



O PAPEL DAS PLATAFORMAS DIGITAIS NA PROTEÇÃO DA INTEGRIDADE ELEITORAL

Edição 2026

Expediente

Este relatório foi desenvolvido pela SAD em parceria das organizações FALA, Instituto Marielle Franco, Data Privacy e InternetLab.

Como citar: SALA DE ARTICULAÇÃO CONTRA DESINFORMAÇÃO. O papel das plataformas digitais na proteção da integridade eleitoral: edição 2026. [S. l.]: SAD, 2026. Disponível em: XX. x

Declaração de uso de IA: Ferramentas de IA foram utilizadas no auxílio de busca de conteúdo quando aplicável e em parte da revisão de texto.

Coordenação do projeto: Andressa Michelotti (Margem/UFMG), Liz Nóbrega (Aláfia Lab) e Tatiana Dourado (Instituto Democracia em Xequê)

Pesquisa e redação: Andressa Michelotti (Margem/UFMG), Camila Tsuzuki e Heloísa Massaro (InternetLab), Gabrielle Martins da Silva e Aron Giovanni Oliveira (Instituto Marielle Franco), Rafael de Pino (FALA/Observatório de Integridade da Informação/Mentira Tem Preço), Carla Rodrigues e Natasha Nóvoa (Data Privacy Brasil).

Revisão e edição: Liz Nóbrega (Aláfia Lab), Tatiana Dourado (Instituto Democracia em Xequê), Nina Santos, Andressa Michelotti (Margem/UFMG), Maria Paula Almada (Aláfia Lab) e Carla Vreche (Conectas Direitos Humanos).

ÍNDICE

INTRODUÇÃO

SUMÁRIO EXECUTIVO

METODOLOGIA

AValiação GERAL

REDES SOCIAIS

INTEGRIDADE ELEITORAL

VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA

DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL

APLICATIVOS DE MENSAGEM

INTEGRIDADE ELEITORAL

DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL e VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA

PLATAFORMAS DE IA

GOVERNANÇA DE PLATAFORMAS DE IA E INTEGRIDADE ELEITORAL

VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA

DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL

RECOMENDAÇÕES

ÀS PLATAFORMAS: REDES SOCIAIS

INTEGRIDADE ELEITORAL

VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA

DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL

ÀS PLATAFORMAS: APLICATIVOS DE MENSAGEM

INTEGRIDADE ELEITORAL

VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA

DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL

ÀS PLATAFORMAS: INTELIGÊNCIA ARTIFICIAL

GOVERNANÇA DE PLATAFORMAS E IA

VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA

DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL

AO ESTADO BRASILEIRO

INTEGRIDADE ELEITORAL

VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA (Instituto Marielle Franco)

DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL

À JUSTIÇA ELEITORAL

À SOCIEDADE CIVIL

REFERÊNCIAS

INTRODUÇÃO

Desde 2022, a Sala de Articulação contra Desinformação (SAD) traz o panorama sobre o papel das plataformas digitais na integridade eleitoral durante as eleições no Brasil. Em meio a discussões mais amplas sobre a sua regulação, que avançou nos últimos anos, as políticas definidas pelas próprias empresas continuam tendo um papel central. Elas representam as regras definidas sobre o que é ou não permitido em suas plataformas e, por isso, têm impacto essencial em como o debate se desenrola nos espaços digitais. É sobre essas políticas que a SAD se debruça neste documento, dando continuidade ao trabalho feito desde 2022.

As eleições gerais de 2026 ocorrem em um ambiente informacional estruturalmente diferente daquele da eleição de 2022. Naquele momento, o debate sobre integridade eleitoral concentrava-se na moderação de conteúdo em redes sociais e na transparência das bibliotecas de anúncios. O lançamento do ChatGPT, em novembro de 2022, e a posterior massificação do uso da inteligência artificial (IA), porém, incorporaram novos desafios à realidade.

Enquanto as eleições municipais de 2024 permitiram observar os primeiros indícios do papel da inteligência artificial na disputa eleitoral¹ — na ocasião ainda de maneira pontual —, 2026 promete um cenário mais complexo, sobretudo pela ascensão e proliferação de plataformas de IA generativa², que oferecem alternativas para o desenvolvimento e a disseminação de desinformação de forma vasta, rápida e de baixo custo, inaugurando questões que o arcabouço regulatório brasileiro ainda não enfrentou em escala.

Em 2024, a Resolução nº 23.732³ do Tribunal Superior Eleitoral (TSE) já havia estabelecido as primeiras diretrizes sobre o tema da IA, como a proibição do uso de *deepfakes*⁴ e a obrigatoriedade de rotular usos de mídia sintética. Em 2026, o TSE

¹ FGV. “Inteligência Artificial No Mundo: Mapeamento Dos Usos Durante as Eleições de 2024 | Mídia e Democracia.” 2024.

<https://midiademocracia.fgv.br/estudos/inteligencia-artificial-no-mundo-mapeamento-dos-usos-durante-eleicoes-de-2024>.

CRUZ, M.; SANTOS, N.; CARREIRO, R.; NÓBREGA, L.; AMORIM, Gabriel. IA no primeiro turno: o que vimos até aqui. Salvador e São Paulo: Aláfia Lab & Data Privacy Brasil, 2024.

<https://observatorioianaseleicoes.com.br/wp-content/uploads/2025/08/RELATORIO-Balanco-Observatorio-IA-nas-Eleicoes.pdf>

² IA generativa são grandes modelos de linguagem que emulam a “estrutura e as características dos dados de entrada para gerar conteúdo sintético derivado” CSRC. “Generative Artificial Intelligence - Glossary | CSRC.” https://csrc.nist.gov/glossary/term/generative_artificial_intelligence.

³ Maiores detalhes:

<https://www.tse.jus.br/legislacao/compilada/res/2024/resolucao-no-23-732-de-27-de-fevereiro-de-2024>

⁴ “A expressão “*deepfake*” é uma combinação de termos em inglês: “aprendizado profundo” (*deep learning*) e “falso” (*fake*). O termo não tem tradução em português porém ele significa “conteúdos falsos produzidos com um alto grau de elaboração” seja em imagens, vídeos ou outros conteúdos sintéticos. BRASIL. [Guia Ilustrado Contra as Deepfakes](#). Supremo Tribunal Federal; Data Privacy.

aprovou a Resolução nº 23.755⁵, atualizando novamente as regras sobre propaganda eleitoral da Resolução nº 23.610/2019, que avança nesse cenário ao instituir seis normas inéditas em relação ao uso de IA e à atuação das plataformas digitais. Alguns exemplos em destaque são: a proibição de publicar ou republicar novos conteúdos sintéticos gerados por IA no período entre 72 horas antes e 24 horas após a votação, o que abrange tanto publicação orgânica quanto impulsionamento pago⁶; a rotulagem obrigatória de conteúdo sintético⁷, a vedação à emissão de opinião, recomendação ou ranqueamento de candidaturas por plataformas de IA generativa⁸; e a proibição expressa de *deepfakes* pornográficos ou de teor íntimo envolvendo candidaturas⁹.

É relevante destacar que, embora a IA tenha concentrado a atenção este ano, os problemas ligados às plataformas de redes sociais e aos aplicativos de mensagem não desapareceram. Por isso, este relatório tem como objetivo lançar luz sobre as medidas e políticas adotadas tanto pelas empresas de redes sociais e aplicativos de mensageria quanto dos sistemas de IA generativa no presente ano eleitoral, identificando suas principais lacunas. Para isso, foram avaliadas as políticas da Meta (Facebook, Instagram e WhatsApp), Kwai, X, LinkedIn, YouTube, TikTok e Telegram, bem como dos principais modelos de IA, tais como Grok, Gemini, ChatGPT, Claude e Llama. A análise foi estruturada a partir dos principais eixos de atuação da SAD: i) integridade eleitoral; ii) meio ambiente e clima; iii) violência política de gênero e raça. Por fim, este relatório também aborda um tema transversal a todas as áreas: a governança de inteligência artificial. Em todos os eixos, o foco central é o combate à desinformação no contexto eleitoral.

Nas próximas seções, são apresentados o sumário executivo, a metodologia e uma avaliação geral dos resultados do relatório. Em seguida, é realizada a análise por tipo de plataforma digital: inteligência artificial, redes sociais e aplicativos de mensagem. Por fim, são apresentadas as recomendações.

SUMÁRIO EXECUTIVO

Plataformas de IA

- As plataformas de IA apresentam nível baixíssimo ou inexistente de políticas específicas para a integridade eleitoral. As plataformas se dividem em três blocos.

Na Resolução Nº 23.610/2019, o artigo 9º-B, § 1º destaca: “É proibido o uso, para prejudicar ou para favorecer candidatura, de conteúdo sintético em formato de áudio, vídeo ou combinação de ambos, que tenha sido gerado ou manipulado digitalmente, ainda que mediante autorização, para criar, substituir ou alterar imagem ou voz de pessoa viva, falecida ou fictícia (*deep fake*).

⁵ Maiores detalhes:

<https://www.tse.jus.br/legislacao/compilada/res/2026/resolucao-no-23-755-de-2-de-marco-de-2026>

⁶ Art. 9º-B, § 3º-A

⁷ Art. 9º-B, § 5º

⁸ Art. 28, § 1º-C

⁹ Art. 28, § 1º-C

- Bloco 1 (Llama e Grok) não possuem política eleitoral específica identificável; documentação remete à empresa controladora (Meta e xAI), sem regra própria sobre uso político da IA;
- Bloco 2 (Gemini) possui política parcial; informação sobre eleições dispersa entre páginas institucionais do Google, sem documento dedicado à governança eleitoral da ferramenta;
- Bloco 3 (OpenAI e Claude) possuem políticas específicas, mas com naturezas diferentes.
- Quando se trata de violência política de gênero e raça e desinformação climática e socioambiental, nenhuma das plataformas de IA traz os temas como categoria própria.
- Há, portanto, uma necessidade urgente de incorporar as plataformas de IA ao debate regulatório eleitoral brasileiro.

Redes sociais

- Em relação às redes sociais, o cenário é mais maduro, especialmente no eixo sobre integridade eleitoral. YouTube, TikTok e Meta (Facebook e Instagram) mantêm páginas que reúnem as políticas e processos de moderação aplicáveis durante o período eleitoral, o que facilita o acesso e a compreensão das pessoas usuárias. As demais, como X e LinkedIn, têm previsões para conteúdos eleitorais, mas em políticas esparsas. A maioria das plataformas analisadas trata sobre desinformação relacionada à data/local de votação, conteúdo de supressão de voto e incitação à violência eleitoral. Existem, contudo, diferenças nos critérios de análise e tratamento dos conteúdos em cada plataforma. Na Meta, por exemplo, alegações de fraude eleitoral só são tratadas quando acompanhadas de ameaça de violência explícita; no X, a política de integridade cívica prevê apenas “Notas da Comunidade”¹⁰ e selos, sem remoção para a maioria dos casos.
- *Em relação às políticas das redes sociais sobre violência política de gênero e raça:* Meta (Facebook, e Instagram), X, TikTok, YouTube e LinkedIn tratam o tema somente por meio de políticas mais abrangentes como discurso de ódio, assédio; Kwai é a exceção, com política específica, porém restrita à violência política de gênero, ignorando outras interseccionalidades.
- *O tema da desinformação climática e socioambiental segue sendo o menos trabalhado:* Facebook, Instagram, X e LinkedIn não possuem categoria própria sobre o tema; YouTube e Kwai tratam do assunto de maneira parcial; TikTok trata o tema explicitamente, inclusive em regras de publicidade.

Aplicativos de mensagem

- Nos três eixos - *integridade eleitoral, violência política de gênero e raça e desinformação climática e socioambiental* -, observa-se o contraste entre

¹⁰ Notas da comunidade são avisos colaborativos no X adicionados por pessoas usuárias abaixo de publicações para fornecer contexto. <https://communitynotes.x.com/guide/pt/about/introduction>

WhatsApp e Telegram. Enquanto o WhatsApp possui uma política para eleições com mecanismos específicos para prevenir o envio em massa de mensagens durante o período (com limite de encaminhamento a cinco conversas, rotulagem de mensagens virais, banimento por comportamento anormal, veto ao uso do WhatsApp Business por campanhas), o Telegram não possui política específica, apenas regras genéricas de spam.

- Observa-se uma lacuna comum aos dois aplicativos de mensageria: nenhum trata da vedação a *chatbots*, avatares ou conteúdo sintético simulando conversa com candidato real, exigência do art. 9º-B, §3º da Resolução 23.610/2019.

Achados transversais

- No geral, os principais desafios referem-se às políticas dispersas, sem páginas dedicadas às políticas adotadas em período eleitoral, difíceis de identificar e sem clareza quanto à data de publicação e atualização. Este é o padrão da maioria das empresas de plataformas digitais, mas se torna ainda mais crítico nas de IA.
- O mesmo desafio é identificado nos canais de denúncia, que exigem navegação por múltiplas páginas de políticas. Além disso, os procedimentos adotados após a realização de denúncias são pouco transparentes para a pessoa usuária comum, elevando o custo para obtenção de informação
- Essa opacidade também impõe limitações ao presente relatório. Outras políticas existentes podem não ter sido mapeadas aqui precisamente pela dificuldade de localizá-las.

METODOLOGIA

Para entender quais as medidas e políticas adotadas pelas plataformas digitais no ano eleitoral no Brasil, identificando as principais lacunas, a análise abrangeu:

Redes Sociais: Facebook (Meta), Instagram (Meta), Kwai (Kuaishou), LinkedIn (Microsoft), TikTok (ByteDance), X (X Corp), YouTube (Google).

Aplicativos de Mensageria: WhatsApp (Meta), Telegram

Plataformas de IA: ChatGPT (OpenAI), Claude (Anthropic), Gemini (Google), Grok (xAI) e Llama (Meta).

A análise foi conduzida em duas etapas. Na primeira, utilizou-se o levantamento feito pelo projeto Ache a Regra (<https://achearegra.internetlab.org.br/>), do InternetLab, por meio da qual foram coletadas — entre março e abril de 2026 — 529 políticas de conteúdo publicadas pelas seguintes plataformas de redes sociais e mensageria: Facebook, Instagram, LinkedIn, WhatsApp, X, TikTok, Telegram, Kwai e YouTube.

Na segunda etapa, essas políticas foram complementadas por verificação manual, totalizando um universo de 564 políticas coletadas. Após a remoção de duplicatas e

políticas não relacionadas aos temas deste relatório, restaram 73 políticas, classificadas nos seguintes temas:

- *Anúncios Digitais*: políticas sobre publicidade política paga, bibliotecas de anúncios e impulsionamento de conteúdo, aqui tratada como parte do eixo de integridade eleitoral.
- *Integridade Eleitoral/Cívica*: políticas sobre proteção do processo eleitoral, confiança institucional e participação cívica.
- *Meio Ambiente*: políticas sobre desinformação climática, desmatamento e pautas socioambientais.
- *Violência Política*: políticas específicas sobre violência política ou violência política de gênero ou raça.
- *Eleitoral (geral)*: demais políticas com aplicação direta ao período de campanha e votação.
- *Desinformação ou misinformation (geral)*: políticas diretamente relacionadas à misinformation ou desinformação, excetuadas aquelas classificadas como políticas de saúde.

Em uma segunda etapa, foram analisadas as políticas sobre eleições e/ou eventos cívicos de cinco plataformas de inteligência artificial generativa (OpenAI, Claude, Gemini, Llama e Grok) por meio de dois passos sucessivos de coleta, adotados para verificar a consistência entre as fontes.

Passo 1: Busca direta nas páginas das plataformas de IA. Consulta manual às páginas oficiais de política de uso e de confiança e segurança de cada uma das cinco plataformas (OpenAI, Claude, Gemini, Llama e Grok) para explorar o contexto das políticas relacionadas à IA.

Passo 2: Busca no Google. Para cada plataforma digital, foi realizada busca manual no buscador Google.com e Google.com.br os seguintes termos:

<i>Termo de busca</i>	<i>Método de busca</i>
<i>"openai policy elections"</i>	Manual
<i>"claude policy elections"</i>	Manual
<i>"gemini policy elections"</i>	Manual
<i>"llama policy elections"</i>	Manual
<i>"grok policy elections"</i>	Manual

Tabela 1. Termos de busca utilizados no levantamento de políticas das plataformas de IA.

No total, foram revisadas 22 URLs contendo políticas relacionadas à IA.

Cada etapa envolveu, no mínimo, duas rodadas de revisão para identificar possíveis lacunas no mapeamento das políticas das plataformas digitais. Essas revisões foram complementadas por verificações manuais diretamente nos *help centers* e nas páginas oficiais de políticas de cada plataforma.

Nem todas as políticas analisadas se relacionam diretamente à desinformação e termos correlatos ao fenômeno. Em muitos casos, exploramos políticas mais amplas que também acionam a governança de desinformação e *misinformation*, tais como comportamento inautêntico, discurso de ódio, contas falsas, entre outras. É importante ressaltar que, na maior parte das vezes, as empresas não agem sobre desinformação em si, pois enquadram os casos em políticas mais abrangentes. Nas seguintes seções apresentamos o debate nos eixos apresentados.

AVALIAÇÃO GERAL

Esta seção apresenta a avaliação geral das plataformas digitais analisadas. As áreas testadas foram três: i) acessibilidade das políticas, ii) disponibilidade do canal de denúncias e iii) transparência em contextos eleitorais no Brasil. É importante ressaltar que não foram analisadas, em nenhum momento, a efetividade dos canais de denúncia nem a aplicação prática das políticas por parte das plataformas digitais.

Em geral, as plataformas digitais apresentam um desafio transversal: as políticas são dispersas, difíceis de localizar e não possuem página dedicada a tratar especificamente delas. Além disso, muitas ainda não informam a data de publicação, o que dificulta o acesso à informação e a compreensão de sua vigência. Esse cenário é ainda mais acentuado nas plataformas de inteligência artificial. Independentemente do conteúdo da política, as pessoas usuárias podem enfrentar dificuldades para se informarem sobre o conteúdo e comportamento vedado ao utilizar esses serviços.

Esse desafio impõe também uma limitação a este relatório, uma vez que políticas existentes podem não ter sido aqui mapeadas precisamente em função da dificuldade de sua localização nos respectivos sites das plataformas digitais.

Problema semelhante ocorre em relação aos canais de denúncias. A pessoa usuária não só precisa entender a política como é exigido dela descobrir como reportar à plataforma digital uma publicação que supostamente a viola. Em muitos casos, cumprir essas etapas exige conhecimento específico sobre governança de plataformas digitais, sobre a navegação interna de cada produto e, no limite, também normas eleitorais. Esse processo se torna inacessível para pessoas leigas ou pouco familiarizadas com essas tecnologias e serviços digitais.

		Acessibilidade das políticas			Canal de Denúncias	Transparência Eleitoral no Brasil	
Tipos de plataformas digitais	Nome das plataformas digitais	Existe política de desinformação?	A página é facilmente acessível?	A página está atualizada e fornece a data de publicação?	É fácil denunciar conteúdo relacionado à desinformação?	Existem informações sobre as eleições de 2026 no Brasil?	TOTAL (média)
Redes Sociais	Kwai	1	1	1	0.5	0	0.7
	Linkedin	1	0	0.5	0	0	0.3
	Meta	1	1	1	1	0	0.8
	TikTok	1	0	1	0	0	0.4
	YouTube	1	1	0.5	1	0.5	0.8
	X	1	1	0.5	0.5	0	0.6
Mensageria	Telegram	0	0	0	0	0	0
	WhatsApp	1	1	0.5	0.5	0	0.6
IA	Claude	0.5	0	1	0.5	0	0.5
	Gemini	1	0	0.5	1	0	0.5
	Llama	1	0	0.5	0	0	0.3
	Grok	0	0	0	0	0	0
	OpenAI	1	0	1	0.5	0	0.5

Tabela 2. Avaliação geral das plataformas digitais organizada por categorias e por ordem alfabética

Metodologia de avaliação:

Avaliação conduzida em julho de 2026.

Notas: 0 → inexistente/falho; 0.5 → insuficiente 5; 1 → existente

Acessibilidade da página das políticas

A página é facilmente acessível?

Conta-se quantos cliques a partir da página inicial do produto são necessários até chegar na página da política. O teste foi efetuado via browser apenas.

Até 3 cliques → 1; Mais de 3 cliques → 0.

A página está atualizada e fornece a data de publicação?

0 → página inexistente da política em questão.

0,5 → página existente, mas sem data de publicação.

1 → página com data de publicação da política e/ou versões de atualizações publicadas.

Canal de Denúncias

É fácil denunciar contas e/ou publicações associadas à desinformação ou problemático?

0 → sem canal facilmente identificável ou inexistente.

0,5 → canal existe, mas exige passo extra que não deveria ser necessário. Exemplo: perguntar ao próprio *chatbot* como denunciar, visitar buscadores externos em vez de ter o caminho dentro do produto.

1 → canal visível, poucos cliques, instrução clara.

Transparência sobre regras relativas às eleições no Brasil

Existem informações sobre as eleições de 2026 no Brasil?

0 → sem página facilmente identificável ou inexistente.

0,5 → páginas gerais sobre conteúdos relacionados à integridade eleitoral ou temas correlatos que mencionem eleições no Brasil de alguma forma.

1 → página dedicada às eleições 2026 no Brasil

REDES SOCIAIS

EIXOS ANALISADOS: INTEGRIDADE ELEITORAL, VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA, DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL

As redes sociais constituem a infraestrutura mais testada de moderação de conteúdo no Brasil. Passaram por quatro ciclos consecutivos de eleições sob escrutínio regulatório e operam hoje sob obrigações consolidadas pela Resolução TSE nº 23.610/2019 e suas atualizações posteriores. O amadurecimento das políticas dessas empresas ocorreu de maneiras distintas entre eixos temáticos: integridade eleitoral com políticas específicas, enquanto temas como desinformação climática e socioambiental e violência política de gênero e raça seguem tratados de forma periférica dependentes de políticas mais amplas.

O levantamento deste relatório mostra que as plataformas de redes sociais se distinguem na profundidade e reconhecimento formal de riscos ligados à desinformação. Em violência política de gênero e raça, a maioria delas - Facebook, Instagram, X, TikTok, YouTube e LinkedIn - trata o fenômeno apenas por empréstimo de outras categorias, sem nomeá-lo ou criar protocolo próprio. O Kwai é a exceção, com política específica para violência de gênero em contexto eleitoral, ainda que limitada a essa dimensão. Em desinformação climática e socioambiental, a divisão é mais gradual: o TikTok reconhece o tema de forma explícita, YouTube e Kwai oferecem tratamento parcial, e Facebook, Instagram, X e LinkedIn dependem inteiramente de políticas mais amplas, sem categoria autônoma para o problema.

INTEGRIDADE ELEITORAL

A maioria das plataformas de redes sociais analisadas aborda, em alguma medida, a análise e o tratamento de conteúdos que afetam a integridade do processo eleitoral. Isso abrange: i) desinformação sobre data, local e hora da votação, candidaturas aptas e requisitos para votar; ii) conteúdos de supressão de voto, como alegações sobre filas, falhas nas urnas ou atividade policial no local de votação para dissuadir o eleitor; e iii) incitação à interferência no pleito, ameaça ou violência em locais de votação ou contra pessoas mesárias, servidoras e membras da Justiça Eleitoral. Há, entretanto, variações nos critérios de moderação, tratamento e localização dessas regras. [YouTube](#), [TikTok](#) e [Meta](#) mantêm páginas dedicadas a reunir diretrizes para as eleições. A página do YouTube, por exemplo, detalha o compromisso da empresa em relação ao tema, as principais políticas aplicáveis e como funcionam, a detecção de conteúdo vedado, as regras de propaganda política e as iniciativas para promover fontes confiáveis. O TikTok

se destaca por prover as nuances na moderação de conteúdo de forma mais detida, explicando as medidas alternativas à remoção, como redução do alcance ou aplicação de rótulos, quando há dúvidas sobre a veracidade do conteúdo.

Embora cada plataforma tenha sua forma de estruturar e organizar as informações às pessoas usuárias, reunir em uma mesma página as políticas, processos e medidas adotadas ofereceria uma visão panorâmica e facilitaria o acesso à informação.

Além das páginas dedicadas ao tema, as três plataformas têm políticas abrangentes que podem ser aplicadas para a proteção à integridade eleitoral. A Meta, por exemplo, possui uma norma específica para [desinformação que possa interferir nas eleições](#) e mantém um registro de alterações no documento, permitindo verificar a norma vigente em períodos anteriores. Também é possível encontrar hipóteses que tenham conexão com a desinformação eleitoral em políticas sobre [comportamento não autêntico](#).

Esse cenário se repete, em maior ou menor medida, nas demais redes. Embora as plataformas mantenham páginas dedicadas às eleições, as normas relacionadas à moderação de conteúdos e comportamentos com impacto sobre a integridade eleitoral nem sempre estão reunidas em um único lugar. Em muitos casos, encontram-se dispersas em diferentes documentos, exigindo que a pessoa usuária navegue por diversas políticas para compreender, de forma integrada, os critérios efetivamente adotados pelas plataformas digitais.

Em relação às políticas de desinformação, cada plataforma emprega abordagens distintas. A Meta adota rótulos informativos, redução de alcance ou a remoção do conteúdo. Esta última medida é aplicada quando o material pode interferir diretamente para o risco de lesão corporal iminente ou quando, no contexto eleitoral, possa contribuir ou contribua diretamente para o “risco de interferência na capacidade das pessoas”¹¹ participarem em processos eleitorais. Já o TikTok proíbe “qualquer forma de desinformação que possa causar danos significativos às pessoas ou à sociedade, independentemente da intenção de quem publica”¹², podendo abranger falsidades sobre data de votação, elegibilidade do eleitor ou sobre o resultado das eleições, além de incentivos à interferência eleitoral, à ameaça a mesários e outras hipóteses. O TikTok prevê medidas como a remoção, a redução do alcance ao torná-lo ineligível ao *feed* “Para você”, e a rotulagem.

Além das diferenças conceituais, há um desafio de compreender como esses critérios são aplicados na prática da moderação de conteúdo, somado ao fato de que as políticas de ambas as plataformas carecem de exemplos do que constituem as hipóteses de desinformação sujeita a restrições. Na Meta, embora existam vários exemplos de

¹¹ Disponível em:

<https://transparency.meta.com/pt-br/policies/community-standards/misinformation/>

¹² Disponível em:

<https://www.tiktok.com/safety/pt-BR/policies-and-engagement/integrity-authenticity>

desinformação sobre saúde sujeitos à moderação de conteúdo, o mesmo não ocorre na política referente à interferência eleitoral ou censitária. Cenário similar é observado na política do TikTok. Ademais, no que concerne à política da Meta, é complexo estabelecer uma conexão direta entre o discurso e o risco de lesão corporal iminente, ou ainda atribuir ao conteúdo uma contribuição direta à interferência na participação eleitoral.

Em regra, as alegações sobre fraude nas eleições não são tratadas na política de desinformação eleitoral da Meta, mas podem ser enquadradas na [política de violência e incitação](#) quando associadas a ameaças de violência. O conteúdo pode ser removido caso haja ameaça de violência relacionada à votação, ao registro de pessoas eleitoras, à apuração ou à divulgação de resultados, bem como a agressões verbais ou físicas dirigidas a funcionários eleitorais, independentemente da credibilidade da ameaça ou da notoriedade da vítima. Cabe notar que a opção por centralizar esses casos na política de violência pode limitar a atuação da plataforma digital em situações nas quais conteúdos falsos sobre fraude eleitoral ou sobre a legitimidade das eleições contribuam para a deterioração da confiança pública, mas não contêm uma ameaça explícita de violência. Adicionalmente, a plataforma digital proíbe conteúdos relacionados [à compra e venda de votos](#) e orientações sobre como participar ilicitamente de processos de votação.

O X também possui uma política de [integridade cívica](#), prevendo a análise e a aplicação de medidas para conteúdo que constitua supressão de voto, desinformação sobre o pleito ou intimidação voltada para coagir pessoas eleitoras. A plataforma prevê, no entanto, apenas dois tipos de medidas aplicáveis: notas da comunidade e os selos de integridade cívica. Enquanto o primeiro mecanismo visa adicionar contexto a publicações potencialmente enganosas, o segundo pode restringir o alcance do conteúdo, excluindo-o de resultados de pesquisa, tendências, das linhas do tempo “para você” ou “seguindo”, ou acarretar a restrição de sua visibilidade apenas ao perfil da pessoa autora. São medidas que podem ser suficientes para mitigar o impacto do conteúdo, mas podem não cumprir o que prevê a Resolução 23.610/2019.

No X, a remoção direta é prevista no caso de ameaça de danos físicos a pessoas ou a infraestruturas cívicas. Não há menção específica, contudo, a conteúdos como ameaças ou incitação à violência em contexto eleitoral, como conteúdo que incite a violência contra membros e servidores eleitorais, à insurreição ou ruptura democrática.

Já no LinkedIn, há uma política para conteúdo falso ou enganoso que estabelece a remoção de alegações apresentadas como fatos, mas comprovadamente falsas ou substancialmente enganosas, que possam causar prejuízos. Caso a publicação não represente um dano direto, seu alcance é apenas restrito à rede da pessoa autora. No contexto eleitoral, isso envolve, por exemplo, informações falsas sobre data e horário da votação. Contudo, não há menção expressa a alegações de fraude nas eleições, na apuração dos resultados eleitorais ou medidas contra incitação à insurreição ou à ruptura

da ordem democrática. Embora haja uma política para lidar com violência, ameaça e assédio, não há uma remissão específica para lidar com as questões acima.

Quanto ao Kwai, há uma [política eleitoral](#) que prevê as medidas adotadas para as eleições, mencionando explicitamente dispositivos da Resolução 23.610/2019. Trata-se de uma política ampla, que versa sobre conteúdos diversos, abrangendo desde desinformação sobre o processo eleitoral a conteúdos considerados propaganda eleitoral antecipada. Em geral, a medida adotada pela plataforma é a remoção, o que inclui conteúdos que alegam fraude no processo eleitoral sem evidências verificáveis ou divulgação de resultados eleitorais falsos, assim como conteúdos de supressão de votos, incentivos à violência e ameaças contra candidaturas, autoridades, jornalistas ou pessoas eleitoras. A política prevê também a remoção de propaganda política antecipada, como pedidos explícitos ou implícitos de voto, ainda que, usualmente, a remoção deste tipo de conteúdo seja feita por plataformas mediante decisão judicial.

Embora seja apresentada de forma bastante ampla, não estão enquadrados no escopo da política do Kwai conteúdos que “não apresentem afirmações verificáveis, como opiniões pessoais, posicionamentos políticos, debates públicos ou especulações sobre razões de voto”¹³. Essa delimitação pode acabar restringindo a aplicação da política, ao afastar do seu alcance conteúdos que, embora não contenham afirmações estritamente verificáveis, ainda podem gerar riscos relevantes ao debate público. Considerando que a moderação de conteúdo é uma atividade altamente contextual, seria desejável que a plataforma previsse uma gradação das medidas de moderação aplicáveis, permitindo respostas que fossem proporcionais aos danos potenciais ou efetivos associados ao conteúdo e ao contexto em que é publicado.

Apesar disso, cabe mencionar que a política da Kwai é uma das únicas a mencionar explicitamente a vedação à publicação ou republicação de conteúdos sintéticos produzidos ou alterados por IA, ainda que rotulados, envolvendo candidaturas ou figuras públicas entre as 72 horas anteriores e as 24 horas posteriores ao pleito, como disposto no artigo 9-B, §3 da Resolução 23.610/2019. Aliás, sobre conteúdos gerados ou manipulados com inteligência artificial, a maioria das plataformas prevê a identificação de conteúdos sintéticos, apresentando ferramentas específicas para apoiar o processo, como é o caso do TikTok e do YouTube.

Em relação a anúncios políticos, no momento da escrita deste relatório, todas as plataformas digitais têm políticas específicas quanto ao conteúdo que pode ser veiculado como publicidade. No último ciclo eleitoral brasileiro, em 2024, porém, apenas a Meta, Youtube e Kwai permitiram anúncios com conteúdo político. Elas, por sua vez, determinam a rotulagem de conteúdos anunciados e um repositório para consulta sobre dados referentes à propaganda veiculada.

¹³ Disponível em: <https://www.kwai.com/pt-BR/safety?id=election>

Por fim, analisou-se também a existência de canais de denúncia e processos para a contestação de decisões de moderação. Todas as empresas de plataformas digitais oferecem canais de denúncia, mas a acessibilidade e o detalhamento do processo de contestação variam entre elas. Por exemplo, o TikTok detalha o fluxo de análise do material e as sanções aplicáveis em caso de violação de suas políticas, disponibilizando também um canal de denúncia e a opção de recorrer das decisões. Comparativamente, o Kwai limita-se a dispor um e-mail de contato para que a pessoa usuária possa denunciar um conteúdo violador em sua plataforma digital.

Outro caso que merece destaque são os canais de denúncia do X. A plataforma digital oferece um formulário para [denúncias de violações eleitorais](#), que também recebe reportes de condutas e atos antidemocráticos, apesar de não ter sido encontrada previsão específica sobre o tratamento conferido a esse tipo de conteúdo. Além disso, a empresa também dispõe de um [formulário para autoridades](#) e de outro voltado à [remoção de conteúdos idênticos](#) já declarados ilícitos por sentença judicial definitiva. Este último remete explicitamente à decisão do Supremo Tribunal Federal no RE 1.037.396, Tema 987, embora caiba lembrar do dever de remoção de conteúdo idêntico previsto no artigo 9-E, VI da Resolução 23.610/2019. Nas políticas das demais plataformas analisadas, não foram identificadas previsões específicas sobre o tratamento de conteúdos idênticos.

VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA

O tratamento da violência política de gênero e raça divide as plataformas em dois blocos.

Bloco 1	Bloco 2
Facebook, Instagram, X, TikTok, YouTube e LinkedIn	Kwai
Sem política específica. Execução das políticas podem ser feitas por políticas mais abrangentes.	Possui política específica. Contexto: violência política contra mulheres em contexto eleitoral

Facebook, Instagram, X, TikTok, YouTube e LinkedIn possuem regras em seus Padrões de Comunidade, Diretrizes de Comunidade e/ou Políticas para Comunidades que podem ser aplicadas a casos de violência política de gênero e raça, especialmente por meio das políticas contra discurso de ódio, assédio, integridade eleitoral e combate à desinformação. Nenhuma dessas plataformas digitais, no entanto, nomeia o fenômeno como uma categoria própria, tampouco estabelece um protocolo específico de prevenção, identificação ou resposta.

Em um segundo bloco está o Kwai, que apresentou o maior avanço na incorporação de políticas voltadas à violência política contra mulheres em contexto eleitoral. A empresa prevê a remoção de conteúdos que humilhem, sexualizem, degradem ou questionem a legitimidade da participação política feminina, incluindo manipulações por IA e ataques à aparência, identidade de gênero e sexualidade. A plataforma digital, porém, ignora outras interseccionalidades, como a de raça.¹⁴ Embora tal política permaneça restrita à dimensão de gênero, demonstra que o reconhecimento da violência política nas políticas de moderação é tecnicamente viável e pode servir de referência para o aprimoramento das demais redes.

DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL

O tratamento da desinformação climática e ambiental divide as plataformas em três blocos.

Bloco 1	Bloco 2	Bloco 3
Facebook, Instagram, X e LinkedIn	YouTube e Kwai	TikTok
Sem política específica. Execução das políticas podem ser feitas por políticas mais abrangentes.	Possui política parcial. YouTube - restrição de monetização e anúncios de conteúdo. Kwai - proibição de falsidades sobre desastres e acidentes ambientais, porém sem políticas mais abrangentes.	Existe política específica. Trata o tema de forma explícita nas regras de publicidade.

Facebook, Instagram, X e LinkedIn formam o primeiro bloco. Nenhuma dessas plataformas reconhece a desinformação climática e ambiental como categoria autônoma. A intervenção depende de políticas mais abrangentes, como fraude, mídia manipulada, assédio e incitação à violência. Não há, tampouco, proteção específica para ativistas, lideranças indígenas e tradicionais, jornalistas e cientistas atacados por sua atuação em defesa do meio ambiente.

¹⁴ O Kwai foca apenas em violência de gênero contra mulheres em contexto eleitoral — mas é um passo importante. Mesmo as políticas eleitorais desenhadas especificamente para o Brasil incorporam pouco dos marcos legais nacionais de enfrentamento à violência política de gênero e raça. O foco continua na integridade do processo eleitoral em abstrato, e na moderação de conteúdos isolados — não nos grupos que essa violência mira de forma sistemática.

Em um segundo bloco, estão YouTube e Kwai, com reconhecimento parcial do problema. Enquanto o YouTube restringe monetização e anúncios de conteúdo que contrarie o consenso científico e pode exibir painéis informativos — sem, contudo, remover os vídeos —, o Kwai proíbe falsidades sobre desastres e acidentes ambientais, mas não tem política climática abrangente.

Por fim, o TikTok trata o tema de forma explícita, inclusive nas regras de publicidade.

APLICATIVOS DE MENSAGEM

EIXOS ANALISADOS: INTEGRIDADE ELEITORAL, VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA, DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL

Assim como as redes sociais, os aplicativos de mensageria estão sujeitos às obrigações consolidadas pela Resolução TSE nº 23.610/2019 e suas atualizações. Sua arquitetura, suas políticas e seu modelo de negócio, entretanto, apresentam características próprias: criptografia, grupos privados, políticas de privacidade e possibilidade de disparo em massa de mensagens impõem desafios específicos ao processo eleitoral.

Embora essas plataformas mantenham políticas voltadas ao modelo de mensageria, elas têm incorporado progressivamente características típicas de redes sociais, tais como comunidades e canais, além de funcionalidades como pagamentos e chatbots de IA. Esse movimento as aproxima do modelo de *super apps*, ampliando tanto seu papel no ecossistema digital quanto os desafios para a governança eleitoral.

Em todos os eixos analisados, os aplicativos de mensageria apresentam características que se distinguem tanto nas políticas quanto nos riscos que representam ao processo eleitoral. Por um lado, o WhatsApp dispõe de políticas mais detalhadas e mecanismos de funcionamento mais transparentes. Por outro, o Telegram apresenta políticas mais limitadas ou inexistentes em alguns temas, além de menor transparência sobre suas funcionalidades, processos de moderação e operações.

Nesta seção, são tratadas as principais plataformas de mensageria no Brasil, WhatsApp e Telegram, a partir dos três eixos deste relatório: integridade eleitoral, violência política de gênero e raça e desinformação climática e socioambiental.

INTEGRIDADE ELEITORAL

O tratamento da integridade eleitoral divide-se em dois blocos.

Bloco 1	Bloco 2
Telegram	WhatsApp
Sem política específica. Execução das políticas podem ser feitas por políticas mais abrangentes.	Possui política específica. Limites de encaminhamento, atualizações de canais para apenas cinco conversas por vez, indicações sobre mensagens que são encaminhadas com frequência, limitação de envio desse tipo de conteúdo rotulado, banimento de contas com comportamento anormal, é possível bloquear e denunciar contas.
Nota: não foram encontradas políticas que atendessem as previsões do artigo 9-B, § 3º da Resolução 23.610/2019.	

No caso das plataformas de mensageria instantânea, buscou-se verificar, especialmente no que concerne às comunicações interpessoais, a existência de mecanismos que aumentem a fricção para o envio e encaminhamento de mensagens em massa, com o objetivo de reduzir o uso indevido dessas aplicações na disseminação de conteúdos problemáticos, enganosos e nocivos.

O Telegram não possui uma política específica para eleições, mas reúne, em uma página de [FAQ](#), orientações sobre diversas situações relacionadas ao uso da aplicação e de seus serviços. Entre elas, há regras voltadas ao combate ao spam, aplicáveis aos casos em que uma pessoa usuária envia mensagens a pessoas desconhecidas ou em grupos e é denunciada por esse comportamento. Essas situações podem envolver anúncios, links de convite a grupos ou canais, imagens, entre outros. A plataforma digital informa que denúncias de spam são averiguadas por pessoas moderadoras humanas, que podem aplicar limitações temporárias à capacidade de envio de mensagens. Em casos de reincidência, as restrições podem durar períodos maiores, chegando até mesmo à

suspensão permanente. Além disso, as pessoas podem recorrer das decisões adotadas pela plataforma.

O WhatsApp, por sua vez, mantém uma [política específica](#) sobre eleições. Entre elas, destacam-se os limites de encaminhamento e atualizações de canais para, no máximo, cinco conversas por vez; identificação de mensagens encaminhadas com frequência, alertando as pessoas usuárias de que seu conteúdo pode não ter sido produzido por um contato próximo; e a limitação do reencaminhamento dessas mensagens para apenas uma conversa por vez. A plataforma também prevê o banimento de contas que apresentem comportamento anormal associado ao envio de mensagens em massa, proíbe a utilização da versão Business por candidaturas e campanhas eleitorais, e oferece ferramentas para que a pessoa usuária controle quem pode adicioná-los a grupos. Por fim, o WhatsApp permite bloquear e [denunciar](#) contas que violem suas políticas. No caso de denúncias, a plataforma recebe até cinco mensagens mais recentes da conversa ou, caso a denúncia se refira a ligações, as últimas cinco chamadas com a conta, além de dados básicos, como quem iniciou a ligação e a duração. Contas denunciadas podem ser bloqueadas, embora a plataforma não informe ao denunciante a decisão adotada. Pessoas bloqueadas podem [contestar](#) eventual banimento determinado pelo WhatsApp.

Cabe comentar que, tanto no WhatsApp como no Telegram, não foram encontrados disposições quanto ao uso de chatbots, avatares e conteúdos sintéticos utilizados para intermediar a comunicação da campanha com pessoas naturais ou providências quando houver a simulação de uma conversa entre a candidatura e uma pessoa real, dispositivos estes previstos no artigo 9-B, § 3º da Resolução 23.610/2019.

DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL e VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA

Questões sobre desinformação climática e socioambiental e violência política e de gênero concentram-se em apenas um bloco.

Bloco 1
WhatsApp e Telegram
Sem política específica. Execução das políticas podem ser feitas por políticas mais abrangentes.

O WhatsApp não tem política climática, socioambiental, de enfrentamento à violência política de gênero e raça ou eleitoral própria. A plataforma atua sobre políticas mais abrangentes que governam a circulação de conteúdo — e não a desinformação — como spam, automação, disparos em massa e fraude.

O Telegram segue a mesma lógica — agindo contra golpes, spam, violência e exposição de dados —, mas não contra desinformação como categoria geral, nem mesmo em canais públicos. Nos dois aplicativos falsidades climáticas, ataques a pessoas defensoras ambientais ou a mulheres com atuação política podem circular livremente, desde que não configurem outra violação já prevista em regra.

PLATAFORMAS DE IA

EIXOS ANALISADOS: GOVERNANÇA DE PLATAFORMAS DE IA E INTEGRIDADE ELEITORAL, VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA, DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL

A consolidação da IA generativa representa uma das principais transformações do ambiente informacional desde as eleições gerais de 2022. Se naquele momento as preocupações regulatórias concentravam-se principalmente nas redes sociais e aplicativos de mensageria, as eleições de 2026 inauguram um novo cenário em que plataformas de IA deixam de ocupar posição periférica e passam a integrar diretamente a infraestrutura digital utilizada para produção, personalização e circulação de conteúdos políticos.

Modelos generativos são atualmente capazes de produzir textos, imagens, vídeos, áudios, códigos e agentes conversacionais em larga escala, reduzindo significativamente os custos de produção de campanhas digitais e ampliando os riscos relacionados à desinformação, fraude de identidade, manipulação audiovisual e automação da comunicação política. Esse novo contexto impõe desafios distintos daqueles tradicionalmente enfrentados pelas plataformas de redes sociais. Diferentemente destas, que exercem papel predominante na distribuição de conteúdos produzidos por terceiros, as plataformas de IA participam diretamente da geração do conteúdo, influenciando tanto sua produção quanto sua forma de apresentação às pessoas usuárias.

A discussão regulatória desloca-se, portanto, da moderação posterior de conteúdos para a governança dos próprios sistemas generativos, seus mecanismos de segurança, suas políticas de uso e seus procedimentos internos de mitigação de riscos. O levantamento realizado neste relatório evidencia que, embora todas as principais plataformas de IA reconheçam a existência de riscos associados ao uso eleitoral de seus modelos, há

diferenças expressivas quanto ao grau de transparência, acessibilidade e detalhamento das políticas disponíveis ao público.

Nesta seção são tratadas as principais plataformas de IA no Brasil: ChatGPT (OpenAI), Claude (Anthropic), Gemini (Google), Grok (xAI) e Llama (Meta). Inicialmente são introduzidos os achados sobre a governança das plataformas de inteligência artificial seguido pelos três eixos deste relatório: governança de plataformas, integridade eleitoral, desinformação climática e socioambiental e violência política de gênero e raça.

O levantamento deste relatório mostra que as plataformas de IA generativa se distinguem na governança eleitoral, mas convergem nos demais eixos.

Em governança de plataformas e IA, a transparência se divide em quatro níveis. A Anthropic apresenta a abordagem mais desenvolvida, por combinar política específica de integridade eleitoral e compromisso já anunciado de banner informativo para as eleições brasileiras de 2026. A OpenAI reúne o conjunto mais abrangente de salvaguardas textuais contra interferência eleitoral, supressão de pessoas eleitoras e conteúdo sintético fraudulento, mas concentra a documentação em restrições à pessoa usuária. O Google dispersa a informação sobre o Gemini entre páginas institucionais, sem documento dedicado à governança eleitoral da ferramenta; e Meta e xAI apresentam elevada opacidade, sem política eleitoral facilmente identificável para Llama e Grok, que remetem à política geral da controladora.

Em violência política de gênero e raça, plataformas como OpenAI, Claude, Gemini, Llama e Grok tratam o fenômeno apenas através do uso de políticas de desinformação eleitoral, *deepfakes* e discurso de ódio, sem reconhecer o tema como categoria própria ou criar mecanismo específico de proteção.

Em desinformação climática e socioambiental, o Claude proíbe o tema de forma mais direta, o Llama estabelece proibição geral de aplicação variável, e ChatGPT, Gemini e Grok oferecem apenas proteção indireta, sem categoria autônoma para o problema.

Em geral, o difícil acesso ou inexistência de políticas acessíveis impede que as pessoas usuárias, pesquisadoras, autoridades públicas e candidaturas saibam quais comportamentos são permitidos e quais mecanismos existem para contestar decisões ou reportar abusos. Em um contexto de crescente utilização de agentes conversacionais para produção de conteúdos políticos, a opacidade regulatória tende a ampliar a assimetria informacional entre plataformas e sociedade.

GOVERNANÇA DE PLATAFORMAS DE IA E INTEGRIDADE ELEITORAL

Questões relacionadas à governança de plataformas de IA e integridade eleitoral são divididas em três blocos.

Bloco 1	Bloco 2
Llama, Grok e Gemini	ChatGPT e Claude
Sem política específica identificável. Documentação pública remete às políticas gerais da empresa controladora — Llama à Meta, Grok à xAI — sem regras próprias sobre uso político da IA.	Possui política específica. OpenAI: política mais abrangente entre as analisadas — restringe interferência eleitoral, supressão de pessoas eleitoras, desinformação sobre votação, campanhas automatizadas em larga escala e conteúdo sintético fraudulento; direciona a fontes oficiais e restringe <i>deepfakes</i> políticos. Documentação concentrada em restrições à pessoa usuária, sem avaliação interna de risco publicada. Claude: abordagem mais desenvolvida — políticas específicas de integridade eleitoral, campanhas enganosas e fraude eleitoral; classificadores automáticos combinados a equipe de threat intelligence; 95%/96% de neutralidade política (Opus 4.7/Sonnet 4.6); banners anunciados para as eleições brasileiras de 2026 — porém como compromisso voluntário e unilateral.

A OpenAI apresenta o conjunto mais abrangente de políticas públicas sobre eleições entre as plataformas analisadas. Suas salvaguardas restringem o uso dos modelos para interferência eleitoral, supressão de pessoas eleitoras, desinformação deliberada sobre processos de votação, campanhas políticas automatizadas em larga escala e conteúdo sintético destinado a fraude ou manipulação política. A empresa indica nas políticas que direciona as pessoas usuárias a fontes oficiais quando consultada sobre procedimentos eleitorais, restringe pedidos de *deepfakes* políticos enganosos e promove transparência sobre conteúdo produzido por inteligência artificial. Essas medidas representam avanço em relação aos ciclos anteriores, por reconhecerem o potencial de uso abusivo da IA em processos democráticos. A documentação pública, porém, se concentra em restrições

impostas às pessoas usuárias, oferecendo poucas informações sobre avaliações sistemáticas de risco realizadas pela própria empresa, metodologias de testes de segurança, indicadores de efetividade das salvaguardas e procedimentos permanentes de auditoria externa.

Quanto ao Gemini, Llama e Grok, não foram identificadas políticas eleitorais específicas dedicadas aos respectivos sistemas de IA generativa. No caso do Gemini, as informações sobre eleições estão dispersas entre páginas institucionais do Google, sem documento próprio sobre governança eleitoral da ferramenta. O Llama e o Grok também não apresentam políticas eleitorais específicas identificáveis, remetendo, respectivamente, às políticas gerais da Meta e da xAI. Embora existam diferenças na documentação pública disponível, os três casos revelam uma lacuna comum que é a ausência de regras compreensíveis e específicas sobre o uso eleitoral desses sistemas generativos, dificultando a compreensão das pessoas usuárias, pesquisadoras e autoridades sobre os mecanismos de prevenção e resposta a riscos.

A Anthropic também possui lacunas na publicação de avaliações de impacto, em métricas agregadas sobre tentativas de abuso identificadas, na transparência sobre revisões independentes e em mecanismos de prestação de contas perante pessoas reguladoras e pesquisadoras. Ainda assim, é a plataforma com a abordagem mais desenvolvida entre as analisadas através da divulgação de políticas específicas sobre integridade eleitoral, o uso do Claude para campanhas enganosas, fraude eleitoral e conteúdo sintético voltado à interferência democrática. Segundo atualização publicada pela empresa em abril de 2026, a governança combina classificadores automáticos com uma equipe dedicada de *threat intelligence*, e seus modelos mais recentes (Opus 4.7 e Sonnet 4.6) pontuaram 95% e 96% em avaliações de neutralidade política. A mesma atualização anuncia planos de *banners* direcionando pessoas usuárias a fontes oficiais durante as eleições brasileiras de 2026.

A identificação de conteúdo sintético expõe outra lacuna. Embora a OpenAI e Anthropic, mencionem iniciativas de rotulagem, estas empresas não contam com padrões interoperáveis que preservem metadados de origem ao longo do compartilhamento entre plataformas. A ausência de mecanismos técnicos padronizados reduz significativamente a efetividade das políticas atualmente existentes, especialmente quando conteúdos produzidos por IA passam a circular em redes sociais, aplicativos de mensageria e serviços de compartilhamento de vídeo.

No geral, as políticas atuais concentram-se sobretudo na prevenção de solicitações ilícitas feitas diretamente pela pessoa usuária, deixando lacunas nos riscos emergentes como personalização política automatizada, integração entre modelos generativos e publicidade digital, produção massiva de conteúdo personalizado e coordenação automatizada de campanhas de influência. Esses desafios extrapolam o escopo das políticas tradicionais de uso aceitável e demandam estruturas permanentes de

governança capazes de acompanhar a rápida evolução tecnológica dos modelos generativos.

Neste contexto, existe uma necessidade urgente de incorporar as plataformas de IA ao debate regulatório eleitoral.¹⁵

VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA

Questões sobre violência política de gênero e raça concentram-se em apenas um bloco.

Bloco 1
OpenAI, Claude, Gemini, Llama e Grok
Sem política específica.
Nenhuma das cinco plataformas a reconhece expressamente como categoria específica de risco.

A análise das plataformas de inteligência artificial generativa — OpenAI, Claude, Gemini, Llama e Grok — mostra que, diferentemente das redes sociais e dos canais de mensageria, suas políticas se concentram na governança, no desenvolvimento seguro e na prevenção do uso indevido dos modelos. As empresas avançaram na regulamentação dos meios pelos quais a violência política pode ser praticada, com políticas voltadas à mitigação de riscos associados à produção de conteúdos sintéticos, *deepfakes*, manipulação de mídia, desinformação eleitoral, operações de influência e discurso de ódio.

Essas medidas contribuem indiretamente para reduzir práticas que podem instrumentalizar a violência política de gênero e raça, mas nenhuma das cinco empresas a reconhece expressamente como categoria específica de risco. A governança permanece centrada na segurança dos modelos e na prevenção de usos indevidos da tecnologia e não em proteger candidaturas, lideranças políticas, defensoras de direitos humanos e demais grupos historicamente marginalizados por meio de políticas, protocolos e mecanismos próprios de prevenção, monitoramento e resposta.

¹⁵ O relatório *Ei, chat: em quem eu voto?* Também reporta uma significativa heterogeneidade entre as plataformas quanto à aplicação de mecanismos de segurança e neutralidade, evidenciando a ausência de padrões mínimos comuns de governança para sistemas de IA utilizados em contexto eleitoral. Esses achados reforçam que a integridade das eleições depende não apenas da regulação das redes sociais, mas também da adoção de deveres estruturais de transparência, gestão de riscos e prestação de contas aplicáveis aos provedores de modelos generativos.

OBSERVATÓRIO IA NAS ELEIÇÕES. *Ei, chat: em quem eu voto? Testando a conformidade de chatbots de IA às resoluções do TSE para eleições 2026*. Salvador: Aláfia Lab; São Paulo: Data Privacy Brasil, 2026. Disponível em: https://observatorioianaseleicoes.com.br/wp-content/uploads/2026/06/20260616_Obs-IA-chatbot-.pdf. Acesso em 10 jul. 2026

DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL

Questões sobre desinformação climática e socioambiental concentram-se em apenas um bloco.

Bloco 1
OpenAI, Claude, Gemini, Llama e Grok
Sem política específica. Nenhuma das cinco plataformas a reconhece expressamente como categoria específica de risco.

Nenhuma das cinco plataformas — ChatGPT, Claude, Gemini, Grok e Llama — apresenta categoria específica sobre desinformação climática e socioambiental em suas políticas centrais, mas a proteção se divide em três níveis.

O Claude proíbe de forma mais direta a criação e disseminação de desinformação como no caso de *temas científicos* — e não especificamente questões climáticas ou ambientais. O Llama estabelece proibições gerais, mas a aplicação depende dos desenvolvedores e serviços que utilizam o modelo.

ChatGPT, Gemini e Grok oferecem proteções principalmente indiretas, relacionadas a fraude, falsificação, manipulação política, assédio, difamação e outras formas de engano. Essas regras podem alcançar casos graves, mas deixam sem tratamento compreensível falsas alegações sobre mudanças climáticas, distorções de dados ambientais, *greenwashing* e ataques a políticas e soluções climáticas. Também não existe um padrão comum que assegure que esses conteúdos sejam recusados, identificados, contextualizados ou corrigidos.

RECOMENDAÇÕES

Nesta seção são apresentadas as recomendações às empresas de plataformas de redes sociais, aos aplicativos de mensagem e às de IA, assim como ao Estado brasileiro, à Justiça Eleitoral e à sociedade civil.

ÀS PLATAFORMAS: REDES SOCIAIS

INTEGRIDADE ELEITORAL

- As políticas devem ser desenhadas segundo os riscos específicos de cada funcionalidade do serviço prestado, garantindo abordagens proporcionais e

salvaguardas para direitos fundamentais, como a liberdade de expressão. Isso inclui informações claras sobre os critérios utilizados para identificar e moderar conteúdos que possam afetar a integridade eleitoral.

- Previsão de políticas específicas para a integridade cívica e eleitoral aplicáveis para além do período eleitoral, abrangendo conteúdos de incitação ou defesa de violência contra a ordem democrática, compra de votos ou conteúdos que promovam apostas sobre o desempenho de candidaturas ou resultados eleitorais.
- A garantia da acessibilidade e do fácil entendimento das políticas aplicáveis, assim como de mecanismos de denúncia, contestação de decisões e sanções, especialmente em períodos eleitorais, é essencial. Reunir as políticas e procedimentos aplicáveis às eleições, assim como exemplos e histórico de alterações das políticas, é necessário para garantir o acesso e compreensão pelas pessoas usuárias dos comportamentos e conteúdos aceitáveis nesse espaço, e como as regras mudaram ao longo do tempo.
- Previsão nas políticas de medidas adotadas para cumprimento dos deveres e obrigações determinadas pelas normas eleitorais, em especial questões como a remoção de conteúdos idênticos àqueles já considerados ilícitos por decisão judicial ou conteúdos sintéticos de candidaturas ou pessoas públicas produzidos ou alterados com inteligência artificial no período entre as 72 anteriores e 24 posteriores do pleito.
- Procedimentos de moderação devem considerar diferentes aspectos para lidar com conteúdos potencialmente danosos, tais como fluxos prioritários, além de considerar nesta avaliação a fonte emissora das mensagens, que pode influenciar o alcance e a ampliação dos riscos que tais publicações podem representar ao processo eleitoral. Além disso, que tais respostas considerem perfis de jornalistas, ativistas, organizações da sociedade civil e outras partes relevantes para o debate público, reduzindo o risco de medidas equivocadas e promovendo maior equilíbrio entre a proteção da integridade eleitoral e a preservação da liberdade de expressão.
- Adoção de protocolos de emergência durante e após as eleições, em especial diante de tentativas de interferência no processo de apuração de resultados eleitorais, transição de poder ou eventos similares que possam afetar a integridade democrática. Nesses contextos, recomenda-se que plataformas tenham protocolos específicos, agindo rapidamente diante de conteúdos falsos, manipulados, descontextualizados ou gerados com inteligência artificial que possam contribuir para a escalada de tensões ou para a mobilização de atos violentos. Além disso, recomenda-se preservar dados necessários à auditoria e à

pesquisa independente, com informações sobre as ações tomadas e impactos alcançados. Por fim, recomenda-se manter canais de comunicação com autoridades eleitorais e instituições públicas competentes para lidar com possíveis impactos decorrentes desses eventos.

VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA

- Reconhecimento expresso da violência política de gênero e raça em seus termos de uso, políticas de integridade eleitoral e políticas de moderação de conteúdo, observando a legislação brasileira e os instrumentos nacionais e internacionais de proteção dos direitos humanos.
- Desenvolvimento de políticas e mecanismos específicos para prevenir, identificar e responder à violência política de gênero e raça, incluindo ameaças, discurso de ódio, desinformação e campanhas coordenadas de assédio dirigidas a grupos historicamente marginalizados.
- Garantia de canais de denúncia que permitam a tipificação da violência política, respostas céleres e transparentes às denúncias e capacitação contínua das equipes de moderação sobre violência política digital, vieses de raça e gênero e dinâmicas de desinformação.
- Observância aos decretos 12.975 e 12.976 e a decisão do STF em relação ao Artigo 19 do Marco Civil da Internet (RE 1.037.396 e RE 1.057.258) especialmente no que tange à remoção de imagens íntimas não consentidas em até 2h após a denúncia e a diminuição de visibilidade de ataques coordenados contra mulheres, especialmente aquelas na vida pública.
- Adoção de medidas que impeçam a monetização, o impulsionamento e a recomendação algorítmica de conteúdos que promovam violência política de gênero e raça, discurso de ódio, assédio ou desinformação contra candidatas, parlamentares, ativistas, pessoas defensoras de direitos humanos e outras lideranças políticas pertencentes a grupos historicamente marginalizados, especialmente durante os períodos eleitorais.
- Criação de mecanismos permanentes de diálogo com organizações da sociedade civil e pessoas pesquisadoras para subsidiar a atualização contínua de suas políticas de integridade, moderação e segurança, bem como cooperar com as autoridades competentes em procedimentos investigativos e judiciais, respeitados os direitos e garantias previstos na legislação aplicável.

DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL

- Adoção de políticas explícitas sobre desinformação climática e socioambiental. As empresas devem definir, a partir de orientações escritas em português e com base em exemplos brasileiros, quais conteúdos estarão sujeitos a contextualização, redução de alcance, desmonetização, restrição publicitária ou remoção. As regras devem contemplar não apenas a negação da ciência climática e ambiental, mas também formas mais sofisticadas de desinformação, como a minimização de riscos e impactos; a distorção de causas, responsabilidades e consensos científicos; o enfraquecimento de políticas e soluções efetivas; a promoção de falsas soluções ou de medidas parciais e insuficientes como respostas adequadas; o *greenwashing*; e as campanhas de deslegitimação dirigidas a instituições públicas, cientistas, jornalistas, povos indígenas, comunidades tradicionais, populações afetadas por crises climáticas e conflitos socioambientais e pessoas defensoras do meio ambiente.
- Interrupção dos incentivos econômicos e da amplificação algorítmica da desinformação. Conteúdos identificados como falsos, enganosos ou manipulados, com base em critérios públicos, consistentes e auditáveis, não devem ser monetizados, promovidos por publicidade paga nem favorecidos por sistemas de recomendação, listas de tendências, sugestões de busca ou reprodução automática. As medidas devem alcançar contas reincidentes, redes e operações coordenadas, anunciantes, pessoas criadoras de conteúdo e empresas que promovam alegações ambientais enganosas ou campanhas de *greenwashing*. Os mesmos critérios devem ser aplicados a agentes políticos, autoridades públicas e grandes anunciantes, sem privilégios ou exceções.
- Adoção de protocolos de emergência e ampliação da transparência. Durante eventos climáticos extremos, desastres socioambientais e outras situações de emergência, as empresas de plataformas devem adotar protocolos reforçados. Isso deve acontecer, especialmente, em períodos eleitorais, quando esses acontecimentos tendem a ser instrumentalizados em disputas políticas, utilizados para difundir desinformação e convertidos em temas de campanha nas redes sociais. Devem priorizar informações verificadas e atualizadas de autoridades competentes, instituições científicas, serviços de emergência e veículos jornalísticos com práticas editoriais reconhecidas; agir rapidamente contra fraudes, falsas campanhas de arrecadação e conteúdos que possam comprometer a segurança das pessoas, dificultar as ações de resposta ou distorcer causas, responsabilidades e medidas de enfrentamento; e reduzir a circulação de imagens antigas, descontextualizadas, manipuladas ou geradas por inteligência artificial apresentadas como registros atuais. As plataformas também devem preservar os dados necessários à auditoria e à pesquisa independente e divulgar relatórios específicos para o Brasil sobre alcance, monetização, publicidade,

impulsionamento, operações coordenadas e medidas de moderação, remoção ou redução de alcance adotadas durante essas situações.

ÀS PLATAFORMAS: APLICATIVOS DE MENSAGEM

INTEGRIDADE ELEITORAL

- As políticas devem ser desenhadas segundo os riscos específicos de cada funcionalidade do serviço prestado, garantindo abordagens proporcionais e salvaguardas para direitos fundamentais, como a privacidade e a liberdade de expressão. Elas devem considerar as diferenças entre funcionalidades de comunicação interpessoal, especialmente aquelas protegidas por criptografia de ponta-a-ponta, e funcionalidades públicas ou abertas. Para prevenir o uso abusivo desses serviços, recomenda-se a adoção de medidas proporcionais ao risco, como o aumento da fricção no envio e no encaminhamento de mensagens em massa em funcionalidades de comunicação interpessoal e o fortalecimento de mecanismos de moderação e transparência em espaços públicos ou abertos, reduzindo seu potencial de disseminação em larga escala.
- Previsão nas políticas as medidas adotadas para cumprimento dos deveres e obrigações determinados pelas normas eleitorais, em especial questões como a remoção de chatbots, avatares e outros conteúdos sintéticos empregados para intermediar a comunicação entre campanhas e pessoas naturais quando não houver identificação compreensível de sua natureza artificial ou quando houver simulação da interlocução da pessoa candidata ou de outra pessoa real.

VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA

- Reconhecimento expresso da violência política de gênero e raça em seus termos de uso e políticas de integridade eleitoral, observando a legislação brasileira e os instrumentos nacionais e internacionais de proteção dos direitos humanos.
- Desenvolvimento de protocolos e mecanismos específicos para prevenir e responder ao uso abusivo de seus serviços para a prática de violência política de gênero e raça, incluindo canais de denúncia acessíveis, respostas céleres e cooperação com as autoridades competentes, respeitados os direitos e garantias previstos na legislação aplicável.
- Instituição de mecanismos permanentes de transparência e diálogo com organizações da sociedade civil e pessoas pesquisadoras para subsidiar o

aprimoramento contínuo de suas políticas de segurança e integridade e fortalecer a prevenção da violência política de gênero e raça no ambiente digital.

DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL

- Contenção da distribuição massiva e responsabilização de canais de grande alcance. Os serviços devem adotar limites e mecanismos de fricção contra a disseminação automatizada ou massiva de desinformação climática e socioambiental, sem comprometer a criptografia das conversas privadas. Canais públicos, comunidades abertas e ferramentas de transmissão devem seguir regras de transparência e moderação proporcionais ao seu alcance.
- Criação de protocolos contra fraudes e desinformação durante emergências. Em eventos climáticos extremos, especialmente no período eleitoral, os serviços devem facilitar denúncias de falsas arrecadações, alertas enganosos, orientações perigosas, conteúdos antigos apresentados como atuais e distorções sobre causas, responsabilidades e medidas de resposta. Dados agregados sobre disparos automatizados, canais reincidentes e providências adotadas devem ser disponibilizados a autoridades eleitorais, pessoas pesquisadoras e organizações da sociedade civil.

ÀS PLATAFORMAS: INTELIGÊNCIA ARTIFICIAL

GOVERNANÇA DE PLATAFORMAS E IA

- Ampliação da transparência de suas políticas eleitorais. Todas as empresas que disponibilizam modelos generativos ao público devem manter documentos específicos, facilmente acessíveis e atualizados, descrevendo suas regras para utilização em contexto eleitoral, seus mecanismos de mitigação de riscos e seus procedimentos de resposta a abusos. A dificuldade identificada para localizar políticas específicas em algumas plataformas demonstra que a acessibilidade das regras constitui requisito básico para sua efetividade.
- Incorporação de mecanismos padronizados de identificação e rastreabilidade de conteúdos sintéticos, assegurando que informações sobre a origem artificial do conteúdo permaneçam preservadas ao longo de seu compartilhamento entre diferentes serviços digitais. A adoção de padrões técnicos interoperáveis fortalecerá a capacidade de pessoas usuárias, jornalistas, pessoas pesquisadoras e autoridades públicas de identificar conteúdos produzidos por IA durante campanhas eleitorais.

- Adoção de medidas técnicas e de governança proporcionais ao risco, como mecanismos de bloqueio ou redução de funcionalidade em cenários de abuso, detecção, identificação e rastreabilidade do conteúdo gerado, e canal efetivo de reporte, correção e revisão. Essas avaliações devem considerar o estado da arte do desenvolvimento tecnológico para gestão de riscos conhecidos e razoavelmente previsíveis, inclusive mediante avaliadores independentes e exercícios de *red-teaming* voltados a medir desempenho e identificar vulnerabilidades antes e durante a oferta das ferramentas — uma abordagem de segurança desde a concepção ("*security by design*").
- Adoção de estruturas permanentes de governança para seus modelos de IA, incluindo avaliações periódicas de impacto, auditorias independentes, mecanismos transparentes de contestação de decisões, programas de red teaming voltados especificamente para cenários eleitorais e publicação regular de relatórios de transparência. Mais do que responder a incidentes específicos, essas medidas permitem construir um modelo preventivo de gestão de riscos, compatível com o papel estrutural que os sistemas de IA passaram a desempenhar na formação da esfera pública digital e na proteção da integridade das eleições.

VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA

- Reconhecimento expresso da violência política de gênero e raça em seus termos de uso, suas políticas de segurança, governança e uso responsável da inteligência artificial, observando a legislação brasileira e os instrumentos nacionais e internacionais de proteção dos direitos humanos.
- Desenvolvimento de políticas, protocolos e mecanismos específicos para identificar, prevenir e mitigar usos da inteligência artificial que possam facilitar, amplificar ou reproduzir a violência política de gênero e raça, considerando os impactos sobre mulheres, pessoas negras, indígenas, quilombolas, LGBTQIA+, pessoas com deficiência, pessoas defensoras de direitos humanos e demais grupos historicamente marginalizados.
- Adequação às obrigações previstas no Decreto 12.976, especialmente no que tange à vedação da geração de imagens íntimas de terceiros via sistemas de IA.
- Incorporar, nos processos de criação e atualização dos modelos de inteligência artificial, critérios específicos para identificar e mitigar riscos de uso dessas tecnologias na produção, amplificação ou reprodução da violência política de gênero e raça, assegurando transparência sobre as medidas adotadas.

- Instituir mecanismos permanentes de diálogo com organizações da sociedade civil, pessoas pesquisadoras e especialistas, capazes de subsidiar a atualização contínua das políticas e dos sistemas de governança, considerando as transformações das dinâmicas de violência política no ambiente digital.

DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL

- Oferta de respostas confiáveis e verificáveis sobre clima e meio ambiente. Os sistemas devem utilizar fontes atualizadas e confiáveis ao responder sobre ciência climática, legislação ambiental, desmatamento, eventos extremos, direitos territoriais e políticas socioambientais, especialmente quando esses temas integrarem debates eleitorais. As respostas devem indicar fontes, distinguir evidências de interpretações e reconhecer limites, incertezas ou ausência de dados.
- Impedimento da produção de desinformação climática e socioambiental em escala. As empresas devem bloquear usos destinados a fabricar falsas evidências de desastres, documentos atribuídos a órgãos ambientais, declarações inventadas, campanhas de *greenwashing*, falsas soluções ou grandes volumes de mensagens que minimizem riscos e atrasem respostas efetivas. As salvaguardas devem ser reforçadas quando houver relação com candidaturas, campanhas, debates legislativos ou disputas territoriais.
- Garantia de procedência, testes e transparência. As empresas devem adotar mecanismos que permitam identificar a origem e a natureza sintética de imagens, vídeos e áudios gerados ou manipulados por inteligência artificial, preservando essas informações quando o conteúdo for compartilhado ou redistribuído. Antes e durante o período eleitoral, os sistemas devem ser submetidos a testes em português e em diferentes contextos regionais brasileiros, com divulgação dos riscos identificados, das salvaguardas adotadas e dos incidentes relevantes.

AO ESTADO BRASILEIRO

INTEGRIDADE ELEITORAL

- Fortalecimento das salvaguardas à liberdade de expressão, ao acesso à informação e ao pluralismo no ecossistema informacional, promovendo um ambiente digital aberto e democrático, com a garantia de uma imprensa livre, plural e independente, bem como a proteção de jornalistas, pessoas pesquisadoras, ativistas e demais atores da sociedade civil que atuam na

produção, circulação e fiscalização de informações de interesse público.

- Promoção da cooperação interinstitucional e multissetorial em prol da integridade eleitoral e fortalecimento democrático, articulando o Judiciário, Executivo, Legislativo, academia, sociedade civil e plataformas digitais, com definição de prioridades e mecanismos de cooperação.
- Aprimoramento de mecanismos regulatórios que possam promover transparência e prestação de contas das empresas de plataformas digitais, estabelecendo obrigações compatíveis com o porte e os riscos dos serviços, especialmente quanto à moderação de conteúdo, anúncios eleitorais, sistemas de recomendação e uso de inteligência artificial.
- Incentivo a pesquisas sobre o impacto das plataformas digitais sobre a integridade eleitoral e democracia, assegurando mecanismos de acesso a dados para pessoas pesquisadoras, observados os direitos fundamentais e a proteção de dados pessoais.
- Fortalecimento de programas permanentes de educação midiática e cidadania digital, integrando competências de letramento digital, combate à desinformação, participação cívica e uso responsável da inteligência artificial.
- Acompanhamento da implementação das obrigações estabelecidas nos decretos 12.975 e 12.976 e na decisão do STF em relação ao Artigo 19 do Marco Civil da Internet (RE 1.037.396 e RE 1.057.258)
- Desenvolvimento de iniciativas de educação midiática e letramento digital:
 - Educação midiática e letramento digital para identificar, prevenir e responder a desinformação, discurso de ódio e usos maliciosos da IA que restrinjam a participação política de grupos marginalizados, incluindo desinformação climática e socioambiental.

VIOLÊNCIA POLÍTICA DE GÊNERO E RAÇA

- Fortalecimento da implementação da Lei nº 14.192/2021 e dos demais instrumentos nacionais voltados à proteção da participação política, incorporando de forma explícita a dimensão digital da violência política de gênero e raça em protocolos, diretrizes e políticas públicas de prevenção e enfrentamento.
- Estabelecimento de diretrizes regulatórias que incentivem plataformas digitais, canais de mensageria e empresas desenvolvedoras de inteligência artificial a

adotar políticas específicas para prevenção, monitoramento e enfrentamento da violência política de gênero e raça, assegurando transparência, prestação de contas e respeito aos direitos humanos.

- Fortalecimento da atuação articulada entre os órgãos responsáveis pela proteção de pessoas defensoras de direitos humanos, autoridades eleitorais, órgãos de segurança pública e instituições de justiça para prevenir, investigar e responsabilizar casos de violência política de gênero e raça praticados no ambiente digital.
- Produção, sistematização e divulgação de dados públicos sobre violência política de gênero e raça no ambiente digital, promovendo pesquisas e mecanismos permanentes de monitoramento que subsidiem a formulação, a implementação e a avaliação de políticas públicas.
- Investimento na formação continuada de agentes públicos, integrantes do sistema de justiça, forças de segurança, autoridades eleitorais e demais órgãos competentes sobre violência política de gênero e raça, desinformação, discurso de ódio, riscos associados à inteligência artificial e seus impactos sobre a democracia e os direitos humanos.
- Incentivo ao desenvolvimento e a adoção de mecanismos de transparência, avaliação e mitigação de riscos relacionados ao uso de sistemas de inteligência artificial em contextos eleitorais, especialmente quanto aos impactos sobre grupos historicamente marginalizados, assegurando que a inovação tecnológica esteja alinhada à proteção dos direitos fundamentais e da integridade democrática.

DESINFORMAÇÃO CLIMÁTICA E SOCIOAMBIENTAL

- Institucionalização da defesa da integridade da informação climática e socioambiental como política de Estado. O Estado brasileiro deve assegurar que a promoção da integridade da informação climática e socioambiental seja incorporada de forma permanente às políticas públicas, com estruturas, responsabilidades, recursos e mecanismos de cooperação que não dependam de governos ou períodos eleitorais específicos. Essa atuação deve articular órgãos ambientais, científicos, de comunicação pública, defesa civil e direitos humanos, considerando os riscos eleitorais e as diferentes realidades regionais.
- Fortalecimento da informação pública e dos protocolos de comunicação durante emergências. Órgãos públicos devem produzir e divulgar dados atualizados, acessíveis e territorializados sobre eventos extremos, prevenção de desastres,

fiscalização, desmatamento, direitos territoriais e políticas socioambientais. Durante o período eleitoral, devem manter canais rápidos e coordenados de esclarecimento sobre falsas orientações, fraudes e distorções de causas, responsabilidades e medidas de resposta.

À JUSTIÇA ELEITORAL

- Reconhecimento da desinformação climática e socioambiental e da violência de gênero e raça nas políticas e estratégias da Justiça Eleitoral.
- Desenvolvimento de mecanismos de transparência e publicação de dados com publicação periódica de métricas agregadas sobre salvaguardas das plataformas (tentativas de violação, bloqueios, denúncias, decisões automatizadas).
- Exigência de transparência em bibliotecas de anúncios para publicidade temática socioambiental (financiadores, valores, segmentação, alcance).
- Preservação de dados sobre anúncios, recomendação, monetização e contas coordenadas.
- Garantia de acesso supervisionado e protegido de pessoas pesquisadoras e organizações da sociedade civil a dados preservados pelas plataformas digitais, incluindo registros relevantes sobre circulação e alcance de conteúdos, sistemas de recomendação, publicidade política, aplicação de políticas e moderação, contas e redes coordenadas e outros dados necessários à compreensão do ambiente informacional eleitoral e supervisão dos planos de conformidade.
- Criação de instância consultiva permanente com especialistas, cientistas, sociedade civil e territórios afetados.
- Continuidade dos protocolos de cooperação com plataformas digitais e empresas de IA, com procedimentos, responsáveis e prazos de resposta definidos.
- Capacitação de agentes públicos com o desenvolvimento de iniciativas de formação continuada sobre violência política de gênero e raça, desinformação (incluindo climática e socioambiental), discurso de ódio e impactos da IA nos processos eleitorais.

À SOCIEDADE CIVIL

- Monitoramento permanente e produção de evidências:
 - Monitoramento de plataformas digitais e sistemas de IA, produzindo evidências sobre os impactos sobre a integridade eleitoral, a violência política de gênero e raça, assim como a desinformação ambiental, seus impactos e a efetividade das políticas das empresas.
 - Criação de uma rede especializada de monitoramento (organizações socioambientais, universidades, jornalistas, agências de checagem, entidades científicas, representantes dos territórios) para identificar narrativas emergentes e preservar evidências, com encaminhamento ao TSE e às empresas.
 - Acompanhamento de conteúdos, anúncios, canais públicos de mensageria e materiais produzidos por IA.
 - Acompanhamento da implementação das obrigações estabelecidas nos decretos 12.975 e 12.976 e na decisão do STF em relação ao Artigo 19 do Marco Civil da Internet (RE 1.037.396 e RE 1.057.258)
- Desenvolvimento de mais incidência e investigação estrutural:
 - Incidência junto a plataformas, empresas de IA e poder público para reconhecimento formal da violência política de gênero e raça como fenômeno específico, com políticas, protocolos e mecanismos de responsabilização.
 - Investigação da cadeia produtiva da desinformação incluindo financiadores, anunciantes, beneficiários de receita, redes de distribuição para compreendê-la como estrutura econômica e política, não apenas como conteúdo isolado.
- Desenvolvimento de redes de apoio e resposta a pessoas afetadas:
 - Criação ou manutenção de redes de apoio e resposta rápida para pessoas e grupos afetados por violência política de gênero e raça: segurança digital, documentação de casos, orientação sobre denúncia, articulação institucional.
 - Respostas preventivas e proteção às pessoas atingidas por desinformação climática e socioambiental incluindo segurança digital, apoio jurídico, preservação de provas, suporte de comunicação a povos, comunidades, cientistas, jornalistas e pessoas defensoras ambientais.
 - Preparação antecipada de narrativas previsíveis e protocolos para eventos extremos.

Signatários

1. ABPEducom - Associação Brasileira de pesquisadores e Profissionais em Educomunicação
2. Aláfia Lab
3. Aliança Brazil Office
4. Alumia - Laboratório de Combate à Desinformação
5. ANDA - Agência de Notícias de Direitos Animais
6. Apuraí - Núcleo de Combate à desinformação (UFJF)
7. Aqualtune Lab
8. ARTIGO 19 Brasil e América do Sul
9. Associação Brazil Office
10. Associação Soluções Inclusivas Sustentáveis (SIS)
11. Bereia - Informação e Checagem de Notícias
12. Business and Human Rights Centre - BHRC
13. Casa Galileia
14. Central Única dos Trabalhadores
15. Centro de Estudos da Mídia Alternativa Barão de Itararé
16. Centro de Defesa dos Direitos Humanos Petrópolis
17. Coding Rights
18. Coletivo de Jornalistas RJ
19. Comissão Pró-Índígenas do Acre (CPI-Acre)
20. COMP - Grupo de Pesquisa em Comunicação Internet e Política da PUC-Rio
21. Conectas Direitos Humanos
22. CPA - Centro Popular de Comunicação e Audiovisual
23. CTRL+Z
24. Cyber Leviathan - Observatório do Mundo em Rede
25. Data Privacy Brasil
26. Digital Action
27. DiraCom - Direito à Comunicação e Democracia
28. Elas no Poder
29. FALA - Estúdio de Impacto
30. Federação Nacional dos Jornalistas - FENAJ
31. Fundo de Apoio ao Jornalismo
32. Grupo de pesquisa Comunicação Socioambiental dos Biomas Brasileiros desde a Amazônia (UFRR)
33. landé Comunicação e Educação
34. Idec - Instituto de Defesa de Consumidores
35. Instituto Brasileiro de Segurança Digital - IBRASEG
36. Instituto de Referência em Internet e Sociedade (IRIS)
37. Instituto Democracia em Xequê

38. Instituto Eledewá
39. Instituto Lamparina
40. Instituto Marielle Franco
41. Instituto Nossa Ilhéus
42. Instituto Novos Paradigmas
43. Instituto Update
44. Instituto Vladimir Herzog
45. Intercom - Sociedade Brasileira de Estudos Interdisciplinares da Comunicação
46. International Center for Information Ethics - ICIE
47. Justiça Global
48. Movimento Mulheres Negras Decidem
49. Movimento Negro Evangélico do Brasil - MNE
50. NJA - Núcleo de Jornalismo e Audiovisual
51. NetLab UFRJ
52. Núcleo de Estudos em Jornalismo de Dados e Computacional (DataJor)
53. OBSERVA - Observatório de Conflitos Online
54. Observatório da Ética Jornalística (objETHOS)
55. Observatório de Desinformação e Intolerância
56. Observatório de Mídia do Curso de Jornalismo da UFRRJ
57. Observatório Paraibano de Jornalismo
58. Open Knowledge Brasil (OKBR)
59. Oxfam Brasil
60. Rede Nacional de Combate à Desinformação (RNCD)
61. Rede Pela Soberania
62. Redes Cordiais
63. Ruídos MediaLab - Laboratório de Pesquisa em Jornalismo, Informação Política e Cultura Democrática da PUC-Rio
64. Sleeping Giants Brasil
65. SOCICOM - Federação Brasileira das Associações Científicas e Acadêmicas de Comunicação (SOCICOM)
66. Território Mídias Brasil