



Medidas da Meta para salvaguardar a integridade dos processos eleitorais do Brasil: Revisão pós-eleitoral

3 de dezembro de 2024

Este relatório é publicado de acordo com os requisitos das Resoluções n.ºs 23.610/2019 e 23.732/2024, do Tribunal Superior Eleitoral do Brasil. O relatório a seguir descreve o impacto das medidas da Meta Platforms, Inc. ("Meta") e do WhatsApp LLC ("WhatsApp")¹ para proteger a integridade das eleições do Brasil no Facebook, Instagram, Threads e WhatsApp antes e durante o período eleitoral de 2024, de acordo com o Art. 9(D). Este relatório é uma continuação do nosso [relatório pré-eleitoral](#), publicado em 15 de agosto de 2024, que descreveu de maneira detalhada as medidas planejadas.

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

Introdução

Proteger a integridade das eleições brasileiras de 2024 em nossos aplicativos era uma prioridade para as Empresas da Meta. Ao longo de muitos anos, a Meta desenvolveu uma abordagem abrangente para as eleições em suas plataformas. Descrevemos essa abordagem em detalhes, e sua relevância, em particular, para as eleições de 2024 no Brasil, em nosso [relatório](#) público [pré-eleitoral](#) disponível em nosso Centro de Transparência.² Remetemos os leitores ao nosso relatório pré-eleitoral para obter um relato completo de nossos esforços para identificar e mitigar os riscos mais urgentes das

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 Meta" ao longo deste relatório.

² Meta, "Medidas da Meta para salvaguardar a integridade do processo eleitoral do Brasil", Meta Transparency Centre, <https://transparency.meta.com/reports/regulatory-transparency-reports/>.

plataformas, apresentados pelas eleições no Brasil e, abaixo, fornecemos uma visão geral resumida das ações adotadas e do impacto de nossos esforços durante esse período.

No Facebook, Instagram e Threads ("Meta", conforme acima), descrevemos essas informações nos seguintes pilares:

- 1) Aplicando os Padrões da Comunidade e os esforços gerais de moderação de conteúdo para remover conteúdo prejudicial e manter as pessoas seguras;
- 2) Nossos processos de gerenciamento de riscos eleitorais;
- 3) Prevenir a interferência e as informações falsas;
- 4) Reduzir a disseminação de informações falsas;
- 5) Medidas de proteção e transparência relacionadas à publicidade eleitoral;
- 6) Ferramentas da Meta para apoiar o engajamento cívico;
- 7) Abordagem responsável para IA Generativa.

Na segunda seção deste relatório, descrevemos essas informações para o WhatsApp, com foco nos três pilares a seguir:

- 1) Manter a natureza privada do WhatsApp;
- 2) Como evitar o uso coordenado indevido do WhatsApp;
- 3) Capacitar os usuários para combater informações falsas.

A menos que especificado de outra forma, os dados fornecidos neste documento abrangem o período entre os dias 1º de agosto e 31 de outubro de 2024, incluindo três meses em torno do período eleitoral.

SEÇÃO 1: META (FACEBOOK, INSTAGRAM, THREADS)

1. Aplicando os padrões da comunidade e os esforços gerais de moderação de conteúdo para remover conteúdo nocivo e manter as pessoas seguras

Nossos [Padrões da Comunidade](#) estabelecem regras rígidas sobre o conteúdo que pode e não pode ser publicado no Facebook, Instagram e Threads.³ Além disso, [os Padrões de](#)

³ Em 12 de novembro de 2024, a Meta lançou os Padrões da Comunidade unificados, reunindo o que antes eram nossos Padrões da Comunidade para o Facebook e nossas Diretrizes da Comunidade para o Instagram e o Threads em um único conjunto de políticas. As regras em si não foram alteradas como resultado dessa unificação. Para obter mais informações, consulte: <https://transparency.meta.com/community-standards-unified>.

[Publicidade da Meta](#) fornecem detalhes da política e orientações sobre os tipos de conteúdo de anúncios que permitimos e os tipos de conteúdo de anúncios que proibimos. Remetemos os leitores ao nosso relatório pré-eleitoral para obter mais detalhes sobre essas políticas e sua relevância no contexto das eleições, bem como nossas práticas de aplicação de moderação de conteúdo.

Aplicando as políticas da Meta em torno das eleições do Brasil:

Conforme demonstrado na tabela a seguir, nos meses em que inserido o período eleitoral, no Facebook, no Instagram e no Threads, a Meta removeu, no Brasil, mais de 2,9 milhões de conteúdos que violaram as nossas políticas de *Bullying* e Assédio, Discurso de Ódio e Violência e Incitação. A grande maioria do conteúdo violador foi removida antes mesmo de ser denunciada pelos usuários.

Ações de remoção:

"Ações de remoção" refere-se ao número de instâncias de conteúdo não pago (como postagens, fotos, vídeos ou comentários) que removemos por violar nossas políticas.

As tabelas abaixo mostram as Ações de Remoção de conteúdo no Brasil no Facebook, Instagram e Threads, abrangendo o período entre os dias 1º de agosto e 31 de outubro de 2024.

| | Ações de remoção para Facebook, Instagram e Threads combinados |
|---|--|
| Ações de remoção de conteúdo de <i>Bullying</i> e Assédio | Mais de 520.000 |
| Ações de remoção de conteúdo de Discurso de Ódio | Mais de 550.000 |
| Ações de remoção de conteúdo de Violência e Incitação | Mais de 1.900.000 |

Taxa proativa:

"[Taxa Proativa](#)" mostra a porcentagem de todos os conteúdos não pagos que encontramos e agimos antes que os usuários os denunciassem.

A tabela abaixo mostra a Taxa Proativa para conteúdos no Brasil, abrangendo o período entre os dias 1º de agosto e 31 de outubro de 2024.

| Política | Taxa proativa para Facebook, Instagram e Threads combinados |
|---|---|
| Padrões da Comunidade sobre <i>Bullying</i> e Assédio | Mais de 95% |
| Padrões da Comunidade sobre Discurso de Ódio | Mais de 99% |
| Padrões da Comunidade sobre Violência e Incitamento | Mais de 99% |

Restrições de conteúdo com base na legislação local:

Canais locais para denúncias:

Conforme descrito no relatório pré-eleitoral, 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.

Como resultado das denúncias enviadas pelo TSE por meio de seu portal e também por indivíduos (usuários ou não) ou entidades no canal de denúncias criado pela Meta, tomamos medidas contra menos de 100 conteúdos e contas no Facebook, Instagram e Threads por violarem as nossas políticas ou a lei eleitoral local. Além disso, encaminhamos para agências de verificação de fatos menos de 10 conteúdos de informações supostamente falsas que não violavam nossas políticas ou a lei eleitoral local.

Contencioso eleitoral

De 1º de agosto a 31 de outubro, os juízos eleitorais dos 5.569 municípios brasileiros proferiram mais de 3.000 decisões judiciais determinando a restrição de conteúdos do Facebook, Instagram ou Threads. Essas decisões foram proferidas em mais de 5.200 processos de natureza eleitoral nos quais as Empresas da Meta foram notificadas como terceiras interessadas ou incluídas como partes.

2. Nossos processos de gerenciamento de riscos eleitorais

Um ano e meio antes das Eleições Municipais do Brasil, a equipe responsável por conduzir os esforços de integridade eleitoral entre as empresas da Meta trabalhou ao lado de um grupo multidisciplinar de especialistas para identificar e desenvolver estratégias de

mitigação para riscos relacionados às eleições na plataforma. Isso inclui inteligência, ciência de dados, engenharia, pesquisa, operações, políticas públicas e equipes jurídicas.

Essa equipe multidisciplinar executa estratégias proativas de mitigação de riscos em coordenação contínua, intensificando as ações à medida que o dia da eleição se aproxima e respondendo a questões à medida que surgem em determinadas áreas problemáticas relevantes para as eleições, identificadas por meio de um processo de avaliação de riscos, incluindo riscos cívicos, discurso hostil, informações falsas, ameaças adversariais e publicidade eleitoral.

Para as Eleições Municipais de 2024 no Brasil, as equipes de integridade da Meta ativaram um Centro de Operações Eleitorais antes dos finais de semana de votação, reunindo especialistas de toda a empresa para identificar possíveis ameaças e implementar mitigações específicas em nossos aplicativos e tecnologias em tempo real. Alguns dos riscos que o Centro de Operações Eleitorais gerenciou ao longo desse ciclo eleitoral foram especialmente focados no combate a ameaças adversariais, como [interferência aos eleitores](#), [fraudes, golpes e práticas enganosas](#), e anúncios não declarados sobre [questões sociais, eleições ou política](#), incluindo conteúdo violador criado por meio da IA Generativa.

Embora algumas dessas políticas, ferramentas e processos tenham sido implementados ao longo de muitos anos com base em experiências anteriores, aprimoramos continuamente essas medidas para garantir que continuem a responder aos riscos à medida que surgem, além de levar em consideração o *feedback* de terceiros interessados.

3. Prevenção da interferência e das informações falsas

A Meta emprega uma série de medidas para evitar que a interferência eleitoral e as informações falsas se manifestem em nossas plataformas. Isso inclui nossa abordagem para evitar o *Coordinated Inauthentic Behavior* (CIB) (Comportamento Inautêntico Coordenado), a remoção de contas falsas e a aplicação proativa das regras, conforme descrito em nosso relatório pré-eleitoral. A seguir, apresentamos outras conclusões relevantes para o período eleitoral.

Comportamento inautêntico coordenado:

Conforme detalhado em nosso relatório pré-eleitoral, a Meta avançou em suas operações de segurança para derrubar campanhas de manipulação e identificar ameaças emergentes. Nossa equipe de investigadores especializados procura ativamente e

derruba redes coordenadas de contas não autênticas que buscam manipular o debate público. A Meta investiga e elimina redes adversárias 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. O trabalho para remover essas redes ajuda a proteger a integridade das eleições contra essas redes adversárias.

A Meta continua a informar sobre os esforços para interromper as ameaças adversárias por meio do Relatório de Aplicação dos Padrões de Comunidade e do Relatório Trimestral de Ameaças Adversárias.⁴ Durante o período eleitoral, não foram identificadas campanhas de CIB direcionadas ao Brasil.

Remoção de contas falsas e falsificação de identidade:

Com nosso compromisso com a autenticidade, não permitimos que as pessoas falseiem as suas identidades, usando contas falsas, fazendo-se passar por outras pessoas ou se envolvendo em comportamentos destinados a permitir outros danos, inclusive 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. 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.

Para as eleições municipais de 2024 no Brasil, nossas equipes concentraram esforços no monitoramento de instâncias de falsificação de identidade direcionadas a figuras públicas, políticos eleitos e candidatos que concorrem às eleições, bem como na aplicação de camadas adicionais de proteção a suas contas contra comentários de discurso hostil e tentativas de invasão de contas.

Aplicação proativa das regras antes das eleições no Brasil:

Antes e durante esse ciclo eleitoral, a Meta adotou proativamente uma abordagem holística para evitar ameaças e danos de interferência em relação a fraudes e golpes, discurso de ódio, contas falsas e organizações e indivíduos perigosos. Para isso, a Meta utilizou tecnologias e protocolos pré-existentes para tratar de questões de alta prioridade nessas áreas durante todo o período eleitoral.

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

4. Reduzir a disseminação de informações falsas

A Meta tem o compromisso de impedir a disseminação de informações falsas durante e fora dos períodos eleitorais. Os esforços da Meta incluem parcerias com nossa rede de verificadores independentes de fatos de terceirizados ("3PFCs"), limitando a distribuição de conteúdo de acordo com nossas diretrizes de distribuição de conteúdo e políticas de recomendações, e não permitindo anúncios que desencorajem a participação eleitoral ou questionem sua legitimidade. Consulte nosso relatório pré-eleitoral para obter uma explicação detalhada dessas medidas e parcerias.

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

Aplicação das políticas nas eleições do Brasil:

Conforme mostrado nas seções a seguir, como resultado de nossas políticas e medidas sobre informações falsas, rotulamos mais de 8,2 milhões de conteúdos no Facebook e no Instagram com verificações de fatos no mês que antecedeu e incluiu o período eleitoral. Em média, 57% das pessoas que começaram a compartilhar conteúdos verificados não concluíram essa ação após receberem um aviso da Meta de que os conteúdos haviam sido verificados, demonstrando o impacto dos esforços de rotulagem na redução da disseminação de informações falsas em ambas as plataformas. Além disso, removemos mais de 43.000 anúncios por violarem nossas políticas de informações falsas.

Conteúdo tratado com rótulos de informações falsas:

A tabela abaixo mostra informações sobre os conteúdos visualizados por usuários no Brasil que foram tratados com rótulos de informação falsa no Facebook e no Instagram, entre os dias 1º de agosto e 31 de outubro de 2024, bem como tentativas de compartilhamento.

| | Facebook e Instagram juntos |
|--|-----------------------------|
| Conteúdos tratados com verificações de fatos após ser classificado pelos 3PFCs | Mais de 8.200.000 |

| | |
|---|-----|
| % de tentativas de compartilhamento que não foram concluídas com conteúdos tratados | 57% |
|---|-----|

Anúncios removidos por violações da política sobre informações falsas relativas às eleições no Brasil:

Entre os dias 1º de agosto e 31 de outubro de 2024, mais de 43.000 anúncios foram removidos por violações da política sobre informações falsas no Facebook e no Instagram, onde a localização inferida do anunciante no momento da ação era o Brasil. Essa contagem inclui o número de anúncios rejeitados durante nosso processo de revisão de anúncios antes de serem publicados, bem como anúncios publicados que foram removidos por violarem nossas políticas de informações falsas.

5. Salvaguardas e esforços de transparência relacionados à publicidade eleitoral

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

Conforme descrito no relatório pré-eleitoral, a Meta utiliza uma definição ampla de publicidade eleitoral e nossa política nesse espaço se aplica a todos os "anúncios sobre questões sociais, eleições ou política" (anúncios "SIEP"), 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. Para veicular anúncios eleitorais no Brasil, os anunciantes devem atender a vários requisitos de autenticidade e transparência, incluindo o uso de rótulos [*disclaimers*]. Se a Meta detectar um anúncio veiculado sem um rótulo, ele será reprovado e adicionado à Biblioteca de anúncios, e o anunciante precisará concluir o processo de autorização. Além disso, nossa política de divulgação de IA exige que os anunciantes divulguem sempre que um anúncio do tipo SIEP contenha uma imagem ou um vídeo fotorrealista ou um áudio realista que tenha sido criado ou alterado digitalmente, inclusive por meio do uso de IA, em determinados cenários específicos. Remetemos os leitores ao relatório pré-eleitoral para obter mais informações e detalhes sobre essas políticas e sua aplicação.

Pausas nos anúncios do tipo SIEP pelo TSE

Como parte de nossas obrigações regulatórias, a Meta preparou e implementou períodos de pausa nos anúncios, impedindo que candidatos e partidos veiculassem anúncios do tipo SIEP no Brasil com base nos períodos designados pelo TSE. Essa medida foi aplicada em ambos os turnos de votação, entre os dias 4 e 7 de outubro e 25 e 28 de outubro.

Durante esses períodos, a Meta monitorou os anúncios veiculados no Brasil, por meio de detecção automatizada e revisão humana, em busca de anúncios que possam ter sido criados em violação à nossa política de anúncios SIEP descrita acima e que burlavam os períodos de pausa dos anúncios. Como resultado, identificamos cerca de 43.000 anúncios do tipo SIEP não declarados, nos quais os usuários não aplicaram o aviso de responsabilidade exigido no momento da criação. Esses anúncios foram controlados de acordo com a nossa política de anúncios SIEP (na maioria dos casos, esses anúncios foram rejeitados antes de entrarem no ar), reforçando a conformidade com a resolução do TSE.

Anúncios do tipo SIEP em torno das eleições

A tabela abaixo mostra informações sobre o número de anúncios aceitos e veiculados com avisos de responsabilidade SIEP, bem como o número de anúncios removidos por não conformidade com a política SIEP da Meta, entre os dias 1º de agosto e 31 de outubro de 2024, em que a localização inferida do anunciante no momento da aplicação da medida foi o Brasil. Isso reflete a aplicação das políticas e medidas mencionadas acima.

| | |
|--|-------------------|
| Número de anúncios SIEP aceitos e rotulados no Facebook e no Instagram juntos | Mais de 1.700.000 |
| Número de anúncios SIEP removidos por não estarem em conformidade com nossa política de anúncios SIEP no Facebook e no Instagram juntos | Mais de 420.000 |

6. Ferramentas da Meta para apoiar o engajamento cívico

Um fator importante da abordagem da Meta na preparação para as eleições em nível global, bem como no Brasil, 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 desenvolvidas sob medida, quando aplicável, para levar em conta o contexto local.

Lembretes sobre o registro de eleitores e o dia da eleição:

Continuamos focados em fornecer aos usuários informações eleitorais confiáveis e, ao mesmo tempo, combater informações falsas em todos os idiomas. É por isso que continuamos a conectar as pessoas com detalhes sobre as eleições por meio de notificações no aplicativo, quando permitido por lei.

Indicamos aos usuários, de forma proativa, informações confiáveis sobre o processo eleitoral por meio de avisos no aplicativo "Registro de eleitores" e "Lembretes do dia da eleição". Esses são avisos na parte superior do feed no Facebook e no Instagram, lembrando as pessoas do dia em que podem votar e redirecionando-as para fontes locais confiáveis sobre como e onde votar. Para as eleições no Brasil, o aviso "Registro de eleitores" foi lançado em 28 de abril e 3 de maio, e o recurso "Informações sobre o dia da eleição" foi exibido em 6 de outubro, de acordo com o procedimento de votação local.

Os usuários no Brasil interagiram com essas notificações no aplicativo cerca de 9,7 milhões de vezes no Facebook e no Instagram. A tabela abaixo mostra o alcance que essas notificações no aplicativo tiveram com os usuários no Brasil:

| | Cliques no Facebook e no Instagram |
|----------------------------------|------------------------------------|
| Registro de eleitores | 6,1 milhões |
| Lembretes sobre o dia da eleição | 3,6 milhões |

Alerta de votação:

Um alerta de votação é uma notificação no Facebook, que vem diretamente de uma página governamental autorizada, destinada a notificar os usuários sobre atualizações importantes referentes às eleições ou votações, como locais, datas e horários de votação, métodos ou requisitos de elegibilidade. No Brasil, a Página do TSE no Facebook (<https://www.facebook.com/TSEJus>) está autorizada a compartilhar um alerta de votação. Para esse ciclo eleitoral, o TSE publicou um alerta de votação em sua Página do Facebook em 3 de outubro de 2024 e recebeu 8,7 milhões de cliques.

Stickers eleitorais:

A Meta lançou adesivos eleitorais – *stickers* – para o Instagram do dia 30 de setembro ao dia 6 de outubro e, depois, do dia 21 ao dia 27 de outubro de 2024. Esse produto ajuda as pessoas a expressar sua empolgação com a eleição e direcionar seus seguidores para informações confiáveis.

Intersticiais de pesquisa:

A Meta criou intersticiais de pesquisa no Facebook. Esses intersticiais ajudam as pessoas a obter acesso a informações pertinentes sobre a votação ao pesquisarem termos específicos, incluindo, entre outros, "eleição", "cédula", "votação", "máquina de votação", "prefeito" (traduzidos no(s) local(is) adequado(s)). Esses intersticiais foram disponibilizados entre os dias 30 de setembro e 29 de outubro de 2024.

Guia para Políticos sobre Violência Política Baseada em Gênero:

As Empresas da Meta acreditam 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 cartilhas com informações para a comunidade eleitoral e um guia de combate à violência on-line contra mulheres na política, também apoiado pela Women's Democracy Network (WDN) - Unidade Brasil, a Safernet e a Secretaria da Mulher da Câmara dos Deputados. O guia inclui dicas e informações úteis para prevenir e denunciar atos de violência política baseada em gênero no Facebook, Instagram e WhatsApp.⁵

Engajamento com candidatos políticos:

Os representantes da Meta têm contato com todos os partidos políticos registrados, e o suporte é dado a todos eles igualmente por meio do canal de especialistas – *Meta Support Pros*. De meados de agosto até o início de novembro, os especialistas da Meta forneceram suporte técnico e treinamento a mais de 20.000 casos de candidatos, partidos políticos e suas equipes.

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

⁵ Disponível em <https://www.facebook.com/government-nonprofits/2024-brasil-elections>.

recebessem informações e educação iguais e consistentes sobre as tecnologias da Meta, por meio de conteúdo, apresentações e sessões de treinamento gravadas.

Por fim, em agosto de 2024, as Empresas da Meta organizaram um encontro *online* – o *web Summit* – que possibilitou a candidatos, partidos políticos, coligações e suas equipes aprenderem sobre ferramentas e recursos, como, por exemplo, como proteger suas contas ou como funciona a transparência da publicidade eleitoral. Todas as sessões foram gravadas e adicionadas à Central Eleitoral. Esse *summit* contou com a presença de mais de 1.900 participantes.

Campanha de educação pública:

As Empresas da Meta lançaram uma campanha educativa no Brasil, entre os dias 20 de agosto e 30 de setembro de 2024, para aumentar a conscientização sobre as iniciativas da Meta e do WhatsApp para proteger a integridade das eleições de 2024 no Brasil. A campanha enfatizou o compromisso das empresas em fornecer ferramentas e recursos para combater informações falsas e destacou as principais colaborações e parceiros nesse esforço.

A campanha incluiu anúncios de alta visibilidade em Brasília, São Paulo e no Rio de Janeiro, incluindo anúncios impressos em jornais e revistas, anúncios de rádio, anúncios de display e vídeo, anúncios de mídia social (Instagram, Facebook, LinkedIn) e outdoors, direcionando o público para recursos eleitorais no [site da Meta](#).

Execução de MOUs com o Tribunal Superior Eleitoral:

Desde 2018, durante todos os ciclos eleitorais, as Empresas da Meta vêm executando, com sucesso, Memorandos de Entendimento (MOUs) com o TSE no Brasil, com o objetivo de promover a colaboração com o Tribunal e fomentar uma variedade de iniciativas para combater informações falsas e avançar o processo democrático. O MOU foi renovado com sucesso para o ciclo eleitoral de 2024.

7. Abordagem responsável para IA generativa (GenAI)

A abordagem da Meta em relação à IA responsável é outra forma de protegermos a integridade das eleições no mundo inteiro.

Conforme descrito no relatório pré-eleitoral, os Padrões da Comunidade e [os Padrões de Publicidade](#) da Meta se aplicam a todo o conteúdo, inclusive ao conteúdo gerado pela IA. O conteúdo gerado pela IA também pode ser revisado e classificado pelos parceiros de

verificação de fatos terceirizados da Meta, cujas [opções de classificação](#) permitem que eles abordem várias maneiras pelas quais o conteúdo da mídia pode enganar as pessoas, incluindo, entre outros, a mídia criada ou editada pela IA.

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

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

Como resultado das políticas e medidas acima, entre os dias 1º de agosto e 31 de outubro de 2024, mais de 26.000 anúncios do tipo SIEP criados por usuários no Brasil no Facebook e no Instagram foram rotulados com o aviso "criado digitalmente" como resultado da auto divulgação, proporcionando maior transparência aos usuários.

SEÇÃO 2: WHATSAPP

O WhatsApp se dedica a ajudar as pessoas a se expressarem livremente e reconhecemos que as instituições democráticas protegem esse direito. Continuamos comprometidos em lidar com o abuso e, ao mesmo tempo, proteger a privacidade das pessoas. Mantemos uma equipe dedicada a evitar abusos, focada especificamente em três linhas de esforço:

- Manter o caráter privado do WhatsApp
- Impedir o uso indevido do WhatsApp
- Empoderar os usuários para combater informações falsas

O WhatsApp não usa algoritmos para priorizar a ordem das mensagens ou das atualizações do canal que as pessoas recebem, nem fornece conteúdo em um feed. Não há pesquisa no aplicativo ou capacidade de descoberta de pessoas ou grupos não conectados. Geralmente, as pessoas usam o WhatsApp para se comunicar com pessoas que elas já conhecem, e é necessário saber o número de telefone para enviar mensagens para alguém.

Antes e durante este ciclo eleitoral, o WhatsApp adotou proativamente uma abordagem holística para evitar ameaças à integridade eleitoral e danos na plataforma. Para isso, o WhatsApp aproveitou-se de tecnologias pré-existentes e protocolos de integridade para identificar e resolver problemas de alta prioridade durante o período eleitoral.

1. Manter o caráter privado do WhatsApp: Limitar a viralidade e reduzir as informações falsas

O WhatsApp utiliza uma série de medidas para evitar danos na plataforma, limitar a viralidade e reduzir informações falsas, mantendo a privacidade do usuário. Isso inclui: criptografia de ponta a ponta; limites de encaminhamento e rotulagem; prevenção de danos coordenados perpetrados por meio de automação, de disparo de mensagens em massa e de abuso de grupo; e restrições de uso político.

Para obter informações mais detalhadas sobre nossa estratégia de integridade eleitoral, [consulte nosso relatório pré-eleitoral](#).

2. Impedir o uso abusivo do WhatsApp

O WhatsApp tem uma tecnologia de ponta para detectar *spam* que funciona 24 horas por dia. Nossa tecnologia detecta contas com comportamento anormal para que elas não possam ser usadas para espalhar spam ou informações falsas. Isso significa que, frequentemente, nossos sistemas automatizados são capazes de impedir abusos antes mesmo que eles sejam denunciados.

Durante o período eleitoral, entre os dias 1º de agosto e 31 de outubro de 2024, o WhatsApp banuiu aproximadamente 4,8 milhões de contas no Brasil por envolvimento em disparo de mensagens em massa ou spam, em violação de nossas políticas (conforme descrito no relatório pré-eleitoral).

Denúncias do TSE:

Além disso, o WhatsApp também aplicou medidas reativas contra contas denunciadas por usuários e encaminhadas a nós pelo TSE. Entre os dias 1º de agosto e 31 de outubro de 2024, o TSE nos encaminhou mais de 1.000 contas ativas exclusivas do WhatsApp. Analisamos todas as contas encaminhadas pelo TSE, com os seguintes resultados:

| | |
|---|-------------|
| Contas únicas banidas proativamente (antes do relatório TSE) | Mais de 70 |
| Contas únicas banidas de forma reativa (mediante denúncia do TSE) | Mais de 20 |
| Contas únicas revisadas, mas não | Mais de 900 |

| | |
|--------------------------|--|
| consideradas em violação | |
|--------------------------|--|

Conforme mostrado na tabela, nos casos em que essas contas estavam envolvidas em mensagens automatizadas ou em massa, a maioria já havia sido banida por nossos sistemas automatizados, antes de serem encaminhadas pelo TSE. A maioria das contas sinalizadas pelo canal de denúncias do TSE não tinha indicação de atividade de mensagens automatizadas ou em massa; a conta denunciada pode ter enviado uma mensagem indesejada, mas não o fez em violação das políticas do WhatsApp.

Medidas além da plataforma - proibição de anúncios:

A Meta realiza esforços contínuos para identificar e remover anúncios que oferecem serviços ilegais de mensagens em massa no WhatsApp, tanto em seus próprios aplicativos quanto em superfícies externas, como mecanismos de pesquisa, incluindo o Google Search e o Microsoft Bing.

Para esse fim, entre os dias 1º de agosto e 31 de outubro de 2024, a Meta tomou providências com relação a mais de 40 anúncios no Facebook e Instagram que promoviam serviços violadores para usuários brasileiros, incluindo mensagens em massa ou mensagens automatizadas via WhatsApp.

Além disso, entre os dias 1º de agosto e 31 de outubro de 2024, como parte de nossa colaboração entre setores, informamos ao Google e à Microsoft a existência de mais de 500 anúncios sendo veiculados no Brasil, nos quais terceiros ofereciam versões não autorizadas do WhatsApp ou serviços ilegais de mensagens em massa.

Contencioso eleitoral:

Entre os dias 1º de agosto e 31 de outubro de 2024, o WhatsApp recebeu mais de 210 liminares de juízos eleitorais civis determinando o fornecimento de dados de usuários. O WhatsApp também foi intimado em mais de 120 liminares determinando restrições às contas dos usuários no mesmo período. O WhatsApp cumpriu em tempo hábil as liminares dos juízos eleitorais, de acordo com suas capacidades técnicas.

3. Capacitação dos usuários para lidar com informações falsas

Suporte para verificação de fatos:

Firmamos uma parceria com a International Fact-Checking Network para disponibilizar a checagem de fatos, diretamente no WhatsApp, inclusive nos Canais. Isso permite que os usuários consultem as organizações de verificação de fatos diretamente por meio de mensagens criptografadas de ponta a ponta e permite que os verificadores de fatos alcancem um público mais amplo por meio dos Canais do WhatsApp - uma ferramenta de transmissão de um para muitos não criptografada. Por meio dessa parceria, organizações de verificação de fatos em quase 50 países usam o WhatsApp para ajudar a conectar usuários com informações confiáveis. No Brasil, temos parceria com os seguintes verificadores de fatos, com os quais os usuários podem interagir pelos telefones listados:

- 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 a capacitar as organizações de verificação de fatos e a garantir que elas tenham os recursos necessários para combater informações falsas e assegurar que os eleitores tenham informações confiáveis sobre como se registrar para votar e onde votar.

Este ano, as Empresas Meta apoiaram a International Fact-Checking Network (IFCN) com um prêmio de US\$ 300.000 em subsídios "*Spread the Facts*" (Divulgue os Fatos) para 10 organizações de verificação de fatos em todo o mundo para combater informações falsas geradas por IA no WhatsApp e aproveitar a IA generativa para otimizar os fluxos de trabalho dos verificadores de fatos. A Agência Lupa e Aos Fatos foram beneficiadas por esse subsídio.

Bloqueio e denúncia:

Ao contrário do SMS tradicional, o WhatsApp oferece uma maneira simples para os usuários bloquearem contas, inclusive diretamente da tela de bloqueio do telefone. Os usuários também podem facilmente fazer denúncias ao WhatsApp se encontrarem mensagens, grupos, contas ou atualizações de canais problemáticos. Nós banimos a grande maioria das contas abusivas por meio de detecção automática. No entanto, as denúncias nos ajudam a identificar se as contas estão envolvidas em mensagens em massa ou abuso coordenado. Assim, podemos conduzir uma investigação mais aprofundada para evitar danos.

Pesquisar na Web:

O WhatsApp oferece uma maneira simples de verificar mensagens pessoais que foram encaminhadas várias vezes. Isso ajuda nossos usuários a encontrar resultados de notícias ou outras fontes de informação sobre o conteúdo que receberam. Esse recurso funciona permitindo que os usuários toquem em uma lupa que faz o *upload* da mensagem por meio do navegador. Fornecemos esse serviço para permitir que os usuários combatam informações falsas; como um serviço de mensagens criptografadas, no entanto, o WhatsApp não rastreia as conversas dos usuários e não teria a capacidade de contar o número de instâncias em que os usuários usaram esse recurso para identificar informações falsas.

Campanhas de educação pública:

O WhatsApp lançou parcerias, campanhas educativas em larga escala e colaborações para lidar com informações falsas em vários países. Para esta eleição, conforme detalhado acima, as Empresas da Meta lançaram uma campanha educacional no Brasil entre os dias 20 de agosto e 30 de setembro de 2024, para aumentar a conscientização sobre as iniciativas da Meta e do WhatsApp para proteger a integridade das eleições de 2024 no Brasil (consulte a página 12 para obter mais detalhes). A campanha promoveu a verificação de fatos no WhatsApp, destacando o papel de nossos parceiros no combate a informações falsas.

Assistente virtual do TSE no WhatsApp:

Para as eleições de 2020, 2022 e 2024 no Brasil, fizemos uma parceria com o TSE para apoiar o lançamento do assistente virtual do TSE no WhatsApp. De acordo com o TSE e seus parceiros, durante o período eleitoral de 2024, o chatbot do TSE acumulou mais de 9 milhões de mensagens (de mais de 1 milhão de usuários únicos) e permitiu que os

eleitores interagissem diretamente com a autoridade eleitoral e recebessem informações relevantes sobre a votação.