



Medidas da Meta para salvaguardar a integridade do processo eleitoral do Brasil

15 de agosto de 2024

Este relatório é publicado de acordo com o previsto nas Resoluções nº 23.610/2019 e nº 23.732/2024, do Tribunal Superior Eleitoral do Brasil. O relatório a seguir descreve as medidas da Meta Platforms, Inc. ("Meta") e do WhatsApp LLC ("WhatsApp")¹ para salvaguardar a integridade das eleições brasileiras no Facebook, Instagram, Threads e WhatsApp, antes e durante o período eleitoral de 2024. Informamos que um novo relatório será publicado após a conclusão do período eleitoral.

Ao longo deste relatório, as referências às políticas e práticas da "Meta" referem-se ao Facebook, Instagram e Threads; a discussão sobre as políticas e práticas do WhatsApp será abordada de forma independente. Nos casos em que as políticas e práticas se estendem a toda a empresa, elas serão atribuídas às "Empresas da Meta" como um todo.

Introdução

Proteger a integridade das eleições brasileiras de 2024 em nossos aplicativos é uma prioridade para as Empresas da Meta. Embora cada eleição seja única, nosso trabalho se baseia nas principais lições aprendidas em mais de 200 eleições em todo o mundo desde 2016. Esse aprendizado nos ajuda a concentrar nossas equipes, tecnologias e investimentos para que tenham o maior impacto possível.

Para cada eleição, as Empresas da Meta avaliam se nossas práticas padrão (que incluem as políticas, ferramentas e processos documentados em uma série de postagens na Sala de Imprensa sobre o assunto) abordam as ameaças específicas da eleição em questão. Se houver algum risco fora do comum, as Empresas da Meta trabalham com várias equipes para implementar as medidas adequadas para mitigar esses riscos.

¹ A Meta Platforms, Inc., o WhatsApp LLC e, quando aplicável, o Facebook Serviços Online do Brasil Ltda., podem ser referidos em conjunto como "as Empresas da Meta" ao longo deste relatório.

No Facebook, Instagram e Threads ("Meta", conforme acima), nossa abordagem para as eleições é informada pela experiência e pelas novas estruturas regulatórias em vigor. Na primeira seção deste relatório, descrevemos a abordagem da Meta nos seguintes pilares, levando em consideração a natureza e a viabilidade técnica de cada aplicativo:

- 1) Utilizar e implementar os Padrões e as Diretrizes da Comunidade, bem como os esforços gerais de moderação de conteúdo, para remover conteúdo nocivo e manter as pessoas seguras
- 2) Prevenir a interferência e a desinformação
- 3) Reduzir a disseminação de informações falsas
- 4) Medidas de proteção e transparência relacionadas à publicidade política
- 5) Ferramentas da Meta para apoiar o engajamento cívico
- 6) Abordagem responsável para a IA generativa

No WhatsApp, nossa abordagem é informada de forma semelhante pela experiência e por nossas obrigações de conformidade regulatória, mas reflete as diferentes medidas apropriadas para um aplicativo de mensagens privadas. Na segunda seção deste relatório, descrevemos as medidas implementadas no WhatsApp, com foco nos três pilares a seguir:

- 1) Manter o caráter privado do WhatsApp
- 2) Impedir o uso indevido do WhatsApp
- 3) Empoderar os usuários para combater a desinformação

SEÇÃO 1: META (FACEBOOK, INSTAGRAM, THREADS)

1. Utilizar e implementar os Padrões da Comunidade e as Diretrizes da Comunidade e os esforços gerais de moderação de conteúdo para remover conteúdo nocivo e manter as pessoas seguras

Padrões e Diretrizes da Comunidade relevantes para as eleições:

Os Padrões da Comunidade no Facebook e as Diretrizes da Comunidade no Instagram e no Threads estabelecem regras rígidas sobre o conteúdo que pode ou não ser publicado. Essas políticas abrangem supressão de voto, fraude eleitoral, violência eleitoral e desinformação, entre outras categorias. Essas políticas foram aperfeiçoadas ao longo de

muitos anos, por meio de parcerias com acadêmicos, sociedade civil e verificadores de fatos, almejando o equilíbrio adequado entre a proteção das pessoas e a proteção da liberdade de expressão. Essas políticas são revisadas periodicamente e são disponibilizadas ao público.²

A Meta remove informações falsas quando é provável que elas contribuam diretamente para o risco de danos físicos iminentes. A Meta faz parcerias com especialistas independentes que possuem conhecimento e experiência para avaliar a veracidade do conteúdo e se é provável que ele contribua diretamente para o risco de dano iminente.

A Meta também remove conteúdos que possam contribuir diretamente para o risco de interferir na capacidade das pessoas de participarem das eleições. Por exemplo, a Meta remove informações incorretas sobre datas, locais, horários e métodos de votação, incluindo números incorretos de candidatos; ofertas de compra ou venda de votos; e informações incorretas sobre quem pode votar, qualificações para votar ou se um voto será contado.

A violência política de gênero

As políticas da Meta, inclusive as que tratam de violência e incitação, discurso de ódio, bullying e assédio, cobrem, em conjunto, os riscos decorrentes da violência política de gênero. Por exemplo, a Meta remove a linguagem que incita ou facilita a violência e ameaças reais à segurança pública ou pessoal. Isso inclui discurso violento direcionado a uma pessoa ou grupo de pessoas com base em suas características protegidas, como sexo, identidade de gênero e orientação sexual, e a glorificação da violência baseada em gênero.³

Mídia manipulada

A abordagem da Meta em relação à "Mídia Manipulada" evoluiu e se desenvolveu junto com as mudanças e os avanços tecnológicos. A mídia, incluindo imagem, áudio ou vídeo, pode ser editada de várias maneiras. Em muitos casos, essas alterações são benignas, como um efeito de filtro em uma foto. Em outros casos, a manipulação não é aparente e pode induzir ao erro; por exemplo, conteúdo de vídeo que mostra alguém dizendo ou fazendo algo que não aconteceu. Todo o conteúdo, independentemente de como foi gerado, está sujeito à revisão de acordo com as políticas da Meta e também pode ser revisado por nossa rede de parceiros de verificação de fatos, que se concentram em classificar a desinformação viral. Mais detalhes sobre o programa de verificação de fatos da Meta são fornecidos na Seção 3 deste relatório.

² <https://transparency.meta.com/en-gb/policies/community-standards/misinformation/>

³ <https://transparency.meta.com/policies/community-standards/violence-incitement/>

Embora a Meta não tenha obrigação legal de rotular proativamente a mídia manipulada, em [maio de 2024 a Meta começou a rotular o conteúdo gerado ou editado por Inteligência Artificial \(IA\)](#). A Meta aplica rótulos a vídeos, áudios e imagens gerados por IA com base na detecção, pela Meta, de sinais compartilhados pela indústria ou com base na autodeclaração das pessoas que estão publicando conteúdo gerado por IA. Se a Meta determinar que imagens, vídeos ou áudios criados ou alterados digitalmente representem um risco particularmente alto de enganar o público de maneira relevante sobre uma questão importante, um rótulo mais proeminente poderá ser adicionado para que as pessoas tenham mais informações e contexto. Mais informações sobre isso e sobre nossa abordagem mais ampla em relação à mídia manipulada estão disponíveis na Seção 6 deste relatório.

Aplicando as políticas da Meta por meio da moderação de conteúdo:

Como parte dos esforços para fazer cumprir seus termos de serviço, políticas, padrões e diretrizes – uma prática conhecida mais amplamente como moderação de conteúdo – a Meta usa uma combinação de revisão humana e tecnológica. Todos os dias, a Meta remove milhões de conteúdos e contas que violam as suas regras. Na maioria dos casos, isso acontece automaticamente, usando tecnologia para detectar, restringir e remover conteúdo e contas que sejam contra às políticas da Meta. Em outros casos, nossa tecnologia seleciona o conteúdo para a análise humana. Nossas equipes de análise revisam uma combinação de conteúdo identificado por meio da nossa tecnologia e por meio de denúncias dos usuários. Nossa tecnologia também apoia as equipes de análise, priorizando o conteúdo mais crítico a ser analisado, com base na gravidade, na viralidade e na probabilidade de violação. O conteúdo de alta gravidade é aquele com potencial para causar danos off-line (por exemplo, publicações relacionadas a terrorismo e suicídio) e o conteúdo viral é aquele que está se espalhando rapidamente e tem potencial para atingir um grande público. Nossa tecnologia prioriza esse conteúdo para ser analisado, a fim de evitar danos ao máximo possível.

Centro de Operações Eleitorais:

Para as próximas eleições de 2024 no Brasil, as Empresas da Meta ativarão um Centro de Operações Eleitorais, uma iniciativa que vem sendo implementada para as eleições no país desde 2018. Esse centro irá reunir especialistas de toda a empresa – incluindo as equipes de inteligência, ciência de dados, engenharia, pesquisa, operações, políticas públicas e jurídica. Eles trabalharão juntos para identificar possíveis ameaças em nossas plataformas e colocar em prática as mitigações específicas em tempo real.

Canais locais para denúncias:

No Brasil, além dos canais de denúncia disponíveis globalmente para os usuários, o Facebook Brasil e as Empresas da Meta possuem canais para facilitar o cumprimento de diversas obrigações criadas pela Resolução nº 23.732/2024 do TSE. Em detalhe, o Facebook Brasil firmou uma parceria com o Tribunal Superior Eleitoral (TSE) para definir canais dedicados das Empresas da Meta para que o tribunal denuncie conteúdo e contas suspeitas de envolvimento em abuso eleitoral. Esses canais datam de 2020 e foram reformulados este ano para atender à Resolução do TSE. Além disso, a Meta estabeleceu um canal que permite que usuários, indivíduos e entidades notifiquem às Empresas da Meta conteúdo que eles entendam ser conteúdo eleitoral ilegal. Esse mecanismo está disponível na [Central de Eleições](#), em "Como Denunciar um Conteúdo".

2. Prevenção da interferência e da desinformação

A Meta emprega uma série de medidas para evitar que a interferência eleitoral e a desinformação se manifestem em nossas plataformas. Isso inclui nossa abordagem para evitar o Comportamento Inautêntico Coordenado [*Coordinated Inauthentic Behavior (CIB)*], a remoção de contas falsas e a aplicação proativa de suas regras.

Comportamento Inautêntico Coordenado:

A Meta avançou em suas operações de segurança para derrubar campanhas de manipulação e identificar ameaças emergentes. As equipes de investigadores especializados da Meta procuram ativamente e derrubam redes coordenadas de contas não autênticas que buscam manipular o debate público. A Meta investiga e derruba redes coordenadas de contas, páginas e grupos no Facebook e no Instagram que buscam enganar as pessoas sobre quem está por trás delas e o que estão fazendo. A Meta também trabalha para dar escala aos seus esforços, alimentando os sistemas de detecção automatizados com as informações obtidas na investigação dessas redes globalmente, para ajudar a encontrar agentes mal-intencionados envolvidos nesses e em comportamentos de violação semelhantes, incluindo as redes que tentam voltar depois de terem sido derrubadas.

Em 2022, um grande marco na aplicação das regras foi o fato de a Meta ter interrompido mais de 200 redes em todo o mundo, desde 2017, por violarem nossa política de CIB. Essas redes eram provenientes de 68 países, incluindo o Brasil, e operavam em pelo menos 42 idiomas. A Meta também continua a atualizar sua política de comportamento inautêntico para melhorar sua capacidade de combater novas táticas e agir mais

rapidamente contra o espectro de práticas enganosas - tanto o CIB quanto outras táticas inautênticas (geralmente usadas por atores com motivação financeira) vistas nas plataformas - sejam elas estrangeiras ou nacionais, estatais ou não estatais. O trabalho para remover essas redes ajuda a proteger a integridade das eleições contra essas redes mal-intencionadas.

A Meta continua a informar sobre os esforços para interromper as ameaças adversárias por meio do Relatório da Aplicação dos Padrões [*Community Standards Enforcement Report*] e do Relatório Trimestral de Ameaças Adversárias [*Quarterly Adversarial Threats Report*].⁴

Remoção de contas falsas:

A Meta não permite contas inautênticas ou que se façam passar por outras pessoas, pois elas podem ser veículos para uma série de conteúdos e comportamentos nocivos, incluindo tentativas de interferência eleitoral.

A capacidade de detectar e remover contas falsas vem melhorando ao longo dos anos, e houve um declínio geral no volume de contas falsas encontradas em nossas plataformas desde 2019. Como exemplo, somente em 2023, a Meta detectou e removeu 2,6 bilhões de contas falsas no Facebook, 98,95% das quais foram detectadas e removidas proativamente pela Meta antes mesmo de serem denunciadas.⁵

Aplicação proativa antes das eleições no Brasil:

Parte da abordagem holística da Meta para evitar a interferência nas eleições envolve a aplicação proativa das regras. Isso inclui a utilização das tecnologias pré-existentes da Meta para lidar proativamente com áreas de alta prioridade, como fraudes e golpes, discurso de ódio e organizações e indivíduos perigosos. Todas essas áreas podem ser relevantes durante o período eleitoral e são descritas em mais detalhes abaixo.

Fraudes e golpes

A Meta tem o compromisso de proteger os usuários contra fraudes e anúncios enganosos em nossas plataformas.

Para isso, a Meta realiza voluntariamente varreduras proativas para identificar e remover os conteúdos que violam nossas políticas, como parte de uma estratégia mais ampla baseada em riscos. Isso inclui os anúncios que apresentam figuras políticas, benefícios

⁴ Meta, '[Community Standards Enforcement Report](#)' e Meta, '[Adversarial Threats Report](#)', Meta Transparency Center.

⁵ Meta, '[Community Standards Enforcement Report](#)', Meta Transparency Center.

falsos e alegações fraudulentas. Esses anúncios são projetados para enganar as vítimas e obter seu dinheiro ou informações pessoais, o que é uma violação das nossas políticas.

Para resolver esse problema, a Meta usa uma combinação de automação de alta precisão, ferramentas de ação contra os conteúdos em massa e uma equipe de revisão especializada para o Brasil. Essa equipe de especialistas tem amplo conhecimento de nossos Padrões da Comunidade, Diretrizes da Comunidade e das políticas de anúncios pagos, o que lhes permite revisar e aplicar as políticas com eficácia.

A Meta trabalha continuamente para aprimorar a aplicação voluntária e proativa para ficar à frente daqueles que tentam tirar vantagem de nossas plataformas. Nossos esforços nessa área são contínuos e a Meta se dedica a oferecer um ambiente seguro e protegido para todos os usuários.

Discurso de ódio

A Meta acredita que as pessoas usam sua voz e se conectam mais livremente quando não se sentem atacadas por serem quem são. É por isso que a Meta não permite discursos de ódio, eis que cria um ambiente de intimidação e exclusão e, em alguns casos, capaz de incitar a violência *offline*. A Meta define discurso de ódio como ataques diretos contra pessoas com base no que chamamos de características protegidas: raça, etnia, nacionalidade, deficiência, afiliação religiosa, casta, orientação sexual, sexo, identidade de gênero e doença grave.

Como parte do trabalho contra o discurso de ódio, a Meta removeu 8,6 milhões de conteúdos de discurso de ódio no Instagram somente no primeiro trimestre de 2024, dos quais 98,2% foram encontrados e acionados antes que alguém os denunciasse. Nosso compromisso de oferecer um ambiente seguro e respeitoso para todos os usuários é inabalável, e a Meta continua a trabalhar incansavelmente nesse esforço.

Organizações e indivíduos perigosos

Não há lugar nas plataformas da Meta para conteúdo ou pessoas que promovam a violência, o ódio ou o terrorismo. É prioridade da Meta manter nossas plataformas e comunidades seguras. A Meta leva esse compromisso muito a sério. Como parte de nosso trabalho contra organizações e indivíduos perigosos, a Meta removeu 8,4 milhões de conteúdos terroristas no primeiro trimestre de 2024, dos quais 99,3% foram encontrados e acionados pela Meta antes mesmo que alguém as denunciasse.⁶

A Meta está constantemente atualizando e aprimorando suas tecnologias e técnicas usadas para combater as organizações perigosas. Entre elas estão as investigações para

⁶ Meta, '[Community Standards Enforcement Report](#)', Meta Transparency Center.

identificar grupos de proliferadores de conteúdo violador e a identificação e designação proativa de novas organizações perigosas à medida que elas surgem.

Medidas fora da plataforma: Ações judiciais contra infratores

A Meta entrou [com ações](#) judiciais contra malfeitores no Brasil que estavam vendendo curtidas, seguidores e visualizações, uma prática conhecida como engajamento falso, que viola os Termos de Uso da Meta. Além de entrar com ações judiciais, a Meta desativou contas e enviou notificações extrajudiciais a empresas localizadas no Brasil que ofereciam serviços semelhantes de engajamento falso para o Facebook e o Instagram.

Ao entrar com esses processos e tomar essas medidas no Brasil, a Meta deixa claro que atividades fraudulentas não são toleradas em nossos serviços e que agiremos para proteger a integridade de nossa plataforma. Essas ações fazem parte dos esforços coordenados de contencioso inter-jurisdições da Meta para fazer cumprir seus termos e proteger os usuários, e continuam em andamento.

3. Reduzir a disseminação de informações falsas

A Meta toma várias medidas para reduzir a disseminação de informações falsas. Essas medidas incluem trabalhar em estreita colaboração com nossa rede de verificadores de fatos terceirizados, limitar a distribuição de conteúdo problemático e não permitir anúncios que desencorajem a participação eleitoral ou questionem sua legitimidade.

Verificadores de fatos (3PFCs) no Brasil:

Conforme descrito na Seção 1, a Meta remove o conteúdo que desencoraje a votação ou interfira na votação. Para o conteúdo não removido por violar os Padrões da Comunidade do Facebook ou as Diretrizes da Comunidade do Instagram e do Threads, a Meta também trabalha com organizações independentes de verificação de fatos para verificar a veracidade das publicações denunciadas. Quando os verificadores de fatos classificam um conteúdo, a Meta aplica um rótulo informativo e poderá reduzir seu alcance no Facebook, Instagram e Threads.

O foco do programa de verificação de fatos da Meta é de identificar e lidar com as informações falsas virais, especialmente as fraudes claras que não têm base em fatos. Os parceiros de verificação de fatos priorizam as alegações provavelmente falsas que são oportunas, em tendência e relevantes. A Meta aplica um rótulo de advertência ao conteúdo considerado como informação falsa pelas organizações de verificação de fatos.

Depois que os verificadores de fatos determinam que um conteúdo tem informações falsas, a Meta usa a tecnologia para identificar versões quase idênticas desse conteúdo no Facebook, Instagram e Threads. Usamos a nossa tecnologia para detectar o conteúdo que seja igual ou quase igual ao classificado pelos verificadores de fatos e adicionamos avisos a esse conteúdo também. Geralmente, não adicionamos avisos ao conteúdo que faz uma afirmação semelhante à avaliada pelos verificadores de fatos, se o conteúdo não for idêntico. Isso é porque pequenas diferenças na forma de uma afirmação ser formulada podem alterar o fato de ela ser verdadeira ou falsa.

As pessoas que ainda virem esse conteúdo em seus feeds o verão com um rótulo e um link direcionando-as para páginas com mais informações do verificador de fatos. Os rótulos de verificação de fatos provaram ser altamente eficazes para evitar o compartilhamento de informações que foram verificadas e classificadas como falsas.

A Meta tem seis parceiros 3PFC no Brasil: Agência Lupa, AFP, Aos Fatos, Estadão Verifica, Reuters Fact Check e UOL Confere. A Meta também tem mais dois parceiros (Polígrafo e Observador) que fornecem *fact-checking* em português globalmente. A Meta está em contato com seus parceiros do 3PFC no Brasil e no mundo, a fim de se manter no topo das tendências e ameaças emergentes, especialmente em relação às eleições.

A Meta continuará a investir em parcerias (inclusive com jornalistas, acadêmicos e verificadores independentes de fatos) para ajudar a reduzir a distribuição de notícias e informações falsas, bem como para informar melhor as pessoas sobre o conteúdo que encontram on-line.

Evitar supressão de voto ou informações falsas eleitorais em anúncios:

A Meta está trabalhando constantemente para proteger a integridade das eleições no Facebook e no Instagram. A Meta estabeleceu medidas pelas quais os anúncios relacionados à votação, em períodos de eleições, estão submetidos a proibições adicionais e que podem ser rejeitados se violarem nossas políticas. A Meta não permite anúncios direcionados ao Brasil que desencorajem as pessoas a votarem na eleição, questionem a legitimidade de uma eleição, futura ou em andamento, ou contenham reivindicações prematuras de vitória eleitoral. Essa política se aplica no Brasil desde 2022.

4. Salvaguardas e esforços de transparência relacionados à publicidade política

A Meta tem uma série de medidas em vigor em relação à publicidade política e à garantia de transparência.

A definição de anúncios políticos:

A Meta utiliza uma definição ampla relacionada à publicidade política e nossa política nesse espaço se aplica a todos os "anúncios sobre questões sociais, eleições ou política", de modo que as obrigações de transparência e outros requisitos possam ser impostos a uma ampla categoria de anúncios que possam influenciar o discurso político, inclusive as eleições.

Qualquer anunciante – tanto político quanto não político – que queira veicular anúncios direcionados a pessoas no Brasil, que tratem de um candidato a cargo público, uma figura política, partidos políticos, eleições (incluindo campanhas de "divulgação do voto"), defesa legislativa ou questões sociais (direitos civis e sociais, crime, economia, educação, política ambiental, armas, imigração, saúde, valores políticos e governança, segurança e política externa) precisará passar pelo processo de autorização e colocar rótulos nos anúncios.

A transparência e os rótulos de anúncios políticos:

Para veicular anúncios políticos no Brasil, os anunciantes devem atender a vários requisitos de autenticidade e transparência: um anunciante deve ter a autenticação de dois fatores ativada; apresentar documentos oficiais de identificação brasileiros para confirmar sua identidade; estar localizado no país; ser um administrador de Página ou anunciante na Página a partir da qual pretende veicular anúncios sobre questões sociais, eleições ou política; criar um rótulo "Pago por" ou "Propaganda Eleitoral" que esteja vinculado a uma conta de anúncios usando reais brasileiros; e declarar os anúncios como políticos durante o processo de criação do anúncio, de modo que o anúncio só possa ser direcionado ao Brasil.⁷

Conforme mencionado nos [Padrões de Publicidade](#) da Meta, a Meta aplica as nossas políticas a todos os anunciantes e, como regra geral, os anunciantes não devem se esquivar ou tentar se esquivar do nosso processo de revisão e das ações de aplicação das regras. Os anúncios políticos devem ter um rótulo com o nome e a entidade que pagou pelos anúncios. Se a Meta detectar um anúncio veiculado sem um rótulo, ele será

⁷ Meta, '[Get authorized to run ads about social issues, elections or politics](#)', Meta Business Help Center.

reprovado e adicionado à Biblioteca de anúncios (descrita abaixo), e o anunciante precisará concluir o processo de autorização. Se a Meta descobrir que uma conta de anúncio, Página, conta de usuário ou conta comercial está evitando o processo de revisão e as ações de aplicação das regras, o anunciante poderá enfrentar restrições na publicidade. Além disso, os anunciantes devem usar contas de usuário autênticas para configurar ativos comerciais e veicular anúncios nas tecnologias da Meta, e não devem gerenciar ativos comerciais que estejam conectados a outros ativos comerciais abusivos ou que apresentem comportamento semelhante ao de ativos comerciais que já foram retirados do ar. Com relação a temas especificamente sociais, eleitorais ou políticos, os anunciantes que repetidamente veicularem tais anúncios sem autorização enfrentarão restrições, que poderão resultar em restrições permanentes da capacidade de anunciar.

Além dos rótulos, a Meta oferece mais transparência em relação aos anúncios políticos por meio da [Biblioteca](#) de Anúncios da Meta. A Biblioteca de Anúncios é um site dedicado que permite pesquisar todos os anúncios atualmente sendo veiculados nas plataformas da Meta. Todos os anúncios que estão sendo veiculados atualmente nas tecnologias Meta mostram: o conteúdo do anúncio; as informações básicas como, por exemplo, quando o anúncio começou a ser veiculado e qual anunciante o está veiculando. Com relação aos anúncios sobre questões sociais, eleições ou política, a Biblioteca de Anúncios mostra: o conteúdo do anúncio, as informações básicas – como quando o anúncio começou a ser veiculado e qual anunciante o está veiculando – e transparência adicional sobre gastos, alcance e entidades financiadoras. Esses anúncios permanecem na Biblioteca de Anúncios por sete anos. Os usuários também podem visualizar o valor total gasto com esses anúncios por um anunciante e as opções de segmentação feitas, divididas por período. Para cumprir ainda mais com os novos requisitos estabelecidos pela Resolução nº 23.732/2024, a Meta incluiu as informações adicionais de segmentação de anúncios (e as informações exatas de alcance divididas por idade, gênero e local) na Biblioteca de Anúncios.

Anúncios políticos e a política de divulgação de IA da Meta:

A Meta anunciou, em [novembro de 2023](#), uma política de divulgação de IA para ajudar as pessoas a entenderem quando um anúncio sobre questões sociais, eleições ou política (SIEP) no Facebook ou Instagram foi criado ou alterado digitalmente, inclusive por meio do uso de IA. Essa política entrou em vigor no início de 2024.

Os anunciantes agora precisam divulgar sempre que um anúncio de questão social, eleitoral ou política contiver uma imagem ou vídeo fotorrealista, ou áudio com som realista, que tenha sido criado ou alterado digitalmente:

- Representar uma pessoa real dizendo ou fazendo algo que ela não disse ou fez; ou

- Representar uma pessoa de aparência realista que não existe ou um evento de aparência realista que não aconteceu, ou alterar a filmagem de um evento real que aconteceu; ou
- Retratar um evento realista que supostamente ocorreu, mas que não é uma imagem, vídeo ou gravação de áudio real do evento.

Se os anunciantes não divulgarem esses cenários especificados, o anúncio poderá ser reprovado. A falha repetida na divulgação poderá resultar em outras penalidades para a conta.

5. Ferramentas da Meta para apoiar o engajamento cívico

Uma parte importante da abordagem da Meta na preparação para as eleições, tanto no Brasil quanto no mundo, é o apoio a uma série de ferramentas e medidas relacionadas ao engajamento cívico. Isso abrange desde o fornecimento de informações confiáveis sobre a própria eleição até o apoio ao engajamento e à educação com figuras políticas e reguladores relevantes, bem como o apoio a processos e parcerias sob medida, quando pertinente, para levar em conta o contexto local.

A Meta utilizará uma série de produtos para garantir que as pessoas tenham acesso a informações pertinentes e confiáveis sobre as eleições brasileiras – por exemplo, quando e como votar – incluindo lembretes. Exemplos de produtos cívicos da Meta que estão sendo usados em 2024 para proteger a integridade das eleições brasileiras são:

Lembretes sobre o dia da eleição:

A Meta exibirá lembretes do dia da eleição no Facebook e no Instagram no Brasil. Esse lembrete é exibido no dia da eleição com o objetivo de ajudar as pessoas a obter informações sobre o voto no dia da eleição. Os usuários podem clicar no botão "Saiba mais" para receber mais informações sobre a eleição.

Adesivos eleitorais (Stickers):

A Meta lançará Stickers eleitorais para o Instagram. Esse produto ajuda as pessoas a expressar sua empolgação com a eleição e direcionar seus seguidores para informações confiáveis.

Intersticial de Pesquisa:

A Meta criará intersticiais de pesquisa no Facebook. Esses intersticiais ajudam as pessoas a ter acesso a informações relevantes sobre votação quando pesquisam termos específicos, incluindo, entre outros, "eleição", "cédula", "votação", "urnas", "prefeito" (traduzido no(s) local(is) relevante(s)).

Guia para Políticos sobre a Violência Política de Gênero:

A Meta acredita que as mulheres têm o direito de participar do processo político de forma aberta e livre de violência.

A parceria com o TSE descrita acima também inclui a divulgação de um [guia de combate à violência on-line contra a mulher na política](#), também apoiado pela Women's Democracy Network (WDN) - Capítulo Brasil, Safernet e Secretaria da Mulher da Câmara dos Deputados.

O guia inclui dicas e informações úteis para prevenir e denunciar atos de violência política de gênero no Facebook, Instagram e WhatsApp.

O material oferece exemplos de ações que podem ser consideradas agressivas, como a distribuição de imagens íntimas sem consentimento, assédio, discurso de ódio e outras ameaças. As mulheres também podem encontrar ferramentas oferecidas nos aplicativos que lhes permitem controlar sua experiência, tais como: aviso de comentários, assistente de moderação e ferramenta de restrição de contas.

É importante ressaltar que nossas políticas são vivas: elas estão sendo constantemente aprimoradas e refinadas para oferecer recursos que acompanhem as mudanças on-line e off-line. As Empresas da Meta contam com o apoio de especialistas de alto nível, com conhecimento em segurança global para mulheres, a fim de oferecer um local onde elas possam exercer seus direitos de participar da política sem medo de sofrer violência.

Engajamento com candidatos políticos:

Os representantes da Meta têm contato com todos os partidos políticos registrados, e apoia todos eles igualmente, por meio do canal de concierge - Meta Support Pros.

Além disso, em junho de 2024, a Meta lançou a [Central de Eleições](#) como guia abrangente, com recursos que os candidatos e partidos políticos podem consultar para obter informações precisas e práticas recomendadas para usar as plataformas com segurança.

Por fim, em Agosto de 2024, as Empresas da Meta organizaram um evento on-line que permitiu a candidatos, partidos políticos, coalizões e suas equipes obterem informações sobre ferramentas e recursos, tais como, por exemplo, como proteger suas contas ou como funciona a transparência da publicidade política. Todas as sessões foram gravadas e adicionadas à [Central de Eleições](#).

Assinatura de MEs com o Tribunal Superior Eleitoral:

Desde 2018, durante cada ciclo eleitoral, as Empresas da Meta vêm firmando com sucesso os Memorandos de Entendimento (MEs) com o TSE no Brasil, cujo objetivo é de promover a colaboração com o Tribunal e fomentar várias das iniciativas para combater a desinformação e avançar o processo democrático. [A parceria foi novamente renovada](#) para o ciclo eleitoral de 2024.

6. Abordagem responsável para a IA generativa

A abordagem da Meta em relação à IA responsável é outra forma de protegermos a integridade das eleições em todo o mundo.

Padrões da comunidade, verificação de fatos e rotulagem de IA:

Os Padrões da Comunidade, as Diretrizes da Comunidade e os [Padrões de Publicidade](#) da Meta aplicam-se a todo o conteúdo, inclusive ao conteúdo gerado pela IA. O conteúdo gerado por IA também é elegível para ser revisado e classificado por parceiros de verificação de fatos da Meta, cujas [opções de classificação](#) permitem que eles abordem várias maneiras pelas quais os conteúdos possam enganar as pessoas, incluindo conteúdos criados ou modificados pela IA.

A Meta rotula as imagens fotorrealistas criadas usando a IA da Meta, bem como as imagens geradas por IA do Google, OpenAI, Microsoft, Adobe, Midjourney e Shutterstock que os usuários publicam no Facebook, Instagram e Threads.

A Meta começou a rotular uma gama mais ampla de conteúdos de vídeo, áudio e imagem ao detectarmos indicadores de imagem de IA padrão do setor ou quando os usuários informem que estão carregando conteúdo gerado por IA. A Meta exigirá que as pessoas usem essa ferramenta de divulgação e rotulagem quando publicarem conteúdo orgânico com vídeo fotorrealista ou áudio com som realista que tenha sido criado ou alterado digitalmente, e poderá aplicar as penalidades se não o fizerem. Se a Meta determinar que o conteúdo de imagem, vídeo ou áudio criado ou alterado digitalmente cria um risco particularmente alto de enganar o público, de maneira relevante, sobre um assunto de

importância, poderemos adicionar um rótulo mais proeminente, para que as pessoas tenham mais informações e contexto.

Conforme descrito na Seção 4, os anunciantes que veiculam anúncios relacionados a [questões sociais, eleições ou política](#) com a Meta, também precisam divulgar se estão utilizando uma imagem ou vídeo fotorrealista, ou áudio com som realista, que tenha sido criado ou alterado digitalmente, inclusive com IA, em determinados casos.

Continuando a promover a transparência da IA por meio da colaboração do setor:

A Meta também tem trabalhado com outras empresas do setor de tecnologia com relação a padrões e diretrizes comuns. A Meta é membro da [Partnership on AI](#), por exemplo, e recentemente assinou [o acordo tecnológico](#) criado para combater a disseminação de conteúdo enganoso de IA nas eleições de 2024 no mundo todo.

SEÇÃO 2: WHATSAPP

O WhatsApp é um serviço dedicado a ajudar as pessoas a se expressarem livremente. Nós acreditamos que as instituições democráticas protegem esse direito. Continuamos comprometidos a combater o abuso e, ao mesmo tempo, proteger a privacidade das pessoas que usam nossos serviços. Temos uma equipe dedicada a prevenir abusos, com base em três máximas:

- Manter o caráter privado do WhatsApp
- Impedir o uso indevido do WhatsApp
- Empoderar os usuários para combater a desinformação

O WhatsApp não usa algoritmos para determinar a ordem das mensagens ou das atualizações de canais recebidas por nossos usuários, nem exibe conteúdo em um *feed*. Também não oferecemos recursos para que as pessoas procurem conteúdo em conversas ou grupos de que não participam. Geralmente, as pessoas usam o WhatsApp para se comunicar com quem elas já conhecem, e é necessário saber o número de telefone para enviar mensagens para alguém.

O uso do WhatsApp é regido por seus [Termos de Serviço](#), [Termos de Serviço Suplementares dos Canais do WhatsApp](#), [Termos de Serviço do WhatsApp Business](#) e [Políticas de Negócios](#), bem como por suas Diretrizes [de Mensagens](#) e [Canais](#).

1. Manter o caráter privado do WhatsApp: Limitar a viralidade e reduzir a desinformação



Criptografia de ponta a ponta para conversas pessoais

O WhatsApp usa a criptografia de ponta a ponta por padrão nas conversas pessoais, a fim de proteger essas conversas contra hackers, criminosos e outras ameaças cibernéticas. Isso significa que ninguém, nem mesmo o WhatsApp, pode ler ou ouvir o conteúdo de suas mensagens, ou ligações pessoais. Especialistas em eleições afirmam que fornecer uma segurança rigorosa é fundamental para proteger [as mensagens que as pessoas enviam e recebem](#). Isso inclui discursos políticos e conversas sobre candidatos e suas campanhas.

Limites de encaminhamento

Definimos [um limite](#) para o encaminhamento de mensagens e atualizações de canais para apenas cinco conversas por vez. Isso faz do WhatsApp uma das poucas empresas de tecnologia a restringir intencionalmente o encaminhamento de conteúdo. Quando introduzida, essa limitação reduziu o encaminhamento de mensagens no WhatsApp em mais de 25%.⁸

Limites mais rígidos e etiquetas

Os limites para mensagens encaminhadas com frequência são ainda mais rígidos. Essas mensagens e atualizações de canais são identificadas com a etiqueta . A etiqueta  indica que, é provável que a mensagem pessoal ou atualização do canal não tenha sido criada por um contato próximo e, portanto, ela deve ser analisada com cuidado. Essas mensagens e atualizações só podem ser encaminhadas para uma conversa por vez, medida que reduziu a viralização desse tipo de mensagem em mais de 70%. Além disso, as atualizações de canais encaminhadas sempre direcionam as pessoas de volta ao canal original para que elas tenham acesso ao contexto de origem da atualização.

Garantia de Privacidade

Desenvolvemos recursos de privacidade para dar controle aos usuários. Esses recursos incluem a possibilidade de sair de grupos silenciosamente, sem notificar os demais

⁸ WhatsApp, [Sobre o WhatsApp e as eleições](#), Central de Ajuda do WhatsApp.

membros, a possibilidade de silenciar ligações de números desconhecidos e muito mais. Quando um usuário recebe uma [mensagem de um número de telefone](#) que não está salvo nos seus contatos, o WhatsApp exibe informações sobre a conta, como grupos em comum, para que o usuário decida como deseja prosseguir. Além disso, o [controle de privacidade](#) permite definir configurações para reforçar a segurança da conta. Esse recurso oferece instruções detalhadas sobre configurações de privacidade importantes para ajudar os usuários a escolher o nível certo de proteção, tudo em um só lugar.

2. Impedir o uso abusivo do WhatsApp

Banimento de mensagens em massa

O WhatsApp tem uma tecnologia de ponta para detectar *spam* que funciona 24 horas por dia. Nossa tecnologia identifica contas com comportamentos anormais para impedir que elas sejam utilizadas para espalhar spam ou desinformação. Nosso sistema impede o uso abusivo do WhatsApp antes mesmo que os usuários denunciem essas contas.

Proibição de anúncios

A Meta realiza esforços contínuos para identificar e remover anúncios que oferecem serviços ilegais de disparos em massa no WhatsApp. Esses esforços se estendem por seus aplicativos e superfícies externas, como mecanismos de buscas. No Facebook e no Instagram, a Meta também pode impedir que as contas criem novos anúncios ou até mesmo banir os usuários de suas plataformas.

Ações judiciais contra infratores

O WhatsApp já entrou com ações judiciais no Brasil contra agentes mal intencionados, que violam deliberadamente suas políticas contra *spam* e envio de mensagens em massa. Também foram enviadas notificações extrajudiciais a empresas localizadas no Brasil para que cessassem a oferta e prestação de serviços violadores.

Ao entrar com esses processos e tomar essas medidas no Brasil, o WhatsApp deixa claro que a atividade infratora não é tolerada e que agiremos para proteger a integridade de nossa plataforma. Essas ações fazem parte dos esforços coordenados de contencioso inter-jurisdições do WhatsApp para fazer cumprir seus termos e proteger os usuários, e continuam em andamento.

Prevenção contra o uso abusivo de grupos

Desenvolvemos uma [configuração de privacidade](#) para que os usuários possam decidir quem pode adicioná-los aos grupos. As opções disponíveis são: "Todos", "Meus Contatos" ou "Meus contatos exceto". Essa configuração impede que as pessoas sejam adicionadas contra a sua vontade a grupos que possam ter sido criados para enviar mensagens em massa. Contamos com a tecnologia de aprendizado de máquina para impedir que as contas tentem criar grupos em massa para enviar mensagens aos usuários. Além disso, desenvolvemos ferramentas para ajudar os administradores a gerenciar seus grupos, incluindo a opção de trancar a imagem e o nome do grupo para que os membros não possam modificá-lo, autorizar a entrada de novos membros e restringir as configurações para que somente administradores possam enviar mensagens em um grupo.

Uso político

O envio de mensagens, automatizadas ou não, por candidatos ou partidos políticos sem o consentimento dos usuários pode resultar no banimento de suas contas. No momento, o uso da [Plataforma do WhatsApp Business](#) por candidatos políticos e campanhas políticas não é permitido. Em muitos países, como o Brasil, o WhatsApp realiza ações educacionais para entidades políticas antes das eleições, para enfatizar nosso comprometimento com a segurança. Também enfatizamos a importância de [usar o WhatsApp com responsabilidade](#).

3. Empoderar os usuários para combater a desinformação

Apoio à checagem de fatos

Firmamos uma parceria com a Aliança Internacional de Checagem de Fatos (International Fact-Checking Network - IFCN, na sigla em inglês) para disponibilizar a [checagem de fatos](#) diretamente no WhatsApp, inclusive nos canais. Assim, as pessoas podem verificar a veracidade das informações por trocas de mensagens protegidas com a criptografia de ponta a ponta e as checagens de fatos podem alcançar mais pessoas nos canais. Graças a essa parceria, organizações de checagem de fatos em cerca de 50 países usam o WhatsApp para ajudar a conectar as pessoas com informações confiáveis. No Brasil, é possível interagir com os checadores de fatos nos seguintes números:

- Aos Fatos: +55 21 99747-2441 (<http://wa.me/5521997472441>)
- Estadão Verifica: +55 11 97683-7490 (<http://wa.me/5511976837490>)

- Agência Lupa: +55 21 99193-3751 (<http://wa.me/5521991933751>)
- UOL Confere: +55 11 97684-6049 (<https://wa.me/5511976846049>)
- AFP Checamos: +55 21 98217-2344 (<http://wa.me/5521982172344>)
- Reuters Fact Check Brasil: +55 11 91599-9278 (<http://wa.me/5511915999278>)

Outros parceiros que apoiam a verificação de fatos:

- Projeto Comprova: +55 11 97045-4984 (<http://wa.me/5511970454984>)

Continuamos empenhados em colaborar com as organizações de checagem de fatos e em garantir que elas tenham as ferramentas necessárias para combater a desinformação e assegurar que os eleitores tenham informações confiáveis sobre as eleições.

Bloqueios e denúncias

Ao contrário do que acontece com as mensagens SMS, no WhatsApp os usuários podem [bloquear](#) contas com facilidade. É possível bloquear uma conta diretamente na tela de bloqueio do celular. Além disso, os usuários também podem fazer denúncias ao WhatsApp com facilidade caso recebam mensagens ou atualizações de canais que sejam abusivas, ou caso encontrem contas ou grupos problemáticos. Nós banimos a grande maioria das contas abusivas usando um sistema de detecção automática. Porém, as denúncias feitas por usuários nos ajudam a identificar contas ou canais que enviam mensagens em massa ou fazem uso abusivo do WhatsApp de outras formas.

Pesquisar na Internet

O WhatsApp oferece uma [maneira simples](#) de verificar na internet a veracidade de mensagens pessoais encaminhadas com frequência. Ao tocar ou clicar na lupa exibida ao lado de uma mensagem, os usuários podem carregar essa mensagem em um navegador e encontrar notícias ou outras fontes de informação sobre o conteúdo recebido.

Parcerias e campanhas educativas

O WhatsApp firmou parcerias e lançou colaborações e campanhas educativas em grande escala para ajudar a combater [informações incorretas](#) em diversos países. No Brasil o WhatsApp lançou em 2022 uma ampla campanha contra desinformação, cujo mote era “Vamos juntos combater as informações falsas”. O objetivo da campanha foi dar visibilidade às parcerias estabelecidas para combater a desinformação e aumentar o acesso a informações confiáveis.

Assistente virtual do WhatsApp

Para as eleições de 2020 e 2022 no Brasil, desenvolvemos, em colaboração com o TSE, um assistente virtual no WhatsApp. O chatbot acumulou mais de 4,7 milhões de usuários registrados e permitiu que os eleitores interagissem diretamente com a autoridade eleitoral e recebessem informações relevantes sobre a votação. Alguns Tribunais Regionais Eleitorais (TREs), como [BA](#), [PB](#), [PR](#), [RO](#), [RS](#) and [TO](#) também desenvolveram seus chatbots, aumentando o alcance das informações oficiais sobre as eleições.