da lei](#).

Atualizamos diversos mecanismos da ONU e relatores especiais sobre nossa abordagem e detalhamos o processo para fazer solicitações de retenção prolongada à Meta. Analisaremos cuidadosamente todas as solicitações recebidas para verificar a consistência com nossas políticas e a legislação aplicável.

Essa é uma área nova, sem práticas recomendadas estabelecidas ou testadas, e existem considerações legais, de privacidade e políticas significativas inerentes a esse trabalho. Esperamos compartilhar mais atualizações sobre nosso trabalho nesta área em nossos [relatórios semestrais sobre o Comitê de Supervisão](#) e em nosso Relatório Anual sobre Direitos Humanos.



Riscos salientes relevantes:

- vida, liberdade e segurança pessoal
- privacidade

Migração nas Américas Central e do Sul

O contrabando de migrantes pode resultar na exploração de pessoas vulneráveis que estão tentando deixar o país de origem delas, geralmente em busca de uma vida melhor. Investimos recursos para evitar e impedir danos, removemos conteúdo que facilite ou coordene a exploração humana, incluindo tráfico de pessoas e contrabando de migrantes.

Em 2023, identificamos possíveis riscos aos direitos humanos relacionados ao contrabando de migrantes nas Américas do Norte, Central e do Sul, inclusive ao longo do Darién Gap, um trecho de selva perigoso e densamente florestado no norte da Colômbia e no sul do Panamá. Essas são áreas de maior risco de atividade comercial ilegal, presença online de organizações perigosas e desinformação para migrantes e refugiados, especialmente quando alguns tentam chegar à fronteira entre os EUA e o México.

Em resposta a esses riscos, aplicamos nossos [Padrões da Comunidade](#) e políticas de negócios sobre [comércio](#) que proíbem a exploração humana, inclusive o contrabando de migrantes. Colaboramos com a sociedade civil e com as autoridades para combater a desinformação e continuamos a monitorar e a aplicar nossas políticas de violação de conteúdo orgânico, pago e comercial. Também facilitamos o trabalho de verificadores de fatos independentes para encontrar e avaliar conteúdos relacionados à fronteira entre os EUA e o México, pois reconhecemos que a agilidade é especialmente importante durante eventos críticos. Usamos a detecção de palavras-chave para agrupar conteúdos relacionados em um único lugar para nossos parceiros de verificação de fatos, que avaliaram o conteúdo em nossa plataforma e publicaram artigos em inglês e espanhol abordando a desinformação.

Também apoiamos comunidades de migrantes e refugiados, assim como organizações não governamentais, por meio de programas de integridade e educação, como o We Think Digital, que prioriza habilidades digitais para identificar desinformação, fraudes e golpes.

Avaliações de impacto sobre os direitos humanos

As metodologias aplicadas às auditorias de direitos humanos estão em constante evolução. Consideramos os riscos potenciais e aprendemos com os esforços anteriores para determinar como podemos melhorar nossa abordagem e fortalecer as iniciativas que tiveram bons resultados.

Realizamos avaliações de impacto sobre os direitos humanos para países, apps, serviços, hardware ou iniciativas estratégicas específicos, conforme orientações da nossa Avaliação Abrangente de Riscos Relevantes a Direitos Humanos. Realizamos essas avaliações para tentar prever possíveis impactos, especialmente na preparação de novos produtos e recursos, como [criptografia de ponta a ponta](#). As avaliações de impacto sobre os direitos humanos envolvem o [engajamento com partes interessadas](#), um componente fundamental da auditoria de direitos humanos. Quando realizamos uma avaliação, cada recomendação geralmente requer várias equipes e fluxos de trabalho para implementar as recomendações em toda a empresa.

Compartilhamos resumos e recomendações de avaliações de impacto sobre os direitos humanos no passado, e eles estão listados no [Anexo](#). Também fornecemos atualizações sobre as ações tomadas em

resposta às avaliações realizadas sobre os possíveis riscos a direitos humanos relacionados às plataformas da Meta nas [Filipinas](#), na [Índia](#), na [criptografia de ponta a ponta](#) e em [Israel e na Palestina](#).

Israel e Palestina: atualização sobre nossas medidas

Pouco depois de publicarmos nossa [atualização da auditoria de direitos humanos sobre Israel e Palestina de setembro de 2023](#), ocorreram os ataques terroristas do Hamas contra Israel em 7 de outubro, seguidos pela resposta militar de Israel em Gaza e pelo envolvimento e escalada adicionais de outros atores regionais. Sabemos que as pessoas na região e em várias partes do mundo se sentiram profundamente afetadas por nossa resposta à situação de conflito em andamento na região.

Imediatamente após o ataque, a Meta classificou a violência no nível mais alto do nosso Protocolo de Política de Crise e implementou medidas de resposta à crise, incluindo uma equipe multidisciplinar dedicada disponível 24 horas por dia, 7 dias por semana. Os princípios básicos de direitos humanos orientaram nossas ações. Consideramos os [Princípios Orientadores da ONU sobre Empresas e Direitos Humanos](#), como base para nossa [Política Corporativa sobre Direitos Humanos](#), visando priorizar e mitigar os riscos mais importantes aos direitos humanos. Também usamos o direito humanitário internacional como uma referência importante.

Inicialmente, compartilhamos os detalhes da nossa resposta em uma publicação no blog em [inglês](#), [árabe](#) e [hebraico](#) no dia

Riscos salientes relevantes:

- vida, liberdade e segurança pessoal
- liberdade de opinião e expressão
- igualdade e não discriminação
- privacidade

13 de outubro de 2023, e fornecemos outras atualizações em 18 de outubro, 5 de dezembro e 8 de dezembro. Continuamos a refinar nossa abordagem para refletir as mudanças na dinâmica, incluindo a crise humanitária em curso em Gaza e os reféns ainda mantidos em cativeiro pelo Hamas.

Equilibrar a segurança e o bem-estar das nossas comunidades e a voz em nossas plataformas é especialmente difícil em situações de conflito, e ainda mais desafiador quando envolve entidades sancionadas pelos EUA, como o Hamas, o Hezbollah e a Jihad Islâmica Palestina, que também são designadas como organizações perigosas de acordo com nossas políticas.

O trabalho que já realizamos em resposta à [auditoria de direitos humanos sobre Israel e Palestina](#) ajudou a fundamentar nossa abordagem. Por exemplo, conseguimos direcionar o conteúdo em árabe de forma mais adequada em nossos sistemas para maior precisão. Também aumentamos nossos recursos de moderação de conteúdo no mercado hebraico. Para ver uma atualização mais detalhada, veja [Atualização da Meta de 2024: Auditoria de direitos humanos sobre Israel e da Palestina](#).

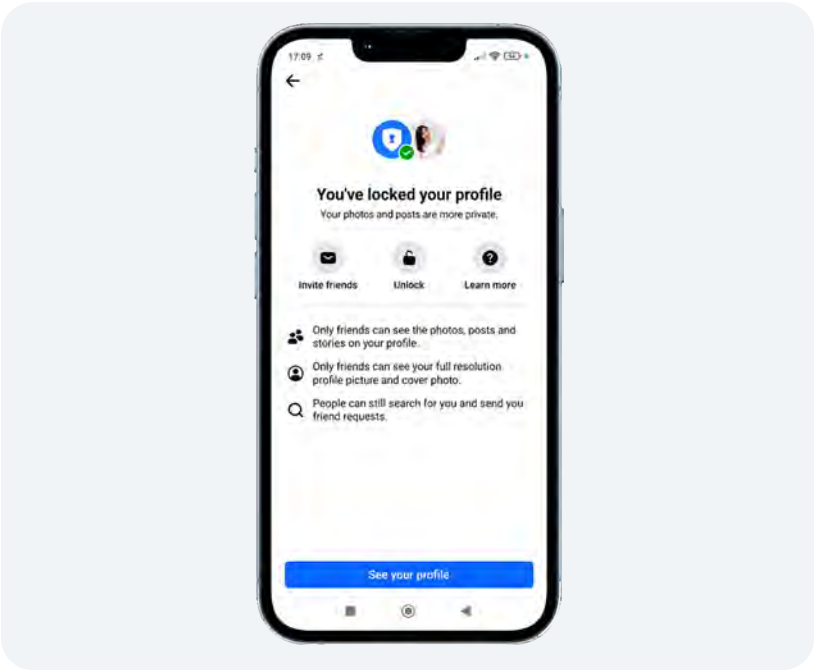
Nossa resposta ao conflito foi além das medidas descritas na [auditoria](#) e também foi guiada pela aplicação do nosso [Protocolo de Política de Crise](#). Adotamos diversas [medidas temporárias em relação a produtos e políticas](#) para mitigar os riscos mais relevantes aos direitos humanos. Sabemos que, em alguns casos, essas medidas (por exemplo, a redução dos limites para a aplicação automatizada), podem restringir inadvertidamente a discussão de eventos mundiais críticos, afetando a liberdade de expressão. Por isso, buscamos adotar medidas que sejam temporárias e proporcionais aos riscos à medida que os reconhecemos.

Medidas temporárias sobre políticas

Usamos como referência nossa Política sobre Organizações e Indivíduos Perigosos e ampliamos nossa Política sobre Violência e Incitação para remover conteúdos que mostrassem reféns identificáveis sendo sequestrados ou mantidos em cativeiro, mesmo quando isso era feito para condenar ou conscientizar sobre a situação deles. Tínhamos como objetivo proteger a privacidade e a dignidade das vítimas, de acordo com os padrões do direito humanitário internacional, e impedir que a propaganda do Hamas apareça em nossas plataformas, conforme nossa Política sobre Organizações e Indivíduos Perigosos.

À medida que a guerra prosseguia, vimos que as pessoas começaram a compartilhar conteúdo sobre reféns para refutar as narrativas emergentes que negavam a ocorrência dos ataques terroristas de 7 de outubro. Adotamos uma abordagem mais sutil e permitimos que as pessoas compartilhassem conteúdo que mostrasse reféns sendo sequestrados para aumentar a conscientização ou condenar os ataques. Continuamos a remover conteúdo e agimos com cautela quando a intenção do usuário ao compartilhar o material não estava clara.

Fizemos outras alterações com o objetivo de proteger a liberdade de expressão. Para algumas áreas de políticas, como certos tipos de [conteúdo violento e explícito](#), removemos alguns conteúdos violadores sem aplicar [penalidades](#), que poderiam resultar em [restrições de contas](#) crescentes, a fim de evitar penalizar ou restringir excessivamente os usuários que tentavam conscientizar sobre os impactos do conflito.



Medidas temporárias sobre produtos

Adotamos várias medidas temporárias sobre produtos para ajudar a manter as pessoas seguras e mitigar os principais riscos aos direitos humanos, conforme descrito em nossa publicação no [blog que descreve nossa resposta ao conflito](#). Isso incluiu ajustes nos limites de confiança para a aplicação automática de conteúdo, o bloqueio de certas hashtags nas buscas e mudanças nos produtos para tratamento de comentários indesejados e problemáticos, além de medidas de segurança como a ferramenta de [bloqueio de perfil](#).

Alteramos temporariamente o critério que inviabiliza a recomendação de conteúdos limítrofes ou potencialmente violadores, como imagens ou vídeos que mostram violência gráfica . Essa medida se aplica a conteúdo não conectado , ou seja, conteúdo de pessoas que alguém ainda não escolheu seguir e que pode aparecer em superfícies como Feed, Pesquisar, Explorar e Reels.

Assim como em nossas medidas políticas temporárias, fizemos outras alterações para proteger a liberdade de expressão. Em resposta a um grande pico no uso dos nossos produtos, ajustamos temporariamente alguns limites de taxa automatizados criados para evitar comportamentos definidos como [spam de acordo com nossas políticas](#), como fazer publicações com alta frequência, para torná-los mais permissivos, reduzindo o risco de restrições a usuários legítimos.

Separadamente, nossas políticas mais amplas sobre [recomendação e rebaixamento](#), que *não* fazem parte de nossas ações de resposta a crises, podem afetar a visibilidade de itens de conteúdo relacionados ao conflito, como conteúdos explícitos e violentos.

A Meta conduziu investigações frequentes para avaliar o desempenho dos nossos sistemas de moderação de conteúdo e verificar se eles estavam aplicando as regras de forma excessiva ou insuficiente. Além disso, identificamos áreas para melhorias contínuas, incluindo a priorização da análise de conteúdo em árabe por dialeto e o monitoramento e supervisão aprimorados dos nossos bancos de imagens e vídeos, para mitigar possíveis riscos de aplicação excessiva de regras e os impactos relacionados à liberdade de expressão. Este trabalho está em andamento.

Para informações mais detalhadas sobre nossa resposta, consulte nossas cartas em resposta às consultas recebidas de organizações da sociedade civil [aqui](#) e [aqui](#). Para mais detalhes sobre nossa atualização da auditoria de direitos humanos sobre Israel e Palestina de 2024, clique [aqui](#).



Riscos salientes relevantes:

- privacidade
- liberdade de opinião e expressão
- liberdade de associação e reunião
- igualdade e não discriminação

Criptografia de ponta a ponta: atualização sobre nossas medidas

Em dezembro de 2023, começamos a implementar a criptografia padrão de ponta a ponta em todos os bate-papos e ligações pessoais no Facebook e no Messenger. Isso representou um marco importante em nossa jornada, lançada em 2019, para expandir a proteção de criptografia padrão de ponta a ponta para todos os nossos apps de mensagens. Também continuamos a trabalhar na proteção da criptografia de ponta a ponta para o Instagram Direct. O WhatsApp oferece criptografia padrão de ponta a ponta desde 2016.

A criptografia de ponta a ponta aumenta a segurança e a privacidade, garantindo que ninguém veja sua mensagem, exceto você e a pessoa com quem você está trocando mensagens. Essa é uma camada extra de segurança que mantém suas mensagens e ligações com familiares e amigos protegidas e significa que ninguém,

incluindo a Meta, pode ver ou ouvir o que está sendo enviado ou dito. Se você denunciar uma conversa criptografada de ponta a ponta, as mensagens recentes dessa conversa serão enviadas de forma segura do seu dispositivo para a nossa Equipe de Ajuda.

A estrutura de direitos humanos tem sido fundamental para orientar nosso trabalho de expansão da criptografia de ponta a ponta. A criptografia de ponta a ponta possibilita diretamente o direito à privacidade, o que, por sua vez, permite a realização de uma grande variedade de outros direitos humanos, incluindo a liberdade de expressão, associação, opinião, movimento e segurança física. Ao mesmo tempo, sempre existe o risco de indivíduos abusarem de qualquer serviço de mensagens, incluindo recursos criptografados, para prejudicar outras pessoas.

Em 2022, compartilhamos integralmente os resultados de uma avaliação independente de impacto sobre os direitos humanos, realizada pela BSR, que encomendamos para examinar os possíveis benefícios e riscos para os direitos humanos em nossos planos de expandir a criptografia de ponta a ponta como padrão. A avaliação fez 45 recomendações à Meta para maximizar os benefícios potenciais e mitigar possíveis impactos adversos. Desde nossa resposta inicial a essas recomendações, continuamos trabalhando para implementar nossos compromissos declarados e fornecemos atualizações públicas abrangentes, incluindo em janeiro e agosto de 2023, além de um documento técnico detalhado sobre segurança.



Destaques de problemas

Riscos salientes relevantes:

- privacidade
- vida, liberdade e segurança pessoal

Combate a ameaças de segurança

Nos últimos seis anos, a Meta tem divulgado relatórios públicos sobre nossos esforços para detectar e combater ameaças à segurança em nossas plataformas. As análises e investigações feitas após incidentes dessa natureza contribuíram para que o Citizen Lab e 45 grupos da sociedade civil apresentassem um caso na Cúpula pela Democracia para proibir a venda de spyware até que medidas de proteção sejam implementadas para evitar abusos dos direitos humanos. Depois disso, o presidente dos EUA, Joe Biden, emitiu uma Ordem Executiva para proibir entidades federais dos EUA de usarem spyware para possibilitar abusos dos direitos humanos.

Alguns exemplos de ameaças que mitigamos em 2023 estão relacionados ao setor de vigilância contratada. Trata-se de empresas oferecem serviços para monitorar e rastrear indivíduos ou organizações para diversos fins. Estão sendo levantadas questões sobre a privacidade e o uso indevido da vigilância contratada por governos autoritários, organizações criminosas e outros agentes mal-intencionados que tentam espionar críticos do regime, figuras de oposição, jornalistas e defensores dos direitos humanos para suprimir a liberdade de expressão.

Em 2023, nós identificamos e removemos seis redes distintas de contas no Facebook e Instagram vinculadas a oito empresas do setor de vigilância contratada, localizadas na Itália, Espanha e nos Emirados Árabes Unidos. Entre as táticas adotadas, estão engenharia social e phishing na internet, perfis falsos se passando por manifestantes, jornalistas e mulheres jovens para levar as pessoas a compartilhar emails e números de telefone, bem como clicar em links maliciosos.

"O crescimento da indústria de spyware privado democratiza o acesso a recursos sofisticados de vigilância e, ao mesmo tempo, oculta os clientes que compram e usam esses serviços. Esse monitoramento indiscriminado e pouco transparente representa uma ameaça significativa para os defensores dos direitos humanos. Por isso, estamos comprometidos em trabalhar com parceiros da sociedade civil e com o setor para responsabilizar as empresas de spyware abusivas."

David Agranovich, diretor, Interrupção de ameaças, Meta



Riscos salientes relevantes:

- melhor interesse da criança
- privacidade
- liberdade de opinião e expressão

Segurança de crianças e jovens

A segurança online para crianças e jovens⁴ é um desafio que permeia toda a internet, exigindo esforço e colaboração entre a indústria, o governo e a sociedade civil. Buscamos ajudar a manter os usuários seguros enquanto promovemos um ambiente para o exercício de uma variedade de direitos humanos, incluindo a liberdade de expressão e o acesso à informação, entre outros.

Passamos uma década trabalhando nessa questão e continuamos adotando uma abordagem baseada em direitos para a elaboração de nossos serviços e política de conteúdo. Nossa estrutura em defesa dos interesses das crianças se baseia na Convenção sobre os Direitos da Criança da ONU, bem como em regulamentações e orientações, como o Código de Design Apropriado para a Idade do Reino Unido, os fundamentos para crianças da Comissão Irlandesa de Proteção de Dados e a recomendação sobre menores da *Comissão Nacional de Informática e Liberdades (CNIL)* da França.

Exploração infantil

A segurança infantil é sempre uma prioridade para nós. Buscamos ajudar a prevenir danos por meio da aplicação das nossas políticas e do desenvolvimento de ferramentas de prevenção. Trabalhamos arduamente para tornar as ferramentas de denúncia de exploração infantil mais

fáceis de encontrar, o que resultou em um aumento de 75% no número de denúncias enviadas por adolescentes para nós no primeiro trimestre de 2023 em comparação ao mesmo período de 2022 no Messenger e no Instagram. Também introduzimos formas adicionais e inovadoras de encontrar, remover ou restringir proativamente contas que possam violar nossas políticas relacionadas à segurança infantil. Por exemplo, lançamos uma tecnologia proativa para detectar e desativar automaticamente contas (e quaisquer contas vinculadas a elas) em caso de envolvimento em atividades e interações na plataforma que demonstrem interesse sexual em crianças. Como resultado desse método, mais de 90 mil contas foram identificadas e automaticamente removidas de 1º de agosto a 31 de dezembro de 2023.



⁴ Não existe uma definição universal para o termo "jovem". A Convenção da ONU sobre os Direitos da Criança define "criança" como todo ser humano abaixo da idade de 18 anos (Artigo 1), o que também abrange adolescentes acima de 13 anos. No presente relatório, usamos os termos "jovem" e "adolescente" como sinônimos.



Trabalhamos para detectar e restringir proativamente contas pertencentes a adultos ao primeiro sinal de interações indesejadas com ou sobre adolescentes, ou tentativas de estabelecer contatos potencialmente inseguros.

Como membro fundador do [Take It Down](#), uma ferramenta criada para impedir proativamente a disseminação de imagens íntimas de jovens online, trabalhamos com o [Centro Nacional para Crianças Desaparecidas e Exploradas](#) para expandi-la para mais idiomas e incentivamos mais empresas do setor a participar, resultando no acesso de milhões de adolescentes.

Também somos membros fundadores do [Lantern](#), um novo programa da Tech Coalition lançado em novembro de 2023 que permite que empresas de tecnologia compartilhem sinais sobre contas e comportamentos predatórios em plataformas online. A Business for Social Responsibility realizou uma [avaliação de impacto sobre os direitos humanos](#) desse programa.

"Prevenir a exploração infantil é um dos desafios mais importantes que nosso setor enfrenta hoje. Predadores online são criminosos determinados que usam vários apps e sites para atacar jovens. Eles também testam as defesas de cada plataforma e aprendem a se adaptar rapidamente. Portanto, agora mais do que nunca, estamos trabalhando arduamente para manter nossa posição de liderança. Além de desenvolver tecnologias que identificam predadores, contratamos especialistas dedicados à segurança infantil online e compartilhamos informações com nossos colegas do setor e com as autoridades policiais."

Antigone Davis, diretora global de segurança, Meta



Segurança de jovens e bem-estar

Queremos que os jovens tenham experiências online seguras e apropriadas para a idade, e respeitamos seus direitos digitais. Nossa abordagem se concentra em prevenir danos desde o início, oferecendo às pessoas maneiras de controlar sua experiência e respondendo rapidamente a violações das nossas políticas. Desenvolvemos uma grande variedade de ferramentas e recursos para apoiar os adolescentes e seus pais, mães ou responsáveis. Este trabalho está em andamento.

Em 2023, continuamos a expandir nosso trabalho envolvendo ferramentas de supervisão dos pais em nossos serviços. Realizamos sessões de cocriação com pais, adolescentes, responsáveis e especialistas por meio dos Laboratórios de Confiança, Transparência e Controle (um esforço multi-industrial para colocar as pessoas no controle da privacidade) e trabalhamos com consultores jovens para entender os comportamentos dos adolescentes e orientar o desenvolvimento dos produtos. Também lançamos contas gerenciadas por pais, mães ou responsáveis para pré-adolescentes de 10 a 12 anos no Meta Quest 2 e 3.

Consultamos regularmente especialistas em desenvolvimento adolescente, psicologia e saúde mental para ajudar a manter nossas plataformas seguras e adequadas à faixa etária dos jovens, incluindo o aprimoramento contínuo da nossa compreensão sobre quais tipos de conteúdo podem ser mais apropriados para adolescentes. Foram realizadas consultas com parceiros de segurança, grupos de jovens e consultores de várias partes do mundo, bem como a UNICEF, para informá-los e envolvê-los em nosso trabalho.

Levando em consideração os pontos de vista de especialistas, desenvolvemos novas proteções para proporcionar experiências mais adequadas à idade em nossos apps. Por exemplo, ocultamos mais resultados de pesquisa, incentivamos adolescentes a atualizar suas configurações de privacidade no Instagram e automaticamente ajustamos suas configurações para a opção de controle de conteúdo mais restritiva no Facebook e Instagram.

Envolvimento das partes interessadas

Realizamos sessões de cocriação por meio dos Laboratórios de Confiança, Transparência e Controle para envolver os usuários, nossa comunidade e especialistas com o objetivo de criar experiências online positivas para os jovens, ouvindo suas necessidades e preocupações, e fundamentar a criação dos nossos produtos.

87 adolescentes

71 responsáveis

221 especialistas

13 países



Envolvimento das partes interessadas



O engajamento ativo com uma ampla variedade de partes interessadas é fundamental para nossa abordagem aos direitos humanos na Meta. Isso traz conhecimento externo, perspectivas diversas e feedback para nossos processos de desenvolvimento de produtos e políticas de conteúdo, além de ajudar a garantir responsabilidade e transparência. Os grupos com os quais interagimos variam de organizações internacionais, diversos grupos da sociedade civil e acadêmicos a comunidades marginalizadas e sub-representadas, além de usuários, como adolescentes, pais e cuidadores. Coletamos perspectivas de todos os lados de uma questão e de todas as regiões do mundo para garantir um equilíbrio.

Além disso, continuamos colaborando com nossos parceiros confiáveis para identificar tendências, entender melhor os desafios dos nossos canais de escalonamento e explorar como podemos tornar esses mecanismos mais eficazes para os parceiros da sociedade civil.

O engajamento ativo com uma ampla variedade de partes interessadas é fundamental para nossa abordagem aos direitos humanos na Meta.

Grupos marginalizados e defensores dos direitos humanos

Reconhecemos a importância de um envolvimento significativo com as partes interessadas de comunidades marginalizadas, levando em conta as variações dentro de uma região e até mesmo de um país, e buscamos conhecer essas diferenças.

Estamos comprometidos em ouvir e buscar aconselhamento de especialistas em direitos humanos e ativistas, acadêmicos e outros sobre

uma ampla gama de questões, além de mantê-los informados sobre os desenvolvimentos relevantes na Meta. Os insights coletados orientam áreas como o desenvolvimento de nossas políticas de conteúdo e a aplicação delas para promover o direito de expressão e a segurança.

Em todo o mundo, defensores dos direitos humanos que usam plataformas digitais como instrumentos para seu ativismo frequentemente enfrentam intimidação, opressão e desafios legais como consequência de seus esforços online destinados à promoção e proteção dos direitos humanos. Isso é especialmente perceptível entre grupos marginalizados, como mulheres e a comunidade LGBTQIA+.

Partindo da nossa [estrutura de inclusão](#), realizamos mesas-redondas, workshops e reuniões individuais para entender questões relevantes envolvendo políticas de conteúdo, identificar lacunas nas políticas e buscar oportunidades de participação no desenvolvimento de políticas. Por exemplo, trabalhamos com o [Grupo de Direitos das Minorias](#) para realizar oito encontros globais em países considerados em risco de conflito armado ou violência social.

Também entramos em contato com mais de 250 partes interessadas de diferentes origens, incluindo mulheres, membros da comunidade LGBTQIA+, minorias religiosas, minorias étnicas e grupos indígenas, para coletar insights valiosos sobre os desafios relacionados a políticas enfrentados por essas comunidades. Com base no feedback de mais de 19 consultas e workshops com organizações LGBTQIA+ e defensores dos direitos humanos no Oriente Médio e no Norte da África, colaboramos com a Jordan Open Source Association (JOSA) para lançar o [Kit de Ferramentas de Segurança Digital para Defensores dos Direitos Humanos](#). Este guia, disponível em [árabe](#) e [inglês](#), identifica as melhores práticas em segurança digital, oferecendo recursos de segurança, dicas e medidas reativas para aprimorar a segurança online de ativistas e defensores dos direitos humanos no Oriente Médio e no Norte da África. Também lançamos uma campanha de influenciadores LGBTQIA+ na África Subsaariana para aumentar a conscientização sobre recursos de segurança para ativistas da causa e destacar nosso compromisso em proteger usuários em alto risco na região.

Após um programa piloto bem-sucedido de dois anos concentrado na região da Ásia-Pacífico, em 2023, expandimos nosso [Fundo de Defesa de Direitos Humanos](#), operado em parceria com [defensores de direitos civis](#), para apoiar defensores de direitos humanos globalmente. Além disso, renovamos o financiamento para aplicação de mais USD \$500 mil em 2024.

AfricanDefenders

Com base nos aprendizados e princípios do nosso Fundo de Defensores de Direitos Humanos da [Ásia-Pacífico](#), e em parceria com a [AfricanDefenders](#), lançamos o [Fundo Africano de Defensores de Direitos Humanos](#), projetado para apoiar defensores de direitos humanos que são alvos de assédio, perseguição e/ou processos judiciais devido às suas atividades online em defesa dos direitos humanos. Entre os envolvidos, estão membros de grupos marginalizados que defendem seus direitos, como ativistas e jornalistas cidadãos, ativistas políticos não violentos, defensoras de direitos humanos e a comunidade LGBTQIA+. Os microfinanciamentos serão usados para apoio em situações de emergência, como aquisição de novos dispositivos e tecnologias de segurança, relocação temporária, além de assistência jurídica e de segurança emergencial.



Parceiros confiáveis

Em janeiro de 2023, nossa rede de [parceiros confiáveis](#) incluía mais de 400 ONGs, agências humanitárias e defensores e pesquisadores dos direitos humanos de 113 países. Além do nosso engajamento contínuo, amplo e diversificado com as partes interessadas, esses parceiros trazem uma rica experiência e conhecimento para ajudar a orientar nossa moderação de conteúdo e o desenvolvimento de políticas eficazes e transparentes. Eles também têm acesso a um canal de escalonamento para denunciar conteúdo potencialmente nocivo e relatar problemas de segurança da conta, em relação aos quais tomamos medidas conforme apropriado para ajudar a manter a segurança dos usuários.

Em 2023, as denúncias de parceiros confiáveis nos ajudaram a remover 49.600 itens de conteúdo que violavam nossas políticas. Os parceiros confiáveis também nos ajudaram a descobrir e dismantelar seis redes enganosas envolvidas em [comportamento inautêntico coordenado](#) nos EUA, na Geórgia, em Burkina Faso e no Togo. Além disso, tomamos medidas contra três operações de [espionagem cibernética distintas no Sul da Ásia](#). Essas redes foram criadas para manipular o debate público para um objetivo estratégico, o que pode representar uma ameaça aos direitos humanos.

Em 2023, realizamos uma análise abrangente dos nossos canais para receber conteúdo de emergência e escalonamentos relacionados à segurança de contas de parceiros da sociedade civil. Uma parte importante desse trabalho foi a organização de uma cúpula global que contou com a participação de especialistas em políticas, operações e assuntos jurídicos de todo a Meta, bem como de 15 parceiros da sociedade civil de 13 países da Ásia-Pacífico, do Oriente Médio, da Europa, da América Latina e da África. Por meio desse envolvimento, conseguimos avaliar de forma holística nossos canais de escalonamento da sociedade civil e formular uma visão compartilhada para seu futuro, melhorando a priorização de relatórios e fortalecendo os processos de resposta a emergências e coleta de insights.





Estudo de caso: apurando novos tipos de danos no Haiti

O dia 7 de julho de 2023 marcou o aniversário de dois anos do assassinato do presidente haitiano Jovenel Moïse. As tensões continuam altas no país, exacerbadas pelo ressurgimento da violência das gangues. Nossos parceiros confiáveis denunciaram conteúdo relacionado a vários tipos de violação, como violência e incitação, organizações e indivíduos perigosos, bullying e assédio e violência explícita. Eles também nos alertaram sobre a presença de um novo movimento de vigilantes, o Bwa Kale, em nossas plataformas. Esses insights permitiram que nossas equipes de operações e investigação utilizassem palavras-chave para identificar de forma proativa conteúdos potencialmente violadores, reduzindo danos no meio físico

Estudo de caso: aumentando a divulgação de informações na Etiópia

Diante das crescentes tensões sociopolíticas e a violência contínua na Etiópia, fazemos reuniões mensais com nossos parceiros confiáveis para

discutir tendências de conteúdo e reforçar práticas de denúncia de conteúdo de alta gravidade por meio do Canal de Parceiros Confiáveis. Também realizamos diversos diálogos sobre a Etiópia a cada dois meses com a sociedade civil da Etiópia e da diáspora. Nossos parceiros na Etiópia continuaram a denunciar conteúdos e fornecer avaliações de danos relacionados a desinformação e denúncias de ameaças veladas. Trabalhando com esses parceiros, conseguimos identificar melhor questões importantes, como imagens fora de contexto, contas falsas disseminando desinformação, contas verificadas usando influência para espalhar informações falsas, discurso de ódio, assédio coordenado e doxing, termos com teor político inicialmente criados por figuras públicas, além de organizações e indivíduos perigosos.

Estudo de caso: mitigando conflitos intra-religiosos em Bangladesh

Em março de 2023, ocorreram tumultos intra-religiosos entre muçulmanos sunitas e muçulmanos ahmaditas em Bangladesh, uma comunidade muçulmana minoritária. Nossos

parceiros confiáveis em Bangladesh denunciaram conteúdo incluindo desinformação prejudicial, discurso de ódio, bem como violência e incitação direcionado à comunidade Ahmadi. Com base em reivindicações prejudiciais previamente designadas e analisadas, conseguimos agir rapidamente em relação às denúncias de desinformação. Os parceiros confiáveis também forneceram sinais locais críticos que nos permitiram remover conteúdos de acordo somente com políticas de escalonamento, incluindo ameaças veladas, além de insights sobre termos locais de discurso de ódio.

Organizações internacionais e com várias partes interessadas

À medida que a tecnologia avança e seu impacto nos direitos humanos aumenta, continuamos a fortalecer nosso engajamento com uma grande variedade de órgãos governamentais internacionais, incluindo:

- [O Alto Comissariado das Nações Unidas para os Direitos Humanos](#)
 - [UNESCO](#)
 - [Escritório para a Prevenção do Genocídio e a Responsabilidade de Proteger](#)
 - [Escritório do Enviado do Secretário-Geral para Tecnologia](#)
- [UNICEF](#)
 - [UNHCR](#)
 - [Fórum de Governança da Internet](#)
 - [Fundação das Nações Unidas](#)
 - [OECD](#)
 - [Fórum Econômico Mundial](#)
 - Várias organizações regionais



Somos membros de várias iniciativas envolvendo diferentes partes interessadas, como o [Fórum Global da Internet para Combater o Terrorismo](#), a [Global Network Initiative](#) e a [Freedom Online Coalition Advisory Network](#). Também participamos proativamente de colaborações em todo o setor para fornecer informações e apoiar políticas governamentais.

Participamos da segunda [Cúpula pela Democracia](#) e compartilhamos publicamente compromissos em resposta ao apelo da cúpula ao setor privado para promover a democracia. Fomos reconhecidos pela nossa coliderança no desenvolvimento de princípios do setor que visam "combater as ameaças crescentes representadas por mercenários cibernéticos" e fornecemos contribuições ao apelo para que o setor privado apoie [tecnologias anticensura que defendem a democracia para combater regimes autoritários](#). Além disso, recebemos reconhecimento pelo nosso trabalho para possibilitar conexões seguras e privadas e combater o abuso da tecnologia digital por parte do governo, como o WhatsApp por proxy e a [vigilância contratada](#).

Participamos de consultas sobre o desenvolvimento do [Compromisso Digital Global](#) com o [Escritório do Enviado do Secretário-Geral para Tecnologia](#) para apoiar a necessidade de uma abordagem global harmonizada para a governança da internet e a proteção dos direitos humanos online, incluindo os direitos à privacidade e à liberdade de expressão. Também participamos de consultas sobre o [Código de Conduta para a Integridade da Informação da ONU](#) e enviamos uma proposta detalhada.

Continuamos a fazer reuniões regulares com autoridades do Escritório do Alto Comissário da ONU para os Direitos Humanos, especialistas independentes e relatores especiais para discutir questões globais e específicas de cada país, além de participar ativamente do [Projeto B-Tech](#). Fizemos um compromisso com os direitos humanos em razão do 75º aniversário da Declaração Universal dos Direitos Humanos. Isso envolveu a reafirmação de medidas como a auditoria de direitos humanos, a divulgação e o engajamento com a sociedade civil e comunidades afetadas, entre outros.

A Meta contribuiu com a iniciativa colaborativa [Internet for Trust da UNESCO](#), participando da Cúpula de fevereiro em Paris e do processo de múltiplas partes interessadas para desenvolver diretrizes centradas nos direitos humanos sobre regulamentações de plataformas digitais.

Em colaboração com o [Fórum Econômico Mundial](#), participamos de um extenso projeto para construir consenso sobre políticas de tecnologia emergente. Ajudamos a desenvolver e apoiamos os [Princípios Globais sobre Segurança Digital: Traduzindo os Direitos Humanos Internacionais para o Contexto Digital](#), que estabelece um marco destinado a auxiliar governos e prestadores de serviços online a promover a segurança digital por meio de uma abordagem de múltiplas partes interessadas. Também fazemos parte da [Aliança de Governança de IA](#) e da [iniciativa Definindo e Construindo o Metaverso](#), que busca orientar o desenvolvimento de um metaverso ético

"Não podemos fazer isso sozinhos. Precisamos estabelecer esses compromissos e parcerias estratégicos para que esse projeto tenha algum impacto."

Lene Wendland, Chefe da Unidade de Negócios e Direitos Humanos, Direitos Humanos da ONU

e responsável. No ano passado, aderimos à [Aliança de Governança de IA](#), que tem como objetivo estabelecer padrões globais para a governança da IA, contribuindo positivamente para os avanços tecnológicos em prol do bem-estar social, orientados por compromissos com a ética e a diversidade.

Estabelecemos uma parceria com a [Comissão de Entorpecentes da ONU](#) para criar a Prevent Alliance, uma iniciativa público-privada entre a Meta, a Snap, o governo dos EUA e o Escritório das Nações Unidas sobre Drogas e Crime, com o objetivo de prevenir o uso indevido de plataformas digitais para atividades ilícitas e prejudiciais relacionadas ao uso não médico de drogas sintéticas.

Firmamos um [compromisso](#) com o [Fórum Global sobre Refugiados](#) em apoio aos direitos dos refugiados, solicitantes de asilo e pessoas deslocadas internamente.



Data for Good na Meta

Nosso programa Data for Good segue trabalhando com agências humanitárias, organizações sem fins lucrativos, pesquisadores e governos em todo o mundo para apoiar populações marginalizadas. Por exemplo, esse programa continua a apoiar o trabalho de organizações sem fins lucrativos, como a IMPACT Initiatives, que ajuda a realocar os desabrigados pela guerra na Ucrânia. Ele também oferece assistência a agências internacionais, incluindo a Organização Internacional para Migrações, na melhoria de métricas globais para auxiliar populações migrantes em todo o mundo e aqueles deslocados

em consequência de desastres naturais. Além disso, a Data for Good na Meta contribui significativamente para o trabalho das instituições de desenvolvimento, incluindo o Banco Mundial, por meio de colaborações em larga escala, como a Development Data Partnership, que permite que grandes instituições multilaterais direcionem melhor os programas que atendem a comunidades carentes. Por exemplo, ferramentas baseadas em IA criadas pela Data for Good, como o Índice de Prosperidade Relativa, ajudaram o Banco Mundial a entender melhor como a poluição do ar afeta as populações mais pobres em países de baixa e média renda.





Transparência e reparação



Mantemos o compromisso com a transparência e a reparação como princípios fundamentais no gerenciamento de riscos relacionados aos direitos humanos. Ambos são elementos centrais dos [Princípios Orientadores das Nações Unidas sobre Empresas e Direitos Humanos](#). Nossos [Relatórios de transparência](#), publicados regularmente, e nosso Relatório Anual sobre Direitos Humanos dão visibilidade sobre como aplicamos os nossos [Padrões da Comunidade](#) e as [Diretrizes da Comunidade](#), respondemos a solicitações governamentais e protegemos a propriedade intelectual. Também trabalhamos para melhorar continuamente a acessibilidade aos nossos Padrões da Comunidade, que estavam disponíveis em 90 idiomas no final de 2023. Acreditamos que o Comitê de Supervisão continua fornecendo acesso a reparações que são únicas na indústria de redes sociais.

Resposta a solicitações governamentais

Riscos salientes relevantes:

- liberdade de opinião e expressão
- privacidade
- direito a um julgamento justo
- direito a uma reparação eficaz

Não tomamos medidas em relação a conteúdos ou divulgamos dados de usuários apenas porque uma entidade governamental solicita. Como membros da [Global Network Initiative](#), estamos comprometidos com seus princípios sobre liberdade de expressão e privacidade, que se baseiam em padrões de direitos humanos reconhecidos internacionalmente. Isso é importante porque os princípios nos orientam sobre como responder a solicitações governamentais e minimizar os impactos negativos na liberdade de expressão e na privacidade dos nossos usuários.

Ao longo dos anos, houve uma evolução nas solicitações governamentais. Os atores governamentais solicitam que tomemos diferentes tipos de medidas, que vão além de restringir o acesso a certos conteúdos em um país. Desde 2020, o número de solicitações governamentais aumentou, assim como o volume de conteúdo restrito pela Meta com base em legislações locais.

10 anos com a Global Network Initiative

"Os princípios da GNI orientam o gerenciamento que fazemos de solicitações de remoção governamentais. Essas solicitações para restringir conteúdos com base em legislações locais frequentemente suscitam dilemas significativos e abrangentes relacionados aos direitos humanos. Quando confrontados com conflitos entre leis locais e nossos compromissos com os direitos humanos, buscamos orientação nos princípios da GNI, em padrões internacionais de direitos humanos e em nossa Política Corporativa sobre Direitos Humanos."

Siobhán Cummiskey, diretora de políticas de conteúdo, Meta

[GNI](#) é uma iniciativa envolvendo várias partes interessadas cujos membros se comprometem a defender os direitos à liberdade de expressão e à privacidade, conforme definido em um importante tratado de direitos humanos, o Pacto Internacional sobre Direitos Civis e Políticos (ICCPR). A GNI estabelece um padrão global para a tomada de decisões responsáveis por parte das empresas, visando promover e impulsionar os direitos à liberdade de expressão e à privacidade em todo o ecossistema tecnológico. No contexto dos Princípios Orientadores da ONU sobre Empresas e Direitos Humanos, os Princípios da GNI funcionam como uma ponte entre as obrigações dos governos de proteger os direitos humanos e a responsabilidade das empresas de respeitar esses direitos.

Em 2023, celebramos o 10º aniversário da adesão da Meta à GNI e do nosso compromisso com os direitos humanos. Aplicamos e somos responsáveis pelos Princípios da GNI sobre como as empresas de tecnologia devem respeitar a liberdade de expressão e os direitos de privacidade de seus usuários no atendimento às solicitações e restrições do governo. Somos regularmente avaliados de forma independente quanto à nossa implementação.

As solicitações governamentais incluem restrições a conteúdos e acesso a dados de usuários. Essas solicitações envolvem os direitos à liberdade de expressão e à privacidade, respectivamente. Quando recebemos uma denúncia para restringir conteúdo ou uma solicitação de acesso a dados de usuários de uma entidade governamental ou de um tribunal, ela é analisada seguindo um processo global consistente para ambas as situações ([aqui](#) e [aqui](#)).

- Às vezes, representante do governo fazem solicitações de dados sobre pessoas que usam nossas plataformas como parte de investigações oficiais. Continuamos a aplicar os [Princípios da GNI](#) em nossa resposta a todas as formas de solicitações governamentais. A grande maioria dessas solicitações refere-se a processos criminais. Em muitos desses casos, essas solicitações do governo buscam informações básicas sobre assinantes, como nome, data de registro e duração do serviço. Outras solicitações também podem pedir registros de endereços IP ou o conteúdo da conta. Adotamos [diretrizes](#) rígidas para lidar com todas as solicitações de dados do governo. Relatamos o número de solicitações recebidas, o número de usuários/contas solicitados e a porcentagem de solicitações em que produzimos alguns dados. Desde 2016, temos [divulgado](#) publicamente esses dados, juntamente com nossas políticas de resposta a solicitações.
- Em alguns casos, as autoridades governamentais solicitam que o conteúdo seja removido por supostamente violar a lei local. Também podemos receber uma ordem de um tribunal exigindo a remoção de conteúdo. Se determinarmos que o conteúdo não viola nossas políticas, podemos restringir o acesso a ele no país onde se afirma que é ilegal, após uma cuidadosa avaliação legal e de direitos humanos realizada em conformidade com nossa [Política Corporativa sobre Direitos Humanos](#) e nossos compromissos como membros da GNI. Também informamos o usuário afetado, caso não haja proibição legal para fazê-lo. Consideramos se

devem ser adotadas medidas de mitigação, como tomar providências para limitar o impacto da solicitação, incluindo o cumprimento com ressalva (o que, em alguns casos, levou à revogação de tais solicitações), ou um recurso legal por parte da Meta em relação à ordem de remoção, além da transparência em tempo real por meio de um estudo de caso em nossa [Central de Transparência](#) e/ou o envio de uma cópia da solicitação ao [banco de dados Lumen](#). Nos casos em que acreditamos que as solicitações governamentais ou as ordens de tribunais carecem de validade jurídica, são excessivamente amplas ou inconsistentes com as normas internacionais de direitos humanos, podemos solicitar esclarecimentos ou não tomar nenhuma medida.

- Ocasionalmente, recebemos solicitações ou ordens legais de governos e tribunais para adotar ações inovadoras com base nos requisitos da legislação local, incluindo a aplicação de um aviso de correção em conteúdos (por exemplo, informando que o conteúdo é considerado desinformação pelas autoridades locais), a limitação do acesso de um usuário a recursos específicos, a restauração de conteúdos que foram removidos anteriormente por violar nossas políticas, a obrigação de restringir automaticamente conteúdos em larga escala e em um país, e a restrição global de conteúdos, mesmo que estes violem apenas as leis locais de um único país. Embora respeitemos as leis dos países onde operamos, quando nos deparamos com conflitos entre as obrigações legais nacionais e nossos compromissos de direitos humanos e de transparência, procuramos cumprir ao máximo possível os princípios dos direitos humanos internacionalmente reconhecidos. Podemos solicitar esclarecimentos, contestar a solicitação ou não tomar nenhuma medida.



- Em alguns casos, podemos estar legalmente proibidos de publicar solicitações de remoção ou certas informações contidas nelas. Quando isso acontece, nós procuramos publicar a maior quantidade possível de informações sobre a ordem judicial e a existência dela, sem violar nossas obrigações legais.

Como interpretar denúncias de restrição de conteúdo

Recebemos denúncias de governos, reguladores, tribunais, entidades não governamentais e membros do público envolvendo conteúdos que supostamente violam legislações locais.

O [Relatório de Restrição de Conteúdos](#), publicado semestralmente desde 2013, detalha os casos em que restringimos o acesso a conteúdos no Facebook e no Instagram com base em legislações locais.

- A seção sobre [dados específicos de países](#) fornece detalhes sobre o volume, natureza e tipo de conteúdo que restringimos a cada seis meses com base nas leis locais. Desde 2023, nossos relatórios também incluem informações sobre onde, em países limitados, somos obrigados, com base nas exigências da legislação local, a restringir automaticamente conteúdo em grande escala e no país, ou a aplicar um aviso de correção sobre certo conteúdo em âmbito nacional.
- A seção sobre [restrições globais](#) detalha as ocasiões em que fomos legalmente obrigados a restringir a disponibilidade

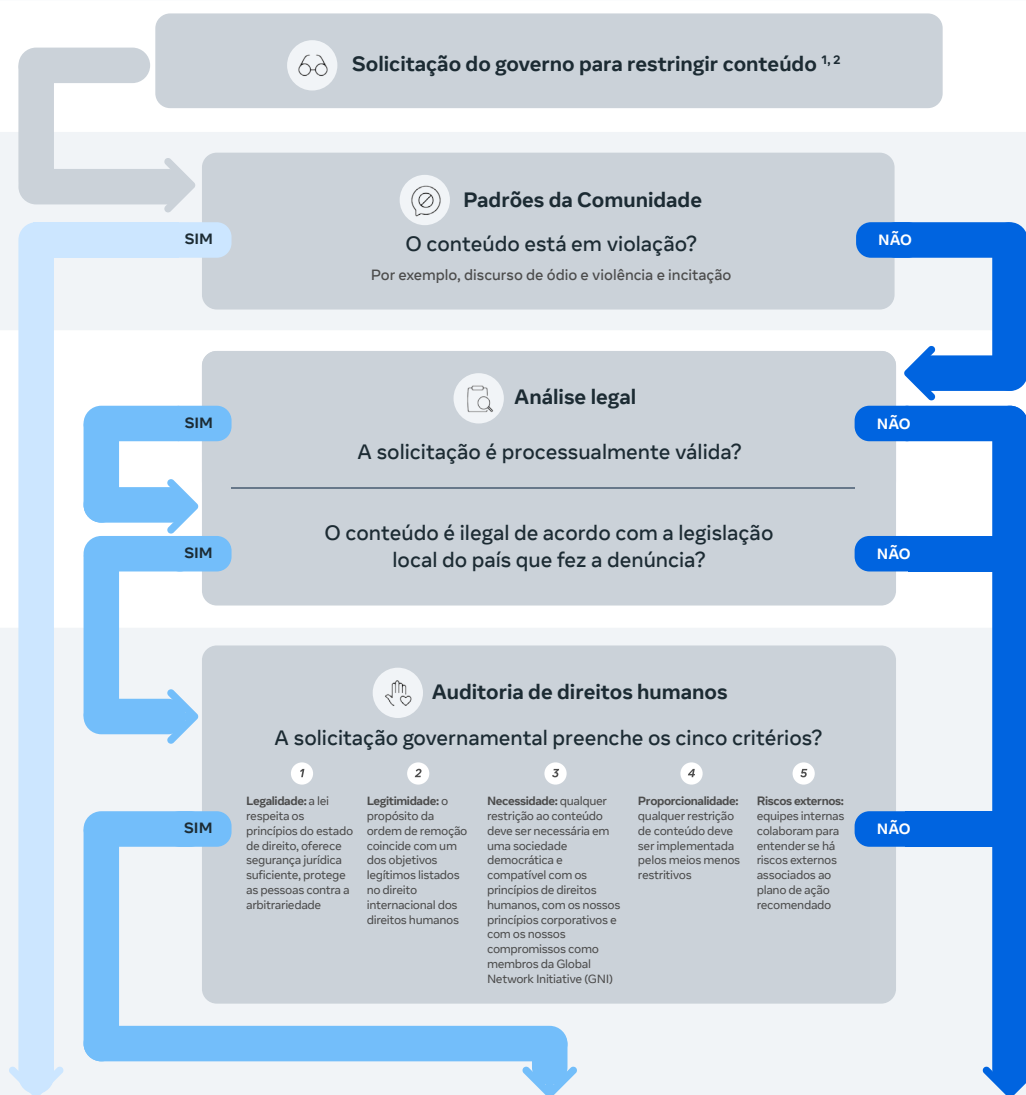
de conteúdo globalmente devido a exigências legais extraterritoriais, embora o conteúdo em questão violasse apenas a legislação local de um país específico.

- A seção sobre [estudos de caso](#) oferece detalhes sobre algumas das solicitações de remoção de conteúdo que recebemos de governos e tribunais. Desde 2023, damos prioridade à publicação exclusiva de estudos de caso de alto interesse público.
- A seção sobre [como avaliamos denúncias de conteúdo que violam legislações locais](#) contém informações sobre nosso processo de análise para solicitações de remoção de conteúdo feitas por governos e as medidas de mitigação que adotamos para limitar o impacto das restrições, em conformidade com os padrões internacionais de direitos humanos.

Publicamos dados sobre restrições de conteúdo com base na lei local para o segundo semestre de 2023 em maio de 2024 (veja [aqui](#)).



Qual é o ciclo de vida de uma solicitação governamental?



Resultado

Podemos implementar medidas de mitigação, como recursos ou o compartilhamento de informações sobre a solicitação governamental para restringir conteúdo externamente, a fim de limitar o impacto da solicitação.



Tomar medidas de acordo com nossas políticas

Quando identificarmos conteúdo que viola os Padrões da Comunidade do Facebook ou as Diretrizes da Comunidade do Instagram, as medidas apropriadas serão tomadas (por exemplo, o conteúdo será removido ou limitado por idade etc.).



Restringir o acesso em jurisdições onde for ilegal

Notificamos as pessoas quando restringimos algo que elas publicaram com base em uma denúncia de que o conteúdo viola a legislação local e também informamos quando elas tentam ver algo restrito no país onde estão em função de uma solicitação governamental para restringir conteúdo. Também notificamos a autoridade governamental relevante de que foram tomadas medidas em resposta à sua solicitação.



Não tomar medidas OU buscar mais informações

Podemos entrar em contato com a autoridade para obter mais informações sobre a solicitação.

¹ Esta ilustração reflete nosso processo de análise e resposta a solicitações governamentais para restringir conteúdo no Facebook e no Instagram, conforme descrito em nossa Central de Transparência. Para obter uma descrição completa, consulte a Central de Transparência (link aqui). Algumas diferenças podem se aplicar a anúncios e outros produtos da Meta.

² Talvez seja necessário desviar do ciclo de vida descrito acima em situações de emergência. Embora incomum, a legislação de um país pode nos obrigar de modo ocasional a restringir automaticamente o acesso a determinados conteúdos (em escala em países específicos), com base nos requisitos das leis locais. Nesses casos, vamos continuar considerando nossos compromissos com a Global Network Initiative e a Política Corporativa sobre Direitos Humanos para orientar nossa abordagem.



Atualização sobre nossa participação no Lumen

Em março de 2022, firmamos o compromisso de [participar do Lumen](#), um projeto de pesquisa independente hospedado pelo Berkman Klein Center for Internet and Society da Universidade de Harvard. O projeto permite que pesquisadores analisem solicitações de remoção de conteúdo feitas por governos e agentes privados em relação a conteúdo online. Em novembro de 2023, enviamos nosso primeiro conjunto de solicitações de remoção da Áustria, do Vietnã, da Singapura, da Índia e do México para o [Lumen](#), sendo possível acessá-lo no respectivo banco de dados. Esse fortalecerá a capacidade da comunidade global de analisar, denunciar e defender os direitos digitais dos usuários da internet. Os princípios que orientam nossa abordagem em relação ao compartilhamento público de informações sobre solicitações de remoção de conteúdo por governos estão detalhados em nosso [Relatório de Restrição de Conteúdos](#).

Melhoria do aviso aos usuários

Para aumentar nossa transparência e em conformidade com os padrões do direito a uma reparação eficaz e ao direito a um julgamento justo, aprimoramos nosso aviso para usuários em 2023 no Facebook e Instagram sobre conteúdos restringidos pela Meta com base em solicitações legais de governos e tribunais. Na maioria dos casos, nossas notificações informam aos usuários qual autoridade estadual enviou a solicitação que resultou na restrição, bem como o país em que o conteúdo foi restringido. Isso não se aplica a conteúdos que removemos devido a violações das nossas políticas.





Riscos salientes relevantes:

- direito a uma reparação eficaz
- direito a um julgamento justo
- liberdade de opinião e expressão
- vida, liberdade e segurança pessoal
- participação pública, direito de votar e de ser eleito
- igualdade e não discriminação
- melhor interesse da criança
- privacidade

Comitê de Supervisão

O Comitê de Supervisão é uma iniciativa líder no setor criada para ajudar a Meta a responder a algumas das perguntas mais desafiadoras sobre liberdade de expressão e segurança online, o que remover, o que deixar ativo e por quê. Trata-se de um órgão independente que pode analisar casos tanto da Meta diretamente quanto de usuários do Facebook, Instagram ou Threads que discordem de nossas decisões de moderação de conteúdo. Além de emitir decisões vinculantes sobre deixar ativo ou remover conteúdo, o Comitê também faz recomendações para aprimorar nossa moderação, assim como emite pareceres consultivos sobre política mediante solicitação nossa. O Comitê de Supervisão aplica padrões globais de direitos humanos para oferecer à Meta insights e recomendações especializadas essenciais, ajudando a garantir que as decisões sobre nossas políticas e produtos estejam alinhadas com os melhores interesses dos usuários.

O ano de 2023 foi emblemático para o Comitê de Supervisão, que superou sua meta de 50 decisões anuais, mais do que triplicando o número de decisões tomadas em 2022.

O Comitê de Supervisão implementou alterações significativas em seu regulamento interno para permitir análises expressas em "circunstâncias excepcionais, inclusive quando o conteúdo pode gerar consequências urgentes no mundo real". Esse processo permite que decisões sobre conteúdos sejam tomadas de forma acelerada, em até 30 dias. O Comitê usou o processo de análise acelerada pela primeira vez em dois casos relacionados ao conflito Israel-Hamas.

As decisões do Comitê abrangeram diferentes questões em diversos países, incluindo o slogan de protesto do Irã "Death to Khamenei" (Morte a Khamenei), o caso do Primeiro-Ministro do Camboja, Hun Sen, temas relacionados a gênero e nudez, prisioneiros de guerra armênios, as eleições no Brasil e protestos de mulheres em Cuba, entre outros. Além de suas decisões, o Comitê fez 60 recomendações à Meta em 2023. Também publicou um parecer consultivo sobre política para a Remoção de desinformações sobre a COVID-19. Para mais informações, veja os relatórios de transparência do Comitê de Supervisão, [aqui](#), [aqui](#) e [aqui](#).



Tolerância a conteúdos interessantes

Em resposta às recomendações do Comitê, fornecemos mais transparência sobre como e quando aplicamos nossa [tolerância a conteúdos interessantes](#) para conteúdos que podem violar nossos [Padrões da Comunidade do Facebook](#) ou as [Diretrizes da Comunidade do Instagram](#), mas que é de interesse público manter visíveis.

- De 1º de junho de 2022 até 1º de julho de 2023, documentamos 69 tolerâncias a conteúdos interessantes.
- Entre as tolerâncias documentadas, nove delas (aproximadamente 13%) foram para publicações de políticos.
- Entre as 69 decisões, documentamos um total de 17 decisões escalonadas, o que significa que cada decisão se aplica a mais de um conteúdo.



Nosso programa de verificação cruzada

Em 2023, também houve uma mudança significativa em nosso programa de verificação cruzada, que começou com nossa resposta inicial às 33 recomendações do Comitê de Supervisão referente a sofisticação técnica, rigor operacional, equidade, governança e transparência do programa. Tomamos medidas para aumentar a participação de parceiros confiáveis e da sociedade civil no desenvolvimento da governança do programa de verificação cruzada, realizando sessões de feedback com parceiros de nove regiões diferentes.



Em 2023, finalizamos 61 recomendações de um total de 122 e continuamos nossos esforços em relação a muitas outras (para mais detalhes, consulte nossas [atualizações semestrais do Comitê de Supervisão](#)). As recomendações que seguimos abrangeram nossas operações, políticas e produtos, contribuindo para melhorias amplas e significativas em toda a empresa e em nossa comunidade global.

Recomendações do Comitê de Supervisão de 2023

Recomendações do
Comitê de Supervisão

66

(91 em 2022)

Avaliação e/ou
implementação da
Meta em andamento*

69

(75 em 2022)

Recomendações
totalmente
implementadas*

61

(14 em 2022)

* Entre as avaliações e/ou implementações em andamento, ou recomendações totalmente implementadas, estão sugestões de anos anteriores (consulte nosso [Relatório sobre Direitos Humanos de 2022](#) para mais detalhes).



Olhar para o futuro

Em 2023, vivenciamos muitos desenvolvimentos rápidos, conforme descrito neste relatório, e várias oportunidades para incluir os direitos humanos nas tomadas de decisão. Buscamos fazer isso. No entanto, os contextos sociopolíticos mais amplos são instáveis, e muitas fontes de dados indicam uma deterioração da liberdade de expressão e das normas democráticas. Tecnologias de inteligência artificial (IA) existem há décadas, mas a conscientização pública sobre as oportunidades e riscos que podem gerar aumentou significativamente em 2023.

O cenário regulatório em rápida evolução oferece tanto promessas quanto desafios para os direitos humanos: pode impulsionar inovações significativas na forma como as empresas aplicam o gerenciamento de riscos relacionados aos direitos humanos, alinhando-se a princípios fundamentais, incluindo os Princípios Orientadores sobre Empresas e Direitos Humanos da ONU. No entanto, as exigências de conformidade também podem limitar o desejo e o espaço para a inovação. Está cada vez mais claro que um gerenciamento sustentável em larga escala de riscos relacionados aos direitos humanos envolve o uso contínuo de dados e ferramentas alinhados aos Princípios Orientadores sobre Empresas e Direitos Humanos da ONU, integrados a processos mais amplos da empresa.

"Estamos rodeados de oportunidades para incorporar questões de direitos humanos nas decisões cotidianas de empresas e governos."

John Ruggie, autor dos Princípios Orientadores das Nações Unidas sobre Empresas e Direitos Humanos

A Meta inovar em suas metodologias de gerenciamento de riscos relacionados aos direitos humanos, incluindo no que diz respeito ao envolvimento com as partes interessadas. Aprendizado experimental, exercícios de simulação e estudos de caso parecem ter um grande potencial, mas ainda não se tornaram comuns. No entanto, buscamos fortalecer nosso trabalho nessas áreas. Também reconhecemos a importância de esforços coordenados e inovações para endereçar os possíveis riscos aos direitos humanos associados à IA generativa, conforme discutido no presente relatório.

Sabemos que ainda temos um longo caminho pela frente. Queremos aproveitar nosso trabalho e as lições aprendidas para contribuir para a evolução do conhecimento e de melhores práticas para o setor de tecnologia e para as comunidades de negócios e direitos humanos.



Estamos em um momento desafiador, mas promissor, para os direitos humanos no mundo da tecnologia. Todos nós, incluindo empresas, sociedade civil, governos, o sistema da ONU, investidores e reguladores, precisaremos trabalhar juntos para desenvolver novas habilidades e conhecimentos, além de novas formas de colaboração, para moldar, compartilhar e aprender como o gerenciamento de riscos aos direitos humanos pode ser aplicado em larga escala.

Novas ondas de imensa criatividade estão se formando: precisaremos de muitas pessoas corajosas, experimentos e intervenções empreendedoras para aproveitá-las. Felizmente, nunca faltou coragem para inovar no movimento pelos direitos humanos. É hora de colocar a mão na massa.



Anexo

Como os direitos humanos são regidos e gerenciados na Meta

Estruturas de gerenciamento e governança bem definidas nos ajudam a promover o respeito aos direitos humanos em nossos programas, serviços e políticas. Nossos especialistas em direitos humanos direcionam a implementação da nossa [Política Corporativa sobre Direitos Humanos](#), que é supervisionada pelo presidente de Assuntos Globais e pelo diretor jurídico.

As atribuições dos especialistas em direitos humanos incluem promover a integração da política nas políticas, programas e serviços existentes e em desenvolvimento, realizar auditorias e oferecer suporte ao treinamento dos funcionários sobre a política. A política orienta a criação de produtos que respeitem os direitos humanos, a resposta a crises emergentes e a atuação com rapidez e agilidade para incorporar os direitos humanos em larga escala.

Nossa Política Corporativa sobre Direitos Humanos nos impõe o compromisso de apresentar relatórios periódicos à Diretoria sobre questões relevantes relacionadas a direitos humanos. O Comitê de Supervisão de Risco e Auditoria da Diretoria supervisiona as diversas exposições a riscos da empresa, inclusive os riscos relacionados a direitos humanos, assim como as medidas que a administração adotou para monitorar ou mitigar essas exposições. O comitê é atualizado periodicamente sobre o trabalho existente e em desenvolvimento realizado pelos especialistas em direitos humanos.

Treinamento dos funcionários da Meta sobre direitos humanos

Na Meta, a forma como criamos é tão importante quanto o que criamos. O treinamento em direitos humanos proporciona aos colaboradores uma compreensão mais abrangente de suas responsabilidades, além do conhecimento e das habilidades necessárias para cumpri-las.

Nosso treinamento *Bigger than Meta: Human Rights* (Maior do que a Meta: direitos humanos) foi lançado em 2022 e continuou ao longo de 2023. O treinamento destaca os impactos potenciais e reais dos nossos serviços, políticas e decisões comerciais sobre os direitos humanos. O objetivo é fomentar uma mentalidade voltada para os direitos humanos em nosso trabalho diário, promovendo o respeito a esses direitos em benefício de todos que usam nossos serviços. O treinamento complementa nossa capacitação em [direitos civis](#), centrada nos princípios da não discriminação, justiça e equidade.

Os objetivos do nosso treinamento em direitos humanos também são reforçados pelo nosso treinamento anual obrigatório em privacidade. A prioridade é aprimorar nossa capacidade coletiva de proteger indivíduos, especialmente aqueles pertencentes a categorias marginalizadas, contra danos resultantes do processamento dos dados das pessoas. Esse treinamento respalda a proteção dos direitos das pessoas à privacidade e à proteção de dados.



Links dos relatórios indicados

- [Relatório de Sustentabilidade de 2024](#)
- [Relatório sobre Práticas Comerciais Responsáveis da Meta de 2024](#)
- [Relatório sobre Direitos Humanos de 2022, Relatório sobre Direitos Humanos de 2021](#)
- [Relatório contra Escravidão e Tráfico de Pessoas de 2023](#)
- [Relatório contra Escravidão e Tráfico de Pessoas de 2024](#)
- [Relatório sobre Minerais de Conflito de 2023](#)
- [Atualização sobre o Progresso dos Direitos Civis de 2023](#)
- [Relatórios de Transparência da Meta](#)
- [Relatórios regulamentares e outros relatórios de transparência](#)
- Avaliações de impacto sobre os direitos humanos: [Criptografia de ponta a ponta](#), [Filipinas](#), [Mianmar](#), [Indonésia](#), [Camboja](#), [Índia](#), [Sri Lanka](#) e [Israel e Palestina](#)

