X Brasil Internet Ltda.

PLANO DE CONFORMIDADE

Eleições 2026

Portaria TSE nº 463, de 27 de julho de 2026, e Resolução TSE nº 23.610/2019

1. INTRODUÇÃO

O presente Plano de Conformidade ("**Plano**") é apresentado por X Brasil Internet Ltda. ("**X BRASIL**"), sociedade limitada inscrita no CNPJ sob nº 16.954.565/0001-48, com sede na Cidade de São Paulo, Estado de São Paulo, na Avenida Imperatriz Leopoldina, nº 1.248, sala 203, BA088, em cumprimento ao artigo 125-B da Resolução nº 23.610/2019, conforme alterada pela Resolução nº 23.755/2026[1] ("**Resolução**"), regulamentado pela Portaria nº 463, de 27 de julho de 2026 ("**Portaria**"), ambas do Tribunal Superior Eleitoral ("**TSE**").

O Plano descreve as medidas técnicas e organizacionais empregadas para prevenir e mitigar riscos à integridade do processo eleitoral e para assegurar o cumprimento das obrigações previstas na Resolução, nos termos dos artigos 3º, incisos I e III, e 4º da Portaria.

Em atenção ao artigo 2º da Portaria, os serviços abrangidos por este Plano são a rede social X e o *chatbot* de inteligência artificial generativa "Grok".

O X é uma plataforma virtual de informação de uso gratuito, alimentada exclusivamente pelos usuários, que permite o compartilhamento em tempo real de publicações sobre assuntos variados – mensagens contendo imagens, vídeos, *links* e textos. Como condição para utilizar a plataforma, o usuário deve criar uma conta por meio do *site* x.com, mediante aceitação dos

[1] Art. 125-B. "Os provedores de aplicação de internet deverão, nos termos, prazos e diretrizes editados pela Presidência do Tribunal Superior Eleitoral, elaborar plano de conformidade destinado à prevenção e à mitigação de riscos à integridade do processo eleitoral e ao efetivo cumprimento das obrigações previstas nos arts. 9º-B, 9º-C, 9º-D, 9º-E, 27-A, 28, 29, 30, 32, 33, 33-A, 33-B, 34, 36, 38, 39 e 40 desta Resolução. § 1º O plano de conformidade deverá conter, no mínimo: I - deveres e medidas de conformidade apontadas para cada disposição desta Resolução; II - critérios e indicadores mensuráveis e periódicos para acompanhamento de sua implementação; III - prazos, resultados esperados e trajetória de alcance. § 2º O cumprimento do disposto neste artigo constitui requisito para o credenciamento de que trata o art. 27-A, § 4º, e para o cadastro previsto no art. 29, § 9º desta Resolução. § 3º A Portaria a que se refere o caput deste artigo será publicada no Diário de Justiça Eletrônico." (Incluído pela Resolução nº 23.755/2026.)

Termos do Serviço[2] e da Política de Privacidade[3], que constituem os contratos de uso da ferramenta, estabelecem direitos e obrigações para as partes e dão aos usuários transparência a respeito das medidas técnicas e organizacionais empregadas na plataforma.

A plataforma X é operada e provida pelas empresas X Corp. e X Internet Unlimited Company ("**Operadoras do X**"), a depender da localização dos usuários. Os usuários localizados nos Estados Unidos, no Brasil e em qualquer outro país fora da União Europeia ou do Espaço Econômico Europeu contratam com a empresa norte-americana X Corp. Os usuários localizados nos demais países contratam com a empresa irlandesa X Internet Unlimited Company.

O Grok constitui modelo de inteligência artificial generativa desenvolvido pela SpaceXAI LLC ("**SpaceXAI**"), sociedade empresária distinta, responsável pelo treinamento, desenvolvimento, atualização e funcionamento desse sistema. Entre as formas de disponibilização do Grok encontra-se a interface "Grok no X", o site grok.com e aplicativos móveis para iOS e Android. Sua utilização por site e aplicativos se dá mediante aceite aos Termos de Serviço[4] da xAI e de sua Política de Uso Aceitável[5]. As medidas de conformidade aplicáveis ao Grok estão descritas na Seção 8 deste Plano.

O X BRASIL é empresa brasileira, sediada em São Paulo/SP, dotada de personalidade jurídica própria, autônoma e independente das Operadoras do X e da SpaceXAI. Justamente por isso, **não** possui meios técnicos ou jurídicos para intervir na operação e no funcionamento do X e do Grok, prerrogativas exclusivas das Operadoras do X[6] e da SpaceXAI.

Sem prejuízo dessas distinções, o X BRASIL atua na interlocução institucional e jurídica necessária ao cumprimento das obrigações eleitorais, em coordenação com as Operadoras do X e com a SpaceXAI.

A elaboração deste Plano observa as premissas interpretativas estabelecidas pelo artigo 4º da Portaria, que orientam tanto a forma quanto o conteúdo das informações apresentadas:

[2] https://x.com/tos
[3] https://x.com/privacy
[4] https://x.ai/legal/terms-of-service
[5] https://x.ai/legal/acceptable-use-policy
[6] As Operadoras do X são consideradas provedoras de aplicações de internet e, portanto, estão submetidas à Lei nº 12.965/2014 ("Marco Civil da Internet"), regulamentada pelo Decreto nº 8.771/2016.

(a) **Autonomia técnica:** as soluções e controles descritos refletem as escolhas técnicas e operacionais das Operadoras do X, compatíveis com o estado da arte, a arquitetura e o funcionamento da plataforma;

(b) **Adequação finalística:** cada medida é descrita em função do resultado concreto que visa alcançar no contexto da proteção da integridade do processo eleitoral e da democracia;

(c) **Transparência:** são fornecidas informações objetivas e suficientes para permitir a verificação de conformidade pelo TSE;

(d) **Proporcionalidade:** as medidas são calibradas de forma a equilibrar a proteção do processo eleitoral com o exercício legítimo da liberdade de expressão e demais direitos fundamentais dos usuários; e

(e) **Verificabilidade:** o Plano indica mecanismos de controle, indicadores e canais que permitem ao TSE aferir o efetivo cumprimento das obrigações assumidas.

O Plano está organizado em três blocos. As Seções 2 apresenta a avaliação de impacto que orienta a abordagem baseada em risco. As Seções 3 a 11 apresentam as medidas de conformidade organizadas por categoria de dever, incluindo as políticas e a arquitetura de moderação que dão suporte às demais medidas. A Seção 12 trata da atualização do Plano e do relatório consolidado, e a Seção 13 apresenta a conclusão.

A correspondência entre as obrigações da Resolução, os dispositivos da Portaria e as seções deste Plano é a seguinte:

Dispositivo da Resolução	Tema	Dispositivo da Portaria	Seção do Plano
Art. 9º-B, Art. 9º-C e 28 §1º-C	Gestão de IA, conteúdo sintético e período de vedação	Arts. 15 e 16	8; 9
Art. 9º-D, caput e I	Deveres de diligência; instrumentos normativos e arquitetura de moderação	Arts. 22 e 23	3
Art. 9º-D, II	Instrumentos eficazes de notificação e de canais de denúncia	Art. 17	7
Art. 9º-D, III e IV	Planejamento e a execução de ações corretivas e preventivas, incluindo o aprimoramento de seus sistemas de recomendação de conteúdo, e transparência	Arts. 13, 14 e 23	3; 4; 5

Art. 9º-D, V	Avaliação de impacto sobre a integridade do processo eleitoral	Art. 20	2
Art. 9-E, caput	Violações eleitorais e ações proativas	Art. 13	3
Art. 9º-E, § 1º, § 2º, 32, 36, 38, § 6º, 39 e 40	Cumprimento de ordens e determinações da Justiça Eleitoral e canais de interação	Art. 12 e 17	6; 7
Art. 27-A, 28, 29, caput e § 11, 30	Publicidade política paga	Arts. 13, 15, 18 e 19	8, 9 e 10
Art. 33, 33-A, 33-B	Proteção de dados pessoais	Art. 21	11
Art. 34	Vedação de propaganda por disparo em massa com tecnologias externas	Art. 14 e 22, parágrafo único	4

2. AVALIAÇÃO DE IMPACTO SOBRE A INTEGRIDADE DO PROCESSO ELEITORAL (Resolução, art. 9º-D, V; Portaria, art. 20)

Medida de conformidade. O X realiza avaliação de impacto sobre a integridade do processo eleitoral, conduzida por equipes multifuncionais das Operadoras do X, com participação das áreas de segurança, engenharia de produto, jurídica e assuntos governamentais. A avaliação é elaborada em anos eleitorais no Brasil e observa três fases: (i) identificação dos riscos decorrentes do desenho, do funcionamento e do uso do serviço por terceiros; (ii) avaliação do risco inerente (probabilidade × severidade), da força de controle e do risco residual; e (iii) inventário dos controles, das melhorias e dos indicadores de mitigação. A severidade desdobra-se em escopo do impacto, escala ou alcance potencial e remediabilidade. Os indicadores considerados incluem sinais quantitativos – volumes e taxas de *enforcement* nas políticas eleitoralmente relevantes, tempos de tratamento e métricas de redução de exposição – e qualitativos – contexto político-eleitoral brasileiro, superfícies do produto, inclusive integrações de inteligência artificial generativa, lições de ciclos anteriores e referências externas de integridade informacional.

Prazo. A avaliação de impacto relativa às Eleições Gerais de 2026 já foi elaborada. Confira-se anexo.

Resultado esperado. Ter uma avaliação com matriz de riscos e indicadores de mitigação aferíveis ao longo do ciclo eleitoral, de modo a orientar o tratamento dos riscos.

Trajetória. Os controles eleitorais específicos descritos abaixo serão ativados progressivamente, com previsão de que todos estejam plenamente operacionais até 4 de outubro de 2026; os indicadores serão monitorados durante o período eleitoral e reportados no relatório consolidado do artigo 28 da Portaria, nos termos do artigo 7º, § 4º.

A avaliação parte do reconhecimento de que a plataforma X constitui um espaço relevante para o debate público e o discurso cívico, inclusive em contexto eleitoral. O X oferece oportunidades de participação nos processos democráticos ao permitir que cidadãos acessem informações, expressem suas opiniões e se organizem no âmbito da sociedade civil.

Contudo, a influência das redes sociais também implica riscos aos processos democráticos, às eleições e ao discurso cívico, caso impactem a confiança pública nas instituições, a capacidade de participação livre dos cidadãos ou o exercício de direitos fundamentais e políticos que são a base de qualquer democracia.

A definição dessa área de risco é abordada por meio de um marco de direitos fundamentais, que analisa o impacto sobre direitos individuais que, se violados em escala, podem representar risco sistêmico. São relevantes, nesse contexto: (1) a liberdade de expressão e de informação, que inclui a liberdade e o pluralismo dos meios de comunicação; (2) o direito de voto; (3) o direito à autodeterminação; e (4) a participação cívica no debate público.

2.1 Vetores de risco identificados

Com base nessa metodologia, a avaliação de impacto identificou doze riscos à integridade do processo eleitoral.

1. **Informações enganosas sobre o processo eleitoral** — conteúdo falso ou enganoso sobre quando, onde ou como participar, ou sobre resultados, com potencial de distorcer decisões cívicas e minar a confiança nas instituições.
2. **Intimidação e supressão da participação cívica, inclusive violência política de gênero** — ameaças, assédio e violência política de gênero capazes de inibir a participação, com impacto particular sobre mulheres e grupos vulneráveis.
3. **Amplificação por contas inautênticas, *spam*, *astroturfing* e atividade coordenada** — condutas destinadas a inflar hashtags, tendências e narrativas e a distorcer o debate público.

4. **Operações de informação** — criação ou aquisição de contas inautênticas, comprometimento de contas estabelecidas e denúncias de má-fé, com potencial de minar a confiança em resultados e instituições.
5. **Pedidos governamentais e vigilância politicamente motivados** — requisições excessivas ou politicamente motivadas que podem resultar em censura, restrição de contas ou cerceamento da participação.
6. **Funcionalidades de design que viabilizam o compartilhamento prejudicial** — posts, Spaces, mensagens diretas, Assuntos do Momento, aba Para Você, notificações e integrações de IA generativa podem ser explorados para espalhar conteúdo prejudicial.
7. **Efeito colateral das políticas de conteúdo sobre a liberdade de expressão** — aplicação excessiva da Política de Integridade Cívica, das regras contra violência e abuso ou da vedação de anúncios políticos pode restringir discurso político legítimo.
8. **Sistemas de recomendação que amplificam conteúdo prejudicial** — recomendações podem elevar conteúdo ainda não sinalizado, reduzir o alcance de fontes pluralistas e formar bolhas de filtro.
9. **Erros dos sistemas de moderação (falsos positivos/negativos, viés e evasão)** — falhas de automação e de revisão humana, vieses de IA, dificuldade com contexto e táticas de evasão.
10. **Anúncios violadores ou enganosos e segmentação sensível** — risco residual de anúncios que violem as políticas, inclusive tentativas de contornar a vedação de anúncios políticos no Brasil.
11. **Mídia sintética e manipulada, inclusive *deepfakes* eleitorais** — áudio ou vídeo falso de candidatos ou autoridades, com potencial de distorcer a informação do eleitor.
12. **Apropriação de identidade e identidades enganosas** — contas que se passam por candidatos, autoridades ou veículos, enganando eleitores e prejudicando a fala institucional autêntica.

2.2 Controles e medidas de mitigação mapeados

Frente aos vetores de risco identificados, a avaliação de impacto mapeou os controles disponíveis – aqueles já implementados e aqueles aplicáveis ao período eleitoral –, na forma do inventário da própria avaliação (art. 20, II). A abordagem tem dois eixos – conteúdos e atores – e quatro premissas: os direitos humanos como fundamento da elaboração e da iteração de políticas e de *enforcement*; a gravidade, a iminência e a probabilidade do dano como fatores que

informam as diretrizes e os processos de aplicação; a justiça procedimental como garantia de aplicação equânime das regras e de resultados justos; e a exigência de que os resultados de *enforcement* sejam necessários e proporcionais. Cada tópico abaixo corresponde ao inventário da avaliação de impacto; o detalhe operacional consta da seção deste Plano em que o controle opera como medida de conformidade.

- **Políticas e sua aplicação (*enforcement*)**. As políticas que mitigam os riscos eleitorais são, em especial: Integridade Cívica, que veda o uso do serviço para manipular ou interferir em eleições, inclusive conteúdo que suprima participação, induza a erro sobre quando, onde ou como participar, ou conduza a violência offline; Conteúdo Não Autêntico, que veda mídia sintética, manipulada ou fora de contexto enganosa e admite rotulagem; Identidades Enganosas, que veda apropriação de identidade e personas fabricadas para enganar; Comportamentos Não Autênticos, que veda ampliação ou supressão artificial e condutas que manipulem a experiência ou as defesas da plataforma; Conteúdo Violento (discurso violento); Conduta de Propagação de Ódio; e Abuso e Assédio. A aplicação observa abordagem "info-first", em que o moderador responde a uma série de perguntas objetivas em vez de decidir de imediato, para reduzir subjetividade e viés. Usuários podem denunciar posts, perfis e mensagens – inclusive por página de recursos e canal específicos para o Brasil – e contestar *enforcement* que considerem equivocado. As medidas graduadas incluem restrição de alcance (exclusão de busca, tendências e recomendações; remoção das *timelines*; limitação ao perfil; rebaixamento em respostas; restrição de engajamento), exigência de remoção do post e suspensão de conta. Termos e políticas relevantes são publicados em português.

- **Autenticidade e uso inautêntico**. Contas devem ser autênticas. É vedada a criação, a operação ou o registro em massa de contas que não sejam legítimas, genuínas e transparentes quanto à origem, à identidade e à popularidade.

- **Design e funcionalidades**. Controles de produto mitigam o uso prejudicial das superfícies da plataforma: verificação de usuário (Premium e, de forma obrigatória, criadores em programas de receita); verificação por assinatura e selos (cinza para governo e organismos multilaterais; dourado para organizações verificadas; afiliados); rotulagem de URLs inseguras; bloqueio, silenciamento, posts protegidos, filtros de notificação e controle de quem pode responder, inclusive ocultação de respostas; preferências de pedidos de mensagem direta; *denylist* de termos nos Assuntos do

Momento, na busca e no *autocomplete*; proteções em Live (detecção de mídia sensível e denúncia do espectador) e em Spaces (detecção de títulos e transcrições tóxicos, encerramento de transmissão, ocultação de gravações); controles em Communities; likes privados; e salvaguardas em camadas do Grok – filtragem de treino, políticas de recusa, filtros de entrada e saída, *red teaming* e monitoramento contínuo.

- **Sistemas de recomendação**. O X aplica requisitos de elegibilidade antes de recomendar conteúdos e contas; exclui das recomendações violações das Regras, Integridade Cívica, mídia sintética, spam e conteúdo gráfico ou adulto; emprega modelos de segurança (safety content models); mantém *denylist* de tendências; oferece opções não personalizadas; e realiza testes A/B, verificação de modelos e implantação incremental, com possibilidade de ligar e desligar componentes. O código-fonte dos principais componentes é público, o que permite escrutínio dos sinais, dos controles de elegibilidade e segurança e da inexistência de fatores explícitos de promoção ou supressão política.

- **Moderação e capacidade eleitoral**. A moderação combina detecção automatizada e revisão humana 24/7 em português, com especialistas de idioma e contexto, garantia de qualidade, revisão semanal de capacidade, escalonamento até equipes especializadas e, quando necessário, à liderança, e protocolo de resposta a incidentes com retrospectiva. No ciclo eleitoral, o X ativa serviços específicos: varreduras proativas de violação a Regras; canais de escalonamento 24/7; *prompts* informativos na *timeline* e na busca; diretrizes de triagem de demandas judiciais e extrajudiciais; etiquetas de rastreamento de ações; e protocolo de resposta a episódios sensíveis e urgentes.

- **Anúncios**. Anúncios políticos são vedados no Brasil. Permanecem ativos os sistemas que detectam e bloqueiam anúncios políticos direcionados ao país e que impedem o onboarding de contas de candidatos para publicidade. Na criação, modelos de machine learning e *denylist* enviam o anúncio a revisão humana; anúncios detectados são interrompidos ou restringidos. Escalonamentos de anúncios tramitam pelo canal prioritário eleitoral.

- **Outras medidas**. O programa Notas da Comunidade permite que a comunidade acrescente contexto a posts sobre o processo eleitoral, declarações de agentes públicos e temas cívicos; notas em imagens podem aparecer automaticamente em posts com a mesma imagem e são enviadas a quem curtiu, republicou ou respondeu. O

botão de análise do Grok permite ao usuário pedir resumo, explicação ou contextualização de um post. O X processa ofícios e ordens de autoridades brasileiras, inclusive eleitorais, e pode aplicar restrições ao conteúdo quando o pedido é válido e o conteúdo não viola os Termos, mas viola lei local.

2.3 Cobertura dos doze riscos

A correspondência entre os vetores e os controles acima é a seguinte:

1. **Informações enganosas sobre o processo eleitoral** — Política de Integridade Cívica, Notas da Comunidade, botão de análise do Grok, abordagem "info-first" e suporte prioritário de eleições.
2. **Intimidação e supressão da participação cívica, inclusive violência política de gênero** — Políticas de Conteúdo Violento, de Conduta de Propagação de Ódio e de Abuso e Assédio, denúncia 24/7, protocolo de resposta a incidentes e controles do usuário.
3. **Amplificação por contas inautênticas, *spam*, *astroturfing* e atividade coordenada** — Política de Comportamentos Não Autênticos, exigência de autenticidade, detecção humana e automatizada, varreduras e *denylist*.
4. **Operações de informação** — os mesmos controles do item 3, acrescidos da Política de Identidades Enganosas, do escalonamento e da resposta estratégica.
5. **Pedidos governamentais e vigilância politicamente motivados** — princípios de necessidade e proporcionalidade, restrição de conteúdo no país apenas mediante pedido válido e processamento de ofícios e ordens.
6. **Funcionalidades de design que viabilizam o compartilhamento prejudicial** — controles de Spaces, Comunidades, mensagens diretas, Assuntos do Momento, Live e Grok.
7. **Efeito colateral das políticas de conteúdo sobre a liberdade de expressão** — princípios de necessidade e proporcionalidade, abordagem "info-first", recursos de contestação, garantia de qualidade e orientação de aplicação (*enforcement*).
8. **Sistemas de recomendação que amplificam conteúdo prejudicial** — requisitos de elegibilidade, modelos de segurança (*safety content models*), *denylist* de tendências, testes e código-fonte público.
9. **Erros dos sistemas de moderação (falsos positivos/negativos, viés e evasão)** — abordagem "info-first", combinação de revisão humana e automação, garantia de qualidade, contestação e revisão de capacidade.

10. **Anúncios violadores ou enganosos e segmentação sensível** — vedação de anúncios políticos no Brasil, detecção, revisão humana e canal prioritário eleitoral.
11. **Mídia sintética e manipulada, inclusive *deepfakes* eleitorais** — Política de Conteúdo Não Autêntico, Notas da Comunidade, rótulos “Feito com IA”, varreduras e salvaguardas do Grok.
12. **Apropriação de identidade e identidades enganosas** — Política de Identidades Enganosas, verificação e selos, e varreduras de falsificação de identidade.

2.4 Matriz de risco

A matriz abaixo consolida, para cada risco, probabilidade (P), severidade (S), risco inerente, força de controle, risco residual e prioridade de tratamento. O alcance potencial integra a severidade, como componente de escala. A descrição qualitativa de cada vetor segue a matriz.

#	Risco	P	S	R Inerente	Controle	Residual	Prioridade
1	Informações enganosas sobre o processo eleitoral	4	3	Médio	Gerenciado (2)	Baixo	1
2	Intimidação / supressão da participação cívica (incl. violência política de gênero)	4	4	Alto	Gerenciado (2)	Baixo	1
3	Amplificação via contas inautênticas, spam, astroturfing e atividade coordenada	4	3	Médio	Gerenciado (2)	Baixo	1
4	Operações de informação	3	2	Baixo	Gerenciado (2)	Desprezível	1
5	Solicitações governamentais e vigilância	3	2	Baixo	Gerenciado (2)	Desprezível	3
6	Recursos de design que possibilitam o compartilhamento prejudicial	3	2	Baixo	Definido (3)	Baixo	2
7	Efeito colateral das políticas de conteúdo sobre a liberdade de expressão	3	2	Baixo	Gerenciado (2)	Desprezível	3
8	Sistemas de recomendação que amplificam conteúdo prejudicial	3	3	Baixo	Definido (3)	Baixo	2
9	Erros do sistema de moderação de conteúdo (FP/FN, viés)	3	2	Baixo	Gerenciado (2)	Desprezível	3
10	Anúncios violadores / enganosos; segmentação sensível	2	2	Desprezível	Gerenciado (2)	Desprezível	2
11	Mídia sintética / manipulada (incl. deepfakes)	4	4	Alto	Gerenciado (2)	Baixo	1
12	Apropriação indevida de identidade / identidades enganosas	3	2	Baixo	Gerenciado (2)	Desprezível	2

Em conclusão, o risco inerente é reconhecido como relevante, especialmente em contexto eleitoral polarizado. O X apresenta, contudo, um conjunto robusto e multicamadas de controles – políticas, produto, moderação, recomendações e medidas eleitorais específicas – destinado a reduzir a probabilidade e o impacto desses riscos, buscando equilíbrio entre liberdade de expressão e integridade dos processos democráticos.

Ademais, como um dos itens de calibração, a avaliação constatou que a participação do X no Brasil é relativamente menor que a de outras plataformas de redes sociais. Dados da Statcounter sobre a participação de mercado de redes sociais no Brasil situam o X (listado como Twitter) em cerca de 4%, bem atrás de Facebook (~45%), Instagram (~28%) e YouTube (~14%).[7] Segundo dados de 2025 do Reuters Institute[8], o X ocupa uma posição baixa como plataforma de compartilhamento de notícias no Brasil (9%), ficando muito atrás de plataformas como YouTube (37%), Instagram (37%), WhatsApp (36%), Facebook (28%) e TikTok (18%). O estudo "Painel TIC – Integridade da Informação 2025", do Cetic.br, também aponta um uso comparativamente baixo do X entre internautas brasileiros com 16 anos ou mais: o uso diário é de cerca de 16% para o X (10ª posição entre as plataformas pesquisadas), contra WhatsApp (91%), Instagram (73%), YouTube (73%), Facebook (57%) e TikTok (50%).[9] Os rankings de tráfego da Semrush para o Brasil corroboram esse padrão: o x.com não figura entre os principais sites em volume de visitas mensais, ficando atrás de YouTube, Instagram, WhatsApp e Facebook.[10]

2.5 Metodologia de aferição das medidas de mitigação (Portaria, art. 20, III)

A aferição das medidas de mitigação utiliza os indicadores abaixo.

Indicador 1 — Acurácia das denúncias de usuários (medida por automação reativa ou reversão em recurso)

O que mede: Acurácia do *enforcement* reativo (manual ou automatizado) em denúncias de usuários e escalonamentos relativos às políticas aplicáveis.

Indicador 2 — Tempo mediano de tratamento

O que mede: Rapidez da revisão e da resposta humanas (filas e escalonamentos).

---

[7] https://www.semrush.com/trending-websites/br/all (Acesso em 14.08.2026)
[8] https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025/brazil (Acesso em 14.08.2026)
[9] https://cetic.br/media/analises/painel_tic_integridade_informacao_apresentacao.pdf (Acesso em 14.08.2026)
[10] https://www.semrush.com/trending-websites/br/all (Acesso em 14.08.2026)

Indicador 3 — Métricas de aplicação para políticas selecionadas (contextual)

O que mede: Volume e composição do *enforcement* no Brasil para políticas relevantes ao período eleitoral — apenas contextual (sem meta de desempenho).

Indicador 4 — Ativação dos controles eleitorais

O que mede: Se os controles específicos da eleição previstos foram ativados até a data prevista e estavam operacionais (não apenas planejados).

Controles previstos (serviços eleitorais)

#	Controle previsto	Ativado e operacional quando
1	**Listas de candidatos, partidos e palavras-chave**	Handles finais de candidatos/partidos e hashtags relevantes carregados para observação
2	**Exercício de mesa**	Simulação pré-eleitoral com as equipes relevantes, prevista para ser concluída até 10 dias antes do primeiro turno
3	**Monitoramento proativo de Segurança**	Monitoramento de informações eleitorais enganosas, intimidação, discurso violento e violações de segurança correlatas ativo conforme o planejado
4	**Proteção contra comprometimento de contas de candidatos de alto perfil**	Medidas de proteção ativadas para o conjunto designado de alto perfil
5	**Canais de escalonamento priorizados 24/7**	*Aliases* jurídicos do Brasil e via interna de escalonamento monitorados; regras de recebimento conhecidas por times jurídico interno e externo e Segurança
6	**Avisos na plataforma (prompts de HTL / Busca)**	*Prompts* eleitorais preparados e ativáveis (ou ativados) conforme o plano de Produto (uma semana antes do primeiro turno e no dia da eleição)
7	**Trabalho prévio jurídico-operacional**	Diretrizes em vigor para triagem/processamento de demandas judiciais e extrajudiciais relacionadas à eleição
8	**Postura de aplicação da proibição de anúncios**	Bloqueio de anúncios políticos e controles de onboarding de contas de anúncios de candidatos confirmados como ativos para o Brasil
9	**Tags de rastreamento das políticas eleitorais**	Tags/notas dedicadas à aplicação no Brasil 2026 por política em uso, para que as ações sejam mensuráveis
10	**Prontidão do protocolo de resposta a prioridades / incidentes**	Via de resposta interfuncional para conteúdo de gravidade extraordinária confirmada como pronta para os fins de semana eleitorais

2.6 Capacidade de intervenção em momentos sensíveis (Portaria, art. 20, parágrafo único)

Nos termos do artigo 20, parágrafo único, da Portaria, a avaliação de impacto inclui análise da capacidade de intervenção sobre conteúdo de extraordinária gravidade em propagação acelerada em momentos sensíveis, em especial os fins de semana do primeiro turno (4 de outubro de 2026) e do segundo turno (25 de outubro de 2026). Nessas ocasiões, as equipes do X operarão com priorização máxima de triagem e *enforcement*, integrando equipe jurídica local, escritórios externos e as áreas globais de Segurança, Operações de Segurança e Relações Governamentais. O acionamento se dá pelos canais de atendimento prioritário a candidaturas, de intimações, de alertas do SIADE e de denúncia na plataforma.

3. POLÍTICAS DE CONTEÚDO POLÍTICO-ELEITORAL, ARQUITETURA DE MODERAÇÃO, MODERAÇÃO PROATIVA (Resolução, art. 9º-D, I; Portaria, art. 22; Resolução, art. 9º-D, caput e III; Portaria, art. 13)

Medida de conformidade. O uso da plataforma X é disciplinado pelos Termos de Serviço[11], pela Política de Privacidade[12] e pelas Regras e Políticas do X[13], conjunto normativo aceito por todo usuário como condição de acesso ao serviço. Dentro desse conjunto, o X mantém políticas diretamente relevantes para o conteúdo político-eleitoral – em especial as Políticas de Integridade Cívica[14], de Autenticidade[15] (abrangendo contas falsas, mídia sintética e manipulada e comportamento de manipulação e *spam*), de Conteúdo Violento[16], de Conduta de Propagação de Ódio[17], de Abuso e Assédio[18] e de Conteúdo Privado (incluindo Nudez Não Consensual)[19] –, aplicadas por arquitetura de moderação que combina detecção automatizada e revisão humana, com gradação de medidas conforme a gravidade e o contexto da violação e com mecanismos de notificação e contestação assegurados aos usuários afetados. As subseções seguintes descrevem, nessa ordem, o processo de elaboração das políticas, os critérios que orientam sua aplicação, o inventário das políticas relevantes, a arquitetura de moderação, o conjunto de medidas disponíveis, o programa Notas da Comunidade[20] e as iniciativas de alfabetização midiática.

3.1 Processo de elaboração e revisão das políticas[21]

O X mantém processo estruturado de elaboração, revisão e aplicação de suas políticas. A criação ou alteração de uma política envolve pesquisas detalhadas sobre tendências de comportamento *online*, o desenvolvimento de linguagem externa clara para definir as expectativas quanto ao que é permitido na plataforma e a criação de orientações internas para as equipes de moderação, dimensionadas para operação em escala global.

---

[11] Termos de Serviço do X. Disponível em: https://x.com/pt/tos.html.
[12] Política de Privacidade do X. Disponível em: https://x.com/pt/privacy.html.
[13] Regras e Políticas do X. Disponível em: https://help.x.com/pt/rules-and-policies.
[14] Política de Integridade Cívica. Disponível em: https://help.x.com/pt/rules-and-policies/election-integrity-policy.
[15] Política de Autenticidade. Disponível em: https://help.x.com/pt/rules-and-policies/authenticity.
[16] Política de Conteúdo Violento. Disponível em: https://help.x.com/pt/rules-and-policies/violent-content.
[17] Política de Conduta de Propagação de Ódio. Disponível em: https://help.x.com/pt/rules-and-policies/hateful-conduct-policy.
[18] Política de Abuso e Assédio. Disponível em: https://help.x.com/pt/rules-and-policies/abusive-behavior.
[19] Política de Informações Privadas. Disponível em: https://help.x.com/pt/rules-and-policies/personal-information. Política sobre nudez não consensual. Disponível em: https://help.x.com/pt/rules-and-policies/intimate-media.
[20] Sobre as Notas da Comunidade no X. Disponível em: https://help.x.com/pt/using-x/community-notes.
[21] https://help.x.com/pt/rules-and-policies/enforcement-philosophy

Os Termos de Serviço e as Regras e Políticas do X são publicados em português; quando os acordos e recursos voltados ao público são atualizados, equipes dedicadas realizam a localização do conteúdo de forma sistemática e baseada em princípios.

A abordagem do X na elaboração de políticas privilegia a contra-argumentação como mecanismo de autocorreção do debate público, incentivando que opiniões divergentes sejam discutidas abertamente na plataforma. O contexto em que o conteúdo é publicado é, por isso, elemento central na avaliação de medidas corretivas.

O processo de revisão de políticas também abrange a identificação de áreas que devem ser cobertas para cumprimento de leis locais, que variam ao redor do mundo. Por meio de estudos da legislação, os times jurídico e de conformidade e segurança identificam possíveis lacunas e adotam as providências para garantia de conformidade local.

3.2 Critérios de aplicação e proporcionalidade

Ao determinar se é necessário adotar uma medida corretiva, o X considera, entre outros, os seguintes fatores: se o comportamento é direcionado a um indivíduo, grupo ou categoria protegida de pessoas; se a denúncia foi registrada pela vítima ou por um espectador; se o usuário já é conhecido por violar as políticas do X; a gravidade da violação; e se o conteúdo pode ser tema de interesse público legítimo.

3.3 Inventário das políticas relevantes para o conteúdo político-eleitoral

Para mitigar os vetores de risco identificados na avaliação de impacto, o X mantém as políticas descritas a seguir. Em cada caso, apresenta-se a função da política no contexto eleitoral e, na extensão em que contribui para demonstrar o conteúdo da medida, o teor das regras correspondentes.

3.3.1 Política de Integridade Cívica[22]

A Política de Integridade Cívica aborda conteúdos enganosos que possam afetar a participação em eleições ou outros processos cívicos, incluindo conteúdo que possa confundir as pessoas sobre como, quando ou onde participar, ou que vise desencorajar a participação eleitoral.

[22] https://help.x.com/pt/rules-and-policies/election-integrity-policy

O X define os processos cívicos como eventos ou procedimentos obrigatórios, organizados e conduzidos pelo órgão governamental e/ou eleitoral de um país, estado, região, distrito ou município para abordar uma questão de interesse comum por meio da participação pública. Eleições políticas, censo, referendos e iniciativas de votação são alguns exemplos de atos cívicos.

Para mitigação dos riscos à integridade desses processos, o X monitora as seguintes categorias de conteúdo:

- **Informações enganosas sobre como participar de uma eleição ou outro processo cívico**, incluindo informações enganosas sobre os procedimentos para participar (por exemplo, que se pode votar por post, mensagem de texto, e-mail ou telefonema em jurisdições onde essas opções não são uma possibilidade); informações enganosas sobre exigências para participação, como identificação ou cidadania; alegações enganosas que causem confusão a respeito de leis, regulamentos, procedimentos e métodos de um ato cívico já estabelecidos, ou sobre as ações de autoridades ou entidades que viabilizam tais atos; e declarações ou informações enganosas sobre a data ou hora oficiais anunciadas de um ato cívico.

- **Informações comprovadamente falsas ou enganosas sobre as circunstâncias que envolvem um processo cívico**, com o intuito de intimidar ou dissuadir as pessoas de participar de uma eleição, incluindo alegações enganosas de que os locais de votação estão fechados, de que a votação terminou ou de que votos não estão sendo contados; sobre atividades policiais ou de aplicação da lei relacionadas à votação, aos locais de votação ou à coleta de informações do censo; e sobre longas filas, problemas com equipamentos ou outras interrupções nos locais de votação.

- **Envolver-se ou promover comportamentos que possam coagir outros a absterem-se de participar de um processo cívico**, incluindo incitar ou promover comportamentos violentos intencionalmente próximo a uma localização onde um processo eleitoral está sendo conduzido (seções eleitorais e locais de apuração); incitar a interrupção ou destruição de procedimentos, infraestrutura ou equipamentos eleitorais necessários para que alguém participe de um processo cívico; incitar outros a assediar eleitores ou trabalhadores das urnas; promover a exposição de armas de fogo perto de locais de votação para intimidar eleitores e funcionários eleitorais; e ameaças aos locais

de votação ou a outros locais ou eventos importantes – hipóteses em que a Política de Discurso Violento pode também se aplicar.

Para apoiar a participação cívica e a transparência, o X utiliza uma combinação de participação da comunidade e moderação da plataforma para promover informações precisas e reduzir a disseminação de conteúdo enganoso. Em determinados casos, reservados a escalonamentos críticos, o X pode aplicar rótulo de Integridade Cívica a publicações que violem a política. As publicações que recebem o rótulo têm seu alcance restringido por meio das medidas de limitação de visibilidade descritas abaixo e exibem aviso informando tanto o autor quanto os leitores de que a visibilidade foi restringida. A etiqueta funciona, assim, simultaneamente como instrumento de notificação e como medida corretiva. Os autores podem recorrer do rótulo caso entendam que seu post foi restringido por engano.

3.3.2 Política de Autenticidade[23]

O X privilegia a autenticidade das experiências dentro da plataforma. De acordo com a Política de Autenticidade, não é permitida nenhuma atividade que tente manipular a plataforma ou interromper os serviços por meio de **contas**, **comportamentos** ou **conteúdos falsos**. A política é estruturada em três dimensões:

(a) Contas falsas. As contas no X devem ser autênticas. Usuários não podem criar, operar ou registrar em massa contas que não sejam legítimas, genuínas e transparentes quanto à sua origem, identidade e popularidade. Isso inclui:

- **Automação não autorizada:** contas automatizadas ou com *script* que não estejam em conformidade com a Política do Desenvolvedor[24]. O usuário é o responsável final pelos aplicativos de terceiros que autoriza a acessar ou usar a sua conta.

- **Personas falsas:** é vedado o uso de identidades fabricadas para comportamentos disruptivos ou enganosos, incluindo o uso de fotos de perfil de banco de imagens, roubadas ou geradas por IA, biografias copiadas ou roubadas e/ou informações de perfil enganosas com o propósito de enganar outras pessoas.

- **Falsa identidade:** não é permitido assumir a identidade de indivíduos, grupos ou organizações para induzir outras pessoas a erro. Embora o usuário não seja obrigado a

[23] https://help.x.com/pt/rules-and-policies/authenticity
[24] https://developer.twitter.com/en/developer-terms/agreement-and-policy

divulgar sua identidade real no perfil, a conta não deve usar informações falsas para se passar por outras pessoas. Contas de paródia, comentário e fã-clube (PCF) são permitidas apenas se o objetivo for discutir, satirizar ou compartilhar informações. Essas contas devem atender, cumulativamente, aos seguintes requisitos: incluir rótulo PCF; não usar avatar idêntico ao da entidade; incluir termos como “paródia”, “falso”, “fã-clube” ou “comentário” no início do nome da conta; e incluir termos dessa natureza na biografia ou descrição de perfil.

- **Múltiplas contas e coordenação:** atividades coordenadas e inautênticas que influenciem artificialmente as conversas ou interrompam o X são proibidas, incluindo: **(i)** a operação de várias contas que interagem com o mesmo conteúdo ou conteúdo substancialmente semelhante, ou com o objetivo de inflar ou manipular a importância de conteúdos e/ou contas – por exemplo, criando várias contas para “impulsionar” tópicos de tendência ou hashtags; interagir com os mesmos posts, contas ou enquetes; usar indevidamente o recurso de menção/resposta; e amplificar uma das próprias contas usando indevidamente os recursos de engajamento; **(ii)** a operação de várias contas que publicam conteúdo substancialmente semelhante ou idêntico entre si, por exemplo, por meio de postagens cruzadas, com conteúdo em várias contas (conteúdo postado em várias contas, mas adaptado em outro idioma, é permitido) e conteúdo semelhante ou duplicado para os mesmos assuntos do momento ou hashtags em destaque; e **(iii)** o emprego de soluções alternativas para ultrapassar os limites técnicos de criação de contas (por exemplo, os limites de número de telefone por conta).

O X permite que os usuários criem e/ou operem até dez contas para fins diferentes e não duplicados. É permitido operar várias contas com identidades, finalidades ou casos de uso distintos que estejam em conformidade com os limites técnicos de criação de contas, como: contas que rastreiam quando objetos no espaço passam sobre um local específico na Terra; contas que compartilham notícias sobre diferentes times de uma mesma liga esportiva; contas para projetos pessoais, *hobbies* ou negócios; contas para entidades de marca específicas para locais ou idiomas exclusivos; o controle de contas em nome de terceiros (por exemplo, gestores de mídias sociais), desde que não haja violação das Regras do X; contas de organizações com comitês ou unidades relacionadas, porém separadas, como uma empresa em vários locais; e a operação de conta pessoal ao lado de contas pseudônimas ou temáticas. Essa delimitação distingue expressamente a atuação coordenada legítima das redes inautênticas.

É vedado tentar contornar as ações de execução ou corretivas no X, por exemplo: criando novas contas; imitando uma conta suspensa para substituí-la; reutilizando uma conta já existente; e tendo outra pessoa operando uma conta em seu nome. Se uma conta for suspensa por violações das Regras X, o X se reserva o direito de também suspender qualquer outra conta que acredite que o mesmo titular da conta ou entidade possa estar operando em violação à nossa suspensão anterior, independentemente de quando a outra conta foi criada.

Em complemento, o acesso não autorizado ou a modificação da conta X de alguém é estritamente proibido. O comprometimento da conta pode ocorrer quando: os dados de login (como credenciais, senhas, tokens, chaves, cookies) são compartilhados com um site ou aplicativos de terceiros potencialmente maliciosos; é feito uso de uma senha fraca ou reutilizada; vírus ou malware foram instalados no computador para coletar senhas; ou é feito login na conta X em uma rede comprometida.

(b) Comportamentos não autênticos. Os usuários não podem se envolver em comportamentos que manipulem o X ou afetem artificialmente a forma como o conteúdo é descoberto e amplificado. Isso inclui:

- **Spam de conteúdo:** os usuários não podem compartilhar ou publicar conteúdo em massa, duplicado, irrelevante ou não solicitado que atrapalhe a experiência das pessoas. Não permitido: o envio de respostas, menções ou mensagens diretas não solicitadas em massa, agressivas e de alto volume; usar *hashtags* populares ou de tendência com a intenção de subverter ou manipular uma conversa ou de direcionar tráfego ou atenção para contas, sites, produtos, serviços ou iniciativas; postar com *hashtags* excessivas e não relacionadas em um único post ou em vários posts; postar ou enviar repetidamente mensagens diretas que consistem em links compartilhados sem comentários, de modo que isso constitua a maior parte de sua atividade de post/mensagem direta; promover conteúdo respondendo com conteúdo irrelevante para o tópico do post original; postar e excluir o mesmo conteúdo repetidamente; postar repetidamente postagens idênticas ou quase idênticas de maneira duplicada, popularmente conhecida como “Copypasta”, ou enviar mensagens diretas idênticas; editar de forma enganosa para amplificar artificialmente o conteúdo ou enganar as pessoas, por exemplo, usando o envolvimento existente de um post, a fim de amplificar um conteúdo substancialmente diferente (por exemplo, editando uma post sobre “O que é melhor? panquecas ou waffles?” para “Milhares de pessoas confiam no meu serviço. Curta meu post e se inscreva no meu canal para receber dicas de investimentos”); e

editar links (URLs) para que a página de destino final mude significativamente, seja no conteúdo ou na localização (por exemplo, domínio, caminho da URL).

- **Spam de engajamento:** o X proíbe o uso não autêntico dos recursos de engajamento para impactar artificialmente o tráfego ou interromper a experiência das pessoas. Não permitido: negociar, comprar, vender (seja por meio de compensação monetária ou virtual) ou solicitar acesso a contas X, incluindo a transferência ou venda temporária ou permanente de contas, nome de usuário ou produtos X (por exemplo, esquemas de "pagamento por afiliação"); coordenar para trocar engajamento em quaisquer recursos X, como curtidas, enquetes, respostas, republicações, listas, visualizações ou seguidores; coordenar com e/ou compensar outros para conduzir a inflação métrica da conta em quaisquer recursos X, incluindo curtidas, enquetes, respostas, republicações, listas, visualizações ou seguidores; envolver-se em "rotatividade de seguidores" – seguir e depois deixar de seguir um grande número de contas em um esforço para inflar sua própria contagem de seguidores; envolver-se em seguir de maneira indiscriminada: seguir e/ou deixar de seguir um grande número de contas não relacionadas em um curto período, especialmente de forma automatizada; interagir com postagens de forma agressiva ou por meio do uso de automação para direcionar tráfego ou atenção para contas, sites, produtos, serviços ou iniciativas; duplicar seguidores de outra conta, principalmente usando automação; usar ou promover serviços de terceiros para realizar qualquer uma das transações descritas nesta política; interagir com os recursos ou formulários do relatório X, seja manualmente ou por meio do uso de automação, para enviar relatórios duplicados ou falsos em grande número, incluindo relatar repetidamente as mesmas postagens ou contas; e coordenar e/ou incentivar outras pessoas a usar indevidamente os recursos de denúncia do X para assediar outras pessoas sob falsos pretextos ou na esperança de restringir ou remover suas contas ou seus posts.

(c) Conteúdo não autêntico. Os usuários não podem compartilhar conteúdos não autênticos no X que possam enganar as pessoas ou causar danos, incluindo:

- **Mídia sintética e manipulada, incluindo mídia fora de contexto**, que possa resultar em confusão generalizada sobre questões públicas, impactar a segurança pública ou causar danos graves (“mídia enganosa”);

- **Mídia compartilhada como autêntica, mas significativa e enganosamente alterada, manipulada ou fabricada** de forma que mude fundamentalmente seu significado e pode resultar em confusão generalizada sobre questões públicas, impactar a segurança pública ou causar danos graves, como mídia substancialmente editada ou pós-processada de maneira que altere sua composição, sequência, tempo ou enquadramento e distorça seu significado; mídia com informações visuais ou auditivas adicionadas, editadas ou removidas (como novos quadros de vídeo, áudio dublado, legendas modificadas) que alterem fundamentalmente sua compreensão; mídia criada, editada ou pós-processada com aprimoramentos ou filtros que alterem fundamentalmente a compreensão, o significado ou o contexto do conteúdo; e mídia que representa pessoa real fabricada ou simulada, especialmente por meio de algoritmos ou inteligência artificial mais ampla;

- **Mídia não manipulada, mas compartilhada de forma enganosa ou fora de contexto**, ou com a intenção de enganar sobre a natureza ou origem do conteúdo e pode resultar em confusão generalizada sobre questões públicas, impacto na segurança pública ou causar danos graves, como mídia apresentada com contexto falso ou enganoso quanto à fonte, ao local, à época ou à autenticidade, quanto à identidade dos indivíduos ou entidades visualmente retratadas na mídia, ou com falsas declarações ou citações do que está sendo dito ou apresentada com declarações fabricadas do fato que está sendo retratado.

- **URLs maliciosos, prejudiciais ou enganosos**, incluindo *links* que possam danificar ou interromper o navegador de outra pessoa (*malware*), comprometer a privacidade (*phishing*) ou redirecionar as pessoas para destinos inesperados ou enganar as pessoas sobre o conteúdo do site.

3.3.3 Política de Conteúdo Violento[25]

A Política de Conteúdo Violento proíbe conteúdos que ameaçam, incitam, glorificam ou expressam desejo por violência ou danos. No contexto eleitoral brasileiro, essa política é relevante para o combate à violência política, à intimidação de eleitores e candidatos e à incitação a atos de violência contra instituições democráticas.

[25] https://help.x.com/pt/rules-and-policies/violent-content

A política define que Conteúdo Violento constitui qualquer conteúdo que contenha Discurso Violento ou Mídia Violenta, com a seguinte divisão e especificação:

(a) Discurso Violento. Abrange todo aquele conteúdo que ameaça, provoca, glorifica ou expressa desejo por violência ou danos. As medidas corretivas aplicadas pelo X incluem:

- **Não é permitido:** o X proíbe o discurso violento quando julga ser de alta severidade, com probabilidade de danos a alguém. Esse conteúdo deve ser removido e violações subsequentes podem fazer com que a conta seja colocada em modo somente leitura ou suspensa. Exemplos incluem declarações explícitas que são:

  – **Ameaças violentas**: ameaças de infligir danos físicos a outras pessoas, o que inclui ameaçar matar, torturar, agredir sexualmente ou ferir alguém de alguma outra forma. Isso também inclui ameaçar danificar casas e abrigos civis, ou infraestrutura essencial para atividades diárias, cívicas, religiosas ou comerciais;

  – **Desejo de prejudicar**: querer, esperar ou expressar desejo de prejudicar alguém. Isso inclui esperar que outros morram, sofram doenças, incidentes trágicos ou sofram outras consequências fisicamente prejudiciais;

  – **Incitação à violência**: incitar, promover ou encorajar outros a cometer atos de violência ou dano, incluindo encorajar outros a se machucarem ou incitar outros a cometer crimes de atrocidade, como crimes contra a humanidade, crimes de guerra ou genocídio;

  – **Glorificação da violência**: glorificar, elogiar ou celebrar atos de violência em que foram causados danos, incluindo expressar gratidão ou elogiar o fato de alguém ter sofrido danos físicos causados por Entidades Violentas. Inclui-se aí exaltar crueldade ou abuso contra animais.

  – O X também proíbe o discurso violento em áreas de alta visibilidade, como vídeos ao vivo, perfil, capa, bio, imagens do banner da Lista ou fotos de capa da Comunidade.

- **Alcance restrito**: o X garante que toma medidas proporcionais com base na gravidade e na probabilidade do dano. Portanto, dependendo do caso, o X pode tornar o conteúdo

violento menos visível com a restrição de alcance no X. Isso inclui avaliar se: o dano é secundário ou não deliberado; o contexto é autodefesa ou conflito militar; o contexto é de indignação ou reação contra autores de danos devastadores; o alvo não está claro; ou uso de linguagem codificada (geralmente referida como "dog whistles", ou apito de cachorro) para provocar violência indiretamente.

- **Permitido**: o X também busca avaliar e compreender as nuances contextuais por trás da conversa antes de tomar providências, e permitimos expressões de discurso violento quando não há contexto abusivo ou violento claro, como:

  - Discurso hiperbólico e consensual entre amigos, ou durante discussões sobre videogames e eventos esportivos;
  - Figuras de linguagem, sátira ou expressão artística quando o contexto expressa um ponto de vista em vez de instigar violência ou dano acionável; ou
  - Citações de livros e filmes, letras de música ou poesias quando o contexto expressa um ponto de vista em vez de instigar atos de violência ou danos.

(b) Mídia Violenta. Abrange todo material visual que retrata conteúdo gráfico, violento ou excessivamente sangrento, incluindo violência sexual. As medidas corretivas aplicadas pelo X incluem:

- **Não é permitido:** o X proíbe a mídia violenta quando julga ser de alta severidade, com probabilidade de danos a alguém. Esse conteúdo deve ser removido e violações subsequentes podem fazer com que a conta seja colocada em modo somente leitura ou suspensa.

Também é proibida a mídia violenta em áreas de alta visibilidade do X, como vídeos ao vivo, perfil, capa, bio, imagens do banner da Lista ou fotos de capa da Comunidade. Ou, então, ter como foco alguém que não consente a mídia violenta indesejada.

3.3.4 Política de Conduta de Propagação de Ódio[26]

A Política de Conduta de Propagação de Ódio proíbe expressamente promover violência, atacar ou ameaçar outras pessoas com base em raça, etnia, nacionalidade, orientação sexual, sexo, identidade de gênero, religião, idade, deficiência ou doença grave, bem como incitar

[26] https://help.x.com/pt/rules-and-policies/hateful-conduct-policy

lesões a outros com base nessas categorias. Nos termos da própria política, o X assume o compromisso de combater o assédio motivado por ódio, preconceito ou intolerância, particularmente aquele que tem o objetivo de silenciar as vozes de quem é historicamente marginalizado, proibindo comportamentos de assédio direcionados a indivíduos ou grupos com base em seu pertencimento a categoria protegida.

Identificada violação, as Operadoras do X adotam providências em relação a comportamentos direcionados a indivíduos ou a toda uma categoria protegida – direcionamento que pode ocorrer, por exemplo, por menções, inclusão de foto da pessoa ou referência ao seu nome completo. Na determinação da medida aplicável são considerados, entre outros fatores, a gravidade da violação e o histórico de violações do usuário, com aplicação gradual das medidas de limitação de visibilidade, solicitação de remoção e suspensão, inclusive a suspensão de contas que violem a política em seus elementos de perfil. Essa política é diretamente relevante para o enfrentamento da violência política de gênero.

3.3.5 Política de Abuso e Assédio[27]

A Política de Abuso e Assédio veda o direcionamento de abuso ou assédio a outras pessoas, bem como o incentivo a que terceiros o façam. A política proíbe comportamentos e conteúdos que assediam, envergonham ou degradam outros, reconhecendo que o assédio pode prejudicar a capacidade de expressão das pessoas atingidas e causar danos físicos e emocionais. Para auxiliar as equipes a compreender o contexto, a plataforma pode precisar ouvir diretamente a pessoa que está sendo alvo antes de adotar medidas de fiscalização apropriadas e proporcionais.

Dentre as condutas vedadas, destacam-se:

- **Assédio direcionado:** o direcionamento malicioso e não recíproco (como mencionar ou marcar) de indivíduos, especialmente quando compartilhado para humilhar ou degradar alguém, incluindo o compartilhamento de várias postagens em curto período ou respostas contínuas com conteúdo malicioso com o objetivo de atacar um indivíduo (inclusive contas dedicadas a assediar uma ou várias pessoas) e a menção ou marcação de usuários com conteúdo malicioso; e

[27] https://help.x.com/pt/rules-and-policies/abusive-behavior

- **Conteúdo sexual indesejado e objetificação gráfica:** condutas que sexualizem um indivíduo sem seu consentimento, incluindo o envio de mídia adulta não solicitada e/ou indesejada (imagens, vídeos e GIFs), discussão sexual não desejada sobre o corpo de alguém, solicitação de atos sexuais e qualquer outro conteúdo que sexualize um indivíduo sem o seu consentimento.

Essas vedações aplicam-se igualmente a respostas e a conteúdo gerado por IA na plataforma. Na determinação da penalidade, as Operadoras do X consideram, entre outros fatores: (i) a gravidade da violação; (ii) se alguém foi alvo (por menção, referência pelo nome completo, foto etc.); e (iii) o histórico anterior de violações do usuário. As medidas aplicáveis seguem o conjunto descrito na Seção 3.5, incluindo a suspensão de contas cujo único propósito seja violar a política de conteúdo sexual indesejado ou dedicadas a assediar indivíduos.

3.3.6 Política de Conteúdo Privado[28] e Nudez Não Consensual[29]

A Política de Conteúdo Privado, da qual faz parte a Política relacionada a Nudez Não Consensual, proíbe estritamente ameaçar expor, incentivar outros a expor ou publicar informações privadas de outras pessoas sem sua autorização expressa (“doxxing”), bem como compartilhar mídias privadas de indivíduos sem o seu consentimento. Quando notificadas pelos indivíduos retratados (ou por representante autorizado) de que não consentiram com o compartilhamento da mídia, as Operadoras do X removem a mídia; nos termos da política, o compartilhamento de mídias sexuais, de nudez ou íntimas sem consentimento expresso viola gravemente a privacidade e a segurança psicológica da pessoa, e a plataforma trabalha para remover esse conteúdo imediatamente.

Nos termos da vertente de Nudez Não Consensual, é vedado compartilhar mídias sexuais, de nudez ou íntimas (fotos/vídeos) sem a permissão da pessoa envolvida, ou que foram tiradas – ou pareçam ter sido tiradas – sem o consentimento das pessoas envolvidas. Isso inclui: conteúdo de câmera escondida com nudez, nudez parcial e/ou atos sexuais; imagens tiradas secretamente por baixo de roupas (*creepshots* ou *upskirts*); imagens ou vídeos que sobrepõem ou manipulam digitalmente o rosto de uma pessoa no corpo nu de outra; e imagens ou vídeos produzidos em ambiente íntimo e não destinados à distribuição pública. Também são vedados: ameaçar expor publicamente mídias íntimas de alguém; compartilhar informações que permitam o acesso não autorizado a mídias íntimas (como credenciais de álbuns privados); e

[28] https://help.x.com/en/rules-and-policies/personal-information
[29] https://help.x.com/pt/rules-and-policies/intimate-media

pedir ou oferecer recompensa financeira em troca de postar – ou de não postar (*sextortion*) – mídias íntimas de alguém.

Em caso de violação, as Operadoras do X podem exigir a remoção do post – com período em modo somente leitura antes de nova publicação, podendo violações subsequentes levar à suspensão da conta – e suspender contas cujo único propósito seja postar informações privadas ou mídias íntimas de terceiros, incluindo contas dedicadas a *creepshots*. O X disponibiliza formulário específico para denúncia de violações à Política de Nudez Não Consensual, que atende a qualquer tipo de veiculação de conteúdo dessa natureza, seja ele natural ou sintético. A relevância dessa política para as vedações do artigo 28, § 1º-C, III e IV, da Resolução está descrita na Seção 8.5.

3.3.7 Violações Eleitorais

A política do X de conteúdo relativo a violações eleitorais no Brasil consiste em um procedimento interno de avaliação e retenção de conteúdo, aplicável exclusivamente no território brasileiro, desenvolvido como parte do projeto de observância à decisão do Supremo Tribunal Federal sobre o artigo 19 do Marco Civil da Internet. Essa política é um reflexo das exigências legais contidas no art. 9-E da Resolução do TSE 23.610, que estabelece "casos de risco" que devem ensejar a indisponibilização imediata de conteúdos, durante o período eleitoral, em hipóteses específicas de risco à integridade do processo eleitoral.

O escopo da política abrange as categorias previstas no referido dispositivo, incluindo a divulgação ou compartilhamento de fatos notoriamente inverídicos ou gravemente descontextualizados que atinjam a integridade do processo eleitoral (votação, apuração e totalização de votos); o uso de conteúdo sintético gerado ou manipulado por inteligência artificial em desacordo com regras de rotulagem ou com potencial de desequilibrar o pleito; a violência política contra a mulher; condutas que configurem abolição violenta do Estado Democrático de Direito ou tentativa de golpe de Estado; ameaças graves e diretas de violência contra membros e servidores da Justiça Eleitoral ou contra a infraestrutura física do Poder Judiciário; e a incitação às Forças Armadas contra poderes constitucionais. A análise é conduzida de forma a averiguar o preenchimento integral de todos os elementos de cada ofensa — inclusive o dolo específico, quando aplicável —, resguardando-se hipóteses de ambiguidade, opinião, crítica, sátira, humor ou discurso protegido.

O fluxo de trabalho interno contempla a verificação de acessibilidade do conteúdo, a análise sequencial de cada categoria de ofensa à luz de critérios objetivos e exclusões expressas e a consideração do contexto completo da conversa. Somente o conteúdo que satisfaça integralmente os requisitos de pelo menos uma ofensa é classificado como violação e torna-se elegível à retenção no Brasil, nos termos da política local de violações eleitorais.

O tema Eleições é tratado na página de Recursos para Segurança Online no Brasil https://help.x.com/pt/rules-and-policies/brazil-resources e violações baseadas na lei podem ser reportadas pelo canal https://help.x.com/pt/forms/br/report, que disponibiliza formulário para denúncias de irregularidades eleitorais sob a legislação brasileira. O referido canal assegura o recebimento e o tratamento adequado de alertas oriundos de usuários. Candidatos, partidos, federações, coligações e autoridades possuem canal próprio para encaminhamento de denúncias (o e-mail juridicoeleicoes_NE@x.com).

3.4 Arquitetura de moderação: detecção automatizada e revisão humana

Do ponto de vista da moderação de conteúdo, o X conta com dois tipos de recursos complementares:

i. **Automação e Aprendizado de Máquina**. Recursos proativos de moderação que consistem na combinação de métodos automatizados de detecção para identificar conteúdos que violem as Regras do X. Esses métodos incluem modelos de aprendizado de máquina, tecnologias de correspondência por *hash* e mecanismos de bloqueio por palavras-chave ou expressões. Os modelos de aprendizado de máquina incluem técnicas de processamento de linguagem natural para análise de texto, modelos de processamento de imagens para conteúdo visual e outros algoritmos sofisticados treinados com base em dados históricos relacionados a violações anteriores, com o objetivo de identificar padrões indicativos de violações, como conduta de ódio, abuso, extremismo violento ou spam. O X também utiliza tecnologias de correspondência por hash, como o PhotoDNA, para identificar material de abuso sexual infantil (CSAM) previamente conhecido. Além disso, emprega um sistema híbrido que combina modelos de aprendizado de máquina e modelos de linguagem de grande escala (“LLMs”) para detectar e moderar conteúdos que violem as Regras do X. Os LLMs são utilizados para classificar e anotar conteúdos na plataforma como parte do sistema de defesa.

ii. **Revisão humana**, composta por mais de mil analistas que trabalham globalmente para monitorar a integridade da conversa pública e aplicar os Termos de Serviço e as Regras e Políticas do X de forma objetiva e consistente. A equipe de revisão humana funciona 24 horas por dia, 7 dias por semana, em todos os idiomas suportados pelo X, o que inclui o português.

Varreduras proativas identificam tendências, para as quais são desenvolvidas ferramentas automatizadas de detecção e aplicação de medidas contra material abusivo. A abordagem do X é adaptativa e iterativa, dado que agentes mal-intencionados persistem em tentar contornar as políticas e os esforços de mitigação de riscos; as diretrizes de moderação são ajustadas sempre que padrões ou eventos novos são identificados. Os agentes humanos são responsáveis, ainda, pela revisão das denúncias realizadas na plataforma: os casos reportados são analisados, e filtros automatizados adicionais identificam novos potenciais casos a serem encaminhados para revisão humana.

Para reduzir o risco de subjetividade e viés nas decisões de moderação, o X vem migrando de uma abordagem decisória tradicional (*decision-first*) – na qual o revisor analisa o conteúdo frente aos critérios da política e decide diretamente pela existência ou não de violação – para uma abordagem orientada por informações (*information-first*), na qual o revisor chega à decisão de *enforcement* respondendo a uma série estruturada de perguntas e comparando o conteúdo analisado com repositório de conteúdos já avaliados e considerados violadores. Essa abordagem reduz a carga cognitiva do revisor, o potencial de viés e a inconsistência na aplicação dos critérios entre revisores distintos, aumentando a uniformidade das decisões.

Esse arranjo – detecção automatizada, revisão humana em português em regime contínuo e canais de denúncia acessíveis a usuários e não usuários – constitui os procedimentos, fluxos decisórios e estruturas de supervisão destinados à efetiva aplicação das regras.

A avaliação será aferida pelos indicadores descritos na Seção 2.5 e na análise de impacto, incluindo a aferição de falsos positivos por reversão em recurso ou avaliação amostral, conforme aplicável.

3.5 Medidas disponíveis para resposta a violações

Em resposta a violações ou riscos, o X dispõe de conjunto graduado de medidas, aplicáveis conforme a gravidade e o contexto aplicável a políticas como Propagação de Ódio, Abuso e

Assédio, Discurso Violento e Integridade Cívica: conteúdos que violem as políticas podem, em determinadas circunstâncias, permanecer acessíveis, mas não se qualificam para amplificação.

(a) Medidas aplicáveis a conteúdos:

- **Limitação da visibilidade do post:** exclusão dos resultados de pesquisa, dos Assuntos do Momento e das notificações recomendadas; remoção das *timelines* "Para Você" e "Seguindo"; restrição da visibilidade ao perfil do autor; rebaixamento nas respostas; limitação das opções de engajamento (curtidas, respostas, reposts, citações, itens salvos, fixações no perfil, contagens de engajamento e edições); exclusão de posts e/ou contas de e-mails e recomendações de produto; e impedimento de exibição de anúncios adjacentes ao conteúdo violador, com marcação que informa o autor e os visualizadores da limitação aplicada. No contexto eleitoral, essa medida é particularmente relevante por permitir a redução do alcance de desinformação sem a remoção integral do conteúdo, preservando o registro para fins de transparência e para eventual exercício do direito de defesa pelo autor;

- **Solicitação de remoção do post:** quando a violação é suficientemente grave, o X exige que o autor remova o conteúdo como condição para voltar a publicar, podendo o usuário permanecer por período em modo somente leitura. O post permanece oculto da visualização pública, com um aviso, durante o processo, e o autor pode recorrer da decisão;

- **Identificação e rotulagem do post:** quando o X determina que um post contém informações enganosas ou questionáveis de acordo com suas políticas, pode adicionar marcação ao conteúdo para fornecer contexto e informações adicionais aos usuários; Notas da Comunidade também podem estar visíveis nos posts; e

- **Aviso de exceção por interesse público:** em situações excepcionais, o X pode determinar que um post que violaria suas Regras permaneça acessível por ser de interesse público[30]. Nesses casos, o post é colocado atrás de um aviso e tem sua visibilidade limitada. A exceção é aplicável, entre outros cenários, a posts de representantes eleitos ou candidatos que contribuam diretamente para a compreensão ou discussão de assuntos de preocupação pública, mediante avaliação da equipe de Segurança considerando o contexto local e o potencial de danos *offline*. Para conteúdos

[30] https://help.x.com/pt/rules-and-policies/public-interest

que violem a Política de Integridade Cívica (por exemplo, desinformação sobre como, quando ou onde votar), dificilmente será concedida essa exceção, dada a gravidade dos danos que podem decorrer de informações capazes de interferir na capacidade dos indivíduos de participar do processo eleitoral.

(b) Medidas aplicáveis a contas, adotadas quando há envolvimento do usuário em violações reiteradas ou em violações que causem risco significativo – como a publicação de conteúdo ilegal, tentativas de manipulação da plataforma, incitação à violência ou operações de influência coordenada:

- **Modo somente leitura:** a conta é temporariamente impedida de publicar, repostar ou interagir com conteúdos, permanecendo em modo de visualização por período determinado;

- **Suspensão temporária ou permanente:** aplicável a violações graves ou reiteradas. A suspensão permanente é aplicada, entre outras hipóteses, a contas envolvidas em operações de influência coordenada, manipulação da plataforma ou ameaças violentas – categorias particularmente relevantes para o contexto eleitoral; e

- **Limitação da amplificação via recomendações:** restrição da elegibilidade da conta para amplificação algorítmica, de modo que seus conteúdos deixem de ser recomendados a usuários que não a seguem, reduzindo o alcance orgânico de contas envolvidas em disseminação de desinformação eleitoral sem necessidade de suspensão integral.

Adicionalmente, no âmbito da Política de Autenticidade (mídia sintética e manipulada e URLs maliciosos), a plataforma aplica tecnologia própria e recebe denúncias por meio de parcerias com terceiros para determinar se a mídia foi manipulada ou apresentada fora de contexto, podendo excluir publicações, colocar URLs em lista de bloqueio ou exibir avisos sobre URLs consideradas inseguras[31], bloquear contas temporariamente, suspender contas permanentemente – inclusive na primeira detecção –, aplicar etiquetas de contexto, restringir o alcance e limitar ou restringir o acesso a recursos ou produtos do X. Exemplos de rótulos aplicados de acordo com a Política de Autenticidade e as medidas de redução de visibilidade.

[31] Mais informações sobre o bloqueio de links estão disponíveis em: https://help.x.com/pt/safety-and-security/phishing-spam-and-malware-links.

(c) Salvaguardas contra intervenções indevidas. Em todos os casos, os usuários afetados recebem notificação sobre a medida aplicada e sobre a regra violada, e podem apresentar recurso caso entendam que a decisão foi incorreta. Somadas aos critérios de proporcionalidade – em especial a exigência, em certas hipóteses, de denúncia pela própria vítima e a exceção de interesse público –, essas salvaguardas destinam-se a evitar intervenções indevidas sobre conteúdos lícitos.

(d) Medidas preventivas contínuas. O X adota abordagem preventiva contínua que compreende:

- **Detecção proativa por aprendizado de máquina:** modelos de *machine learning*, classificadores, análise de sinais preditivos e tecnologias de correspondência por códigos *hash* identificam conteúdos potencialmente violadores independentemente de denúncias de usuários. Esses sistemas são continuamente aprimorados com base em novas tendências de comportamento e padrões de disseminação de desinformação eleitoral;

- **Varreduras direcionadas:** as equipes de Segurança realizam varreduras proativas de conteúdos e contas quando identificam riscos específicos – por meio de monitoramento interno, parceiros externos, autoridades ou reportagens da imprensa. Essas varreduras podem envolver a busca por palavras-chave, URLs ou hashtags específicas associadas a conteúdos de alto risco no contexto eleitoral;

- **Diretrizes de *enforcement* para eventos específicos:** diretrizes de aplicação são desenvolvidas em resposta a fatores ou eventos externos, conhecidos (proativas) ou não (reativas), com o propósito de preparar as equipes globais de *enforcement* para revisar conteúdo relacionado a evento ou tema específico.

- **Resposta a incidentes:** o X opera protocolo estruturado de resposta a situações de urgência, com abordagem escalonada: as situações são avaliadas segundo fatores que incluem o risco de dano e a urgência.

- **Proteções específicas para transmissões ao vivo, Spaces e Comunidades:** o X mantém medidas permanentes para mitigar o risco de distribuição de conteúdo violador nessas superfícies. Nas transmissões ao vivo, há detecção proativa por *machine learning* de mídia sensível e, atingidos determinados limiares, as transmissões são

encaminhadas para revisão por moderadores de conteúdo; proteções de produto permitem, ainda, que o titular da transmissão bloqueie autores de comentários violentos e que os espectadores os denunciem, com revisão humana reativa. Nos Spaces, detecções proativas por *machine learning* identificam títulos tóxicos, conteúdo tóxico no texto de transcrição e Spaces associados a usuários considerados de alto risco, com encaminhamento para revisão manual e possibilidade de encerramento da transmissão, ocultação dos conteúdos ao vivo e das gravações criadas pelo usuário violador, impedimento de novos ouvintes ou espectadores e impedimento de criação de novas transmissões. Nas Comunidades, os posts estão sujeitos aos controles de segurança da plataforma – em alguns casos reforçados, como a ocultação por *machine learning* de mídia sensível não rotulada pela própria Comunidade –, além da atuação de administradores e moderadores da própria Comunidade e da possibilidade de denúncia por qualquer usuário do X, membro ou não;

- **Bloqueio de termos nos Assuntos do Momento e na busca:** o X mantém listas de palavras-chave e expressões bloqueadas de aparecer nos Assuntos do Momento (*Trending Topics*), bloqueia resultados de busca para determinados termos e aplica proteções no "Autocompletar" da barra de busca, restringindo as sugestões exibidas conforme o usuário digita, a fim de impedir a amplificação de conteúdos nocivos; e

- **Ajuste adaptativo de diretrizes:** a abordagem do X é adaptativa e iterativa, considerando que agentes mal-intencionados persistem em tentar contornar as políticas de aplicação. As diretrizes de moderação são ajustadas sempre que padrões ou eventos novos são identificados, garantindo que as medidas de prevenção acompanhem a evolução das ameaças à integridade eleitoral.

3.6 Notas da Comunidade[32]

Notas da Comunidade é um sistema colaborativo que adiciona contexto útil e informativo a posts que possam conter informações enganosas, constituindo o principal instrumento de notificação e contextualização do X para conteúdos relativos à integridade cívica e eleitoral – como dados incorretos sobre como, quando ou onde votar, ou alegações infundadas sobre a legitimidade de resultados eleitorais.

[32] https://communitynotes.x.com/guide/pt/about/introduction

As notas são escritas e avaliadas por colaboradores, não pelo X: não são usadas para impor as Regras do X, não representam o ponto de vista da plataforma e não podem ser editadas ou modificadas pelas equipes de moderação. Quando um número suficiente de colaboradores com perspectivas distintas classifica uma nota como útil, ela aparece diretamente no post. Diferentemente de mecanismos de moderação tradicionais, as notas não acarretam remoção de conteúdo, aplicação de rótulos pela plataforma ou redução de distribuição: sua função é complementar a informação disponível, permitindo que a própria comunidade adicione contexto e referências que auxiliem os demais usuários a avaliar criticamente o conteúdo.

O programa existe desde janeiro de 2021 e foi projetado, desde a concepção, para mitigar os riscos inerentes a sistemas colaborativos[33], por meio das salvaguardas descritas a seguir.

(i) Prevenção de tentativas coordenadas de manipulação. Todas as contas devem atender aos critérios de elegibilidade para se tornarem colaboradoras[34] – por exemplo, dispor de número de celular exclusivo e verificado –, critérios concebidos para evitar a criação de grandes volumes de contas falsas ou *sockpuppets* destinadas a avaliações inautênticas. O programa não funciona como sistemas de avaliação por engajamento, nos quais o conteúdo mais popular ganha visibilidade e votações em massa podem ser coordenadas: utiliza um algoritmo de *bridging*, pelo qual uma nota somente é exibida se tiver sido considerada útil por pessoas com tendência a ter discordado em avaliações anteriores[35]. Pesquisas acadêmicas indicam que o ranqueamento baseado em *bridging* ajuda a identificar conteúdos mais saudáveis e de qualidade mais elevada, além de reduzir o risco de polarização. O programa mantém, ainda, um sistema de reputação[36], no qual colaboradores ganham pontos pela utilidade de contribuições consideradas úteis por pessoas de ampla gama de pontos de vista – o que confere mais influência a quem tem histórico de contribuições de alta qualidade e menor influência a contas novas –, e monitora métricas que alertam a equipe em caso de atividade suspeita, com proteções e procedimentos para identificar quedas de qualidade abaixo de limites definidos[37], permitindo a detecção proativa de tentativas de coordenação.

(ii) Prevenção de resultados enviesados. Além do algoritmo de *bridging*, o programa pode buscar proativamente avaliações de colaboradores com probabilidade de apresentar perspectiva diferente com base no próprio histórico de avaliação – hoje por meio da aba

[33] https://communitynotes.x.com/guide/pt/about/challenges
[34] https://communitynotes.x.com/guide/pt/contributing/signing-up
[35] https://communitynotes.x.com/guide/pt/contributing/diversity-of-perspectives
[36] https://communitynotes.x.com/guide/pt/under-the-hood/contributor-scores.html
[37] https://communitynotes.x.com/guide/pt/under-the-hood/guardrails.html

“Precisa da sua ajuda”[38] –, e implementou pseudônimos não associados publicamente às contas dos colaboradores no X, favorecendo a participação diversa[39]. O X também pesquisa regularmente amostras representativas de usuários não colaboradores para avaliar se pessoas com diferentes pontos de vista consideram o contexto das notas útil, e usuários não inscritos no programa podem dar *feedback* sobre as notas que veem – indicadores contínuos da utilidade das notas para uma ampla série de perspectivas, e não para um grupo específico.

(iii) Prevenção do assédio. Todos os colaboradores recebem pseudônimo novo e gerado automaticamente ao ingressar no programa, o que inibe a identificação pública e o assédio; mantêm linha de comunicação aberta com a equipe do programa (por Mensagem Direta via @CommunityNotes[40]), com gerente de comunidade dedicado; e todas as contribuições permanecem sujeitas às Regras do X, aos Termos de Serviço e à Política de Privacidade, podendo notas em violação ser denunciadas pelo menu da própria nota ou pelo formulário Denunciar uma Nota da Comunidade[41].

(iv) Redução do impacto de colaborações de baixa qualidade. O programa implementou sistema de integração que exige que novos colaboradores comecem avaliando notas antes de poderem escrevê-las[42]; permite o bloqueio temporário da habilidade de redação de colaboradores que escrevam muitas notas avaliadas como não úteis, com alertas prévios e critérios de desbloqueio baseados na repetição do processo de avaliação; e utiliza o sistema de reputação para reduzir a influência de novas contas que ainda não demonstraram histórico de contribuições úteis, dificultando que uma conta inunde o sistema com avaliações de baixa qualidade ou *spam*.

O programa integrou recentemente funcionalidades de inteligência artificial para auxiliar na redação de notas, acelerando a adição de contexto a posts potencialmente enganosos, mantida a supervisão pela comunidade de colaboradores. Ademais, usuários que tenham interagido com um post (por resposta, curtida ou repost) recebem notificação quando uma Nota da Comunidade é posteriormente exibida naquele post, ampliando o alcance da contextualização. As notas podem, ainda, ser redigidas sobre imagens e vídeos e, nesse caso, passam a aparecer automaticamente em todos os posts que contenham a mesma mídia correspondente, estendendo a contextualização para além do post originalmente anotado.[43]

[38] https://communitynotes.x.com/guide/pt/contributing/rating-notes
[39] https://communitynotes.x.com/guide/pt/contributing/aliases
[40] https://x.com/communitynotes
[41] https://help.x.com/en/forms/community-note
[42] https://communitynotes.x.com/guide/pt/contributing/writing-ability
[43] https://communitynotes.x.com/guide/pt/contributing/notes-on-media

O programa opera, por fim, com elevado grau de transparência: o algoritmo é de código aberto e está publicamente disponível no GitHub[44], acompanhado dos dados subjacentes, permitindo que qualquer pessoa audite, analise ou sugira melhorias ao sistema. As métricas de eficácia externas – redução de 61% nas republicações (com estudo independente adicional apontando queda de aproximadamente 50%) e aumento de aproximadamente 80% na exclusão voluntária de posts após o recebimento de notas – evidenciam a efetividade do instrumento como mecanismo de contextualização de desinformação.

3.7 Alfabetização midiática e recursos informativos

Adicionalmente à moderação proativa e reativa de conteúdo, o X adota plano de ação com medidas voltadas a capacitar os usuários a avaliar criticamente os conteúdos encontrados na plataforma:

- **Rótulos de conteúdo sensível:** avisos de conteúdo que exigem ação deliberada do usuário para visualizar o material, promovendo decisão informada sobre o engajamento;

- **Rótulos "Feito com IA":** rotulagem para auxiliar os usuários na identificação de conteúdo sintético;

- **Funcionalidade Grok:** integração com a ferramenta Grok, que permite aos usuários solicitar análise, resumo ou informações de contexto gerados por inteligência artificial diretamente sobre qualquer post, por meio de ícone dedicado;

- **Central de Ajuda:** recurso educacional abrangente, contendo orientações sobre como denunciar conteúdos, descrição das políticas da plataforma e materiais de apoio à navegação responsável.

3.8 Ferramentas de controle e proteção disponíveis aos usuários

Além das medidas aplicadas pela plataforma, o design do X confere aos próprios usuários um conjunto de controles que mitigam riscos de abuso, assédio e intimidação – condutas aptas a suprimir a participação cívica, inclusive em contexto eleitoral: (i) o bloqueio de contas, que impede que a conta bloqueada interaja com os posts do usuário ou o siga, deixando os posts de contas bloqueadas de aparecer na timeline de quem bloqueou; (ii) o silenciamento de

[44] https://github.com/twitter/communitynotes

contas, palavras, conversas, frases, emojis e *hashtags*; (iii) os posts protegidos e a possibilidade de tornar a conta privada, restringindo a visibilidade dos conteúdos a seguidores aprovados pelo usuário; (iv) o controle de respostas, que permite definir, no momento da publicação ou depois dela, quem pode responder a cada post, inclusive com a desativação integral das respostas; (v) a ocultação de respostas aos próprios posts, que confere ao usuário maior controle sobre as conversas e reduz a visibilidade de conteúdo potencialmente nocivo; (vi) os filtros de notificações, que permitem silenciar notificações de categorias de contas (como contas novas, contas sem número de telefone ou e-mail confirmado, contas com foto de perfil padrão, contas que o usuário não segue ou que não o seguem); (vii) as preferências de solicitação de Mensagens Diretas, que limitam quem pode enviar mensagens ao usuário; e (viii) a privacidade das curtidas, que impede que terceiros visualizem os conteúdos curtidos por um usuário. No contexto eleitoral, essas ferramentas são particularmente relevantes para candidatas, candidatos e demais participantes do debate público expostos a campanhas de assédio.

4. COMPORTAMENTO INAUTÊNTICO COORDENADO E VEDAÇÃO DE DISPARO EM MASSA (Portaria, arts. 14 e 22, parágrafo único; Resolução, arts. 9º-D, caput e III, e 34)

Medida de conformidade. A plataforma X veda, por meio da Política de Autenticidade – em particular de sua vertente de Comportamentos Não Autênticos –, a amplificação artificial e a atividade coordenada inautêntica, e mantém sistemas automatizados de detecção integrados às equipes humanas de moderação para identificar, investigar e desarticular redes de comportamento inautêntico coordenado.

4.1 Parâmetros de identificação e distinção da coordenação legítima

Os parâmetros de identificação decorrem da Política de Autenticidade: são vedadas (i) a operação de várias contas que interagem com o mesmo conteúdo ou conteúdo substancialmente semelhante, ou com o objetivo de inflar ou manipular a importância de conteúdos e contas; (ii) a operação de várias contas que publicam conteúdo substancialmente semelhante ou idêntico entre si; e (iii) o emprego de soluções alternativas para ultrapassar os limites técnicos de criação de contas. A mesma política delimita expressamente as hipóteses de operação legítima de múltiplas contas – com identidades, finalidades ou casos de uso

distintos e em conformidade com os limites técnicos da plataforma –, estabelecendo a distinção entre a atuação coordenada legítima e as redes inautênticas.

4.2 Mecanismos de detecção

Do ponto de vista técnico, a plataforma conta com sistemas automatizados de detecção baseados em aprendizado de máquina que monitoram continuamente padrões de comportamento indicativos de atividade automatizada ou coordenada, tais como volume anômalo de publicações em curto espaço de tempo, padronização de conteúdo entre múltiplas contas e interações que desviem dos padrões regulares de uso. Esses sistemas operam de forma integrada com as equipes humanas de moderação, que atuam na investigação e na desarticulação de redes de comportamento inautêntico coordenado.

4.3 Medidas aplicáveis

Identificadas contas ou redes envolvidas em manipulação da plataforma ou operações de influência coordenada, aplicam-se as medidas descritas na Seção 3.5, incluindo a limitação da amplificação via recomendações, o bloqueio temporário e a suspensão permanente – esta última expressamente aplicável, entre outras hipóteses, a contas envolvidas em operações de influência coordenada. Os resultados dessas ações são consolidados, quando aplicável, no relatório previsto na Seção 12, sem prejuízo das demais iniciativas de transparência adotadas.

O mecanismo de notificação ao Tribunal Superior Eleitoral de operações de comportamento inautêntico coordenado com potencial de risco, caso venham a ser identificados, consistirá em envio de e-mail à Presidência do TSE e à Assessoria Especial de Enfrentamento à Desinformação.

4.4 Remoção de perfis falsos, automatizados ou robôs por iniciativa própria (Portaria, art. 23, II; Resolução, art. 38-A, § 2º)

Quanto à faculdade de remoção de perfis falsos, automatizados ou robôs por iniciativa própria, o X informa que a exerce: contas falsas, inautênticas ou automatizadas em desconformidade com a Política de Autenticidade estão sujeitas às medidas descritas na Seção 3.5, incluindo a suspensão permanente, segundo os critérios e procedimentos de aplicação já detalhados acima.

4.5 Vedação de propaganda por disparo em massa com tecnologias externas (Resolução, art. 34)

As obrigações previstas no artigo 34 da Resolução – que veda a propaganda por disparo em massa de mensagens instantâneas sem consentimento da pessoa destinatária ou a partir da contratação de tecnologias ou serviços não fornecidos pelo provedor e em desacordo com seus termos de uso – não se aplicam ao X BRASIL, uma vez que a plataforma não permite a utilização de tecnologias ou ferramentas externas que viabilizem o disparo massivo de mensagens para fins de propaganda eleitoral. Tal vedação decorre do design e das próprias políticas de uso da plataforma, que proíbem expressamente e impedem a automação voltada ao envio de conteúdo em larga escala, tornando inviável a prática que o dispositivo normativo busca coibir.

Sem prejuízo da inaplicabilidade, a plataforma dispõe de políticas e mecanismos técnicos aptos a identificar e registrar padrões de atividade anômala ou automatizada, na forma do artigo 22, parágrafo único, da Portaria. A Política de Comportamentos Não Autênticos proíbe expressamente (i) a utilização de automação para envio de conteúdo massivo; (ii) a criação de contas falsas ou automatizadas para amplificação artificial de mensagens; e (iii) a coordenação de atividade inautêntica que vise manipular o debate público. Os sistemas de detecção e os fluxos de apuração e resposta a denúncias aplicam-se integralmente a essas hipóteses. O X BRASIL reafirma que os mecanismos existentes de combate à manipulação e ao *spam* atendem à finalidade do dispositivo e serão mantidos durante o período eleitoral.

5. SISTEMAS DE RECOMENDAÇÃO: CONTROLES, TRANSPARÊNCIA E GESTÃO DA EXPERIÊNCIA (Resolução, arts. 9º-D, III, IV e VI, e 28, § 1º; Portaria, art. 23, I e § 2º)

Medida de conformidade. O X aplica restrições ao seu sistema de recomendação de conteúdo com o objetivo de impedir a amplificação algorítmica de conteúdos potencialmente prejudiciais à integridade do processo eleitoral[45], sob o princípio orientador de que “liberdade de expressão é um direito humano fundamental – mas a liberdade de ter esse discurso amplificado pelo X não é”.

Em descrição funcional: o sistema de recomendação “Para Você” seleciona conteúdos – provenientes tanto de contas seguidas pelo usuário quanto de um conjunto mais amplo de

[45] https://help.x.com/pt/rules-and-policies/recommendations

conteúdos elegíveis publicados na plataforma – com base na probabilidade de gerarem interação por parte de cada usuário específico, considerando sinais positivos (curtir, responder, republicar, compartilhar, permanecer no conteúdo, seguir o autor) e sinais negativos (indicar desinteresse, denunciar, bloquear ou silenciar). As probabilidades estimadas são combinadas para formar uma pontuação de relevância individualizada e, antes da apresentação do conteúdo, o sistema aplica controles relacionados, entre outros aspectos, à elegibilidade, à segurança, à repetição de conteúdos, à diversidade de autores e ao histórico de conteúdos já apresentados ao usuário. A arquitetura é predominantemente orientada à previsão de relevância individual, e não à aplicação de classificação política ou editorial uniforme para toda a base de usuários. Os componentes do sistema não utilizam filiação partidária, posição ideológica ou apoio a determinado grupo político como fator direto de pontuação dos conteúdos. Como consequência, o mesmo conteúdo pode receber classificações distintas para usuários diferentes, conforme seus respectivos históricos de interação e contextos de uso – o que reduz o risco de que o sistema funcione, por sua própria lógica central, como mecanismo uniforme de promoção ou supressão de partidos, candidatos ou opiniões políticas em toda a plataforma.

A arquitetura personalizada reduz, ainda, a dependência de classificação única baseada exclusivamente na popularidade global de cada conteúdo, limitando formas de viralização indiferenciada nas quais o volume agregado de interações seria suficiente para impulsionar conteúdo de maneira uniforme em toda a plataforma. Conteúdos com maior probabilidade de gerar rejeição (desinteresse, denúncia, bloqueio) podem receber pontuação inferior, fazendo com que a reação negativa dos usuários influencie a distribuição futura. O sistema incorpora, ademais, controles para reduzir a repetição excessiva de conteúdos e autores em uma mesma experiência de recomendação, incluindo filtros para conteúdos duplicados, republicações repetidas, publicações já visualizadas ou recentemente apresentadas, conteúdos que excedam os limites temporais aplicáveis e atenuação progressiva quando diversas publicações do mesmo autor aparecem entre os conteúdos candidatos – o que contribui para reduzir a saturação e a concentração de conteúdos provenientes de uma mesma fonte.

Antes da implantação de novos componentes nos sistemas de recomendação, o X realiza testes A/B e processos de avaliação de risco, incluindo a possibilidade de ativar e desativar funcionalidades para avaliar seu impacto sobre conteúdo violador.

5.1 Conteúdos e contas não elegíveis para recomendação

O X se empenha em manter as seguintes categorias de conteúdo fora de suas recomendações (isto é, fora da *timeline* “Para Você”, da aba Explorar, das notificações por *push* e do topo das respostas):

- informações enganosas e prejudiciais, inclusive violações da Política de Integridade Cívica e da Política de Autenticidade;

- conteúdo que viole as Regras do X, mas que tenha sido mantido na plataforma por exceção de interesse público;

- conteúdo de contas de mídia afiliada ao Estado ou de governos conhecidos por restringir o acesso à internet aberta durante conflitos armados[46];

- conteúdo identificado como marginalmente abusivo segundo as políticas de Comportamento Abusivo e de Conduta de Propagação de Ódio;

- conteúdo que os sistemas automatizados tenham determinado como potencialmente violador das Regras do X, mas que ainda não tenha sido revisado por revisores humanos; e

- conteúdo que inclua informações obtidas por meio de *hacking*, exceto quando acompanhado de cobertura jornalística ou explicação contextual.

O conteúdo excluído das recomendações permanece acessível na plataforma a quem segue o autor do post e no perfil de quem o publicou – a exclusão das recomendações não equivale à remoção.

Da mesma forma, o X se empenha em excluir das recomendações as seguintes categorias de contas: contas que tenham violado recentemente as Regras do X; contas que compartilhem informações enganosas e prejudiciais, inclusive em violação às Políticas de Integridade Cívica e de Conteúdo não autêntico contas que se acredite estarem envolvidas em comportamento de *spam*; contas marcadas como mídia afiliada ao Estado, ou pertencentes a países que limitem o acesso a informações livres e estejam envolvidos em conflitos armados; e contas que

[46] https://help.x.com/pt/rules-and-policies/government-media-labels

contenham imagens explícitas de violência ou propagação do ódio nos elementos do perfil. Durante o período eleitoral brasileiro, aplica-se ainda a exclusão, da aba “Para Você”, das contas oficiais de candidaturas informadas à Justiça Eleitoral e dos conteúdos nelas postados.

5.2 Transparência algorítmica

O algoritmo de recomendação do X é de código aberto (*open source*), disponibilizado publicamente em repositório no GitHub (https://github.com/xai-org/x-algorithm)[47] para análise de qualquer interessado – pesquisadores, reguladores e sociedade civil. Trata-se de iniciativa ainda pouco comum entre plataformas digitais de grande porte. A publicação permite a inspeção externa da lógica geral do sistema e reduz a assimetria de informações sobre o funcionamento dos mecanismos de recomendação.

As Operadoras do X atualizam periodicamente esse código e publicam as atualizações na plataforma, assegurando que o sistema de recomendação permaneça aberto à revisão independente. Essa medida é particularmente relevante no contexto eleitoral, ao permitir a verificação independente de que os componentes divulgados do sistema de recomendação não contêm sinais expressos destinados a favorecer ou desfavorecer posições político-eleitorais específicas. Isso não significa que sistemas de aprendizado de máquina sejam imunes a vieses ou que seus resultados sejam necessariamente neutros; a abertura do código, contudo, aumenta a possibilidade de identificação, análise e questionamento público desses riscos. O filtro aplicável às eleições brasileiras de 2026, conforme explicado abaixo, integra essa publicação, o que permite a verificação pública, no próprio código, do critério e do alcance da exclusão.

5.3 Exclusão de perfis e conteúdos informados à Justiça Eleitoral

Como parte da medida de conformidade, o X aplicará, do início do período eleitoral e até o segundo turno, filtro específico no sistema de recomendação “Para Você”, de modo a excluir dos resultados as contas oficiais de candidatas e candidatos informadas à Justiça Eleitoral e as publicações dessas contas.

A medida entrou em vigor em 14 de agosto de 2026 e será atualizada com a lista final de candidatos conforme disponível em https://dadosabertos.tse.jus.br/dataset/candidatos-2026, assim que possível a finalização, já que há possibilidade de registro de novos candidatos até a

[47] https://github.com/xai-org/x-algorithm

véspera do início do período eleitoral. As contas e os posts abrangidos não são recomendados na aba “Para Você”, salvo se o usuário seguir explicitamente a conta.

O filtro, denominado *Brazil2026ElectionFilter*, está publicado no repositório de código aberto do algoritmo, o que permite à Justiça Eleitoral e ao público em geral inspecionar o critério aplicado. A implementação foi comunicada publicamente pela conta institucional do X no Brasil e pela atualização do código aberto.[48]

5.4 Controle do usuário sobre as recomendações

O X disponibiliza ferramentas que permitem aos usuários gerenciar sua experiência de recomendação, incluindo: indicação de desinteresse em posts ou tópicos específicos (“Não tenho interesse neste post” / “Não tenho interesse neste Tópico”), sinais utilizados pelo algoritmo para ajustar recomendações futuras; a funcionalidade “Soneca” (*Snooze*), que permite restringir temporariamente posts sobre temas específicos, como “Política”, do *feed* “Para Você”; opções de silenciamento de contas, palavras, frases, hashtags e emojis específicos; e a alternância entre as *timelines* “Para Você” (recomendada) e “Seguindo” (cronológica), permitindo ao usuário optar por experiência sem amplificação algorítmica.

De modo mais geral, os usuários dispõem, para cada sistema de recomendação, de opções de interação com conteúdo não perfilado, no qual o conteúdo exibido é tipicamente o mais recente ou o mais popular, sem considerar informações personalizadas, ou provém estritamente de contas que o usuário optou por seguir.

6. CUMPRIMENTO DE ORDENS E DETERMINAÇÕES DA JUSTIÇA ELEITORAL (Resolução, arts. 9º-D, §§ 3º e 5º; 9º-E, § 1º; 32; 36; 38, § 6º; 39; e 40; Portaria, art. 12)

Medida de conformidade. O X BRASIL mantém estrutura dedicada para recebimento, análise, escalonamento e comprovação de cumprimento de ordens judiciais e determinações eleitorais. Essa estrutura, compartilhada entre serviços e jurisdições, combina a atuação da equipe jurídica local, dos escritórios externos Pinheiro Neto Advogados e Pinheiro Guimarães Advogados, das equipes globais das Operadoras do X e, conforme o caso, das áreas responsáveis por Segurança, segurança da informação e operação de produto. A cooperação entre X BRASIL, Operadoras do X e SpaceXAI aplica-se integralmente a esta medida.

[48] https://x.com/XBR/status/2088341967864320507 e https://x.com/XOpenSource/status/2088373226887889087

6.1 Fluxo de recebimento, triagem, escalonamento e execução

O procedimento padrão de cumprimento inicia-se com o recebimento da ordem pelo canal dedicado, seguido de análise jurídica inicial para verificação de autenticidade, identificação do tipo de determinação, delimitação do escopo material e territorial, confirmação do prazo aplicável e registro interno para acompanhamento.

Quando a ordem exigir esclarecimento, envolver limitações técnicas ou demandar medida processual, a equipe jurídica local e os escritórios externos avaliam a estratégia de resposta e, se necessário, interagem com a autoridade competente para viabilizar cumprimento adequado, proporcional e tecnicamente exequível – procedimento que corresponde ao escalonamento interno para ordens que exijam dados complementares, na forma do artigo 12, II, da Portaria.

Confirmado o escopo da determinação, a demanda é encaminhada às equipes operacionais competentes para execução da providência determinada, que pode incluir remoção ou indisponibilização de conteúdo, restrição de conta, preservação ou fornecimento de dados, ou outra medida compatível com a ordem judicial. Concluída a execução, a confirmação é retornada à equipe jurídica local e aos escritórios externos para adoção das medidas processuais cabíveis e comprovação perante a autoridade emissora.

6.2 Priorização em períodos sensíveis

Nos períodos de maior sensibilidade eleitoral, especialmente nos fins de semana de primeiro e segundo turno, a estrutura de resposta contará com priorização máxima de triagem e execução, integrando equipe jurídica interna, escritórios externos, equipes de Segurança e de Relações Governamentais. Essa configuração permite tratar, de forma coordenada, ordens judiciais, ofícios, alertas do SIADE, comunicações de autoridades e escalonamentos relativos a conteúdos de gravidade extraordinária ou propagação acelerada.

O resultado esperado é o cumprimento tempestivo e integral das ordens judiciais validamente recebidas, observados o escopo e o prazo fixado pela autoridade competente.

7. CANAIS DE DENÚNCIA, NOTIFICAÇÃO, CONTESTAÇÃO E ATENDIMENTO PRIORITÁRIO A CANDIDATURAS (Portaria, art. 17; Resolução, arts. 9º-D, II; 9º-E; 28, §§ 4º e 4º-B; 30; 32)

Medida de conformidade. O X BRASIL disponibiliza e opera, conforme a natureza da demanda, um conjunto de canais dedicados e complementares, voltados ao processamento tempestivo de ordens judiciais, notificações regulatórias e extrajudiciais, denúncias de usuários e comunicações prioritárias de candidatas, candidatos, partidos políticos, federações e coligações – canais especialmente relacionados às hipóteses previstas no artigo 9º-E da Resolução, incluindo condutas antidemocráticas, fatos notoriamente inverídicos ou gravemente descontextualizados, ameaças graves, discurso de ódio, conteúdo sintético não rotulado ou proibido, reprodução de conteúdo já removido e violência política de gênero:

- o **canal prioritário juridicoeleicoes_NE@x.com**, voltado ao atendimento de candidatas, candidatos, partidos políticos, federações e coligações – a solução específica exigida pelo artigo 9º-E, § 2º, da Resolução;

- o **canal juridicoeleicoes@x.com**, destinado ao recebimento de intimações, ordens judiciais e comunicações formais da Justiça Eleitoral;

- os **mecanismos de denúncia de conteúdo ilegal no Brasil, inclusive violações eleitorais** (https://help.x.com/pt/rules-and-policies/brazil-resources > https://help.x.com/pt/forms/br/report); e

- alertas e denúncias encaminhados pelo TSE no âmbito do Memorando de Entendimento-TSE nº 23/2026, de 16 de julho de 2026, inclusive por meio do SIADE, quando aplicável, ao qual foi indicado o e-mail eleicoesbr@x.com.

Fluxo de triagem. As denúncias e notificações recebidas pelos canais gerais da plataforma são direcionadas às equipes responsáveis pela moderação e aplicação das Regras do X. Já as comunicações eleitorais sensíveis, os alertas institucionais e as comunicações de candidaturas e partidos são triados pela estrutura jurídica local, com apoio dos escritórios externos Pinheiro Neto Advogados e Pinheiro Guimarães e, quando necessário, das equipes globais das Operadoras do X, para definição do tratamento adequado, do nível de prioridade e do caminho operacional de execução.

8. GOVERNANÇA DE SISTEMAS DE INTELIGÊNCIA ARTIFICIAL: O GROK (Resolução, arts. 9º-B, § 5º; 9º-C, § 1º; 28, § 1º-C; Portaria, art. 15)

8.1 Natureza do serviço e responsabilidade

O Grok constitui modelo de inteligência artificial generativa desenvolvido e operado pela SpaceXAI, disponibilizado aos usuários por meio da plataforma X, do site grok.com e de aplicativos móveis para iOS e Android. Embora a implementação de filtros de segurança, o aperfeiçoamento de salvaguardas e o funcionamento do modelo sejam de competência exclusiva da SpaceXAI, o X BRASIL, as Operadoras do X e a SpaceXAI atuam em regime de cooperação para assegurar a observância das obrigações regulatórias aplicáveis ao período eleitoral.

8.2 Descrição funcional (Portaria, art. 5º, § 1º, I)

Em linguagem acessível: *chatbots* baseados em grandes modelos de linguagem (*large language models* – LLMs), categoria à qual pertence o Grok, não operam como bancos de dados de fatos verificados, tampouco como repositórios de “verdades” indexadas sob controle editorial humano. Trata-se de sistemas probabilísticos treinados para prever, a partir de padrões estatísticos aprendidos em textos, a sequência de palavras (*tokens*) mais verossímil em resposta a um comando formulado pelo usuário, produzindo texto linguisticamente coerente e plausível.

O Grok pode, adicionalmente, considerar conteúdo público recente (publicações de usuários e demais informações disponíveis na rede) para contextualizar a resposta. Nessa hipótese, o modelo atua precipuamente como sintetizador reativo do que já circula no ambiente público digital, e não como fonte primária de informação. O Grok não possui banco de dados de informações ou publicações, não realiza previsões ou inferências sobre usuários individuais, não cria perfis personalizados e não é utilizado para personalização de conteúdo ou publicidade. O contexto de cada conversa limita-se à sessão em andamento, e o modelo não acessa conversas anteriores de outros usuários durante o processo de geração de respostas.

8.3 Salvaguardas estruturais de segurança

A utilização do Grok por site e aplicativos está sujeita ao aceite aos Termos de Serviço[49] da xAI e de sua Política de Uso Aceitável[50]. A SpaceXAI implementa múltiplas camadas de salvaguardas (*guardrails*) destinadas a prevenir respostas inadequadas do modelo, incluindo:

(a) filtragem de dados de treinamento em múltiplos estágios, por meio de filtragem heurística baseada em palavras-chave e listas de exclusão, seguida de filtragem por classificador de documentos treinado;

(b) técnicas de deduplicação e *fingerprinting* para identificação de potenciais vazamentos de dados pessoais;

(c) destilação de contexto de segurança para impedir que o Grok apresente respostas inadequadas;

(d) filtros de visibilidade (*Visibility Filters*) aplicáveis a categorias sensíveis de conteúdo;

(e) avaliação e *feedback* humano por tutores de IA; e

(f) exercícios contínuos de *red team*, nos quais conjuntos de *prompts* de teste são executados antes de qualquer implantação do modelo, para assegurar que não produza respostas inadequadas.

Adicionalmente, o *Frontier Artificial Intelligence Framework*[51] da SpaceXAI e os *Model Cards* do Grok[52] documentam publicamente avaliações de segurança, proteção e comportamento do modelo.

8.4 Salvaguardas contra recomendação, endosso e favorecimento político-eleitoral (Resolução, art. 28, § 1º-C, I e II; art. 9º-C, § 1º; Portaria, art. 15, I)

Medida de conformidade. Para atender a essas vedações, a SpaceXAI está implementando conjunto integrado de salvaguardas técnicas e operacionais, que se somam à infraestrutura de

[49] https://x.ai/legal/terms-of-service
[50] https://x.ai/legal/acceptable-use-policy
[51] https://data.x.ai/2025-12-31-xai-frontier-artificial-intelligence-framework.pdf
[52] https://media.x.ai/v1/website/card-4p6-4cd2dc57.pdf

segurança já existente no modelo Grok, incluindo classificadores de recusa, filtros de conteúdo, mecanismos de rotulagem e protocolos de monitoramento.

O Grok conta com mecanismos de classificação e recusa (*classifiers* e *refusal systems*) que atuam no momento da geração de respostas, reduzindo o risco de que o modelo produza conteúdos que configurem as condutas vedadas pela legislação eleitoral. Esses mecanismos operam de forma integrada com as regras internas de segurança do modelo e com as políticas de conteúdo aplicáveis à plataforma X, reduzindo o risco de que o Grok dê recomendação, sugestão ou priorização de candidatos, indique preferência eleitoral, ou recomende voto a algum candidato específico.

Adicionalmente, em atendimento ao artigo 9º-C, § 1º, da Resolução, os mesmos mecanismos de recusa reduzem o risco de geração, pelo Grok, de conteúdo sintético em áudio, vídeo ou ambos (*deepfake*) apto a prejudicar ou beneficiar candidatura.

O Grok conta, ainda, com mecanismo de *feedback* dos usuários como medida contínua de aprimoramento: a SpaceXAI considera comentários dos usuários sobre as respostas do Grok para auxiliar no ajuste fino do desempenho e dos mecanismos de proteção.

8.5 Salvaguardas contra conteúdo íntimo não consensual e violência política de gênero (Resolução, art. 28, § 1º-C, III e IV; art. 9º-E; Portaria, art. 15, II)

Medida de conformidade. O Grok incorpora salvaguardas de geração de conteúdo específicas que ajudam a bloquear **(i)** a criação ou promoção de alterações em fotografia, vídeo ou outro registro audiovisual que contenha cena de sexo, nudez ou pornografia envolvendo pessoas reais e **(ii)** a formulação de conteúdos que representem ato de violência, implementadas conforme as salvaguardas estruturais explicadas acima.

Essas salvaguardas técnicas de geração são complementadas, no âmbito da plataforma X, pelo conjunto de políticas descrito acima – aplicável igualmente ao conteúdo sintético gerado por usuários:

- a **Política de Conduta de Propagação de Ódio** proíbe promover violência, atacar ou ameaçar pessoas com base em sexo e identidade de gênero, entre outras categorias protegidas, com compromisso expresso de combate ao assédio dirigido a silenciar vozes historicamente marginalizadas;

- a **Política de Abuso e Assédio** proíbe o assédio direcionado com a intenção de humilhar ou degradar indivíduos, incluindo conteúdo sexual indesejado e objetificação gráfica que sexualize um indivíduo sem seu consentimento – vedação aplicável também a respostas e a conteúdo gerado por IA;

- a **Política de Autenticidade** proíbe o compartilhamento de mídia inautêntica que possa enganar ou confundir pessoas e causar danos, representando camada adicional de proteção contra a disseminação de conteúdos manipulados ou fabricados, inclusive os de natureza sexual; e

- a **Política de Conteúdo Privado e Nudez Não Consensual** proíbe estritamente o compartilhamento de mídias sexuais, de nudez ou íntimas sem consentimento, incluindo expressamente imagens ou vídeos que sobrepõem ou manipulam digitalmente o rosto de uma pessoa no corpo nu de outra – hipótese que alcança diretamente o conteúdo sintético de cunho sexual.

A aplicação dessas políticas segue os critérios de gravidade, direcionamento e histórico de violações e o conjunto graduado de medidas descritos acima, incluindo a suspensão de contas dedicadas às condutas vedadas, inclusive na primeira detecção. As violações podem ser denunciadas pelos fluxos e formulários descritos, incluindo o formulário específico de Nudez Não Consensual, apto a receber denúncias de conteúdo natural ou sintético, com passo a passo disponível na versão aplicativo e no computador.

Esse conjunto normativo e operacional demonstra que a plataforma dispõe de estrutura própria de prevenção, moderação e sanção de condutas abusivas, alinhada à legislação aplicável e especialmente rigorosa na proteção de mulheres.

Prazo, resultado e monitoramento. As medidas descritas nesta seção são empregadas na plataforma X há longa data.

8.6 Resistência a tentativas de contorno (Portaria, art. 15, III e V)

Os mecanismos destinados a prevenir, detectar, bloquear e mitigar tentativas de contornar, neutralizar ou explorar indevidamente as salvaguardas de segurança compreendem, no estado atual: os classificadores de recusa e filtros de conteúdo descritos acima, que atuam no

momento da geração; e os exercícios contínuos de *red team*, nos quais conjuntos de *prompts* de teste são executados antes de qualquer implantação do modelo.

8.7 Rotulagem de conteúdo sintético e identificação de IA (Resolução, art. 9º-B; Portaria, art. 15, IV)

Medida de conformidade. O X permite que usuários apliquem rótulos de identificação a conteúdos gerados ou substancialmente modificados por inteligência artificial. O X também possui mecanismos que podem identificar e rotular automaticamente conteúdos gerados por IA quando as ferramentas de IA incluem certas marcas d'água como a C2PA[53]. A rotulagem "*Feito com IA"* é exibida nas publicações orgânicas na plataforma, conferindo transparência aos usuários quanto à natureza sintética do conteúdo. Uma vez ativado, o rótulo é exibido de forma visível e permanente junto ao conteúdo, permitindo que os demais usuários identifiquem sua origem sintética.

Por sua vez, todos os conteúdos de mídia gerados por usuários do X pelo Grok são automaticamente marcados com o rótulo "Feito com IA" que identifica sua origem sintética. A etiqueta é aplicada de forma automática no momento da geração do conteúdo, independentemente de qualquer ação do usuário, conferindo rastreabilidade e transparência quanto à natureza do conteúdo desde a sua criação.[54]

Caso conteúdos sintéticos violem as disposições dos artigos 9º-B ou 9º-C da Resolução, eles podem ser denunciados pelos canais descritos acima. A remoção de conteúdo observa estritamente as obrigações legais e a Resolução, com atuação coordenada entre as equipes de moderação, a equipe jurídica local e os escritórios externos. Registre-se, por fim, que não há possibilidade de utilização de inteligência artificial para impulsionamento de conteúdo político-eleitoral, uma vez que a plataforma X não permite impulsionamento dessa natureza no Brasil.

[53] https://c2pa.org/

[54] No caso específico do @grok, todo o conteúdo gerado pela conta é automatizado. Por essa razão, a rotulagem voluntária pelo usuário não se aplica a essa conta, sendo a identificação de origem sintética assegurada pelos próprios mecanismos de geração e pelas salvaguardas técnicas do modelo. Mídias geradas nessa interação atualmente possuem marca d'água.

9. GESTÃO DO PERÍODO DE VEDAÇÃO ELEITORAL (Portaria, art. 16; Resolução, arts. 9º-B, §§ 3º-A e 4º; 29, § 11)

Medida de conformidade. Medidas de restrição temporal na criação de mídia de candidatos pelo Grok serão ativadas no início do período de vedação e desativadas ao fim do período de vedação. A medida envolve mecanismos para ajudar a restringir a publicação de imagens dessa natureza pelo Grok no país. Já os canais de denúncia informados anteriormente darão prioridade máxima a denúncias de circulação de conteúdos feitos com IA.

Quanto ao desligamento automático de anúncios pagos (Resolução, art. 29, § 11), a obrigação não incide sobre a plataforma em seu núcleo, uma vez que o X não veicula anúncios políticos no Brasil; os sistemas que detectam e bloqueiam anúncios políticos direcionados ao Brasil permanecem ativos de forma contínua, inclusive durante o período de vedação.

10. PUBLICIDADE POLÍTICA PAGA: INAPLICABILIDADE E DILIGÊNCIA RESIDUAL (Portaria, arts. 18 e 19; Resolução, arts. 27-A, 28 e 29)

10.1 Inaplicabilidade do núcleo das obrigações

Os artigos 27-A, 28 e 29 da Resolução, regulamentados pelos artigos 18 e 19 da Portaria, disciplinam deveres de transparência, rastreabilidade e integridade aplicáveis ao impulsionamento de conteúdo político-eleitoral, incluindo a manutenção de biblioteca de anúncios, a identificação de anunciantes, a rastreabilidade do conteúdo impulsionado em compartilhamentos, a vedação à propaganda negativa e a contratação de impulsionamento exclusivamente por candidaturas, partidos, federações e coligações.

Medida de conformidade. No caso do X, tais obrigações devem ser interpretadas à luz de uma premissa operacional central**:** a plataforma não oferece serviço de impulsionamento de conteúdo político-eleitoral no Brasil, razão pela qual as obrigações específicas relativas à contratação, veiculação e biblioteca de anúncios políticos não incidem sobre o X BRASIL em seu núcleo principal**.**

Essa restrição está imposta na Política de Conteúdo Político do X[55], que permite anúncios de conteúdo político e anúncios de campanha política apenas em países determinados, entre os quais o Brasil não está listado. A política define “anúncios de conteúdo político” como anúncios

[55] https://business.x.com/pt/help/ads-policies/ads-content-policies/political-content

que fazem referência a candidato, partido político, oficial de governo eleito ou indicado, eleição, referendo, contagem de votos, legislação, regulamentação, diretiva ou resultado judicial, e "anúncios de campanha política" como aqueles que defendem ou criticam candidato ou partido, pedem voto, solicitam apoio financeiro para eleição. A política esclarece, ainda, que, em países nos quais anúncios políticos são proibidos, os anúncios não podem fazer referência a referendos, medidas submetidas a voto, projetos de lei, legislação, regulamentação, diretivas, resultados judiciais ou equivalentes específicos do país – o que reforça a vedação aplicável ao Brasil.

Sem prejuízo da inaplicabilidade do núcleo das obrigações, o X mantém controles residuais para prevenir tentativas de contratação irregular, classificação incorreta ou burla de suas políticas de anúncios, na forma exigida pelo artigo 19, § 1º, da Portaria, conforme descrito a seguir.

10.2 Descrição funcional da proibição e dos controles de política

No plano funcional, anúncios que se enquadrem nas categorias de conteúdo político ou campanha política não são aprovados para segmentação no Brasil – vedação que abrange anúncios com referência a eleições, candidatos, partidos, agentes públicos eleitos ou indicados, referendos, legislação, regulamentação, diretivas ou resultados judiciais, bem como anúncios que defendam ou critiquem candidatos ou partidos, peçam voto ou solicitem apoio financeiro eleitoral. Assim, conteúdos desse tipo são detectados e barrados, o que reduz o risco de contorno da vedação por meio de formatos alternativos de publicidade paga.

O X Ads possui requisitos gerais de qualificação[56] aplicáveis a todos os anunciantes. Para participar do X Ads, as contas precisam atender a critérios de elegibilidade, e os anunciantes permanecem responsáveis por todo o conteúdo promovido, inclusive pela conformidade com leis e regulamentos aplicáveis. A conformidade da conta considera elementos como biografia, nome de usuário, identificador, imagem de perfil, imagem de cabeçalho e URL da biografia, além de eventuais requisitos adicionais previstos em políticas específicas. Para ser elegível, a conta deve estar verificada – pelo programa de Organizações Verificadas, no caso de empresas e entidades governamentais, ou pelo X Premium, no caso de indivíduos –, seus posts e informações de anunciante devem ser públicos, e a conta deve observar requisitos de qualidade, incluindo imagens de perfil e capa que não sejam GIFs e URL funcional na biografia que represente com precisão a marca e o produto ou serviço promovido. No contexto eleitoral

[56] https://business.x.com/pt/help/ads-policies/campaign-considerations/about-eligibility-for-twitter-ads

brasileiro, esses requisitos reduzem o risco residual de uso indevido dos produtos de publicidade de maneira não autorizada

Por fim, o X adota política específica para mídia afiliada ao Estado[57], segundo a qual veículos de mídia afiliados ao Estado – bem como pessoas físicas que respondam por tais entidades ou estejam diretamente afiliadas a elas – não podem comprar anúncios, sendo igualmente proibida a promoção de conteúdo de mídia afiliada ao Estado. Essa política reduz o risco de uso de estruturas de mídia sob controle estatal para promoção paga de mensagens com potencial impacto no debate político-eleitoral.

10.3 Controles de detecção, revisão e escalonamento

A diligência residual do X combina requisitos prévios de qualificação de anunciantes, controles de política, detecção automatizada, revisão humana e escalonamentos prioritários para prevenir tentativas de veiculação de publicidade política no Brasil ou de burla às políticas de anúncios da plataforma:

- os sistemas de anúncios contam com controles técnicos e operacionais destinados a detectar anúncios violadores no momento de criação da campanha, incluindo modelos de aprendizado de máquina, lógicas de negócio e termos em listas de bloqueio;

- quando um termo é incluído em lista de revisão de anúncios, conteúdo promovido que mencione o termo ou expressão correspondente pode ser automaticamente colocado em estado de retenção para revisão, exigindo análise humana antes de prosseguir;

- as contas e os conteúdos de anunciantes passam por revisão de qualidade e segurança quando optam por promover conteúdo pelo X Ads, combinando algoritmos de aprendizado de máquina e revisão humana para verificar aderência às políticas do X;

- o X pode remover anunciantes da plataforma de anúncios quando a utilizam em violação às políticas aplicáveis;

- anúncios denunciados por usuários que atinjam determinados critérios podem ser reenviados automaticamente para revisão humana secundária, mesmo que inicialmente aprovados; e

[57] https://business.x.com/pt/help/ads-policies/ads-content-policies/state-media

- durante o período eleitoral brasileiro, os sistemas que detectam e bloqueiam anúncios políticos direcionados ao Brasil permanecem ativos, assim como os sistemas que impedem o *onboarding* de contas de candidaturas para fins de publicidade. A lista final de candidatos conforme disponível em https://dadosabertos.tse.jus.br/dataset/candidatos-2026 será utilizada para tal monitoramento e o *denylist* preventivo implementado conforme plano de controle.

Eventuais escalonamentos relacionados a anúncios serão tratados por meio dos canais prioritários de suporte eleitoral e dos fluxos de atendimento institucional, permitindo que tentativas de veicular ou contornar a proibição de publicidade política sejam rapidamente revisadas e interrompidas.

10.4 Transparência, denúncias e informações ao usuário

O X disponibiliza mecanismos de transparência sobre a seleção e a entrega de anúncios, incluindo a funcionalidade "*Por que este anúncio?*", que fornece aos usuários informações sobre como anúncios são selecionados e exibidos. Usuários podem denunciar anúncios potencialmente violadores diretamente na *timeline* ou por meio de formulário da Central de Ajuda, e os conteúdos denunciados são priorizados e revisados conforme os fluxos de *enforcement* aplicáveis. A Política de Segmentação de Categorias Sensíveis prevê a análise das violações denunciadas e a adoção de providências apropriadas, que podem incluir a remoção de anunciantes e anúncios da plataforma X Ads. Além disso, Notas da Comunidade podem ser adicionadas a conteúdos promovidos, ajudando a contextualizar alegações de anunciantes e oferecendo aos usuários acesso a informações adicionais.

10.5 Canal para atendimento de requisições do TSE e retenção de dados (Portaria, art. 19, § 2º)

Nos termos do artigo 19, § 2º, da Portaria, provedores que não prestem serviço de impulsionamento de conteúdos político-eleitorais, inclusive sob a forma de priorização de resultados de busca, devem prever fluxo e canal para atendimento de requisições do TSE relativas a anúncios ou impulsionamentos veiculados durante o período eleitoral e aos respectivos anunciantes, armazenados na forma do artigo 16-M do Decreto nº 12.975/2026.

Para esse fim, o X BRASIL utilizará os canais e fluxos institucionais descritos na Seção 9 para recebimento, triagem e resposta a requisições do TSE, com apoio da equipe jurídica local, dos

escritórios externos e das equipes globais responsáveis por anúncios, dados, segurança e operações, conforme necessário. As informações responsivas serão fornecidas dentro dos prazos legais aplicáveis e nos formatos tecnicamente disponíveis, observadas as limitações decorrentes da inexistência de impulsionamento político-eleitoral no Brasil e das capacidades técnicas dos sistemas de anúncios do X.

10.6 Resultado esperado, indicadores e trajetória

O resultado esperado desta medida é a manutenção de um volume residual mínimo (entre 0 e 50 ocorrências por ciclo eleitoral) de impulsionamentos político-eleitorais contratados ou veiculados irregularmente no Brasil por meio dos produtos de publicidade do X, em razão da proibição da política aplicável e dos controles técnicos e operacionais que bloqueiam anúncios políticos direcionados ao país. Caso alguma tentativa escape aos bloqueios automatizados, espera-se que, pelo menos, 80% sejam identificados e tratados de forma proativa, ou seja, antes de denúncias de usuários.

Como indicadores de acompanhamento, o X BRASIL poderá considerar: **(i)** a manutenção da política de bloqueio de anúncios políticos no Brasil; **(ii)** a manutenção dos requisitos de elegibilidade de anunciantes aplicáveis ao X Ads; **(iii)** o número de tentativas identificadas, bloqueadas ou escaladas de anúncios potencialmente políticos; **(iv)** o número de denúncias de anúncios relacionadas a conteúdo político-eleitoral; **(v)** o tempo de triagem e resposta a escalonamentos prioritários; e **(vi)** a taxa de resposta tempestiva a requisições do TSE.

A trajetória de alcance não depende da implementação progressiva de biblioteca de anúncios políticos ou de mecanismos de contratação de impulsionamento eleitoral, pois tais funcionalidades não são oferecidas pelo X no Brasil.

10.7 Justificativa de inaplicabilidade das regras de biblioteca de anúncios políticos (Portaria, art. 4º, § 2º)

As regras relativas à biblioteca de anúncios políticos pressupõem a prestação de serviço de impulsionamento político-eleitoral, inclusive com identificação de anunciantes, manutenção de ferramenta de consulta, rastreabilidade e mecanismos de declaração de uso de inteligência artificial no fluxo de contratação. Como o X não oferece impulsionamento de conteúdo político-eleitoral no Brasil, a criação de biblioteca de anúncios políticos brasileira não é aplicável ao núcleo da operação da plataforma.

A diligência proporcional exigível ao X BRASIL, portanto, consiste em documentar a proibição aplicável, demonstrar os controles que impedem a contratação ou veiculação de anúncios políticos no Brasil, manter os requisitos gerais de elegibilidade para uso do X Ads, preservar canais para atendimento de requisições do TSE e monitorar riscos residuais de tentativa de burla, classificação incorreta ou veiculação indevida – exatamente como descrito nas subseções anteriores.

11. PROTEÇÃO DE DADOS PESSOAIS EM CONTEXTO ELEITORAL (Portaria, art. 21; Resolução, arts. 33, 33-A e 33-B)

Medida de conformidade. A plataforma X não oferece aos anunciantes funcionalidades de perfilamento ou de microdirecionamento de propaganda eleitoral no Brasil. Não há tratamento de dados pessoais com a finalidade de direcionar propaganda eleitoral a segmentos específicos de eleitores com base em características individuais, preferências políticas ou comportamentos inferidos. Tampouco o Grok é utilizado para criar perfis de usuários, realizar previsões ou inferências individualizadas, ou personalizar conteúdo ou publicidade exibido aos usuários do X. Ante o exposto, as obrigações relativas a perfilamento e microdirecionamento de propaganda eleitoral previstas nos artigos 33, 33-A e 33-B da Resolução não se aplicam ao X, uma vez que a plataforma não oferece tais funcionalidades no Brasil.

Sem prejuízo, o X mantém operacionais os mecanismos de conformidade com a LGPD e com as disposições aplicáveis da Resolução:

- **Transparência ao titular:** todas as informações sobre o tratamento de dados pessoais estão disponíveis, de forma clara e acessível, por meio dos Termos de Serviço e da Política de Privacidade. Adicionalmente, a Central de Ajuda disponibiliza orientações específicas sobre o funcionamento do Grok e sobre as configurações de privacidade disponíveis aos usuários, incluindo a possibilidade de exercer o direito de oposição ao uso de seus dados para treinamento do modelo de IA;

- **Controles conferidos aos titulares:** o X oferece aos usuários a possibilidade de optar por não ter seus Dados de Postagens Públicas e/ou Dados de *Prompt* utilizados para o treinamento e aprimoramento do Grok, por meio de configuração acessível no caminho: Configurações e Suporte → Configurações e Privacidade → Privacidade e Segurança → Compartilhamento de Dados e Personalização → Grok;

- **Exercício de direitos dos titulares:** o X disponibiliza canais específicos e acessíveis para que usuários e não usuários possam, de forma simples, fazer solicitações relacionadas ao tratamento de seus dados pessoais, por meio de formulários específicos disponíveis na plataforma[58]; e

- **Segurança da informação e resposta a incidentes:** o Programa de Segurança da Informação do X utiliza controles provenientes de estruturas de melhores práticas do setor, especificamente a ISO 27001/2, abrangendo criptografia de dados, restrições de acesso baseadas em autenticação multifator, monitoramento contínuo de atividades e procedimentos de retenção e descarte seguro de dados, mantendo protocolos de resposta a incidentes de segurança.

Registre-se, por fim, na linha do artigo 21, parágrafo único, da Portaria, que os aspectos relacionados à interpretação, à fiscalização e à aplicação da LGPD permanecem sujeitos às competências legalmente atribuídas à ANPD.

12. ATUALIZAÇÃO DO PLANO E RELATÓRIO CONSOLIDADO (Portaria, arts. 24 e 28)

Nos termos do artigo 28 da Portaria, o X BRASIL encaminhará ao TSE, até o dia 19 de dezembro de 2026, relatório consolidado com informações relativas à efetiva implementação deste Plano de Conformidade. O relatório será apresentado de forma simples e estruturada, de modo a permitir a compreensão, a conferência e a comparação dos dados divulgados, inclusive entre ciclos eleitorais distintos (artigo 28, § 5º). As informações que não forem classificadas como de acesso restrito serão divulgadas na camada de acesso público (artigo 28, § 3º). Os compromissos de monitoramento assumidos ao longo deste Plano serão consolidados nesse relatório, na forma do artigo 11, parágrafo único, da Portaria.

Independentemente do relatório consolidado, o X BRASIL comunicará à Presidência do TSE, em até 3 (três) dias após a constatação, a ocorrência de risco extraordinário à integridade do processo eleitoral, assim entendido aquele que, por sua escala, novidade ou potencial de dano, exija resposta que não possa aguardar o relatório único (artigo 28, § 1º). A comunicação informará a natureza e o alcance estimado do risco, as medidas de mitigação já adotadas ou em curso e, quando cabível, a previsão de medidas adicionais e o respectivo cronograma estimado de implementação (artigo 28, § 2º).

[58] https://help.x.com/pt/forms

O X BRASIL atualizará este Plano de Conformidade sempre que ocorrer alteração substancial de risco ou de funcionalidades que afete materialmente as medidas nele descritas, comunicando a alteração à Presidência do TSE em até 48 (quarenta e oito) horas, nos termos do artigo 24, parágrafo único, da Portaria. Alterações rotineiras, testes ou ajustes editoriais sem impacto relevante sobre a integridade do processo eleitoral não ensejam dever de atualização.

13. CONCLUSÃO

O presente Plano de Conformidade demonstra o compromisso do X BRASIL, das Operadoras do X e da SpaceXAI com a integridade do processo eleitoral brasileiro. As medidas aqui descritas foram estruturadas para atender às exigências da Portaria nº 463/2026 e da Resolução nº 23.610/2019, observadas as premissas de autonomia técnica, adequação finalística, transparência, proporcionalidade e verificabilidade estabelecidas pelo artigo 4º da Portaria.

As medidas proativas previstas neste Plano são orientadas por risco e direcionadas às categorias específicas de dano eleitoral identificadas no arcabouço regulatório aplicável, não constituindo compromisso de monitoramento de todo e qualquer conteúdo publicado na plataforma nem garantia de identificação de toda atividade potencialmente ilícita ou violadora das políticas da plataforma (artigo 14, parágrafo único, da Portaria).

O X BRASIL permanece à disposição do TSE para prestar quaisquer esclarecimentos que se façam necessários.