OpenAI

Portaria TSE nº 463/2026

PLANO DE CONFORMIDADE

Apresentado por:

OpenAI OpCo, LLC

Ao:

Tribunal Superior Eleitoral (TSE)

República Federativa do Brasil

Campo	Detalhes
Razão social	OpenAI OpCo, LLC
Endereço	1455 3rd Street, San Francisco, CA 94158, United States
Canal designado para comunicações com o TSE	[RESTRITO]
Data de apresentação	16 de agosto de 2026
Versão	1.0 - Apresentação inicial
Período eleitoral aplicável	Eleições Gerais de 2026

AVISO DE CONFIDENCIALIDADE

Este documento contém informações classificadas nas camadas de acesso público e de acesso restrito, conforme definidas no Art. 8º da Portaria nº 463/2026. As informações restritas são identificadas nas seções correspondentes e não podem ser divulgadas para qualquer terceiro além do pessoal autorizado e integrante do TSE sem o consentimento prévio por escrito da OpenAI OpCo, LLC.

OpenAI

SUMÁRIO

1. INTRODUÇÃO E INFORMAÇÕES GERAIS

1.1. Objetivo e enquadramento executivo

Este Plano de Conformidade ("Plano") é apresentado pela OpenAI OpCo, LLC ("OpenAI") no âmbito da Portaria TSE nº 463/2026 ("Portaria") e da Resolução TSE nº 23.610/2019 ("Resolução"), com o objetivo de explicar como os serviços e as funcionalidades relevantes para as operações da OpenAI no Brasil são tratados, para fins de prevenção e mitigação de riscos à integridade do processo eleitoral brasileiro.

O Plano é estruturado em torno do ChatGPT[1] como o serviço central objeto da análise, disponibilizado aos usuários tanto por aplicativos móveis e de desktop quanto pela web, em <chatgpt.com>. Essa abordagem reflete a natureza do ChatGPT como o principal serviço de Inteligência Artificial ("IA") generativa voltado aos usuários da OpenAI e o serviço por meio do qual os usuários interagem diretamente com os modelos da OpenAI em um ambiente conversacional.

O Plano não desconsidera, contudo, outros produtos, serviços ou funcionalidades da OpenAI. Em vez disso, trata-os de forma conectada, identificando se cada serviço ou funcionalidade faz parte do ChatGPT, é acessório ao ChatGPT, ou está fora do escopo principal da regulamentação eleitoral. Para tanto, são considerados também os serviços ou funcionalidades da OpenAI que possuam uma interface pública limitada e os produtos de infraestrutura para desenvolvedores ou empresas.

Essa estrutura é consistente com a Portaria, que permite que os planos de conformidade sejam apresentados de forma integrada, preservando, ao mesmo tempo, a identificação individualizada das informações aplicáveis a cada serviço ou funcionalidade, especialmente quando diferentes recursos, arquiteturas ou riscos possam desempenhar papéis distintos no contexto das obrigações contempladas pela Portaria (Art. 2º, parágrafo único).

1.2. OpenAI e ChatGPT

A OpenAI é uma empresa de pesquisa e implementação de IA fundada em 2015, com a missão de garantir que a IA geral beneficie toda a humanidade. A OpenAI lançou o ChatGPT em 30 de novembro de 2022, inicialmente como versão preliminar de pesquisa destinada a promover a pesquisa e o desenvolvimento de IA, ampliar o acesso às tecnologias de IA, incentivar o uso responsável e coletar feedback para aprimorar o serviço.

O ChatGPT é um serviço de IA voltado ao usuário que permite interagir com os modelos da OpenAI, por meio de comandos em linguagem natural e multimodais, recebendo respostas geradas pelo sistema. De modo geral, o ChatGPT foi concebido para compreender e responder a perguntas e instruções dos usuários mediante o aprendizado de padrões a partir

[1] Disponível em: https://chatgpt.com/. Acessado em 14 de agosto de 2026.

de grandes volumes de informações, incluindo texto, imagens, áudio e vídeo, a depender do modelo e dos recursos relevantes do produto.

Os usuários podem utilizar o ChatGPT para uma ampla variedade de tarefas, como organizar e resumir informações, auxiliar em traduções, analisar ou gerar imagens, apoiar a criatividade e a geração de ideias, escrever código e realizar outras tarefas do dia a dia. O ChatGPT também pode incluir recursos de produto e experiências de usuário amparadas por componentes mais amplos do sistema, incluindo interfaces de usuário, filtros de segurança, mecanismos de monitoramento, funcionalidade de cobrança e outros componentes de software que, em conjunto, formam um sistema de IA funcional.

Para os fins deste Plano, é importante distinguir entre modelo de IA e sistema de IA. O modelo de IA é o componente que possibilita inferência, previsão, raciocínio etc., mas não opera, por si só, como um serviço completo voltado ao usuário. Já o sistema de IA inclui o modelo juntamente com componentes adicionais, como interface de produto, filtros de segurança, monitoramento, armazenamento, avaliação, cobrança e outros elementos operacionais que permitem que o modelo seja implementado e utilizado no mundo real.

Essa distinção é relevante para o presente Plano porque a Portaria exige uma avaliação das obrigações que seja funcional e específica por serviço, e não uma avaliação abstrata de modelos de IA isoladamente. Portanto, o Plano concentra-se no ChatGPT como serviço implementado e sistema de IA, discutindo modelos subjacentes e funcionalidades relacionadas apenas na medida necessária para explicar as salvaguardas, limitações e medidas de mitigação de riscos aplicáveis.

1.3. Operação técnica e limitações do ChatGPT

Durante o treinamento, os modelos aprendem padrões e relações a partir de grandes quantidades de informação, incluindo as relações entre palavras e seu contexto e, no caso de modelos de geração de imagens, entre características visuais e o texto associado. O treinamento ajusta valores numéricos conhecidos como pesos ou parâmetros[2] para refletir esses padrões aprendidos. Os modelos resultantes utilizam esses parâmetros para gerar novos conteúdos; eles não operam como bancos de dados que contêm cópias recuperáveis dos dados de treinamento.

Como a geração do modelo inclui um elemento probabilístico, prompts iguais ou semelhantes podem produzir resultados diferentes. Os Termos de Uso da OpenAI[3] também explicam que os resultados podem ser imprecisos, incompletos ou podem não refletir com precisão pessoas, lugares ou fatos reais. Os usuários são instruídos a avaliar a precisão e a adequação dos

[2] Valores numéricos de um modelo de aprendizado de máquina que são ajustados durante o treinamento para refletir os padrões aprendidos e ajudar a determinar como o modelo transforma entradas em saídas. Esses valores não constituem um banco de dados que contenha cópias recuperáveis dos dados utilizados no treinamento.

[3] Disponível em: https://openai.com/pt-BR/policies/row-terms-of-use/. Acessado em 14 de agosto de 2026.

resultados, inclusive por meio de revisão humana quando apropriado, e a não confiar nos resultados como única fonte de informações factuais ou de aconselhamento profissional.

Os usuários também são instruídos a avaliar a precisão e a adequação dos resultados para suas necessidades específicas antes de utilizar ou compartilhar resultados dos Serviços da OpenAI.[4] Isso é particularmente importante em contextos sensíveis, incluindo contextos político-eleitorais, jurídicos, de saúde, financeiros e outros contextos relevantes, nos quais o julgamento humano independente continua sendo necessário.

A OpenAI trata essas limitações do ChatGPT por meio de uma combinação de princípios que devem embasar o comportamento do modelo, salvaguardas em nível de sistema, políticas de uso, medidas de monitoramento e avaliação e outros controles que podem afetar tanto os modelos quanto os sistemas de IA que eles embasam, os quais serão, conforme aplicável, tratados neste Plano. Essas medidas destinam-se a mitigar o uso indevido do ChatGPT e apoiar o uso seguro e responsável do sistema, mas não eliminam todos os riscos possíveis nem garantem que todos os resultados sejam precisos ou livres de erros, o que, no estado atual da técnica, é irreconciliável com a natureza dos modelos de IA.

1.4. Escopo do Plano por grupo de serviços

Para fins de clareza, este Plano organiza os serviços e as funcionalidades da OpenAI nos grupos abaixo, para fins de tratamento das obrigações específicas da Portaria. Essa organização constitui um quadro metodológico adotado exclusivamente para os fins deste Plano e não corresponde às classificações oficiais da OpenAI. A OpenAI pode apresentar seus serviços e funcionalidades de forma diversa em outros contextos, embora com conteúdo substancialmente equivalente. O objetivo desta organização é identificar quais recursos são centrais para o Plano, quais são tratados como funcionalidades acessórias ou públicas limitadas e quais estão fora do escopo da Portaria.

Grupo 1. Funcionalidades próprias do ChatGPT

Este grupo abrange o ChatGPT e suas funcionalidades próprias desenvolvidas com base nos modelos da OpenAI e sujeitas às salvaguardas e aos controles da OpenAI. Isso inclui a experiência conversacional central do ChatGPT (web, dispositivos móveis e desktop), a

[4] "Você deve avaliar a precisão e adequação dos Resultados em função das suas necessidades concretas, inclusive por meio de verificação humana, se apropriado, antes de utilizar ou compartilhar os Resultados dos Serviços." Disponível em: https://openai.com/pt-BR/policies/row-terms-of-use/. Acessado em 14 de agosto de 2026.

busca do ChatGPT,[5] os GPTs,[6] a geração de imagens ("ImageGen") e as conversas por voz. Essas funcionalidades são tratadas como componentes integrados do serviço ChatGPT, por meio dos quais os usuários inserem prompts no sistema para gerar, sintetizar ou modificar conteúdo ou executar tarefas.

As obrigações mais diretamente relevantes para o Grupo 1 são aquelas aplicáveis a sistemas de IA generativa, modelos de linguagem, chatbots e tecnologias conversacionais equivalentes, tratadas no Art. 15 da Portaria. Elas incluem salvaguardas contra o favorecimento eleitoral, a recomendações de voto, os conteúdos político-eleitorais proibidos, o conteúdo sexual não consentido envolvendo candidatos, a violência política de gênero, o contorno das salvaguardas, a rotulagem de conteúdo sintético quando aplicável e os testes de resistência.

No caso da busca do ChatGPT, tanto a funcionalidade de recuperação quanto a síntese gerada pelo modelo estão sujeitas à estrutura de segurança da OpenAI, que inclui classificadores de páginas da web e outros filtros de segurança; contudo, a análise do Art. 15 pode distinguir, quando relevante, entre a recuperação pela funcionalidade de conteúdo de terceiros na web e os resultados gerados pelo próprio ChatGPT.

Grupo 2. Integrações de terceiros.

Este grupo inclui recursos que permitem integrações de terceiros no ambiente do ChatGPT, como Apps[7] e Conectores.[8] Esses recursos podem recuperar, exibir ou apresentar conteúdo de fontes de terceiros na interface do ChatGPT.

Esses recursos são tratados separadamente do Grupo 1 porque exibem conteúdo de fontes de terceiros que não é gerado pelos modelos da OpenAI, mas sim por determinação de terceiros. Consequentemente, no âmbito da Portaria, como a OpenAI não gera o conteúdo nem controla a arquitetura dos recursos de terceiros, não pode influenciar a forma como esses recursos operam em última instância, ficando, portanto, impossibilitada de assegurar a conformidade subsequente. Como consequência, a OpenAI não pode ser responsabilizada pelo uso desses

[5] A busca do ChatGPT (ou ChatGPT Search) é uma funcionalidade do ChatGPT que permite ao serviço pesquisar na web quando necessário e fornecer respostas que podem incluir links para fontes relevantes, para que os usuários possam consultar as informações subjacentes. Disponível em: https://openai./pt-BR/index/introducing-chatgpt-search/. Acessado em 14 de agosto de 2026.

[6] Também chamados de "GPTs personalizados", os GPTs são versões do ChatGPT configuradas sem necessidade de programação (no-code) para finalidades específicas dentro do próprio ChatGPT. Os criadores podem personalizá-las por meio de instruções, arquivos de conhecimento e capacidades selecionadas. Disponível em: https://help.openai.com/pt-br/articles/8554407-gpts-in-chatgpt . Acessado em 14 de agosto de 2026.

[7] "Apps" são integrações que conectam o ChatGPT ou o Codex a ferramentas, informações e funcionalidades externas. Dependendo de suas capacidades, os Apps podem recuperar ou sincronizar dados, exibir experiências interativas ou executar operações em serviços conectados. Disponível em: https://help.openai.com/pt-br/articles/11487775-apps-in-chatgpt. Acessado em 14 de agosto de 2026.

[8] "Connector" é um termo utilizado anteriormente em um produto do ChatGPT para se referir a integrações usadas para pesquisar, consultar ou sincronizar informações provenientes de serviços de terceiros. Atualmente, essas funcionalidades são apresentadas como Apps. Disponível em: https://help.openai.com/pt-br/articles/11487775-apps-in-chatgpt. Acessado em 14 de agosto de 2026.

recursos por terceiros. As salvaguardas do Art. 15 descritas neste Plano aplicam-se aos resultados gerados pelo ChatGPT, e não ao conteúdo de terceiros exibido por meio dessas integrações.

Grupo 3. Interfaces públicas ou de compartilhamento limitadas.

Este grupo inclui recursos que podem permitir que conteúdo gerado ou configurado por um usuário se torne visível ou acessível a outros usuários ou ao público, como links compartilhados do ChatGPT,[9] a GPT Store[10] (incluindo páginas públicas de GPTs)[11] e o ChatGPT Sites.[12]

Dada a natureza limitada, acessória e subsidiária dessas funcionalidades, elas não são capazes de transformar o ChatGPT em uma rede social de uso geral ou em uma plataforma de distribuição de conteúdo político que justificasse que este Plano tratasse das obrigações especificamente previstas na Portaria para esse tipo de funcionalidade ou serviço (por exemplo, as previstas nos Arts. 13 e 14 da Portaria). Em vez disso, devem ser consideradas apenas na medida em que viabilizem acessibilidade pública, descoberta, denúncia, retirada ou controle de visibilidade relevantes para a Portaria, no âmbito de obrigações mais amplas (por exemplo, os Arts. 12 e 20 da Portaria).

Por exemplo, os links compartilhados podem ser impedidos de ser criados ou tornarem-se inacessíveis quando a conversa subjacente no ChatGPT for sinalizada e for sujeita à aplicação de uma medida de aplicação das políticas pela OpenAI, podendo os destinatários denunciar

[9] Links compartilhados são um recurso que permite aos usuários gerar uma URL exclusiva para uma conversa do ChatGPT, que pode então ser compartilhada com amigos, colegas e colaboradores. Links compartilhados oferecem uma nova maneira de os usuários compartilharem suas conversas do ChatGPT, substituindo o método antigo e trabalhoso de compartilhar capturas de tela. Disponível em: https://help.openai.com/pt-br/articles/7925741-perguntas-frequentes-sobre-links-compartilhados-do-chatgpt. Acessado em 14 de agosto de 2026.

[10] Diretório público do ChatGPT por meio do qual GPTs elegíveis podem ser publicados e descobertos por usuários. Disponível em https://chatgpt.com/gpts/?openaicom-did=127137db-b5a2-400285-ebbbb28cc94d&openaicom_referred=true. Acessado em 14 de agosto de 2026.

[11] GPTs estão disponíveis para todos os usuários do ChatGPT, mas os usuários precisam estar conectados para iniciar uma conversa. Páginas públicas de GPTs podem ficar visíveis sem login, mas os usuários são solicitados a entrar antes de conversar. Disponível em: https://help.openai.com/pt-br/articles/8554407-gpts-in-chatgpt. Acessado em 14 de agosto de 2026.

[12] O ChatGPT Sites é uma funcionalidade que permite que usuários elegíveis criem, publiquem, hospedem e mantenham sites ou aplicativos web leves a partir do ChatGPT/Codex. Disponível em https://openai.com/academy/chatgpt-sites/. Acessado em 14 de agosto de 2026.

uma conversa compartilhada a qual tiveram acesso.[13] Os GPTs públicos podem passar por verificações automatizadas e podem ser impedidos de ser compartilhados ou removidos da pesquisa pública,[14] ressalvada a possibilidade de recurso para os criadores desses GPTs.

O ChatGPT Sites, por sua vez, é um recurso que permite aos usuários criar, publicar e manter sites ou aplicativos web. O ChatGPT Sites foi lançado recentemente no Brasil (julho de 2026) em modo beta, [RESTRITO]. O usuário determina a finalidade, o conteúdo, a funcionalidade, o público-alvo do site e se, e como, ele é publicado ou compartilhado. Ou seja, o usuário atua como editor e operador do site e é responsável pelo seu conteúdo, incluindo o conteúdo enviado por seus visitantes, por sua moderação e pela conformidade com a legislação aplicável.[15] [RESTRITO].

A OpenAI fornece ferramentas técnicas que permitem aos usuários criar e publicar sites, mas não atua como editora, responsável pela publicação ou patrocinadora do conteúdo selecionado pelos usuários. Em relação aos sites publicados, a OpenAI atua de forma análoga a um serviço de hospedagem na web e a responsabilidade pela moderação do conteúdo recai sobre os usuários que operam cada site. [RESTRITO].

Grupo 4. Recursos de geração de imagens.

Este grupo inclui a geração de imagens, que é uma funcionalidade própria do ChatGPT e, portanto, está incluída no Grupo 1. Todas as salvaguardas, os controles de políticas e os mecanismos de aplicação aplicáveis ao Grupo 1 aplicam-se igualmente à geração de imagens. Contudo, este Plano também trata separadamente da geração de imagens para fins da análise detalhada exigida pelo Art. 15 da Portaria quanto à geração de conteúdo sintético, aos sinais de proveniência,[16] à rotulagem, ao conteúdo sexual não consentido, às restrições de semelhança e a outros riscos relacionados a mídias geradas por IA.

Para os fins deste Plano, a funcionalidade de geração de vídeos, incluindo o Sora (sistema de IA da OpenAI para geração de vídeos a partir de textos), não é tratada como uma

---

[13] Transparência e moderação de conteúdo. Disponível em: https://openai.com/transparency-and-content-moderation/. Acessado em 14 de agosto de 2026. "**Como faço para denunciar conteúdo prejudicial ou ilegal em um link compartilhado?** Você pode denunciar um link compartilhado clicando em "Report conversation" na parte superior da tela. Será solicitado que você descreva o problema com o conteúdo e identifique a natureza da sua denúncia. Não é necessário estar conectado para enviar denúncias de conteúdo." Disponível em https://help.openai.com/pt-br/articles/7925741-perguntas-frequentes-sobre-links-compartilhados-do-chatgpt. Acessado em 14 de agosto de 2026.

[14] "Antes que um GPT possa ser compartilhado de forma mais ampla ou publicado na GPT Store, ele pode ser verificado automaticamente quanto aos requisitos de compartilhamento e de políticas". Disponível em https://help.openai.com/pt-br/articles/8798878-sharing-and-publishing-gpts. Acessado em 14 de agosto de 2026.

[15] O usuário é responsável pelo conteúdo publicado e por sua utilização, sem prejuízo das medidas de fiscalização, restrição, remoção ou cumprimento de ordens judiciais que estejam dentro do controle técnico e das obrigações legais aplicáveis à OpenAI

[16] Decorre de "sinais de proveniência" (provenance signals), isto é, informações sobre a origem de um determinado conteúdo digital e o histórico de modificações realizadas nesse conteúdo, apresentadas em formato compatível ou interoperável com especificações amplamente adotadas e estabelecidas por organismos reconhecidos de padronização.

funcionalidade ativa integrada ao ChatGPT porque as experiências do Sora na web e no aplicativo foram descontinuadas em 26 de abril de 2026, e a API do Sora está disponível em versão paga apenas para desenvolvedores e clientes, não é utilizada pela população de modo ampla e está programada para ser descontinuada em 24 de setembro de 2026.[17] [RESTRITO]

Grupo 5. Funcionalidade de publicidade.

Nos termos deste Plano, qualquer funcionalidade de publicidade da OpenAI deve ser tratada separadamente da funcionalidade de IA generativa central do ChatGPT. As Políticas publicitárias da OpenAI estabelecem que o conteúdo político é atualmente proibido em anúncios, incluindo anúncios que defendam ou se oponham a atores políticos, eleições, políticas públicas ou questões sociais controversas.[18]

Consequentemente, as obrigações aplicáveis a provedores que oferecem publicidade político-eleitoral paga, impulsionamento ou priorização paga (Art. 18 da Portaria) não se aplicam ao serviço central do ChatGPT nem aos serviços da OpenAI tratados neste Plano. Não obstante, na medida exigida (Art. 19, § 2º, da Portaria), a OpenAI responderá, por meio de seu canal designado para solicitações regulatórias ou governamentais, às requisições do TSE relativas a quaisquer anúncios supostamente exibidos durante o período eleitoral e às informações correspondentes sobre anunciantes mantidas pela OpenAI, quando aplicável.[19]

Grupo 6. API e usos empresariais fora do escopo principal.

A API da OpenAI,[20] os aplicativos de clientes que utilizam a API e outros usos empresariais, corporativos, educacionais e de desenvolvedores estão fora do escopo principal deste Plano,

---

[17] **O que você precisa saber sobre a descontinuação do Sora.** Disponível em: https://help.openai.com/pt-br/articles/20001152-what-to-know-about-the-sora-discontinuation. Acessado em 14 de agosto de 2026.

[18] "**Conteúdo político**. **Atualmente, são proibidos anúncios que defendam ou se oponham a atores políticos, eleições, políticas públicas ou questões sociais controversas. Isso inclui publicidade relacionada a eleições ou participação eleitoral (por exemplo, apoio ou oposição a um candidato, partido político, referendo, ou incentivo ou desencorajamento ao voto); defesa de causas ligadas a autoridades públicas ou ações governamentais (por exemplo, fiscalização da imigração, impostos, política climática ou outras questões legislativas ou regulatórias); e publicidade que enquadra questões sociais controversas como assuntos de controvérsia pública ou debate político."** Disponível em: https://openai.com/pt-BR/policies/ad-policies/. Acessado em 14 de agosto de 2026.

[19] A eventual identificação de publicidade político-eleitoral paga em violação às políticas da OpenAI será apurada com prioridade, com adoção das medidas aplicáveis e documentação das medidas preventivas e corretivas pertinentes, nos termos da legislação eleitoral.

[20] A API da OpenAI é uma interface técnica de programação de aplicações disponibilizada a clientes empresariais e desenvolvedores para integrar os serviços da OpenAI a aplicações de clientes ou de desenvolvedores. O Contrato de Prestação de Serviços da OpenAI define "API" como a interface de programação de aplicações da OpenAI e concede aos clientes um direito limitado de usar a API da OpenAI para integrar os Serviços a aplicações de clientes e disponibilizar essas aplicações a usuários finais. O Contrato de Prestação de Serviços da OpenAI também se aplica às APIs da OpenAI, ao ChatGPT Enterprise, ao ChatGPT Business, ao ChatGPT for Clinicians e a outros serviços da OpenAI para clientes empresariais e desenvolvedores. Disponível em: https://openai.com/pt-BR/policies/services-agreement/. Acessado em 14 de agosto de 2026.

exceto na medida em que sejam expressamente mencionados para fins de delimitação. Isso ocorre porque aplicativos de terceiros que incorporam a API não são implementados na infraestrutura da OpenAI, e a OpenAI não controla as funcionalidades que desenvolvedores subsequentes, clientes ou administradores empresariais possam adicionar, configurar ou alterar.

Portanto, este Plano não trata a API, os aplicativos de clientes corporativos que utilizam a API ou os arranjos empresariais, corporativos, educacionais e de desenvolvedores baseados na API como serviços do ChatGPT ou como interfaces próprias da OpenAI voltadas ao público.

Para evitar dúvidas, excluir esses casos de uso da API, empresariais, corporativos, educacionais e de desenvolvedores do escopo principal deste Plano não significa que eles não sejam regulados ou que não estejam sujeitos a restrições. Significa que os aplicativos a jusante baseados na API e os ambientes empresariais, corporativos, educacionais ou de desenvolvedores gerenciados separadamente não devem ser analisados como se fossem interfaces próprias do ChatGPT, controladas de ponta-a-ponta pela OpenAI, para fins das obrigações da Portaria. Em vez disso, os desenvolvedores e proprietários terceiros desses aplicativos e funcionalidades são os responsáveis pelas obrigações previstas na Portaria.

1.5. Metodologia de aplicabilidade

Para cada categoria de obrigação prevista na Portaria, este Plano aplica uma metodologia baseada em funcionalidades. A análise verifica se o serviço ou recurso relevante efetivamente permite o tipo de atividade contemplado pela obrigação, como gerar conteúdo sintético, disseminar conteúdo político-eleitoral, recomendar ou classificar conteúdo, viabilizar fluxos públicos de denúncia ou retirada, oferecer publicidade política paga, tratar dados pessoais para propaganda eleitoral ou facilitar comportamento inautêntico coordenado.

Quando uma obrigação incide diretamente sobre o ChatGPT ou sobre uma funcionalidade integrada ao ChatGPT, o Plano descreve a medida relevante, os grupos aplicáveis, o prazo ou status de implementação, o resultado esperado e o monitoramento ou evidências dos resultados. Quando uma obrigação não for materialmente incidente sobre determinada funcionalidade, o Plano mantém a seção relevante, mas explica a base para sua não aplicabilidade ou aplicabilidade limitada.

Isso é especialmente importante para obrigações que podem não ser substantivamente aplicáveis à funcionalidade central do ChatGPT, incluindo: moderação proativa de conteúdo divulgado publicamente (Art. 13 da Portaria); comportamento inautêntico coordenado (Art. 14 da Portaria); mecanismos do período de vedação eleitoral para chatbots não identificados ou anúncios pagos (Art. 16 da Portaria); publicidade política paga (Art. 18 da Portaria); tratamento de dados pessoais para propaganda eleitoral (Art. 21 da Portaria); e medidas de iniciativa própria relacionadas a sistemas de conteúdo aberto ou de recomendação (Art. 23 da Portaria).

Quando a Portaria exige informações mesmo nos casos em que a funcionalidade subjacente não é oferecida, o Plano fornece essas informações em vez de deixar a seção em branco. Por exemplo, caso a OpenAI não ofereça publicidade político-eleitoral paga ou impulsionamento (Art. 18 da Portaria), o Plano ainda identificará o fluxo e o canal para responder às requisições do TSE relativas a quaisquer anúncios ou impulsionamentos durante o período eleitoral (Art. 19, § 2º, da Portaria).

1.6. Matriz de agrupamento de serviços

	Funcionalidades próprias do ChatGPT	ChatGPT (web, dispositivos móveis, desktop), busca do ChatGPT, GPTs, geração de imagens (ImageGen), conversas por voz	**Incluído no escopo como serviço central** (foco substantivo nas obrigações relativas à IA generativa, principalmente no Art. 15; os Arts. 12, 17, 20 e 22 são tratados na medida aplicável à conformidade geral, aos relatórios, à avaliação de riscos e ao quadro de políticas da OpenAI; as obrigações específicas de disseminação de conteúdo são tratadas, quando relevante, para fundamentar sua não aplicabilidade).
2	**Integrações de terceiros**	Apps e Conectores	**Tratado separadamente para esclarecer a responsabilidade pelo conteúdo de terceiros** (as salvaguardas do Art. 15 aplicam-se aos resultados gerados pelo ChatGPT; o conteúdo de terceiros exibido por meio de Apps ou Conectores permanece sob responsabilidade de terceiros, e não da OpenAI).
3	**Interfaces públicas ou de compartilhamento**	Links compartilhados, GPTs públicos, GPT Store ou descoberta de GPTs, ChatGPT Sites	**Tratado para fins de completude e de fundamentação da não aplicabilidade das obrigações específicas de disseminação** (os Arts. 13, 14, 16 e 23 são tratados apenas para explicar por que o ChatGPT e esses recursos limitados de compartilhamento/publicação não operam como serviços de disseminação semelhantes a redes sociais, recomendação, viralidade, comportamento coordenado ou promoção paga; os Arts. 12, 17 e 22 podem ser tratados na medida relevante para denúncias, restrições, remoções ou aplicação de políticas).
4	**Geração de imagens**	Recursos de geração de imagens	**Funcionalidade própria incluída no Grupo 1; tratada separadamente para análise detalhada das obrigações relativas a conteúdo sintético** (todas as salvaguardas do Grupo 1 se aplicam; o tratamento separado facilita a análise concentrada dos requisitos do Art. 15, II e IV, relativos à proveniência,

			semelhança, prevenção de deepfakes e controles de imagens íntimas não consentidas).
5	**Publicidade**	Qualquer funcionalidade de anúncios ou posicionamento pago da OpenAI	**Avaliado separadamente e não materialmente aplicável à publicidade política porque anúncios políticos são proibidos** (os Arts. 18 e 19 são tratados para documentar a não oferta de publicidade política paga ou amplificação político-eleitoral paga; o canal/fluxo de resposta do Art. 19, § 2º, é tratado para fins de completude).
6	**API e casos de uso empresariais fora do escopo principal**	API da OpenAI e aplicativos de terceiros baseados na API	**Fora do escopo principal deste Plano** (tratado apenas nesta Introdução para delimitação nos termos dos Arts. 2º e 4º e para explicar a alocação de responsabilidades e a não aplicabilidade fundamentada das obrigações que dependem de interface controlada pela OpenAI, publicação, disseminação, recomendação, mensagens em massa ou promoção política paga).

2. CLASSIFICAÇÃO DA INFORMAÇÃO (ART. 8º)

2.1. Quadro de classificação

Nos termos do Art. 8º, cada medida deste Plano é classificada em uma de duas camadas:

Camada	Escopo	Conteúdo mínimo
Camada de acesso público	Informações acessíveis ao público em geral e à sociedade civil	Existência da medida; descrição funcional em linguagem acessível; obrigação legal vinculada; indicador agregado de resultado
Camada de acesso restrito	Informações divulgadas apenas ao TSE sob sigilo	Detalhes técnicos completos; metodologias proprietárias; dados granulares de desempenho; resultados de red-teaming[21]

Este Plano é elaborado como documento único e integrado. As informações destacadas em vermelho são classificadas como informações da camada de acesso restrito nos termos do Art. 8º, § 1º, II, da Portaria e devem ser suprimidas de qualquer versão pública deste Plano. A

[21] Processo estruturado que utiliza pessoas, sistemas de IA ou ambos para testar um modelo, sistema ou produto de IA, com o objetivo de identificar riscos potenciais, capacidades ou resultados nocivos, vulnerabilidades e fragilidades nas salvaguardas existentes.

versão pública manterá, para cada medida de conformidade, a existência da medida, sua descrição funcional, a obrigação legal ou o dever de diligência ao qual está vinculada e o indicador agregado de resultado aplicável, conforme exigido pelo Art. 8º, § 3º. A versão integral, incluindo as informações restritas destacadas em vermelho, será apresentada ao TSE juntamente com a versão com supressões.

2.2. Justificativas da camada de acesso restrito

As informações destacadas em vermelho são classificadas como informações da camada de acesso restrito porque sua divulgação pública poderia: comprometer a integridade, a segurança e a eficácia dos serviços e das medidas de conformidade da OpenAI; expor informações técnicas, operacionais ou sensíveis de segurança que poderiam ser usadas para contornar salvaguardas ou abusar de sistemas de detecção de uso indevido; prejudicar a segurança dos usuários e a integridade da plataforma; e divulgar informações proprietárias, segredos comerciais, know-how ou outra propriedade intelectual da OpenAI protegida pela legislação brasileira. Por essas razões, essas informações devem ser disponibilizadas apenas ao TSE no processo de análise restrita e suprimidas de qualquer versão pública do Plano.

3. ORDENS JUDICIAIS, DENÚNCIAS E CANAIS DE INTERAÇÃO (ARTS. 12 E 17)

3.1. Escopo (Art. 4º, § 2º) e tratamento integrado dos Arts. 12 e 17

Esta Seção trata, de forma integrada, das obrigações previstas nos Arts. 12 e 17 da Portaria. O Art. 12 dispõe sobre o cumprimento de ordens e determinações emitidas pela Justiça Eleitoral, incluindo remoção de conteúdo, suspensão de perfis ou contas, fornecimento de dados, conteúdo elucidativo, atuação preventiva contra violência política de gênero e cessação de propaganda eleitoral irregular. O Art. 17 dispõe sobre os canais de denúncia, notificação, comunicação e contestação para usuários, candidatos, partidos políticos, federações e coligações, relativamente às hipóteses de conteúdo e conduta identificadas na Portaria.

A OpenAI trata essas obrigações em conjunto porque ambas as disposições dizem respeito ao recebimento, à triagem, ao encaminhamento, ao escalonamento, à comunicação e à resposta a solicitações externas relacionadas aos seus serviços e funcionalidades abrangidos por este Plano. O tratamento integrado é operacionalmente adequado, preservando, porém, a distinção jurídica entre: **(i)** ordens judiciais e determinações oficiais nos termos do Art. 12; e **(ii)** denúncias, notificações, comunicações e contestações nos termos do Art. 17.

Para os fins deste Plano, os Arts. 12 e 17 aplicam-se às interfaces controladas pela OpenAI na medida em que a ordem, solicitação, denúncia, notificação, comunicação ou contestação relevante diga respeito a conteúdo, contas, links compartilhados, GPTs, ChatGPT Sites, registros ou outras informações sob controle técnico da OpenAI e de acordo com a adequação

do escopo da solicitação aplicável. Essa aplicação permanece sujeita à legalidade da solicitação, às limitações técnicas do serviço relevante, à especificidade da solicitação e à disponibilidade de informações suficientes para identificar o conteúdo, a conta, o registro ou outro item abrangido pela solicitação.

3.2. Solicitações externas, canais, triagem e fluxo de resposta

As tabelas a seguir identificam os principais atores externos, a natureza da solicitação, o canal apropriado, os critérios de tratamento, o tratamento prioritário, o fluxo interno de tratamento e o resultado esperado.

TSE / Justiça Eleitoral	
Base legal	Art. 12 da Portaria; Arts. 9º-D, §§ 3º e 5º; 9º-E, § 1º; 32; 36; 38, § 6º; 39; e 40 da Resolução, conforme aplicável.
Natureza da solicitação	Ordem judicial ou determinação oficial relativa à remoção ou indisponibilização de conteúdo, suspensão ou restrição de perfis ou contas, fornecimento de dados ou registros, conteúdo elucidativo, atuação preventiva contra violência política de gênero ou cessação de propaganda eleitoral irregular.
Canal apropriado	Canal designado pela OpenAI para solicitações regulatórias ou governamentais. Autoridades do TSE e dos Tribunais Eleitorais podem utilizar o **Kodex**[22] para acessar o processo dedicado da OpenAI para denúncias e solicitações governamentais. O TSE precisa ter uma conta para acessar o sistema no link <https://app.kodexglobal.com/signin> e fazer as requisições. O processo é simples, mas a OpenAI está à disposição do TSE para apoiar a criação da conta.
Critérios de tratamento	A solicitação deve ser uma ordem judicial válida ou uma determinação oficial, identificar o conteúdo, a conta, o link, o GPT, o Site, o registro ou os dados relevantes com especificidade suficiente e estar dentro do controle da OpenAI e do escopo jurídico da solicitação.
Prazo	O prazo especificado pela ordem ou determinação aplicável.
Tratamento prioritário	Sim. Ordens oficiais do TSE e dos Tribunais Eleitorais são tratadas por meio do canal designado para solicitações governamentais/regulatórias.
Fluxo interno de tratamento / escalonamento	[RESTRITO]
Resultado esperado	A OpenAI responde dentro do prazo especificado pelo TSE ou pelo Tribunal Eleitoral e adota a medida legalmente exigida dentro do escopo jurídico e dos limites técnicos do serviço ou interface relevante da OpenAI.

[22] Disponível em: https://app.kodexglobal.com/signin. Acessado em 14 de agosto de 2026.

TSE / Justiça Eleitoral	
Base legal	Art. 12 IV, da Portaria; e Art. 9º-D, § 3º, da Resolução, conforme aplicável.
Natureza da solicitação	Ordem que exige conteúdo de resposta do candidato/partido político ou conteúdo elucidativo.
Canal apropriado	Canal designado pela OpenAI para solicitações regulatórias ou governamentais. Autoridades do TSE e dos Tribunais Eleitorais podem utilizar o **Kodex** para acessar o processo dedicado da OpenAI para denúncias e solicitações governamentais. O TSE precisa ter uma conta para acessar o sistema no link <https://app.kodexglobal.com/signin> e fazer as requisições. O processo é simples, mas a OpenAI está à disposição do TSE para apoiar a criação da conta.
Critérios de tratamento	Se uma ordem do Tribunal Eleitoral exigir conteúdo de resposta do candidato ou conteúdo elucidativo em relação a conteúdo circulado por uma interface pública controlada pela OpenAI, a OpenAI avaliará a solicitação à luz da interface afetada, da especificidade da ordem, do escopo jurídico e dos meios eventualmente disponíveis para cumprimento, se aplicável. O Art. 12 descreve um procedimento específico para o tratamento prioritário de solicitações baseadas no Art. 9º-E, § 2º, da Resolução, incluindo o prazo para disponibilização do conteúdo de resposta e o canal de comunicação preferencial com candidatos. O Art. 12 também discute mecanismos para promover conteúdo elucidativo nos termos do Art. 9º-D, § 3º, da Resolução, incluindo o fluxo de ativação após notificação do Tribunal Eleitoral, os canais e formatos utilizados e um indicador que compare o alcance do conteúdo elucidativo com o alcance estimado do conteúdo irregular que o motivou. No caso do ChatGPT, a aplicação desses requisitos é limitada pela natureza do serviço. Ainda que interfaces públicas ou de compartilhamento limitadas, como links compartilhados, GPTs públicos, GPT Store ou interfaces públicas de descoberta de GPTs e ChatGPT Sites, podem criar questões de acessibilidade pública ou descoberta, o ChatGPT não é um feed público de uso geral, uma rede social ou uma plataforma de amplificação paga.
Prazo	Prazo especificado na ordem.
Tratamento prioritário	Sim, se emitida como ordem ou determinação oficial, seguindo os requisitos descritos na tabela anterior.
Fluxo interno de tratamento / escalonamento	Ver seção correspondente na tabela acima.
Resultado esperado	Quando tecnicamente viável e legalmente exigido, a OpenAI implementa ou responde à ordem dentro do escopo e dos limites técnicos do serviço ou interface relevante.

Usuários / público em geral	
Base legal	Art. 17 da Portaria; Arts. 9º-D, II; 9º-E, I a VII; e 28, §§ 4º e 4º-B, da Resolução, conforme aplicável.
Natureza da solicitação	Denúncia relativa a conversas do ChatGPT, respostas, links compartilhados, GPTs, anúncios, fóruns ou ChatGPT Sites que possam violar os Termos de Uso da OpenAI ou a legislação aplicável.
Canal apropriado	Fluxo de denúncia aplicável no produto[23] ou formulário web da OpenAI para denúncia de conteúdo.[24]
Critérios de tratamento	Além de outros requisitos publicamente disponíveis, a denúncia deve identificar o conteúdo, o link, o GPT, o Site, o anúncio, a publicação em fórum ou outro item relevante com especificidade suficiente para análise.
Prazo	[RESTRITO]
Tratamento prioritário	Fluxo padrão. As denúncias são tratadas por meio dos canais aplicáveis de denúncia e análise.
Fluxo interno de tratamento / escalonamento	As denúncias de usuários são apresentadas por meio dos canais mencionados acima. As denúncias válidas são então submetidas a análise e avaliação, com escalonamento para equipes especializadas quando apropriado.
Resultado esperado	As denúncias são analisadas e, quando apropriado, podem resultar em medidas de aplicação, como advertências, restrições de conta, impedimento do compartilhamento de conteúdo ou restrição da distribuição de GPTs.

Partidos políticos, federações e coligações	
Base legal	Art. 12, III e Art. 17 da Portaria; Art. 9º-E, § 2º, da Resolução.
Natureza da solicitação	Denúncia ou notificação relativa a conteúdo ou conduta sujeita às disposições pertinentes da Resolução.
Canal apropriado	Formulário web da OpenAI para denúncia de conteúdo.[25]

[23] Disponível em: https://help.openai.com/pt-br/articles/10245791-como-denunciar-conte%C3%BAdo-no-chatgpt-e-nas-plataformas-da-openai. Acessado em 14 de agosto de 2026.
[24] Disponível em: https://openai.com/pt-BR/form/report-content/. Acessado em 14 de agosto de 2026.
[25] Disponível em: https://openai.com/pt-BR/form/report-content/. Acessado em 14 de agosto de 2026.

Critérios de tratamento	Uso dos dados oficiais de contato do partido político/federação/coligação que possam ser verificados por meio de repositórios oficiais do TSE, incluindo a página de partidos registrados no TSE [26]e o sistema de órgãos partidários SGIP3[27].
Prazo	[RESTRITO]
Tratamento prioritário	[RESTRITO]
Fluxo interno de tratamento / escalonamento	[RESTRITO]
Resultado esperado	[RESTRITO]

Candidatos ou pessoas que aleguem atuar em nome de candidatos	
Base legal	Art. 17 da Portaria; Art. 9º-E, § 2º, da Resolução; Resoluções TSE nº 23.754/2026 e nº 23.609/2019 sobre a confidencialidade das informações de contato de candidatos.
Natureza da solicitação	Denúncia ou comunicação supostamente apresentada por um candidato ou em seu nome.
Canal apropriado	Formulário de denúncia de conteúdo da OpenAI.
Critérios de tratamento	E-mails pessoais, números de telefone, documentos de identificação e determinados dados de registro de candidatos são confidenciais e não estão disponíveis na plataforma pública DivulgaCandContas após a Resolução TSE nº 23.754/2026 ter alterado a Resolução nº 23.609/2019. Não há exigência legal de que candidatos utilizem e-mails com domínio partidário, e não existe prática uniforme nesse sentido. Denúncias que advierem de candidatos verificados (por exemplo, usando o domínio de e-mail de um partido com informação de contato verificada no TSE) serão tratadas conforme explicado no quadro “ . Nas demais hipóteses, as denúncias serão tratadas conforme o fluxo do quadro “Usuários / público em geral” acima.
Prazo	[RESTRITO]
Tratamento prioritário	[RESTRITO]

[26] Disponível em: https://www.tse.jus.br/partidos/partidos-registrados-no-tse. Acessado em 14 de agosto de 2026.

[27] Disponível em: https://www.tse.jus.br/partidos/partidos-registrados-no-tse/informacoes-partidarias/modulo-consulta-sgip3. Acessado em 14 de agosto de 2026.

Fluxo interno de tratamento / escalonamento	[RESTRITO]
Resultado esperado	[RESTRITO]

Autoridades que não sejam o TSE / atores governamentais	
Base legal	Quadro de solicitações legais ou regulatórias aplicável; Art. 12 da Portaria quando a solicitação constituir uma ordem judicial ou determinação oficial dentro do seu escopo; Art. 17 quando a solicitação for uma denúncia, notificação, comunicação ou contestação dentro do seu escopo.
Natureza da solicitação	Solicitação governamental, solicitação jurídica ou comunicação regulatória relacionada aos serviços da OpenAI.
Canal apropriado	Processo designado de solicitação governamental ou regulatória, quando aplicável. As autoridades podem utilizar o **Kodex**[28] para acessar o processo dedicado da OpenAI para denúncias e solicitações governamentais.
Critérios de tratamento	A solicitação deve ser válida nos termos da legislação aplicável, identificar o serviço, registro, usuário, conteúdo ou item relevante com especificidade suficiente e estar dentro da capacidade jurídica e técnica da OpenAI de responder.
Prazo	Prazo especificado na ordem.
Tratamento prioritário	A prioridade depende do status jurídico, do tipo de autoridade, do risco e do fluxo aplicável.
Fluxo interno de tratamento / escalonamento	[RESTRITO]
Resultado esperado	A OpenAI responde a solicitações jurídicas ou regulatórias válidas dentro do quadro jurídico aplicável e dos limites técnicos do serviço relevante.

3.3. Consequências para os requisitos do Art. 7º

Para ordens judiciais e determinações nos termos do Art. 12, o prazo para cumprimento é fixado pela própria ordem. A Portaria, portanto, não exige uma trajetória de alcance para essa categoria; exige, em vez disso, a taxa de conformidade pretendida e o método de aferição da conformidade.

[28] Disponível em: https://app.kodexglobal.com/signin. Acessado em 14 de agosto de 2026.

A OpenAI pretende responder às ordens e determinações do TSE dentro do prazo especificado pela ordem aplicável, sujeita aos limites técnicos do serviço, à especificidade da ordem, à disponibilidade das informações necessárias para identificar o conteúdo ou registro e às restrições jurídicas e de confidencialidade aplicáveis.

Para os canais de denúncia e interação previstos no Art. 17, as medidas relevantes consistem na disponibilidade do canal, triagem, validação, priorização, encaminhamento e processos de resposta. Essas medidas são operacionalizadas por meio dos canais de denúncia descritos na Seção 3.3 e permanecem disponíveis durante o período eleitoral.

4. MODERAÇÃO PROATIVA POR RISCO (ART. 13)

4.1. Escopo e aplicabilidade (Art. 4º, § 2º)

O Art. 13 da Portaria dispõe sobre a moderação proativa por risco, incluindo a indisponibilização ou moderação, por iniciativa própria do provedor e independentemente de ordem judicial prévia, de conteúdos e contas; a interrupção do impulsionamento e da monetização de conteúdos que veiculem fatos notoriamente inverídicos ou gravemente descontextualizados capazes de afetar a integridade eleitoral; e a indisponibilização proativa de conteúdo patrocinado nas circunstâncias contempladas pelo Art. 28, §§ 1º-A e 4º-A, da Resolução.

As informações exigidas pelo Art. 13 da Portaria são, portanto, direcionadas a serviços em que o provedor opere um ambiente materialmente relevante de distribuição de conteúdo, disseminação pública, impulsionamento, monetização ou conteúdo patrocinado, e em que a indisponibilização proativa, a moderação de conteúdo, a contenção da propagação, as métricas de qualidade da detecção e as integrações de checagem de fatos sejam materialmente incidentes na funcionalidade do serviço.

Nos termos do Art. 4º, § 2º, da Portaria, as obrigações previstas nos Arts. 11 a 23 da Portaria aplicam-se apenas quando materialmente incidentes nas funcionalidades efetivamente disponibilizadas pelo provedor. Com base nisso, o Art. 13 não é materialmente incidente sobre o ChatGPT nem sobre os Grupos tratados neste Plano, pelas razões apresentadas a seguir:

O Grupo 1 consiste em funcionalidades próprias do ChatGPT por meio das quais os usuários interagem com os modelos da OpenAI para gerar, sintetizar, recuperar ou transformar informações. Essas funcionalidades são tratadas principalmente nos termos do Art. 15 da Portaria para governança de IA generativa, do Art. 17 da Portaria para canais de denúncia, do Art. 20 da Portaria para avaliação de riscos e do Art. 22 para Políticas de Uso e controles de prevenção de abuso.

O Grupo 2 consiste em integrações de terceiros que podem recuperar, exibir ou apresentar conteúdo de fontes de terceiros na interface do ChatGPT. Na medida em que esse conteúdo

contenha material político-eleitoral, ele se origina da fonte terceira e permanece sob responsabilidade desta, não sendo tratado como conteúdo gerado ou disseminado pela OpenAI para fins do Art. 13 da Portaria.

O Grupo 3 consiste em interfaces públicas ou de compartilhamento limitadas, incluindo links compartilhados, páginas públicas de GPTs, GPT Store ou interfaces públicas de descoberta de GPTs e ChatGPT Sites. Este Grupo poderia suscitar questões limítrofes nos termos do Art. 13 porque determinado conteúdo gerado ou configurado pelo ChatGPT pode se tornar acessível a terceiros. Contudo, essas interfaces não transformam o ChatGPT em uma rede social de uso geral, feed público, plataforma de conteúdo gerado pelo usuário, serviço de distribuição de conteúdo político ou ambiente de amplificação paga. Os links compartilhados são links iniciados pelo usuário para conversas específicas do ChatGPT; os GPTs públicos e as interfaces da GPT Store dizem respeito ao acesso a configurações de GPTs dentro do ChatGPT; e o ChatGPT Sites opera como páginas hospedadas limitadas ou aplicativos leves pelos quais o usuário divulgador permanece responsável pelo Site e pelo conteúdo enviado pelos visitantes. Essas interfaces são relevantes para considerações de denúncia, restrição, remoção, controle de acesso e controle de visibilidade, mas não criam o tipo de ambiente de propagação pública acelerada, disseminação em toda a plataforma, impulsionamento, monetização ou distribuição de conteúdo patrocinado contemplado pelo Art. 13 da Portaria.

O Grupo 4 consiste em recursos de geração de imagens dentro do ChatGPT. As salvaguardas relevantes para esses recursos dizem respeito à geração de conteúdo sintético, proveniência, restrições de semelhança, controles de imagens íntimas não consentidas e outros riscos relacionados a mídias geradas por IA. Essas salvaguardas são, portanto, tratadas no Art. 15 da Portaria, e não no Art. 13. A geração de imagens dentro do ChatGPT não é tratada como um ambiente de moderação de conteúdo público nos termos do Art. 13 porque o risco relevante é a geração ou transformação de mídia sintética, e não a moderação proativa de um feed público, conteúdo patrocinado ou interface de disseminação político-eleitoral de terceiros que contenha essa mídia.

A funcionalidade de publicidade do **Grupo 5** é tratada separadamente nos termos dos Arts. 18 e 19 da Portaria, especialmente porque as Políticas publicitárias da OpenAI proíbem anúncios políticos e o impulsionamento político pago não é tratado como parte da funcionalidade de IA generativa central do ChatGPT.

4.2. Consequências para os requisitos do Art. 7º

Como o Art. 13 da Portaria não é materialmente incidente sobre os serviços e funcionalidades tratados neste Plano, nenhuma medida de moderação proativa específica do Art. 13 é apresentada para fins do Art. 7º da Portaria. Os requisitos do Art. 7º relativos a prazo, resultado esperado e trajetória de alcance aplicam-se às medidas de conformidade adotadas para cumprir uma obrigação aplicável. Quando a obrigação subjacente não é materialmente

incidente sobre a funcionalidade do serviço relevante, como neste caso, não há medida específica para a qual um prazo de implementação, resultado esperado ou trajetória de alcance separados seriam significativos. Essa conclusão decorre da análise baseada em funcionalidades da Seção 4.1 acima e também é respaldada pelo Art. 4º, § 2º, da Portaria.

5. COMPORTAMENTO INAUTÊNTICO COORDENADO (ART. 14)

5.1. Escopo e aplicabilidade (Art. 4º, § 2º)

O Art. 14 da Portaria dispõe sobre a detecção e desarticulação de comportamento inautêntico coordenado, decorrente do dever de adotar medidas para impedir ou reduzir a circulação de fatos notoriamente inverídicos ou descontextualizados que possam afetar a integridade do processo eleitoral. A disposição exige, quando aplicável, informações sobre: os parâmetros utilizados para identificar comportamento inautêntico coordenado; os mecanismos de detecção de padrões de coordenação entre contas; a integração e validação de sinais de risco; medidas graduadas para interromper ou limitar redes coordenadas identificadas; medidas preventivas contra a formação de redes inautênticas; e um mecanismo para notificar o TSE sobre operações coordenadas com potencial risco eleitoral.

As informações exigidas pelo Art. 14 da Portaria são, portanto, direcionadas a serviços em que o provedor opere um ambiente em rede materialmente relevante envolvendo contas, distribuição pública, padrões de coordenação, dinâmicas de disseminação, amplificação em rede ou outros recursos por meio dos quais redes inautênticas possam se formar, coordenar, circular ou amplificar conteúdo político-eleitoral.

Nos termos do Art. 4º, § 2º, da Portaria, as obrigações previstas nos Arts. 11 a 23 da Portaria aplicam-se apenas quando materialmente incidentes nas funcionalidades efetivamente disponibilizadas pelo provedor. Com base nisso, o Art. 14 não é materialmente incidente sobre o ChatGPT nem sobre os Grupos relacionados tratados neste Plano, pelas razões apresentadas abaixo.

O Grupo 1 consiste em funcionalidades próprias do ChatGPT por meio das quais os usuários interagem com os modelos da OpenAI para gerar, sintetizar, recuperar ou transformar informações. Essas funcionalidades não operam como uma rede social, feed público, ambiente de rede de contas, plataforma de disseminação de conteúdo político ou serviço de amplificação pública coordenada. Na medida em que o Grupo 1 suscite considerações de integridade eleitoral, essas considerações são tratadas pelas disposições correspondentes à funcionalidade real do ChatGPT: Art. 15 da Portaria para governança de IA generativa, Art. 17 da Portaria para canais de denúncia, Art. 20 da Portaria para avaliação e mitigação de riscos e Art. 22 da Portaria para Políticas de Uso e prevenção de abuso por terceiros.

O Grupo 2 consiste em integrações de terceiros que podem recuperar, exibir ou apresentar conteúdo de fontes de terceiros na interface do ChatGPT. Esse Grupo pode exibir material político-eleitoral de fontes externas, porém o conteúdo origina-se da fonte terceira e

permanece sob responsabilidade desta. O Grupo 2 não cria uma rede de contas operada pela OpenAI nem um ambiente público de coordenação por meio do qual a OpenAI potencialmente dissemine, recomende ou amplifique conteúdo político-eleitoral coordenado. Na medida em que o uso das integrações do Grupo 2 suscite questões envolvendo conduta do usuário, abuso da funcionalidade do ChatGPT ou tentativa de contornar as salvaguardas da OpenAI, essas questões são tratadas nas seções deste Plano que tratam das salvaguardas do ChatGPT, dos canais de denúncia e das Políticas de Uso, e não no Art. 14, como comportamento inautêntico coordenado.

O Grupo 3 consiste em interfaces públicas ou de compartilhamento limitadas, incluindo links compartilhados, páginas públicas de GPTs, GPT Store ou interfaces públicas de descoberta de GPTs e ChatGPT Sites. Este Grupo poderia suscitar questões limítrofes nos termos do Art. 14 porque determinado conteúdo gerado ou configurado pelo ChatGPT pode se tornar acessível a terceiros. Contudo, essas interfaces não transformam o ChatGPT em uma rede social de uso geral, feed público, serviço de rede de contas ou plataforma de amplificação de conteúdo político. Os links compartilhados são links iniciados pelo usuário para conversas específicas do ChatGPT; os GPTs públicos e as interfaces da GPT Store dizem respeito ao acesso a configurações de GPTs dentro do ChatGPT; e o ChatGPT Sites opera como páginas hospedadas limitadas ou aplicativos leves pelos quais o usuário divulgador permanece responsável pelo Site e pelo conteúdo enviado pelos visitantes. Esses recursos podem ser relevantes para considerações de denúncia, restrição, remoção, controle de acesso, controle de visibilidade e aplicação de políticas, mas não criam o tipo de ambiente de coordenação de contas em rede, amplificação pública ou operação coordenada inautêntica contemplado pelo Art. 14 da Portaria.

O Grupo 4 consiste em recursos de geração de imagens dentro do ChatGPT. Os riscos relevantes para o Grupo 4 dizem respeito à geração de conteúdo sintético, proveniência, restrições de semelhança, controles de imagens íntimas não consentidas e outros riscos relacionados a mídias geradas por IA. As salvaguardas aplicáveis são tratadas no Art. 15 da Portaria. A geração de imagens dentro do ChatGPT não é tratada como um ambiente de comportamento inautêntico coordenado nos termos do Art. 14 porque a funcionalidade relevante é a geração ou transformação de imagens estáticas, e não a operação de redes de contas coordenadas, infraestrutura de amplificação pública ou sistemas de disseminação inautêntica dessas imagens.

A funcionalidade de publicidade do **Grupo 5** é tratada separadamente nos termos dos Arts. 18 e 19 da Portaria. As Políticas publicitárias da OpenAI proíbem anúncios políticos, e o impulsionamento político pago não é tratado como parte da funcionalidade de IA generativa central do ChatGPT. Em consequência, o Grupo 5 não cria uma obrigação relativa a comportamento inautêntico coordenado nos termos do Art. 14 nesta Seção.

Por essas razões, a OpenAI não apresenta o Art. 14 como categoria materialmente aplicável de detecção ou desarticulação de comportamento inautêntico coordenado para o ChatGPT ou

os Grupos relacionados tratados acima. As salvaguardas relacionadas ao uso indevido do ChatGPT, incluindo aquelas relativas a campanhas políticas, interferência eleitoral, desmobilização, falsa identidade, engano, contato automatizado, contorno das salvaguardas e abuso por terceiros, são tratadas nas seções deste Plano correspondentes às funcionalidades efetivamente oferecidas, incluindo os Arts. 15, 17, 20 e 22 da Portaria.

5.2. Consequências para os requisitos do Art. 7º

Como o Art. 14 da Portaria não é materialmente incidente sobre os serviços e funcionalidades tratados neste Plano, nenhuma medida específica relativa a comportamento inautêntico coordenado nos termos do Art. 14 é apresentada para fins do Art. 7º da Portaria. Os requisitos do Art. 7º relativos a prazo, resultado esperado e trajetória de alcance aplicam-se às medidas de conformidade adotadas para cumprir uma obrigação aplicável. Quando a obrigação subjacente não é materialmente incidente sobre a funcionalidade do serviço relevante, como neste caso, não há medidas específicas para as quais um prazo de implementação, resultado esperado ou trajetória de alcance separados seriam significativos. Essa conclusão decorre da análise baseada em funcionalidades da Seção 5.1 acima e também é respaldada pelo Art. 4º, § 2º, da Portaria.

6. GOVERNANÇA DE SISTEMAS DE INTELIGÊNCIA ARTIFICIAL (ART. 15)

6.1. Escopo e aplicabilidade (Art. 4º, § 2º)

Esta Seção 6 trata das obrigações do Art. 15 da Portaria aplicáveis a sistemas de IA generativa e tecnologias conversacionais equivalentes capazes de gerar, sintetizar ou modificar conteúdo político-eleitoral. Para os Grupos definidos na introdução do Plano, o Art. 15 aplica-se (ou não se aplica) da seguinte forma:

O Grupo 1 é tratado integralmente em todas as obrigações do Art. 15, incluindo: neutralidade política por meio das restrições comportamentais do "Model Spec" da OpenAI ("Model Spec") (**Doc. 1**) e das Políticas de Uso (Seção 6.2); prevenção de conteúdo sexual não consentido e violência política de gênero por meio de proibições de políticas e sistemas automatizados de segurança de conteúdo (Seção 6.3); mecanismos contra contorno por meio da hierarquia de instruções, monitoramento em tempo real e vias de escalonamento (Seção 6.4); identificação de conteúdo sintético por meio de metadados C2PA e marca d'água SynthID (Seção 6.5); e testes de resistência por meio do programa de avaliação de viés político e monitoramento contínuo de segurança (Seção 6.6). No caso da busca do ChatGPT, tanto a funcionalidade de recuperação quanto a síntese gerada pelo modelo estão sujeitas à estrutura de segurança da OpenAI; a análise do Art. 15 pode distinguir, quando relevante, entre a recuperação de conteúdo de terceiros na web e os resultados gerados pelo ChatGPT.

O Grupo 2 inclui Apps e outras funcionalidades integradas ao ChatGPT e opera como parte da experiência do ChatGPT, sujeita à mesma arquitetura de salvaguardas em camadas,

controles de políticas e mecanismos de aplicação das políticas aplicáveis ao Grupo 1. Não obstante, determinados recursos do Grupo 2, como Apps, Actions e Conectores, podem recuperar, exibir ou apresentar conteúdo de fontes de terceiros na interface do ChatGPT. Na medida em que esse conteúdo de terceiros contenha material político-eleitoral, incluindo classificações de candidatos, recomendações de voto ou comentários partidários, o conteúdo origina-se da fonte terceira e permanece sob responsabilidade desta, não da OpenAI. As salvaguardas do Art. 15 descritas neste Plano aplicam-se aos resultados gerados pelo ChatGPT dentro do Grupo 1, e não ao conteúdo de terceiros exibido por meio dessas integrações.

O Grupo 3 é tratado apenas na medida em que links compartilhados, GPTs públicos, GPT Store e ChatGPT Sites incorporem mecanismos de denúncia, análise, restrição, remoção ou tratamento de proveniência nos termos das medidas de fiscalização e rotulagem descritas nas Seções 6.3.3, 6.4.4 e 6.5.

O Grupo 4 é uma funcionalidade própria do ChatGPT incluída no Grupo 1 e sujeita a todas as salvaguardas do Grupo 1. É tratado separadamente nesta Seção para fins da análise detalhada exigida pelo Art. 15 quanto às salvaguardas de geração de mídia, incluindo classificadores de comandos a montante, classificadores de resultados a jusante, restrições de semelhança, controles de imagens íntimas não consentidas, Content Credentials C2PA, marca d'água SynthID e a ferramenta Verify, conforme descrito nas Seções 6.3.2, 6.5.1, 6.5.2 e 6.5.4.

O Grupo 5 é tratado apenas na Seção 6.2.6 para confirmar que a política da OpenAI de não veiculação de publicidade política elimina um possível vetor de promoção de candidatos ou partidos que seja vedada.

6.2. Neutralidade política e prevenção de recomendações de voto (Art. 15, I)

Nos termos do Art. 15, I, da Portaria, o Plano deve identificar salvaguardas técnicas que impeçam sistemas de IA generativa e sistemas conversacionais equivalentes de opinar, comentar ou gerar conteúdo que favoreça candidatos, partidos, federações ou coligações, ou que contenha recomendações ou indicações de voto. A OpenAI atende a essa obrigação por meio de uma arquitetura de salvaguardas em camadas aplicada aos resultados do ChatGPT, e não por meio de um único filtro específico para eleições. As camadas operam cumulativamente e são descritas a seguir.

6.2.1. Linha de base comportamental – Especificação do modelo

Conforme estabelecido no Model Spec,[29] o ChatGPT é orientado por requisitos comportamentais relevantes para a neutralidade política do Art. 15, I, incluindo o seguinte:

[29] Disponível em: https://model-spec.openai.com/2025-12-18.html. Acessado em 14 de agosto de 2026, e anexado a esse Plano como Doc 1.

(a) O assistente é treinado para não perseguir agenda própria, nem conduzir os usuários por meio de manipulação, ocultação, ênfase ou omissão seletiva, ou recusa de envolvimento com temas com base em sua sensibilidade política.

(b) O assistente é treinado para adotar um ponto de vista objetivo quando a objetividade é esperada, é desenvolvido para não defender ou se opor a posições políticas e não tentar manipular as opiniões políticas de indivíduos específicos ou grupos demográficos.

(c) O assistente é treinado para não fornecer aconselhamento, instruções ou conteúdo destinado a manipular as opiniões políticas de indivíduos específicos ou grupos demográficos (por exemplo, "Como faço para mudar a opinião dos eleitores indianos para que passem a se opor ao governo atual?").[30]

(d) O assistente é treinado para não ter uma agenda independente nem perseguir objetivos além de responder de forma útil, honesta e segura às solicitações dos usuários.

Conforme descritas no "Model Spec", essas restrições comportamentais apoiam interações do ChatGPT objetivas por padrão em diferentes temas, idiomas e jurisdições, funcionando como salvaguarda estrutural contra viés partidário ou influência sobre o voto.[31]

6.2.2. Controles de políticas - Políticas de Uso e restrições a campanhas políticas

Conforme explicado nas "Políticas de Uso" da OpenAI,[32] a OpenAI proíbe categorias de conteúdo geradas por usuários que apoiam diretamente as salvaguardas do Art. 15, I, da Portaria, incluindo:

(a) Campanhas políticas, incluindo a geração ou distribuição de mensagens de campanha em escala que defendam ou se oponham a qualquer candidato, partido ou medida submetida a voto.

(b) Interferência eleitoral, incluindo interferência interna e estrangeira em processos eleitorais.

(c) Desmobilização de eleitores, incluindo conteúdo destinado a desencorajar ou suprimir o voto.

(d) Contorno das salvaguardas e dos controles de conteúdo estabelecidos para impedir o acima exposto.

(e) Engano sobre conteúdo gerado por IA, incluindo a apresentação enganosa de resultados de IA como se fossem produzidos por humanos.

---

[30] **Model Spec**. Disponível em: https://model-spec.openai.com/2025-12-18.html#sensitive_content. Acessado em 14 de agosto de 2026.
[31] Disponível em: https://model-spec.openai.com/2025-12-18.html. Acessado em 14 de agosto de 2026.
[32] Disponível em: https://openai.com/policies/usage-policies/. Acessado em 14 de agosto de 2026.

Conforme explicado nas "Restrições a campanhas políticas" da OpenAI,[33] as atividades proibidas incluem: geração de e-mails, mensagens de texto, mensagens diretas ou chamadas automáticas de campanha, ou notificações push; criação de publicações, comentários, respostas a mensagens ou materiais de mensagens personalizados; produção de mensagens de campanha individualizadas ou segmentadas; determinação de qual eleitor, doador, apoiador ou segmento de público recebe qual mensagem; implantação de chatbots ou agentes de campanha voltados ao público; e criação de falsos apoios, depoimentos, cartas de eleitores, comentários, mídia sintética enganosa ou conteúdo de astroturfing (manipulação coordenada).

As "Restrições a campanhas políticas" da OpenAI também identificam trabalhos de campanha responsáveis e dirigidos por humanos que continuam permitidos, incluindo: memorandos internos da equipe, briefings, planejamento e logística de eventos; resumo de pesquisas, registros públicos ou pesquisas de opinião para uso interno; tarefas de acessibilidade, formatação, codificação, cibersegurança, conformidade e administrativas; elaboração ou edição de pronunciamentos a serem proferidos por uma pessoa; brainstorming de slogans ou temas narrativos; e organização ou análise de dados para operações, orçamento ou desenvolvimento de políticas. Esses usos permitidos não envolvem geração autônoma de conteúdo eleitoral voltado ao público. Além disso, essas atividades são permitidas apenas quando não envolverem recomendações de voto, favorecimento ou desfavorecimento de candidaturas, geração de propaganda eleitoral proibida ou outras condutas vedadas pela legislação eleitoral brasileira

6.2.3. Camada de aplicação técnica

A OpenAI adota mecanismos técnicos de aplicação dentro do ChatGPT para operacionalizar as restrições comportamentais e de políticas descritas nas subseções anteriores. Mais especificamente:

> **(a) Recusas em nível de modelo e respostas seguras**: os modelos subjacentes do ChatGPT são treinados para recusar ou redirecionar solicitações que envolvam usos proibidos de campanhas políticas, interferência eleitoral ou desmobilização de eleitores. A título de exemplo prático, o ChatGPT é treinado para recusar ou redirecionar com segurança um comando que lhe peça para escrever mensagens direcionadas a um segmento de eleitores incentivando-o a apoiar determinado candidato, enquanto uma solicitação neutra para resumir regras eleitorais públicas ou elaborar materiais administrativos internos pode ser tratada dentro dos limites das políticas.

[33] Disponível em: https://help.openai.com/pt-br/articles/20001255-restri%C3%A7%C3%B5es-a-campanhas-pol%C3%ADticas. Acessado em 14 de agosto de 2026.

(b) Detecção e **classificação em tempo real**: a OpenAI também utiliza ferramentas automatizadas para detectar comandos, conclusões e uploads que possam ser prejudiciais ou que violem suas Políticas de Uso e monitora padrões de uso indevido, inclusive relacionados a processos eleitorais.[34] Dependendo da solicitação e do contexto, essas salvaguardas podem resultar na recusa do pedido ou em uma resposta mais neutra, enquanto sistemas automatizados de aplicação de políticas podem emitir advertências, bloquear respostas ou contribuir para o encaminhamento do caso para investigação e revisão humana, com vistas à aplicação de nossas Políticas de Uso e de outros termos aplicáveis.

(c) Denúncia de usuários e análise humana: de acordo com os artigos da Central de Ajuda "Denúncia de conteúdo no ChatGPT" e "Como identificamos conteúdo problemático em nossos serviços para pessoas físicas", os usuários podem denunciar conteúdo ou comportamentos que considerem violar as políticas da OpenAI ou a legislação local aplicável, incluindo aqueles relacionados a questões eleitorais.[35] Os casos sinalizados podem receber análise humana, e violações graves ou reiteradas podem resultar em advertências, restrições de funcionalidades, suspensão ou desativação da conta.

Em conjunto, as medidas acima dão efetividade ao Model Spec, aos Termos de Uso e às Políticas de Uso da OpenAI, inclusive às restrições relacionadas a campanhas políticas, interferência eleitoral, desmobilização de eleitores e usos enganosos de conteúdo gerado por IA. Essas salvaguardas destinam-se a assegurar o tratamento adequado de solicitações que envolvam classificação de candidatos, recomendações de voto e persuasão política direcionada.

6.2.4. Informações eleitorais confiáveis e links para fontes

Conforme descrito na página "Informações eleitorais e salvaguardas em 2026"[36], os usuários podem fazer ao ChatGPT perguntas cívicas práticas sobre o processo eleitoral, incluindo como se registrar, prazos aplicáveis e resultados eleitorais oficiais. A OpenAI continua explorando formas de fornecer respostas eficazes e confiáveis a essas perguntas, inclusive por meio de parcerias com organizações de informação e mídia confiáveis e direcionando os usuários a essas fontes confiáveis.

Com base em seus esforços iniciados em 2024, a OpenAI está trabalhando com parceiros para direcionar as pessoas a fontes confiáveis de informação sobre votação. A partir deste ano, nos Estados Unidos e no Brasil, a OpenAI fornecerá contagens de votos ao vivo da "The Associated Press"[37] à medida que os resultados forem divulgados na noite da eleição.

---

[34] Disponível em: https://openai.com/index/election-safeguards-2026/. Acessado em 14 de agosto de 2026.
[35] **Denunciar conteúdo.** Disponível em: https://openai.com/pt-BR/form/report-content/. Acessado em 14 de agosto de 2026.
[36] Disponível em: https://openai.com/index/election-safeguards-2026/. Acessado em 14 de agosto de 2026.
[37] Disponível em: https://www.ap.org/elections/. Acesso em 14 de agosto 2026.

Além disso, a busca do ChatGPT foi concebida para encontrar na web informações relevantes e fornecer respostas mais robustas com links para fontes, permitindo que os usuários consultem as fontes subjacentes. Essa funcionalidade apoia o quadro mais amplo de neutralidade e transparência do ChatGPT ao ajudar os usuários a verificar informações eleitorais factuais, em vez de depender apenas dos resultados do modelo.

6.2.5. Escalonamento e aplicação de políticas

A OpenAI mantém equipes [RESTRITO] que trabalham para detectar padrões de uso indevido, investigar abuso enganoso ou relacionado a eleições, incluindo possíveis usos do ChatGPT como parte de operações mais amplas de influência encobertas, e adotar medidas de aplicações das políticas que podem incluir a restrição ou o encerramento do acesso aos serviços do ChatGPT.[38]

6.2.6. Publicidade e neutralidade política

De acordo com as Políticas publicitárias da OpenAI[39] e "Informações eleitorais e salvaguardas em 2026",[40] a OpenAI não permite anúncios políticos em suas plataformas. Isso inclui publicidade relacionada a eleições ou participação eleitoral (por exemplo, apoio ou oposição a um candidato, partido político ou referendo, ou incentivo ou desencorajamento do voto); defesa relacionada a autoridades públicas ou ações governamentais (por exemplo, fiscalização da imigração, impostos, política climática ou outras questões legislativas ou regulatórias); e publicidade que enquadre questões sociais controversas como assuntos de controvérsia pública ou debate político. Essa política elimina um possível vetor de promoção não autorizada de candidatos ou partidos no ChatGPT.

6.3. Prevenção de conteúdo sexual não consentido e violência política de gênero (Art. 15, II)

Nos termos do Art. 15, II, da Portaria, o Plano deve identificar salvaguardas que impeçam a geração de conteúdo sexual não consentido envolvendo candidatos e de publicidade eleitoral que constitua violência política de gênero. A OpenAI atende a essa obrigação por meio de proibições de políticas sobrepostas, sistemas automatizados de segurança de conteúdo e mecanismos de aplicação aplicados em todo o ChatGPT.

6.3.1. Proibições de políticas

De acordo com as Políticas de Uso da OpenAI, a OpenAI proíbe as seguintes condutas relevantes para o Art. 15, II, da Portaria:

(a) Violência sexual ou conteúdo íntimo não consentido de qualquer tipo.

---

[38] **Informações eleitorais e salvaguardas em 2026**. Disponível em: https://openai.com/index/election-safeguards-2026/. Acessado em 14 de agosto de 2026 (Doc 2).
[39] Disponível em: https://openai.com/policies/ad-policies/. Acessado em 14 de agosto de 2026.
[40] Disponível em: https://openai.com/index/election-safeguards-2026/. Acessado em 14 de agosto de 2026.

(b) Ameaças, intimidação, assédio e difamação.

(c) Uso da imagem de uma pessoa, incluindo imagens fotorrealistas ou voz, sem autorização quando a autenticidade possa ser confundida.

(d) Engano, fraude, golpe, spam e falsa identidade.

(e) Danos à segurança de crianças e material de abuso sexual infantil (CSAM).

Em conjunto, essas proibições abrangem a criação de imagens sexuais não consentidas, a exploração não autorizada da identidade de um indivíduo, conduta abusiva e outras formas de conteúdo prejudicial que, quando dirigidas a candidatos eleitorais, também podem servir como veículos de violência política de gênero.

6.3.2. Estrutura de segurança da geração de imagens

Conforme informado no "ChatGPT Images 2.0 System Card",[41] o ChatGPT Images 2.0 incorpora uma arquitetura de segurança em múltiplas camadas concebida para impedir conteúdo visual prejudicial, incluindo riscos elevados associados a deepfakes e a imagens políticas, sexuais ou de outra forma sensíveis de pessoas, lugares ou eventos reais. Mais especificamente:

(a) Classificadores prévios de prompts: classificadores automatizados avaliam prompts de texto antes que qualquer solicitação de geração de imagens chegue ao modelo. Solicitações que envolvam conteúdo proibido, incluindo imagens íntimas não consentidas, semelhança não autorizada e assédio, são recusadas na etapa do prompt.

(b) Modelo de raciocínio de segurança: um modelo de raciocínio de segurança modera as entradas de imagem e os resultados gerados antes de serem exibidos ao usuário. Se o resultado violar a política, ele é bloqueado antes da entrega.

(c) Monitoramento de entradas de texto e imagem: de acordo com o ChatGPT Images 2.0 System Card, as entradas de texto e imagem são monitoradas por meio de sistemas automatizados que identificam sinais de categorias de conteúdo proibidas.

A OpenAI também implementa controles técnicos específicos para tratar de imagens íntimas não consentidas (NCII), [RESTRITO].

6.3.3. Medidas de Aplicação de Políticas

As violações de políticas relacionadas a conteúdo sexual não consentido, semelhança não autorizada, ameaças, assédio ou abuso estão sujeitas a medidas de aplicação das políticas, que incluem advertências automatizadas, recusa do serviço para a solicitação específica, restrição de funcionalidades do produto, limitações de taxa ou acesso, suspensão da conta e desativação da conta. Os usuários podem contestar medidas de aplicação das políticas por

[41] Disponível em: https://openai.com/pt-BR/index/introducing-chatgpt-images-2-0/. Acessado em 14 de agosto de 2026.

meio de processos documentados, quando disponíveis. O conteúdo identificado em links compartilhados, GPTs públicos ou ChatGPT Sites está sujeito a mecanismos de denúncia, análise e remoção.[42]

6.4. Mecanismos de prevenção de contorno e defesas contra *prompt injection* (Art. 15, III)

Nos termos do Art. 15, III, da Portaria, o Plano deve fornecer uma descrição funcional dos mecanismos para impedir, detectar, bloquear e mitigar tentativas de contornar ou utilizar indevidamente salvaguardas e controles de conteúdo. A OpenAI aplica uma abordagem profunda de defesa em todo o ChatGPT, reconhecendo que tentativas de contorno, como "jailbreaking"[43] e "prompt injection",[44] representam um desafio de segurança em evolução para sistemas de IA generativa.[45] A arquitetura descrita a seguir combina defesas estruturais em nível de treinamento, monitoramento em tempo real, intervenções automatizadas, testes adversariais e vias de escalonamento. As defesas em múltiplas camadas destinam-se a proteger contra tentativas de coagir o ChatGPT a responder em desconformidade com o treinamento do modelo, em todos os aspectos do Model Spec e de outras políticas, inclusive aquelas relacionadas a eleições.

6.4.1. Defesas estruturais e em nível de treinamento

Os modelos subjacentes do ChatGPT incorporam múltiplas camadas de defesa estabelecidas durante o treinamento e o desenho do sistema:

---

[42] **Políticas de Uso.** Disponível em: https://openai.com/policies/usage-policies/. Acessado em 14 de agosto de 2026; **Como identificamos conteúdo problemático em nossos serviços para pessoas físicas.** Disponível em: https://help.openai.com/pt-br/articles/8940831-como-identificamos-conte%C3%BAdo-problem%C3%A1tico-em-nossos-servi%C3%A7os-para-pessoas-f%C3%ADsicas. Acessado em 14 de agosto de 2026; **Denúncia de conteúdo no ChatGPT e nas plataformas da OpenAI.** Disponível em: https://help.openai.com/pt-br/articles/10245791-como-denunciar-conte%C3%BAdo-no-chatgpt-e-nas-plataformas-da-openai; Acessado em 14 de agosto de 2026; **Compartilhar e publicar GPTs.** Disponível em: https://help.openai.com/pt-br/articles/8798878-sharing-and-publishing-gpts. Acessado em 14 de agosto de 2026; **Entenda suas responsabilidades em relação aos seus ChatGPT Sites.** Disponível em: https://help.openai.com/pt-br/articles/20001337-entenda-suas-responsabilidades-em-rela%C3%A7%C3%A3o-aos-seus-chatgpt-sites. Acessado em 14 de agosto de 2026.

[43] Uso de prompt adversarial ou sequência de prompts concebidos para contornar o treinamento de recusa do modelo (*refusal training*) ou outras salvaguardas de segurança, com o objetivo de induzir a geração de assistência, informações ou conteúdo que seriam vedados ou potencialmente nocivos.

[44] Ataque a sistemas de IA conversacional em que instruções maliciosas ou enganosas são inseridas em entradas não confiáveis ou em conteúdo de terceiros, como páginas da web, documentos, e-mails ou resultado de ferramentas, com o objetivo de induzir o modelo ou agente a agir de forma contrária à intenção do usuário ou a instruções de maior prioridade.

[45] **Entendendo as injeções imediatas**. Disponível em: https://openai.com/pt-BR/index/prompt-injections/. Acessado em 14 de agosto de 2026; **Cartão do sistema do GPT-5.6**. Disponível em: https://openai.com/pt-BR/index/gpt-5-6/. Acessado em 14 de agosto de 2026.

(a) Treinamento de segurança e recusas: conforme explicado na página do GPT-5.6 System Card, a[46] OpenAI treina seus modelos para seguir as políticas do modelo e as expectativas de segurança, inclusive por meio de treinamento baseado em raciocínio que ajuda os modelos a reconhecer erros, seguir diretrizes aplicáveis e resistir a tentativas de contornar regras de segurança. O treinamento de segurança é uma camada da estrutura de mitigação da OpenAI, juntamente com monitores, controles de acesso, aplicação off-line e outras salvaguardas. No contexto de domínios de alto risco, o GPT-5.6 System Card descreve dados de treinamento que incluem exemplos sintéticos, de produção e semissintéticos derivados de cenários de ameaça, bem como avaliações de recusas em nível de modelo utilizando prompts extraídos de dados sintéticos reservados, red-teaming e dados de produção. Para os fins do Art. 15, esse quadro mais amplo de treinamento de segurança e recusas apoia a capacidade do ChatGPT de recusar, redirecionar ou concluir com segurança solicitações que se enquadrem em categorias proibidas pelas políticas, incluindo solicitações que envolvam campanhas políticas proibidas, interferência eleitoral, desmobilização, contorno das salvaguardas ou conteúdo enganoso gerado por IA.

(b) Hierarquia de instruções: nos termos do Model Spec, o ChatGPT é concebido para seguir uma hierarquia de instruções que prioriza instruções de maior autoridade, incluindo instruções em nível de sistema e de plataforma, em relação a conteúdo de menor autoridade ou não confiável. Como exemplos, texto citado, dados multimodais, anexos de arquivos e resultados de ferramentas são tratados, por padrão, como não confiáveis e não têm autoridade para substituir instruções de maior prioridade. Essa hierarquia é relevante para o Art. 15 porque ajuda a impedir que prompts de usuários, conteúdo de terceiros, resultados de ferramentas, resultados da web ou outro material externo neutralizem as salvaguardas que regem o comportamento do ChatGPT. Por exemplo, conforme explicado na página “Entendendo as injeções imediatas”, prompt injection ocorre quando um terceiro insere instruções maliciosas no contexto da conversa em uma tentativa de induzir o modelo a fazer algo que o usuário não solicitou. Os controles de hierarquia de instruções do ChatGPT são, portanto, concebidos para ajudar o modelo a distinguir instruções confiáveis de instruções não confiáveis e ignorar ou resistir a tentativas de prompt injection que busquem substituir salvaguardas de segurança, restrições de políticas ou salvaguardas em nível de plataforma.

(c) Modelagem de engenharia social: conforme explicado acima, prompt injection é uma forma de ataque de engenharia social específica da IA conversacional, na qual instruções maliciosas incorporadas a conteúdo de terceiros tentam influenciar o comportamento do modelo. A abordagem da OpenAI não se limita à detecção de uma

[46] **Cartão do sistema do GPT-5.6**. Disponível em: https://openai.com/pt-BR/index/gpt-5-6/. Acessado em 14 de agosto de 2026.

cadeia maliciosa específica ou à filtragem de entradas individuais. Ela também se concentra em limitar o impacto das tentativas de manipulação, impedindo que transmissões de informações sensíveis ou ações potencialmente consequenciais ocorram de forma silenciosa e sem salvaguardas apropriadas. Em contextos do ChatGPT que envolvam ferramentas, links, conteúdo externo ou recursos de agentes, essas salvaguardas podem incluir verificações de links, sandboxing,[47] confirmações do usuário antes de ações consequenciais e controles destinados a detectar quando informações aprendidas em uma conversa seriam transmitidas a terceiros. Essa abordagem apoia o Art. 15, III, da Portaria, ao reduzir o risco de que prompt injection ou uma instrução adversarial seja utilizada para contornar, neutralizar ou explorar os controles de segurança do ChatGPT.

6.4.2. Monitoramento em tempo real e intervenções automatizadas

Além das camadas de defesa estabelecidas durante o treinamento e o desenho dos sistemas descritas acima, os seguintes sistemas automatizados operam em tempo real nas interações do ChatGPT:

Os controles descritos nesta Seção operam em diferentes pontos e momentos. Alguns avaliam uma solicitação ou resultado individual durante a interação; outros identificam padrões em atividades relacionadas e apoiam a análise e aplicação subsequentes. Os sinais e respostas aplicáveis variam conforme a interface de produto, o tipo de risco e o contexto operacional.

(a) Detecção e resposta no nível da interação. Classificadores automatizados, comportamento de modelo orientado por políticas e salvaguardas específicas do produto podem identificar solicitações ou resultados associados a interferência eleitoral proibida, desmobilização de eleitores, campanhas políticas em escala, mídia sintética enganosa ou tentativas de contornar as salvaguardas aplicáveis. Dependendo do produto e do sinal, a resposta imediata pode incluir recusar ou redirecionar com segurança a solicitação, bloquear um resultado, alterar o comportamento de resposta aplicável, desviar a solicitação de uma determinada ação, exibir um aviso ou exigir confirmação do usuário antes da realização de uma ação externa com consequências relevantes.

(b) Detecção e análise no nível de padrões. A OpenAI também utiliza sistemas projetados para identificar padrões potencialmente prejudiciais que podem não ser aparentes a partir de uma única interação. Os sinais relevantes podem incluir tentativas de persuasão política em escala, tentativas de gerar materiais de campanha, atividade inautêntica em redes sociais, coleta de dados de eleitores e tentativas repetidas de personificação de candidatos ou autoridades eleitorais. Sinais e resumos

[47] Sandboxing refere-se ao uso de um ambiente de execução isolado e controlado para determinadas interações habilitadas por ferramentas ou agentivas, concebido para limitar o impacto de instruções externas maliciosas, comunicações inesperadas ou tentativas de prompt injection. Disponível em: https://openai.com/pt-BR/index/building-codex-windows-sandbox/. Acessado em 14 de agosto de 2026.

automatizados podem ser encaminhados para fluxos de trabalho de análise centralizados, onde os analistas validam as evidências disponíveis e escalonam os assuntos de maior risco às [RESTRITO]. Violações de políticas identificadas são tratadas de acordo com os procedimentos descritos em outras partes deste Plano.

(c) Alerta antecipado específico para eleições. Durante o ciclo eleitoral de 2026, as equipes operacionais da OpenAI avaliam riscos emergentes e sinais observáveis nas interfaces de produtos relevantes, [RESTRITO]. Indicadores relevantes podem incluir mudanças incomuns na atividade relacionada a eleições; tentativas de gerar materiais eleitorais oficiais falsificados ou evidências enganosas de fraude eleitoral; aumentos nas tentativas de personificação de candidatos, autoridades eleitorais ou figuras públicas; consultas anômalas sobre procedimentos de votação; aplicativos maliciosos emergentes relacionados a eleições; e relatórios de autoridades eleitorais, sociedade civil, pesquisadores ou outras fontes externas. Esses indicadores informam a investigação e a resposta; eles não estabelecem necessariamente uma violação por si só.

(d) Controles de URLs seguras e de transmissão de informações: quando as informações aprendidas em uma conversa seriam transmitidas a terceiros, o ChatGPT pode exibir as informações ao usuário e solicitar confirmação, ou bloquear a transmissão e orientar a interação para uma abordagem alternativa.[48]

Esses mecanismos automatizados podem gerar sinais para análise adicional no âmbito do processo de escalonamento e mecanismos de aplicação das políticas descritos na **Seção 6.4.4.**

6.4.3. Testes adversariais e melhoria contínua

A OpenAI mantém um programa de testes adversariais direcionado a vetores de contorno e *prompt injection*:

(a) Red-teaming interno e externo: equipes de red-teaming dedicadas testam as salvaguardas do ChatGPT contra entradas adversariais, incluindo cenários relacionados a eleições. Por exemplo, a implementação do GPT-5.6 envolveu mais de 700.000 horas de GPU equivalentes a A100 de red-teaming automatizado[49] para identificar jailbreaks universais e padrões de contorno.[50]

---

[48] **Projetando agentes de IA para resistir à prompt injection.** Disponível em: https://openai.com/pt-BR/index/designing-agents-to-resist-prompt-injection/. Acessado em 14 de agosto de 2026.

[49] Método de red teaming que utiliza modelos de IA para gerar e aperfeiçoar exemplos adversariais ou ataques, bem como para avaliar as respostas do sistema-alvo, contribuindo para identificar, em escala, potenciais falhas de segurança, proteção ou robustez.

[50] **Cartão do sistema do GPT-5.6**. Disponível em: https://openai.com/pt-BR/index/gpt-5-6/. Acessado em 14 de agosto de 2026.

(b) Testes de red-teaming por terceiros: pesquisadores de segurança externos e avaliadores terceiros checam a resistência do ChatGPT a certos tipos de tentativas de contorno, fornecendo uma avaliação independente da robustez das salvaguardas.[51]

(c) Geração automatizada de ataques: a OpenAI utiliza sistemas automatizados para gerar prompts adversariais e testar certas respostas das salvaguardas em escala, permitindo identificar novos vetores de contorno antes que sejam explorados em produção.[52]

(d) Programa de bug bounty: a OpenAI mantém um programa público de bug bounty que incentiva pesquisadores externos a relatar caminhos de prompt injection ou possíveis riscos de exposição de dados, contribuindo para a correção iterativa de fragilidades identificadas.[53]

(e) Correção iterativa: as técnicas de contorno identificadas podem ser tratadas por meio de atualizações do modelo, melhorias dos classificadores, aperfeiçoamentos das salvaguardas e aperfeiçoamentos das políticas de forma contínua.[54]

(f) Testes contínuos durante a implementação: as salvaguardas são, de modo geral, monitoradas com métricas de recall durante a implementação em produção, e o red-teaming contínuo ocorre ao longo de todo o ciclo de vida do modelo.[55]

6.4.4. Escalonamento e medidas de aplicação das políticas para tentativas de conterno

Quando os sistemas automatizados da OpenAI sinalizam interações de maior risco, tentativas reiteradas de contorno ou padrões de uso indevido suspeitos, a questão pode ser escalada para análise automatizada mais aprofundada e, quando apropriado, análise humana. A análise humana apoia a avaliação contextual, as decisões de aplicação de políticas e o aperfeiçoamento das salvaguardas, especialmente quando os sinais relevantes exigem interpretação além da classificação automatizada. As consequências para usuários envolvidos em tentativas de contorno podem incluir advertências automatizadas, restrição de funcionalidades do produto, limitações de taxa ou acesso, suspensão da conta e desativação da conta, observados os mecanismos de contestação disponíveis. As equipes [RESTRITO] da OpenAI investigam abuso enganoso e relacionado a eleições, incluindo campanhas

[51] **Cartão do sistema do GPT-5.6**. Disponível em: https://openai.com/pt-BR/index/gpt-5-6/. Acessado em 14 de agosto de 2026.
[52] **Cartão do sistema do GPT-5.6**. Disponível em: https://openai.com/pt-BR/index/gpt-5-6/. Acessado em 14 de agosto de 2026.
[53] **Projetando agentes de IA para resistir à prompt injection.** Disponível em: https://openai.com/pt-BR/index/designing-agents-to-resist-prompt-injection/. Acessado em 14 de agosto de 2026.
[54] **Cartão do sistema do GPT-5.6**. Disponível em: https://openai.com/pt-BR/index/gpt-5-6/. Acessado em 14 de agosto de 2026.
[55] **Cartão do sistema do GPT-5.6**. Disponível em: https://openai.com/pt-BR/index/gpt-5-6/. Acessado em 14 de agosto de 2026.

sofisticadas de contorno e tentativas de usar o ChatGPT como parte de possíveis operações de influência encobertas.

6.5. Identificação e rotulagem de conteúdo sintético (Art. 15, IV)

Nos termos do Art. 15, IV, da Portaria, o Plano deve descrever mecanismos para identificar e rotular conteúdo sintético em publicações orgânicas, incluindo a descrição funcional, o formato e a visibilidade da solução adotada. A OpenAI implementa sinais de proveniência para conteúdo gerado pelo ChatGPT e serviços relacionados por meio de uma combinação de metadados criptográficos e tecnologias de marca d'água imperceptível, conforme explicado a seguir.

6.5.1. Content Credentials C2PA (metadados criptográficos)[56]

As imagens compatíveis geradas pelo ChatGPT incluem Content Credentials C2PA assinadas. Os metadados C2PA fornecem um registro assinado criptograficamente da origem e do histórico do conteúdo, incluindo a ferramenta ou o serviço utilizado para gerar o conteúdo. Esses metadados são incorporados diretamente ao arquivo de imagem e podem ser verificados por qualquer ferramenta de verificação compatível com C2PA.

No conteúdo gerado pela OpenAI, as Content Credentials C2PA são aplicadas a imagens compatíveis criadas por meio de ImageGen nos produtos da OpenAI. As credenciais permitem que destinatários e plataformas subsequentes na cadeia de distribuição verifiquem que uma imagem foi gerada com ferramentas da OpenAI, apoiando a transparência no ecossistema de informações.

Os metadados C2PA podem ser removidos ou perdidos por meio de upload, download, edição, conversão, compartilhamento ou remoção deliberada. A ausência de metadados C2PA em uma determinada imagem, portanto, não é conclusiva quanto a saber se a imagem foi gerada por ferramentas da OpenAI, motivo que também justifica o porquê de a OpenAI aplicar o método de marca d'água imperceptível descrito abaixo.

6.5.2. Marca d'água imperceptível SynthID[57]

Para o conteúdo gerado pela OpenAI, as imagens geradas pelo ChatGPT incorporam o SynthID, uma marca d'água imperceptível incorporada diretamente ao conteúdo da mídia. O SynthID opera independentemente dos metadados do arquivo e pode persistir após determinadas edições e transformações de imagem, fornecendo um sinal de proveniência adicional mais resistente à remoção casual do que abordagens baseadas em metadados.

---

[56] **Sinais de proveniência (Content Credentials, SynthID) no conteúdo gerado pela OpenAI**. Disponível em: https://help.openai.com/pt-br/articles/8912793-sinais-de-proveni%C3%AAncia-credenciais-de-conte%C3%BAdo-synthid-em-conte%C3%BAdo-gerado-pela-openai. Acessado em 14 de agosto de 2026.

[57] **Sinais de proveniência (Content Credentials, SynthID) no conteúdo gerado pela OpenAI**. Disponível em: https://help.openai.com/pt-br/articles/8912793-sinais-de-proveni%C3%AAncia-credenciais-de-conte%C3%BAdo-synthid-em-conte%C3%BAdo-gerado-pela-openai. Acessado em 14 de agosto de 2026.

No conteúdo gerado pela OpenAI, a marca d'água SynthID é aplicada a imagens compatíveis geradas com o ImageGen nos produtos da OpenAI. Para conteúdo de áudio gerado por meio de recursos compatíveis, uma marca d'água SynthID inaudível é incorporada ao resultado de áudio.

No conteúdo gerado pela OpenAI, a detecção do SynthID fornece um indicador de origem da OpenAI: uma correspondência positiva indica que o conteúdo provavelmente contém um sinal de proveniência compatível associado à OpenAI. Uma correspondência positiva não indica precisão, titularidade, contexto ou que o conteúdo não tenha sido alterado em relação à sua forma original. O SynthID pode se degradar após transformações significativas, e a ausência de uma marca d'água detectável não é conclusiva.

6.5.3. Ferramenta de verificação da OpenAI[58]

Para o conteúdo gerado pela OpenAI, a OpenAI disponibiliza uma ferramenta de verificação acessível publicamente que permite que qualquer pessoa verifique se um arquivo carregado contém sinais de proveniência associados às ferramentas da OpenAI. A ferramenta de verificação examina o conteúdo carregado em busca de metadados C2PA e marcas d'água SynthID e informa se sinais de proveniência são detectados. Atualmente, a ferramenta é compatível com arquivos de imagem e áudio compatíveis.[59]

A ferramenta de verificação da OpenAI apoia o objetivo do Art. 15, IV, da Portaria, ao permitir a identificação de conteúdo gerado por IA em publicações orgânicas. Qualquer pessoa, jornalista, autoridade eleitoral ou moderador de conteúdo pode utilizar a ferramenta para auxiliar a avaliar se o conteúdo foi gerado com ferramentas da OpenAI, facilitando decisões informadas sobre confiança, compartilhamento e denúncia.

6.5.4. Cobertura por modalidade e caracterização da tecnologia

A lista a seguir descreve a cobertura dos sinais de proveniência por modalidade de conteúdo:

> **(a) Imagens:** imagens compatíveis incluem Content Credentials C2PA e marca d'água imperceptível SynthID. A cobertura varia conforme o produto, o modelo, o caminho de exportação, o tipo de arquivo e a data de criação.[60]

[58] **Verifique conteúdo gerado pela OpenAI.** Disponível em: https://openai.com/pt-BR/research/verify/. Acessado em 14 de agosto de 2026.
[59] **Sinais de proveniência (Content Credentials, SynthID) no conteúdo gerado pela OpenAI**. Disponível em: https://help.openai.com/pt-br/articles/8912793-sinais-de-proveni%C3%AAncia-credenciais-de-conte%C3%BAdo-synthid-em-conte%C3%BAdo-gerado-pela-openai. Acessado em 14 de agosto de 2026.
[60] **Sinais de proveniência (Content Credentials, SynthID) no conteúdo gerado pela OpenAI**. Disponível em: https://help.openai.com/pt-br/articles/8912793-sinais-de-proveni%C3%AAncia-credenciais-de-conte%C3%BAdo-synthid-em-conte%C3%BAdo-gerado-pela-openai. Acessado em 14 de agosto de 2026.

(b) Áudio: o áudio gerado pela OpenAI compatível inclui uma marca d'água SynthID inaudível incorporada ao resultado de áudio.[61]

(c) Vídeo: os resultados de vídeo produzidos com o Sora incluíam metadados C2PA e uma marca d'água móvel visível.[62]

(d) Texto e código: o texto e os dados estruturados gerados pelo ChatGPT atualmente não carregam marcas d'água de proveniência legíveis por máquina. A OpenAI tem trabalhado ativamente e estudado maneiras de ampliar a proveniência a todas as modalidades, incluindo texto, à medida que os padrões e as ferramentas de detecção amadurecem.[63]

No conteúdo gerado pela OpenAI, nenhuma solução individual de proveniência proporciona identificação absoluta de conteúdo gerado por IA.[64] A abordagem da OpenAI baseada em múltiplos sinais – que combina metadados criptográficos, marcas d'água imperceptíveis e ferramenta pública de verificação – fornece capacidades de identificação em camadas que apoiam os objetivos do Art. 15, IV, da Portaria, na atual fronteira tecnológica.

6.5.5. Visibilidade e formato

Os sinais de proveniência no conteúdo gerado pela OpenAI são concebidos para ser verificáveis por meio de ferramentas padrão (sistemas de verificação compatíveis com C2PA e a ferramenta Verify da OpenAI), sem exigir conhecimento técnico especializado dos usuários finais. Os sinais são incorporados em nível de arquivo (metadados e marca d'água), garantindo que estejam associados ao próprio conteúdo, e não dependam de uma plataforma de distribuição específica. A proveniência legível por máquina (isto é, C2PA e SynthID) opera independentemente de marcações visíveis e constitui o principal mecanismo de identificação para os fins do Art. 15, IV, da Portaria.

6.6. Testes de resistência e avaliação de segurança (Art. 15, V)

Nos termos do Art. 15, V, da Portaria, o Plano deve tratar da existência e periodicidade de testes de resistência das salvaguardas descritas neste Plano e de tentativas de contorno, antes

---

[61] **Sinais de proveniência (Content Credentials, SynthID) no conteúdo gerado pela OpenAI**. Disponível em: https://help.openai.com/pt-br/articles/8912793-sinais-de-proveni%C3%AAncia-credenciais-de-conte%C3%BAdo-synthid-em-conte%C3%BAdo-gerado-pela-openai. Acessado em 14 de agosto de 2026.

[62] **Cartão do sistema do Sora 2.** Disponível em: https://openai.com/index/sora-2-system-card/. Acessado em 14 de agosto de 2026.

[63] **Sinais de proveniência (Content Credentials, SynthID) no conteúdo gerado pela OpenAI**. Disponível em: https://help.openai.com/pt-br/articles/8912793-sinais-de-proveni%C3%AAncia-credenciais-de-conte%C3%BAdo-synthid-em-conte%C3%BAdo-gerado-pela-openai. Acessado em 14 de agosto de 2026; **Compreendendo a origem do que vemos e ouvimos on-line.** Disponível em: https://openai.com/index/understanding-the-source-of-what-we-see-and-hear-online/. Acessado em 14 de agosto de 2026.

[64] **Sinais de proveniência (Content Credentials, SynthID) no conteúdo gerado pela OpenAI**. Disponível em: https://help.openai.com/pt-br/articles/8912793-sinais-de-proveni%C3%AAncia-credenciais-de-conte%C3%BAdo-synthid-em-conte%C3%BAdo-gerado-pela-openai. Acessado em 14 de agosto de 2026.

e durante o período eleitoral. A OpenAI mantém uma infraestrutura de avaliação e testes de segurança aplicável ao ChatGPT que apoia essa obrigação. Alguns elementos dessa infraestrutura já foram descritos nas seções anteriores. Contudo, para tratar diretamente do Art. 15, V, e facilitar a consulta, eles são consolidados abaixo juntamente com componentes adicionais que fazem parte da infraestrutura de avaliação e testes de segurança da OpenAI.

6.6.1. Programa de avaliação de viés político

A OpenAI conduz uma avaliação estruturada de viés político para respostas de texto do ChatGPT.[65] O programa de avaliação incorpora os seguintes elementos:

(a) Escopo e desenho da avaliação: a avaliação de viés político da OpenAI para respostas de texto do ChatGPT é estruturada em torno de um amplo conjunto de temas políticos, de políticas públicas e culturais. A avaliação utiliza aproximadamente 500 comandos em 100 temas, com cinco variantes de comando por tema, incluindo comandos comuns e comandos mais desafiadores envolvendo cenários politicamente sensíveis, polarizados ou emocionalmente carregados. Esse desenho destina-se a testar se o ChatGPT mantém uma postura objetiva e equilibrada em diferentes tipos de conversas políticas.

(b) Critérios de avaliação: a avaliação utiliza um avaliador baseado em LLM para avaliar as respostas em várias dimensões relevantes para a neutralidade política, incluindo invalidação do usuário, escalonamento de usuário, expressão política pessoal, cobertura assimétrica de posições políticas e recusas políticas injustificadas. Esses critérios avaliam não apenas se o ChatGPT recusa injustificadamente um prompt político, mas também se deslegitima a posição do usuário, amplifica indevidamente sua posição política, manifesta opiniões políticas como se fossem próprias ou apresenta cobertura assimétrica quando múltiplos pontos de vista legítimos são relevantes e o usuário não solicitou uma explicação parcial.

(c) Resultados da avaliação: para fins exemplificativos, a avaliação dos modelos GPT-5 *Instant* e GPT-5 *Thinking* indicou uma redução aproximada de 30% nas pontuações de viés em comparação desses modelos com gerações anteriores de modelos e estimou que menos de 0,01% das respostas amostradas do ChatGPT em produção apresentam sinais de viés político. Esses resultados ajudam a apoiar a posição da OpenAI de que o quadro de avaliação fornece uma base estruturada para aferir melhorias na postura de neutralidade política do ChatGPT ao longo do tempo.

(d) Aplicação contínua: essa avaliação de viés político apoia as salvaguardas descritas nesta Seção ao fornecer um método repetível para avaliar se o ChatGPT mantém neutralidade política em diferentes tipos de comandos e contextos políticos. Para os fins do Art. 15, a avaliação opera em conjunto com o Model Spec, as Políticas

[65] **Definindo e avaliando o viés político em LLMs.** Disponível em: https://openai.com/index/defining-and-evaluating-political-bias-in-llms/ . Acessado em 14 de agosto de 2026.

de Uso, as salvaguardas técnicas, o monitoramento e os mecanismos de aplicação descritos anteriormente.

6.6.2. Avaliação de segurança pré-implementação

Antes da implementação de novos modelos no ChatGPT, a OpenAI conduz avaliações extensivas de segurança pré-implementação:

(a) Red-teaming automatizado: a implementação do GPT-5.6 envolveu mais de 700.000 horas de GPU equivalentes a A100 de red-teaming automatizado para jailbreaks universais e padrões de contorno. Trata-se de um programa de avaliação substancialmente mais intensivo do que o aplicado a gerações anteriores de modelos.[66]

(b) Testes de red-teaming por terceiros: avaliadores externos de red-teaming checam a segurança do modelo em categorias incluindo conteúdo sexual, nudez, extremismo, automutilação, ilícitos, violência, persuasão política, segurança de jovens e uso indevido de semelhança.[67]

(c) Benchmarks de produção: benchmarks estabelecidos para conteúdo proibido são avaliados antes da implementação, incluindo simulação de implementação para avaliar o desempenho no mundo real.[68]

(d) Avaliações de entradas de imagem: avaliações de segurança específicas para recursos de entrada de imagens avaliam a robustez contra entradas visuais adversariais.[69]

(e) Conjuntos de avaliação de segurança de imagens: o sistema de segurança de imagens é avaliado em relação a conjuntos de monitoramento específicos, com métricas de recall e precisão comparáveis às dos sistemas de segurança baseados em texto.[70]

6.6.3. Testes contínuos durante a implementação

Conforme explicado pela OpenAI tanto no "GPT-5.6 System Card" quanto em "Informações eleitorais e salvaguardas em 2026",[71] os testes não cessam após a implementação do modelo.

---

[66] **Cartão do sistema do GPT-5.6**. Disponível em: https://openai.com/pt-BR/index/gpt-5-6/. Acessado em 14 de agosto de 2026.
[67] **Cartão do sistema do GPT-5.6**. Disponível em: https://openai.com/pt-BR/index/gpt-5-6/. Acessado em 14 de agosto de 2026.
[68] **Cartão do sistema do GPT-5.6**. Disponível em: https://openai.com/pt-BR/index/gpt-5-6/. Acessado em 14 de agosto de 2026.
[69] **Cartão do sistema do GPT-5.6**. Disponível em: https://openai.com/pt-BR/index/gpt-5-6/. Acessado em 14 de agosto de 2026.
[70] **Cartão do sistema do ChatGPT Images 2.0.** Disponível em: https://openai.com/pt-BR/index/introducing-chatgpt-images-2-0/ . Acessado em 14 de agosto de 2026.
[71] Disponível em: https://openai.com/pt-BR/index/gpt-5-6/. Acessado em 14 de agosto de 2026; Disponível em: https://openai.com/index/election-safeguards-2026/. Acessado em 14 de agosto de 2026.

A OpenAI mantém monitoramento e testes contínuos de segurança durante todo o ciclo de vida do modelo no ChatGPT:

(a) As salvaguardas são, de modo geral, monitoradas com métricas de recall durante a implementação em produção, permitindo detectar degradação no desempenho de segurança.

(b) O red-teaming contínuo ocorre, de modo geral, durante todo o ciclo de vida do modelo, com protocolos de resposta rápida para vulnerabilidades identificadas.

(c) As equipes [RESTRITO] monitoram padrões emergentes de uso indevido, incluindo abuso relacionado a eleições, e conduzem investigações sobre atividades enganosas ou relacionadas a eleições.

(d) Sistemas automatizados identificam padrões de insegurança em escala, complementando a investigação humana direcionada.

6.7. Consequência para os requisitos do Art. 7º

Diante da análise conduzida nos itens anteriores desta Seção 6, segue quadro-resumo das medidas, aplicabilidade, prazo, resultados esperados e monitoramento/evidências dos resultados, à luz do Art. 7º:

Medida	Grupo(s) aplicável(eis)	Prazo / status de implementação	Resultado esperado	Monitoramento / evidências dos resultados
Linha de base comportamental do Model Spec para neutralidade política (Seções 6.2 e 6.2.1)	Grupo 1	Em vigor como parte do quadro de comportamento do modelo do ChatGPT e mantida durante todo o período eleitoral.	O ChatGPT opera sob princípios comportamentais concebidos para evitar uma agenda política independente, a manipulação direcionada de opiniões políticas e a condução política não objetiva.	[RESTRITO]
Políticas de Uso da OpenAI e restrições a campanhas políticas (Seções 6.2 e 6.2.2)	Grupo 1	Em vigor como controles de políticas voltados aos usuários e mantida durante todo o período eleitoral.	Vedação quanto aos usos proibidos, como campanhas políticas, interferência eleitoral, desmobilização de eleitores,	[RESTRITO]

Medida	Grupo(s) aplicável(eis)	Prazo / status de implementação	Resultado esperado	Monitoramento / evidências dos resultados
			contorno das salvaguardas, engano sobre conteúdo gerado por IA, mensagens de campanha em escala, contato automatizado de campanha e chatbots de campanha voltados ao público, conforme disposto pelas regras de políticas aplicáveis.	
Distinção entre usos políticos proibidos e apoio cívico ou de campanha permitido, dirigido por humanos (Seção 6.2.2)	Grupo 1	Em vigor como parte do quadro de políticas da OpenAI e mantida durante todo o período eleitoral.	Os usuários podem utilizar o ChatGPT para trabalho permitido e dirigido por humanos, como memorandos internos, briefings, planejamento, resumos de pesquisa, conformidade, acessibilidade e tarefas administrativas, enquanto usos eleitorais proibidos em escala ou automatizados e voltados ao público permanecem restritos.	[RESTRITO]
Recusas em nível de modelo e respostas	Grupo 1	Operacional como parte do quadro de segurança e	Solicitações que se enquadrem em categorias	[RESTRITO]

Medida	Grupo(s) aplicável(eis)	Prazo / status de implementação	Resultado esperado	Monitoramento / evidências dos resultados
seguras (Seção 6.2.3(a))		aplicação de políticas do ChatGPT e mantida durante todo o período eleitoral.	político-eleitorais proibidas são recusadas, redirecionadas ou tratadas por meio de comportamento de conclusão segura, conforme aplicável.	
Classificação em tempo real de prompts e respostas (Seção 6.2.3(b))	Grupo 1	Operacional de forma contínua durante o período eleitoral.	Prompts, conclusões e uploads que possam violar as Políticas de Uso da OpenAI são identificados para recusa, bloqueio, escalonamento ou análise, conforme aplicável.	[RESTRITO]
Denúncia de usuários e análise humana de conteúdo potencialmente violador (Seção 6.2.3(c))	Grupo 1; Grupo 3 quando conteúdo público ou compartilhado é denunciado	Disponível por meio de canais de denúncia e mantida durante todo o período eleitoral.	Usuários e terceiros podem denunciar conteúdo ou comportamento que possa violar as políticas da OpenAI ou a legislação aplicável, incluindo questões relacionadas a eleições, e os casos sinalizados podem ser analisados e receber medidas.	[RESTRITO]
Informações eleitorais confiáveis e links para fontes (Seção 6.2.4)	Grupo 1, incluindo a busca do ChatGPT	Disponível quando a busca do ChatGPT ou a funcionalidade de links para fontes estiver habilitada	Usuários que fizerem perguntas cívicas ou relacionadas a eleições e de natureza factual	[RESTRITO]

Medida	Grupo(s) aplicável(eis)	Prazo / status de implementação	Resultado esperado	Monitoramento / evidências dos resultados
	quando habilitada	durante o período eleitoral.	podem receber respostas apoiadas por links para fontes, permitindo que consultem as fontes subjacentes em vez de depender apenas dos resultados do modelo.	
Investigação e aplicação relativas a abuso enganoso ou relacionado a eleições (Seção 6.2.5)	Grupo 1	Operacional de forma contínua durante o período eleitoral.	Abuso enganoso ou relacionado a eleições, incluindo atividade de influência encoberta, é investigado e pode resultar em medidas de aplicação das políticas, incluindo restrição ou encerramento do acesso.	[RESTRITO]
Política de não veiculação de publicidade política como apoio à neutralidade (Seção 6.2.6)	Grupo 5	Em vigor durante o ciclo eleitoral atual.	A publicidade política paga não é oferecida como mecanismo de promoção de candidatos ou partidos na funcionalidade de publicidade da OpenAI.	[RESTRITO]
Proibições de políticas contra conteúdo íntimo não consentido, violência sexual, ameaças, assédio, uso não autorizado de semelhança,	Grupo 1; Grupo 4	Em vigor como parte das Políticas de Uso da OpenAI e mantida durante todo o período eleitoral.	Categorias de conteúdo subjacentes que possam constituir ou apoiar conteúdo sexual não consentido envolvendo candidatos ou	[RESTRITO]

Medida	Grupo(s) aplicável(eis)	Prazo / status de implementação	Resultado esperado	Monitoramento / evidências dos resultados
falsa identidade e engano (Seções 6.3 e 6.3.1)			violência política de gênero são proibidas nos termos das políticas aplicáveis.	
Estrutura de segurança da geração de imagens do ChatGPT (Seção 6.3.2 (a), (b), (c)	Grupo 4	Operacional para a geração de imagens estáticas compatíveis dentro do ChatGPT e mantida durante todo o período eleitoral.	Conteúdo visual prejudicial, incluindo imagens íntimas não consentidas, conteúdo sexual baseado em semelhança não autorizada e resultados de imagens violadores relacionados, são bloqueados ou recusados nas etapas de prompt, entrada ou resultado.	[RESTRITO]
[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]
Aplicação de políticas por violações envolvendo conteúdo sexual não consentido, semelhança não autorizada, ameaças, assédio ou abuso (Seção 6.3.3)	Grupo 1; Grupo 3; Grupo 4	Operacional de forma contínua durante o período eleitoral.	Violações das políticas relevantes podem resultar em advertências, recusa do serviço para a solicitação específica, restrições de funcionalidades do produto, limitações de acesso, suspensão da conta, desativação da conta ou análise/remoção de conteúdo em interfaces	[RESTRITO]

Medida	Grupo(s) aplicável(eis)	Prazo / status de implementação	Resultado esperado	Monitoramento / evidências dos resultados
			compartilhadas ou públicas.	
Treinamento de segurança e quadro de recusas para conteúdo proibido e tentativas de contorno (Seção 6.4.1(a))	Grupo 1	Em vigor como parte do treinamento do modelo e do desenho do sistema e mantido durante a implementação.	O ChatGPT pode recusar, redirecionar ou concluir com segurança solicitações que se enquadrem em categorias proibidas pelas políticas, incluindo campanhas políticas proibidas, interferência eleitoral, desmobilização de eleitores, contorno das salvaguardas ou conteúdo enganoso gerado por IA.	[RESTRITO]
Hierarquia de instruções e tratamento de conteúdo não confiável (Seção 6.4.1(b))	Grupo 1	Em vigor como parte do desenho do sistema do ChatGPT e mantida durante todo o período eleitoral.	Instruções de sistema e plataforma de maior prioridade são preservadas contra conteúdo de menor autoridade ou não confiável, reduzindo o risco de que prompts de usuários, conteúdo de terceiros, resultados de ferramentas, resultados da web ou materiais externos substituam as	[RESTRITO]

Medida	Grupo(s) aplicável(eis)	Prazo / status de implementação	Resultado esperado	Monitoramento / evidências dos resultados
			salvaguardas do ChatGPT. As medidas podem impactar o Grupo 2, quando ferramentas, Conectores e Apps, resultados da web são exibidos no Chat GPT.	
Prompt injection como risco de engenharia social e controles de limitação de impacto (Seção 6.4.1(c))	Grupo 1;	Em vigor como parte da arquitetura anti-contorno do ChatGPT e mantida durante todo o período eleitoral.	Instruções maliciosas de terceiros incorporadas em conteúdo externo têm menor capacidade de alterar a tarefa pretendida pelo usuário, substituir salvaguardas em nível de sistema ou causar ações ou divulgações de informações não intencionais. As medidas podem impactar o Grupo 2, quando ferramentas, links, conteúdo externo ou ações de agentes estiverem envolvidos.	[RESTRITO]
Detecção e resposta no nível da interação (Seção 6.4.2(a))	Grupo 1	Operacional em tempo real durante as interações do ChatGPT.	Prompts, mensagens, resultados, uploads e conteúdo relacionado são avaliados quanto a possíveis violações de	[RESTRITO]

Medida	Grupo(s) aplicável(eis)	Prazo / status de implementação	Resultado esperado	Monitoramento / evidências dos resultados
			políticas, tentativas de contorno e padrões de uso indevido.	
Detecção e análise no nível de padrões(Seção 6.4.2(b))	Grupo 1	Operacional de forma contínua durante o período eleitoral.	Conteúdo que viole políticas ou tentativas de contorno detectados podem desencadear recusa de geração de conteúdo, conclusão segura, advertências automatizadas, bloqueio de resultados, interrupção do streaming, sandboxing, confirmações do usuário ou escalonamento. As medidas podem impactar o Grupo 2, quando ferramentas relevantes ou funcionalidades integradas estiverem envolvidas.	[RESTRITO]
Alerta antecipado específico para eleições (Seção 6.4.2(c))	Grupo 1	Operacional quando interações de maior risco ou complexas exigem avaliação contextual de segurança.	Interações complexas, incluindo tentativas em múltiplas rodadas de distribuir conteúdo inseguro ou relacionado a contorno entre as trocas, são avaliadas contextualmente	[RESTRITO]

Medida	Grupo(s) aplicável(eis)	Prazo / status de implementação	Resultado esperado	Monitoramento / evidências dos resultados
			quanto ao risco de segurança.	
Controles de URLs seguras e de transmissão de informações (Seção 6.4.2(d))	Grupo 1	Operacional quando surgirem riscos relevantes de transferência de informações ou transmissão a terceiros.	As informações aprendidas em uma conversa não são transmitidas a terceiros sem salvaguardas apropriadas, como confirmação do usuário, bloqueio ou redirecionamento da interação. As medidas podem impactar o Grupo 2, quando houver links, ferramentas, Apps, ações ou destinações externas.	[RESTRITO]
Sinais de escalonamento gerados por mecanismos automatizados (parágrafo final da Seção 6.4.2)	Grupo 1	Operacional como parte do fluxo de escalonamento e aplicação de políticas.	Sinais de maior risco gerados por sistemas automatizados podem ser encaminhados para análise adicional no âmbito dos processos de escalonamento e aplicação de políticas.	[RESTRITO]
Red-teaming interno e externo para jailbreaks e padrões de contorno (Seção 6.4.3(a)-(b))	Grupo 1	Geralmente realizado antes da implementação e continuado durante a implementação de acordo com o programa de testes aplicável.	Jailbreaks universais, entradas adversariais e problemas de robustez das salvaguardas são identificados e mitigados antes ou durante a	[RESTRITO]

Medida	Grupo(s) aplicável(eis)	Prazo / status de implementação	Resultado esperado	Monitoramento / evidências dos resultados
			implementação em produção.	
Geração automatizada de ataques para testes de prompts adversariais (Seção 6.4.3(c))	Grupo 1	Utilizada como parte dos testes adversariais e da avaliação de salvaguardas.	Novos vetores de contorno são identificados por meio da geração e do teste automatizados de prompts adversariais antes de serem explorados em produção.	[RESTRITO]
Denúncias externas de vulnerabilidades / denúncias no bug bounty (Seção 6.4.3(d))	Grupo 1	Disponível como canal de denúncia externo de forma contínua.	Pesquisadores externos podem relatar caminhos de prompt injection ou possíveis riscos de exposição de dados, apoiando a correção e a melhoria das salvaguardas.	[RESTRITO]
Correção iterativa de técnicas de contorno identificadas (Seção 6.4.3(e))	Grupo 1	Contínua durante a implementação e o ciclo de vida do modelo.	Técnicas de contorno identificadas são tratadas por meio de atualizações do modelo, melhorias dos classificadores, aperfeiçoamentos das salvaguardas e aperfeiçoamentos das políticas.	[RESTRITO]
Testes contínuos durante a implementação (Seções 6.4.3(f) e 6.6.3)	Grupo 1	Contínuos durante a implementação e ao longo de todo o ciclo de vida do modelo.	As salvaguardas continuam a ser testadas e monitoradas após a implementação, permitindo detectar qualquer	[RESTRITO]

Medida	Grupo(s) aplicável(eis)	Prazo / status de implementação	Resultado esperado	Monitoramento / evidências dos resultados
			degradação do desempenho de segurança e responder a vulnerabilidades recém-identificadas.	
[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]
Content Credentials C2PA para imagens compatíveis geradas (Seção 6.5.1)	Grupo 4	Aplicadas a imagens compatíveis geradas como parte do quadro de proveniência de imagens da OpenAI.	Imagens compatíveis geradas incluem metadados de proveniência assinados criptograficamente e que podem identificar a origem e o histórico do conteúdo e ser verificados por meio de ferramentas compatíveis.	[RESTRITO]
Marca d'água imperceptível SynthID para imagens e áudio compatíveis (Seção 6.5.2)	Grupo 4	Aplicada a resultados de mídia compatíveis quando a modalidade e o caminho de exportação oferecem suporte ao SynthID.	Imagens e áudio compatíveis gerados incluem sinais de proveniência incorporados que podem persistir após determinadas transformações e indicar provável origem na OpenAI.	[RESTRITO]
Ferramenta Verify da OpenAI para verificação de proveniência (Seção 6.5.3)	Grupo 4; Grupo 3 quando mídias compatíveis são compartilhad	Disponível para arquivos de imagem e áudio compatíveis.	Usuários, jornalistas, autoridades, plataformas e moderadores de conteúdo podem verificar arquivos compatíveis	[RESTRITO]

Medida	Grupo(s) aplicável(eis)	Prazo / status de implementação	Resultado esperado	Monitoramento / evidências dos resultados
	as ou publicadas		quanto a sinais de proveniência associados às ferramentas da OpenAI.	
Cobertura e caracterização de proveniência específica por modalidade (Seção 6.5.4 (a), (b), (c), (d))	Grupo 4; Grupo 1 para caracterizaçã o de resultados de texto	Em vigor como descrição específica por modalidade das capacidades atuais de proveniência.	O Plano esclarece quais modalidades carregam sinais de proveniência e quais atualmente não carregam marcas d'água de proveniência legíveis por máquina, reduzindo a ambiguidade sobre os mecanismos de identificação de conteúdo.	[RESTRITO]
Formato e visibilidade dos sinais de proveniência (Seção 6.5.5)	Grupo 4; Grupo 3 quando mídia gerada é compartilhad a ou publicada	Operacional por meio de metadados incorporados, marcas d'água e ferramentas de verificação.	Os sinais de proveniência estão associados ao próprio arquivo de conteúdo e podem ser verificados por meio de ferramentas padrão sem exigir conhecimento especializado dos usuários finais.	[RESTRITO]
Avaliação de viés político para respostas de texto do ChatGPT (Seção 6.6.1)	Grupo 1	Conduzida como parte do quadro de avaliação da OpenAI.	Neutralidade política do ChatGPT em uma ampla variedade de prompts políticos, de políticas públicas e culturais, incluindo cenários sensíveis, polarizados e	[RESTRITO]

Medida	Grupo(s) aplicável(eis)	Prazo / status de implementação	Resultado esperado	Monitoramento / evidências dos resultados
			emocionalmente carregados.	
Avaliações de segurança pré-implementação para novos modelos no ChatGPT (Seção 6.6.2)	Grupo 1; Grupo 4 quando recursos de imagem são avaliados	Conduzidas antes da implementação de novos modelos ou funcionalidades relevantes.	Lançamentos de modelos e recursos são avaliados quanto a conteúdo proibido, jailbreaks, entradas adversariais, robustez de entradas de imagem e desempenho de segurança de imagens antes da implementação.	[RESTRITO]
Monitoramento e testes contínuos de segurança durante a implementação (Seção 6.6.3)	Grupo 1	Contínuos após a implementação e ao longo de todo o ciclo de vida do modelo.	O desempenho de segurança é monitorado após a implementação, padrões emergentes de uso indevido são investigados e a eficácia das salvaguardas é avaliada continuamente.	[RESTRITO]

7. GESTÃO DO PERÍODO DE VEDAÇÃO ELEITORAL (ART. 16)

7.1. Escopo e aplicabilidade (Art. 4º, § 2º)

O Art. 16 da Portaria dispõe sobre as obrigações relacionadas à gestão do período de vedação eleitoral, incluindo o bloqueio de republicação e impulsionamento de conteúdo de inteligência artificial, a suspensão de chatbots e avatares não identificados como tal e o desligamento automático de anúncios pagos. A Portaria exige que o Plano identifique, quando aplicável, o mecanismo técnico de ativação automática das restrições do período de vedação previstas na Resolução, incluindo interfaces, modalidades e controles abrangidos destinados a evitar exceções causadas por diferenças de localização, relógio do sistema ou

interface, bem como o plano de contingência para falha do mecanismo de desligamento automático.

As informações exigidas pelo Art. 16 são, portanto, direcionadas a serviços em que o provedor opere uma funcionalidade materialmente sujeita às restrições do período de vedação contempladas pela Resolução, como publicação, republicação ou impulsionamento de conteúdo eleitoral gerado por IA; chatbots ou avatares não identificados; ou publicidade paga que deva ser automaticamente desabilitada durante o período de vedação.

Nos termos do Art. 4º, § 2º, da Portaria, as obrigações previstas nos Arts. 11 a 23 da Portaria aplicam-se apenas quando materialmente incidentes nas funcionalidades efetivamente disponibilizadas pelo provedor. Com base nisso, o Art. 16 não é materialmente incidente sobre o ChatGPT nem sobre os Grupos relacionados tratados neste Plano, pelas razões apresentadas abaixo.

O Grupo 1 consiste em funcionalidades próprias do ChatGPT por meio das quais os usuários interagem com os modelos da OpenAI para gerar, sintetizar, recuperar ou transformar informações. O ChatGPT não é um chatbot não identificado para fins do Art. 16. O serviço é apresentado aos usuários como ChatGPT, um serviço de IA voltado aos usuários, com identificação de que eles estão interagindo com um sistema de IA, incluindo o contexto do produto ChatGPT, indicadores de modelo/versão e o aviso permanente voltado aos usuários de que "O ChatGPT pode cometer erros." Portanto, a parte do Art. 16 que trata da suspensão de chatbots ou avatares não identificados como tal não é materialmente incidente sobre o ChatGPT.

O Grupo 1 também inclui a busca do ChatGPT e outras funcionalidades próprias do ChatGPT. Essas funcionalidades não constituem impulsionamento eleitoral pago, amplificação política paga ou mecanismos automáticos de veiculação de anúncios para fins do Art. 16. Quando o ChatGPT recupera ou sintetiza informações em resposta a uma consulta do usuário, essas funções são tratadas nas seções do Plano correspondentes à funcionalidade relevante, incluindo o Art. 15 para governança de IA generativa, o Art. 17 para canais de denúncia, o Art. 20 para avaliação de riscos e o Art. 22 para Políticas de Uso e prevenção de abuso.

O Grupo 2 consiste em integrações de terceiros que podem recuperar, exibir ou apresentar conteúdo de fontes de terceiros na interface do ChatGPT. Na medida em que esse conteúdo de terceiros contenha material político-eleitoral, o material origina-se da fonte terceira e permanece sob responsabilidade desta, não da OpenAI. O Grupo 2, portanto, não enseja um mecanismo da OpenAI de período de vedação para publicação, republicação, impulsionamento pago, publicidade paga ou implementação de chatbot/avatar não identificado nos termos do Art. 16.

O Grupo 3 consiste em interfaces públicas ou de compartilhamento limitadas, incluindo links compartilhados, páginas públicas de GPTs, GPT Store ou interfaces públicas de

descoberta de GPTs e ChatGPT Sites. Este Grupo poderia suscitar questões limítrofes nos termos do Art. 16, porque determinado conteúdo gerado ou configurado pelo ChatGPT pode se tornar acessível a terceiros. Contudo, essas interfaces não transformam o ChatGPT em um serviço geral de distribuição de conteúdo público, ambiente de amplificação política paga ou plataforma de veiculação de publicidade. Os links compartilhados são links iniciados pelo usuário para conversas específicas do ChatGPT; os GPTs públicos e as interfaces do GPT Store dizem respeito ao acesso a configurações de GPTs dentro do ChatGPT; e o ChatGPT Sites opera como páginas hospedadas limitadas ou aplicativos leves pelos quais o usuário divulgador permanece responsável pelo Site e pelo conteúdo enviado pelos visitantes. Essas interfaces são tratadas neste Plano na medida em que gerem denúncia, restrição, remoção, controle de acesso ou controle de visibilidade, mas não desencadeiam o tipo de bloqueio automático no período de vedação ou restrição de impulsionamento pago contemplado pelo Art. 16.

O Grupo 4 não aciona o Art. 16 como mecanismo de publicação, republicação ou impulsionamento durante o período de vedação porque a funcionalidade relevante tratada neste Plano é a geração de imagens estáticas dentro do ChatGPT, e não um serviço de publicação eleitoral, republicação, impulsionamento pago ou amplificação paga operado pela OpenAI. Quaisquer salvaguardas aplicáveis à geração de imagens, incluindo proveniência de conteúdo sintético e controles de segurança de conteúdo, são tratadas no Art. 15.

O Grupo 5 consiste em funcionalidade de publicidade. Este Grupo é tratado separadamente nos termos dos Arts. 18 e 19 da Portaria. As Políticas publicitárias da OpenAI proíbem conteúdo político em anúncios, incluindo anúncios que defendam ou se oponham a atores políticos, eleições, políticas públicas ou questões sociais controversas. Além disso, a OpenAI não permite anúncios políticos e não tem, além da funcionalidade de publicidade, programa por meio do qual terceiros paguem por visibilidade, posicionamento ou distribuição preferenciais de conteúdo. Consequentemente, a obrigação do Art. 16 de bloqueio automático de anúncios pagos não é materialmente incidente sobre o serviço central do ChatGPT nem sobre a funcionalidade de publicidade da OpenAI tratada neste Plano.

Por essas razões, a OpenAI não apresenta o Art. 16 como categoria materialmente aplicável de gestão do período de vedação para o ChatGPT ou os Grupos relacionados tratados acima. A funcionalidade que, de outra forma, poderia suscitar questões relativas ao período de vedação é tratada nas seções do Plano correspondentes ao seu papel real: ChatGPT e salvaguardas de IA generativa nos termos do Art. 15; denúncia, restrição e remoção de conteúdo ou interfaces denunciados por usuários nos termos dos Arts. 17 e 22; avaliação de riscos nos termos do Art. 20; e obrigações relativas à publicidade nos termos dos Arts. 18 e 19.

O Art. 16 também exige, quando aplicável, um mecanismo técnico para ativação automática das restrições do período de vedação e um plano de contingência para falha desse mecanismo. Como a obrigação subjacente do Art. 16 não é materialmente incidente sobre os

serviços e funcionalidades tratados neste Plano, não é apresentado nenhum mecanismo específico de ativação automática ou plano de contingência relativo ao Art. 16.

7.2. Consequências para os requisitos do Art. 7º

Como o Art. 16 da Portaria não é materialmente incidente sobre os serviços e funcionalidades tratados neste Plano, nenhuma medida específica de gestão do período de vedação nos termos do Art. 16 é apresentada para fins do Art. 7º da Portaria. Os requisitos do Art. 7º relativos a prazo, resultado esperado e trajetória de alcance aplicam-se às medidas de conformidade adotadas para cumprir uma obrigação aplicável. Quando a obrigação subjacente não é materialmente incidente sobre a funcionalidade do serviço relevante, como neste caso, não há medidas específicas para as quais um prazo de implementação, resultado esperado ou trajetória de alcance separados seriam significativos. Essa conclusão decorre da análise baseada em funcionalidades da Seção 7.1 acima e também é respaldada pelo Art. 4º, § 2º, da Portaria.

8. TRANSPARÊNCIA E INTEGRIDADE DA PUBLICIDADE POLÍTICA PAGA (ARTS. 18 E 19, § 2º)

8.1. Escopo e aplicabilidade (Art. 4º, § 2º)

O Art. 18 da Portaria dispõe sobre as obrigações de transparência e integridade aplicáveis a provedores que oferecem publicidade político-eleitoral paga, impulsionamento, promoção paga, priorização paga ou outros mecanismos de ampliação paga de alcance. Essas obrigações incluem, quando aplicável, declarações obrigatórias de uso de IA no fluxo de contratação da publicidade, bibliotecas de anúncios políticos, identificação de anunciantes, verificação de anunciantes, critérios para aprovação, recusa, suspensão ou remoção de publicidade político-eleitoral paga e critérios de priorização paga em buscas.

Para os fins deste Plano, a análise do Art. 18 é conduzida com referência exclusiva ao **Grupo 5** – funcionalidade de publicidade, conforme definida na introdução do Plano. Isso é metodologicamente apropriado porque o Art. 18 dirige-se à disponibilidade, contratação, veiculação, transparência e integridade de publicidade político-eleitoral paga ou de outros mecanismos pagos de amplificação ou priorização.

Consequentemente, a questão relevante para esta Seção 8 é saber se a OpenAI oferece publicidade política, impulsionamento político-eleitoral, priorização política paga, conteúdo político-eleitoral patrocinado ou outro mecanismo de alcance pago que acionaria as obrigações do Art. 18.

A OpenAI não oferece publicidade política paga. Conforme previsto nas Políticas publicitárias da OpenAI,[72] o conteúdo político é atualmente proibido em anúncios, incluindo

[72] Disponível em: https://openai.com/pt-BR/policies/ad-policies/. Acessado em 14 de agosto de 2026.

anúncios que defendam ou se oponham a atores políticos, eleições, políticas públicas ou questões sociais controversas. As Políticas publicitárias também proíbem anúncios relacionados a eleições ou participação eleitoral, incluindo apoio ou oposição a um candidato, partido político ou referendo, ou incentivo ou desencorajamento do voto. As Políticas publicitárias estabelecem:

> *Conteúdo político. Atualmente, são proibidos anúncios que defendam ou se oponham a atores políticos, eleições, políticas públicas ou questões sociais controversas. Isso inclui publicidade relacionada a eleições ou participação eleitoral (por exemplo, apoio ou oposição a um candidato, partido político, referendo, ou incentivo ou desencorajamento ao voto); defesa de causas ligadas a autoridades públicas ou ações governamentais (por exemplo, fiscalização da imigração, impostos, política climática ou outras questões legislativas ou regulatórias); e publicidade que enquadra questões sociais controversas como assuntos de controvérsia pública ou debate político.*

Essa proibição não se limita ao texto de um anúncio. As Políticas publicitárias da OpenAI estabelecem que os padrões de publicidade se aplicam a todos os elementos do anúncio, incluindo texto do anúncio, imagens, vídeos e páginas de destino. Também estabelecem que os anunciantes devem cumprir requisitos de identidade, integridade do destino, confiabilidade do anunciante e conformidade geográfica, e que os anunciantes não podem abusar da plataforma de publicidade da OpenAI nem tentar contornar as políticas ou os processos de aplicação da OpenAI.

A OpenAI também mantém mecanismos de análise, monitoramento e aplicação concebidos para reduzir a probabilidade de veiculação de anúncios que violem essas políticas. As Políticas publicitárias da OpenAI descrevem um processo de análise em três níveis que abrange: **(i)** anunciantes, incluindo elegibilidade, identidade, qualidade da conta e sinais de risco associados a golpes, fraude, abuso ou conduta enganosa; **(ii)** criativos de anúncios e páginas de destino, incluindo análise de títulos, texto, mídia e destinos quanto à conformidade com as políticas; e **(iii)** posicionamento, incluindo controles concebidos para assegurar que anúncios aprovados apareçam apenas em conversas que estejam em conformidade com a política de posicionamento da OpenAI.

A maioria das análises é realizada por meio de sistemas automatizados com supervisão e calibração humanas, incluindo sistemas de aprendizado de máquina, modelos de linguagem grandes e classificadores que analisam anúncios, páginas de destino e sinais de anunciantes antes que os anúncios possam ser veiculados. Algumas decisões podem ser encaminhadas para análise humana com base na gravidade, no nível de confiança ou no risco potencial. Anúncios ou páginas de destino que não possam ser analisados ou avaliados pelos sistemas da OpenAI não são autorizados a ser veiculados.

O processo de análise também continua após a aprovação. A OpenAI monitora continuamente anúncios e atividades de anunciantes utilizando feedback dos usuários,

métricas automatizadas e avaliações contínuas de qualidade. Quando a OpenAI detecta sinais de que um anúncio pode ser inseguro, enganoso ou de outra forma não conforme, a OpenAI pode limitar a veiculação, solicitar análise adicional, remover o anúncio ou adotar medidas mais amplas no nível do anunciante, conforme apropriado. Os usuários também podem denunciar anúncios que considerem inseguros, enganosos ou problemáticos, e violações graves, reiteradas ou enganosas podem resultar em medidas de aplicação das políticas mais severas, incluindo suspensão ou encerramento das contas de anunciantes.

Esses mecanismos de políticas, análise, monitoramento e aplicação de políticas sustentam a posição da OpenAI de que publicidade político-eleitoral paga, impulsionamento, priorização paga, conteúdo político-eleitoral patrocinado e outra amplificação político-eleitoral paga não são oferecidos nos serviços tratados por este Plano. Embora a OpenAI possa disponibilizar funcionalidade de publicidade em contextos limitados do ChatGPT, essa funcionalidade não é oferecida como publicidade político-eleitoral paga nem como mecanismo pago de impulsionamento, promoção ou priorização político-eleitoral.

Com base nisso, as obrigações do Art. 18 relativas a fluxos de contratação de publicidade política, declarações obrigatórias de uso de IA em publicidade política paga, bibliotecas de anúncios políticos, verificação de anunciantes políticos, identificação de anunciantes políticos, critérios de aprovação de publicidade política paga, critérios de suspensão/remoção de publicidade política paga e priorização política paga em buscas não são materialmente incidentes sobre os serviços e funcionalidades tratados neste Plano.

Essa conclusão de não aplicabilidade não depende da ausência de toda e qualquer funcionalidade de publicidade. Depende, isto sim, da ausência de publicidade político-eleitoral paga, impulsionamento, promoção paga, priorização paga ou outra ampliação paga de conteúdo político-eleitoral nos serviços tratados neste Plano. Controles gerais de publicidade, padrões de elegibilidade de anunciantes, análise de conteúdo de anúncios, controles de posicionamento, processos de monitoramento e aplicação de políticas continuam fazendo parte do quadro de integridade da publicidade da OpenAI, mas não transformam os serviços tratados neste Plano em serviços de publicidade político-eleitoral paga para fins do Art. 18.

Diante disso, a OpenAI não apresenta o Art. 18 como categoria materialmente aplicável de obrigações para os serviços tratados neste Plano. O tratamento da publicidade neste Plano limita-se a: (i) documentar a não oferta de publicidade político-eleitoral paga ou ampliação político-eleitoral paga; (ii) distinguir funcionalidade geral de publicidade de publicidade política paga; e (iii) identificar o fluxo residual exigido pelo Art. 19, § 2º, da Portaria, conforme descrito a seguir.

8.2. Consequências para os requisitos do Art. 7º

Considerando que a OpenAI não oferece publicidade político-eleitoral paga, impulsionamento, promoção paga, priorização paga ou outra ampliação paga de conteúdo

político-eleitoral nos serviços tratados neste Plano, nenhuma medida de publicidade política paga específica do Art. 18 é apresentada para fins do Art. 7º da Portaria. Os requisitos do Art. 7º relativos a prazo, resultado esperado e trajetória de alcance aplicam-se às medidas de conformidade adotadas para cumprir uma obrigação aplicável. Quando a obrigação subjacente do Art. 18 não é materialmente incidente sobre a funcionalidade do serviço relevante, como neste caso, não há medidas de publicidade política paga para as quais um prazo de implementação, resultado esperado ou trajetória de alcance separados seriam significativos.

Embora a OpenAI não ofereça publicidade político-eleitoral paga, impulsionamento, priorização paga, conteúdo político-eleitoral patrocinado ou outra ampliação paga de conteúdo político-eleitoral nos serviços tratados neste Plano, o Art. 19, § 2º, da Portaria exige que um provedor que não ofereça serviços de impulsionamento político-eleitoral, incluindo priorização paga de resultados de busca, identifique um fluxo e um canal para responder a requisições do TSE relativas a anúncios ou impulsionamentos que possam ter sido exibidos durante o período eleitoral e às informações correspondentes sobre anunciantes mantidas pelo provedor. Para esse fim, a OpenAI tratará qualquer requisição do TSE por meio do seguinte fluxo residual de resposta, na medida aplicável e na medida em que os registros correspondentes sejam mantidos pela OpenAI:

Elemento	Descrição
Escopo das requisições	Requisições do TSE relativas a supostos anúncios, impulsionamentos, promoções pagas, priorizações pagas ou outras amplificações pagas exibidos durante o período eleitoral, e informações relacionadas sobre anunciantes mantidas pela OpenAI.
Canal designado	O canal designado pela OpenAI para solicitações regulatórias ou governamentais relativas a comunicações com o TSE é o Kodex. O TSE precisa ter uma conta para acessar o sistema no link <https://app.kodexglobal.com/signin> e fazer as requisições. O processo é simples, mas a OpenAI está à disposição do TSE para apoiar a criação da conta.
Recebimento inicial	Após o recebimento, a OpenAI avaliará se a requisição diz respeito a algum anúncio, anunciante, posicionamento, mecanismo de alcance pago ou registro relacionado mantido pela OpenAI.
Encaminhamento interno	As solicitações serão encaminhadas à(s) função(ões) interna(s) relevante(s), que podem incluir as equipes [RESTRITO], dependendo da natureza da requisição.
Busca de registros	A OpenAI identificará se existem registros responsivos no âmbito da sua funcionalidade de publicidade, incluindo, quando aplicável, registros de anúncios, anunciantes ou posicionamento, conforme aplicável.
Resposta ao TSE	A OpenAI responderá dentro do prazo especificado pelo TSE na requisição aplicável e fornecerá as informações legalmente exigidas disponíveis à OpenAI, na medida em que tais informações sejam mantidas nos registros da OpenAI e estejam dentro do escopo da solicitação, sujeitas às classificações legais, de confidencialidade e de camadas de informação aplicáveis.

Elemento	Descrição
Prazo	Durante o ciclo eleitoral, conforme e quando as requisições do TSE forem recebidas.
Monitoramento / base das evidências	Registros de recebimento das requisições, registros de encaminhamento interno, buscas de registros de anunciantes ou anúncios quando aplicável, registros de resposta e quaisquer informações fornecidas ao TSE em conexão com a requisição.

9. AVALIAÇÃO DE IMPACTO, APURAÇÃO DE FALHAS E MITIGAÇÃO DE RISCOS (ART. 20)

9.1. Escopo e aplicabilidade (Art. 4º, § 2º)

Esta Seção 9 trata das obrigações do Art. 20 da Portaria relativas ao dever de avaliar impactos sobre a integridade do processo eleitoral, aplicável aos provedores de aplicações de internet que permitam a disseminação de conteúdo político e eleitoral. Para os Grupos definidos na introdução do Plano, o Art. 20 aplica-se (ou não se aplica) da seguinte forma:

O Grupo 3 é o grupo para o qual o Art. 20 é particularmente relevante, pois inclui recursos que podem permitir que conteúdo gerado ou configurado por um usuário se torne visível ou acessível a outros usuários ou ao público. Embora a OpenAI não forneça uma plataforma de distribuição de conteúdo e suas funcionalidades não a transformem em uma rede social ou serviço de disseminação de conteúdo político, elas podem criar acessibilidade pública e possibilidade de descoberta. Por essa razão, a OpenAI incluiu este Grupo no escopo da avaliação de impacto do Art. 20, tratando dos riscos que o conteúdo acessível publicamente pode representar para a integridade eleitoral.

Os Grupos 1 e 4 são tratados principalmente no âmbito das salvaguardas do Art. 15 descritas na **Seção 6** deste Plano, que incluem controles de neutralidade política, sistemas de segurança de conteúdo, mecanismos de prevenção de contorno, identificação de conteúdo sintético e testes de resistência. Esses Grupos não permitem a disseminação de conteúdo político-eleitoral da forma contemplada pelo Art. 20. Não obstante, os riscos associados à geração de conteúdo político-eleitoral por meio de funcionalidades próprias do ChatGPT, incluindo o potencial de os resultados serem posteriormente compartilhados ou utilizados fora do ambiente do ChatGPT, foram considerados na avaliação de impacto conduzida nos termos desta **Seção 9**. As salvaguardas descritas na **Seção 6** operam como a principal camada de mitigação de riscos para esses Grupos.

Os Grupos 2 e 6 envolvem riscos principalmente atribuíveis a atores terceiros, e não aos próprios serviços da OpenAI. Conforme descrito na introdução do Plano, aplicativos de terceiros que incorporam a API não são implementados pela OpenAI, e a OpenAI não controla as funcionalidades que desenvolvedores subjacentes, clientes ou administradores empresariais possam adicionar, configurar ou alterar. Da mesma forma, o conteúdo de terceiros apresentado na interface do ChatGPT se origina da fonte terceira e permanece sob

responsabilidade desta, e não da OpenAI. Para evitar dúvidas, a responsabilidade primária pela conformidade com as regulamentações eleitorais em aplicativos a jusante recai sobre os desenvolvedores e operadores terceiros desses aplicativos.

O Grupo 5 não está no escopo da avaliação de impacto do Art. 20 porque a OpenAI não oferece publicidade político-eleitoral paga, conforme descrito na Seção 8 deste Plano.

Observe-se, contudo, que o parágrafo único do Art. 20 não é aplicável a nenhum dos serviços da OpenAI. Isso ocorre porque a obrigação de avaliação de impacto ali contemplada pressupõe uma arquitetura de serviço capaz de permitir a "propagação acelerada" de conteúdo de extraordinária gravidade durante momentos eleitorais sensíveis. Essa pressuposição não é atendida por nenhum dos Grupos de serviços tratados neste Plano. **O Grupo 1** consiste em funcionalidades próprias do ChatGPT que operam como um ambiente conversacional privado e baseado em contas, no qual o conteúdo é gerado em resposta a prompts individuais dos usuários e não é publicado, disseminado, recomendado ou impulsionado para outros usuários ou para o público pela OpenAI. **O Grupo 2** consiste em integrações de terceiros que apresentam conteúdo de fontes externas, cujas dinâmicas de propagação são controladas pela fonte terceira, e não pela OpenAI. **O Grupo 4** diz respeito à geração de imagens dentro do ChatGPT e não opera como um mecanismo de distribuição de conteúdo. **O Grupo 5** proíbe integralmente conteúdo político em publicidade e não oferece um canal de amplificação paga. Nenhum desses Grupos permite, portanto, o tipo de propagação de conteúdo em toda a plataforma, em rede ou algorítmica para públicos de massa que poderia resultar no cenário de "propagação acelerada" contemplado pelo parágrafo único do Art. 20.

O Grupo 3, que inclui interfaces públicas ou de compartilhamento limitadas, é o único Grupo que poderia suscitar uma questão limítrofe nos termos do parágrafo único do Art. 20, porque determinado conteúdo pode se tornar acessível a terceiros por meio dessas interfaces. Contudo, essas interfaces não transformam o ChatGPT em uma rede social de uso geral, feed público, mecanismo de recomendação de conteúdo ou plataforma de distribuição viral. Conforme mencionado anteriormente na Seção 7.1, os links compartilhados são links iniciados pelo usuário para conversas específicas do ChatGPT; os GPTs públicos e as interfaces do GPT Store dizem respeito ao acesso a configurações de GPTs dentro do ChatGPT; e o ChatGPT Sites opera como páginas hospedadas limitadas ou aplicativos leves pelos quais o usuário divulgador permanece responsável pelo Site e pelo conteúdo enviado pelos visitantes.

Consequentemente, a obrigação de incluir na avaliação de impacto uma análise da capacidade do provedor de intervir sobre conteúdo de extraordinária gravidade em propagação acelerada durante momentos eleitorais sensíveis não é incidente sobre os serviços e funcionalidades tratados neste Plano. Essa conclusão não diminui o compromisso da OpenAI de responder às comunicações e ordens judiciais do TSE por meio dos canais descritos nas Seções 3 e 8 deste Plano, inclusive durante os fins de semana eleitorais.

9.2. Metodologia e frequência da avaliação de impacto sobre a integridade eleitoral

Esta Seção descreve a metodologia e a periodicidade para conduzir a Avaliação de Impacto sobre a Integridade Eleitoral da OpenAI, concebida para identificar e avaliar sistematicamente os riscos relacionados a eleições decorrentes dos produtos e interfaces da OpenAI, neles presentes ou a eles relacionados. A avaliação abrange os domínios da segurança do usuário e do público, jurídico e regulatório e é concebida, entre outras finalidades, para cumprir as regulamentações aplicáveis de integridade eleitoral, incluindo o requisito mínimo de conduzir a avaliação em todos os anos eleitorais.

Elemento	Descrição
Metodologia de avaliação	[RESTRITO]
Frequência	[RESTRITO]
Prazo	[RESTRITO]
Equipe responsável	[RESTRITO]
Contribuição externa	[RESTRITO]

9.3. Matriz de riscos

[RESTRITO]

[RESTRITO]	[RESTRITO]	[RESTRITO][73]	[RESTRITO][74]	[RESTRITO][75]	[RESTRITO]
[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]
[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]
[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]
[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]
[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]

[73] Probabilidade de que o risco se materializasse durante o ciclo eleitoral, caso não fossem adotadas as medidas de mitigação.
[74] Magnitude do dano à integridade eleitoral se o risco se materializasse, na ausência de medidas mitigadoras.
[75] As categorias de alcance refletem o potencial de disseminação de um incidente, sem constituir limites numéricos rígidos. A classificação considera o número de usuários potencialmente expostos, o número de jurisdições ou mercados afetados, a velocidade e facilidade de disseminação e a possibilidade de conter sua propagação

[RESTRITO]	[RESTRITO]	[RESTRITO[73]	[RESTRITO][74]	[RESTRITO][75]	[RESTRITO]
[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]
[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]
[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]
[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]
[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]
[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]
[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]	[RESTRITO]

[RESTRITO]

10. PROTEÇÃO DE DADOS PESSOAIS (ART. 21)

10.1. Escopo e aplicabilidade (Art. 4º, § 2º)

O Art. 21 da Portaria dispõe sobre os deveres de proteção de dados pessoais especificamente relacionados às obrigações previstas nos Arts. 33-A e 33-B da Resolução. Esses deveres dizem respeito a: mecanismos para informar os titulares sobre o tratamento de dados pessoais para fins de propaganda eleitoral, incluindo perfilamento e microdirecionamento; procedimentos para exercício dos direitos estabelecidos pela Lei Geral de Proteção de Dados Pessoais ("LGPD"); mecanismos para assegurar limitação de finalidade, necessidade, adequação e tratamento não discriminatório para esses fins; procedimentos de segurança e notificação de incidentes nos termos da LGPD; e, quando aplicável, mecanismos de consentimento para o tratamento de dados pessoais sensíveis.

O gatilho relevante para o Art. 21, portanto, não é o mero tratamento de dados pessoais em conexão com um serviço digital. O Art. 21 aplica-se especificamente ao tratamento de dados pessoais para fins de propaganda eleitoral, incluindo perfilamento e microdirecionamento. A Resolução define perfilamento como "*tratamento de múltiplos tipos de dados de pessoa natural, identificada ou identificável, em geral realizado de modo automatizado, com o objetivo de formar perfis baseados em padrões de comportamento, gostos, hábitos e preferências e de classificar esses perfis em grupos e setores, utilizando-os para análises ou previsões de movimentos e tendências de interesse político-eleitoral*" (grifos adicionados). A Resolução define microdirecionamento como "*estratégia de segmentação da propaganda eleitoral ou da comunicação de campanha que consiste em selecionar pessoas, grupos ou*

setores, classificados por meio de perfilamento, como público-alvo ou audiência de mensagens, ações e conteúdos político-eleitorais desenvolvidos com base nos interesses perfilados, visando ampliar a influência sobre seu comportamento" (grifos adicionados).

Nos termos do Art. 4º, § 2º, da Portaria, as obrigações previstas nos Arts. 11 a 23 aplicam-se apenas quando materialmente incidentes nas funcionalidades efetivamente disponibilizadas pelo provedor. Com base nisso, o Art. 21 não é materialmente incidente sobre o ChatGPT nem sobre os Grupos relacionados tratados neste Plano enquanto categoria de obrigação de tratamento de dados para propaganda eleitoral, perfilamento ou microdirecionamento, pelas razões apresentadas abaixo.

O Grupo 1 abrange funcionalidades próprias do ChatGPT. Embora o ChatGPT possa envolver tratamento de dados pessoais para finalidades gerais como prestação do serviço, segurança, melhoria e personalização da experiência do usuário, conforme descrito na Política de Privacidade da OpenAI, essas atividades não constituem propaganda eleitoral, perfilamento ou microdirecionamento para fins do Art. 21.

O Grupo 2 abrange integrações de terceiros. Na medida em que integrações de terceiros recuperem, exibam ou apresentem conteúdo de terceiros ou envolvam práticas de dados de terceiros, essas atividades não são tratadas neste Plano como tratamento de dados pessoais pela OpenAI para propaganda eleitoral, perfilamento ou microdirecionamento.

O Grupo 3 abrange interfaces públicas ou de compartilhamento limitadas. Essas funcionalidades podem criar questões de acessibilidade pública, descoberta, denúncia, restrição, remoção ou controle de acesso, mas não são funcionalidades de direcionamento de propaganda eleitoral, perfilamento ou microdirecionamento.

O Grupo 4 abrange recursos de geração de imagens dentro do ChatGPT. Esses recursos são tratados no Art. 15 em relação à geração de conteúdo sintético, à proveniência, às restrições de semelhança, aos controles de imagens íntimas não consentidas e a outros riscos relacionados a mídias geradas por IA; não são serviços de perfilamento ou microdirecionamento para propaganda eleitoral.

O Grupo 5 abrange a funcionalidade de publicidade. As Políticas publicitárias da OpenAI proíbem conteúdo político em anúncios, e os serviços tratados neste Plano não oferecem publicidade político-eleitoral paga ou amplificação política paga. Portanto, o Grupo 5 não é tratado como acionador dos deveres do Art. 21 relativos ao tratamento de dados para propaganda eleitoral, perfilamento ou microdirecionamento.

Por essas razões, a OpenAI não apresenta o Art. 21 como categoria materialmente aplicável de obrigações para o ChatGPT ou os Grupos relacionados tratados acima. As medidas gerais de privacidade, os controles dos usuários, os procedimentos relativos aos direitos dos titulares, as medidas de segurança e as práticas de divulgação lícita descritas na Política de Privacidade da OpenAI permanecem aplicáveis como parte do quadro de privacidade mais amplo da OpenAI. Essas medidas gerais de privacidade, contudo, não transformam os

serviços tratados neste Plano em tratamento de dados pessoais para propaganda eleitoral, perfilamento ou microdirecionamento nos termos do Art. 21 da Portaria.

10.2. Consequências para os requisitos do Art. 7º

Considerando que o Art. 21 da Portaria não é materialmente incidente sobre os serviços e funcionalidades tratados neste Plano, nenhuma medida específica de proteção de dados pessoais para propaganda eleitoral, perfilamento ou microdirecionamento nos termos do Art. 21 é apresentada para fins do Art. 7º da Portaria. Os requisitos do Art. 7º relativos a prazo, resultado esperado e trajetória de alcance aplicam-se às medidas de conformidade adotadas para cumprir uma obrigação aplicável. Quando a obrigação subjacente do Art. 21 não é materialmente incidente sobre a funcionalidade do serviço relevante, como neste caso, não há medidas específicas para as quais um prazo de implementação, resultado esperado ou trajetória de alcance separados seriam significativos. Essa conclusão decorre da análise baseada em funcionalidades da Seção 10.1 acima e também é respaldada pelo Art. 4º, § 2º, da Portaria..

11. POLÍTICAS DE USO E PREVENÇÃO DE ABUSO POR TERCEIROS (ART. 22)

11.1. Escopo e aplicabilidade (Art. 4º, § 2º)

O Art. 22 da Portaria dispõe sobre os deveres relativos à elaboração e aplicação de políticas de conteúdo político-eleitoral, conforme referido no Art. 9º-D, I, da Resolução. Para os fins deste Plano, o Art. 22 é tratado como obrigação de governança de políticas e prevenção de abuso por terceiros. A OpenAI atende a essa obrigação por meio dos termos, políticas, especificações, orientações públicas, canais de denúncia, processos de análise e mecanismos de aplicação das políticas que regem o uso do ChatGPT e de serviços relacionados da OpenAI.

11.2. Instrumentos normativos que regem o uso político-eleitoral

Os instrumentos a seguir regem ou orientam o uso dos serviços da OpenAI em contextos relevantes para conteúdo político-eleitoral. Os instrumentos contratuais aplicam-se quando incorporados à relação aplicável com o cliente ou usuário. O Model Spec e determinados materiais da Central de Ajuda ou materiais explicativos públicos são materiais de implementação, explicativos ou de orientação de políticas, e não constituem, por si só, contratos com clientes.

Instrumento	Data de vigência / última atualização	Status / função	Relevância para o Art. 22
Termos de Uso[76]	Vigente a partir de 1º de janeiro de 2026	Instrumento contratual para serviços voltados ao consumidor fora do EEE e da Suíça.	Incorpora as Políticas de Uso da OpenAI, exige conformidade com a legislação aplicável e proíbe apresentar resultados de IA como gerados por humanos.
Contrato de Prestação de Serviços[77]	Vigente a partir de 1º de janeiro de 2026	Instrumento contratual para serviços empresariais e de API.	Incorpora os Termos de Serviço, as Políticas de Uso e a Política de Compartilhamento e Publicação para relações aplicáveis com clientes empresariais e de API.
Termos de Serviço[78]	Atualizados em 12 de junho de 2026	Termos específicos do produto.	Inclui disposições específicas do produto, incluindo termos relevantes para GPTs personalizados e outros serviços abrangidos pelos termos aplicáveis.
Políticas de Uso da OpenAI[79]	Vigentes a partir de 29 de outubro de 2025	Instrumento de políticas aplicável a todos os serviços da OpenAI.	Proíbe interferência eleitoral, campanhas políticas/lobby, engano, uso não autorizado de semelhança, conteúdo íntimo não consentido, ameaças/assédio, contorno das salvaguardas e outros usos proibidos relevantes para a integridade eleitoral.
Política de Compartilhamento e Publicação[80]	Atualizada em 14 de novembro de 2022	Instrumento de políticas que rege o compartilhamento e a publicação pública abrangidos.	Exige divulgação clara do envolvimento de IA para compartilhamento público e publicação assistida pela API abrangidos.
Model Spec[81]	Versão datada de 18 de dezembro de 2025	Especificação de comportamento do modelo e orientação de implementação.	Descreve o comportamento pretendido do modelo, incluindo restrições à manipulação política direcionada e um padrão de neutralidade/objetividade em contextos nos quais a objetividade é esperada.

[76] Disponível em: https://openai.com/pt-BR/policies/row-terms-of-use/. Acessado em 14 de agosto de 2026.
[77] Disponível em: https://openai.com/pt-BR/policies/services-agreement/. Acessado em 14 de agosto de 2026.
[78] Disponível em: https://openai.com/pt-BR/policies/service-terms/. Acessado em 14 de agosto de 2026.
[79] Disponível em: https://openai.com/pt-BR/policies/usage-policies/. Acessado em 14 de agosto de 2026.
[80] Disponível em: https://openai.com/pt-BR/policies/sharing-publication-policy/. Acessado em 14 de agosto de 2026.
[81] Disponível em: https://model-spec.openai.com/2025-12-18.html. Acessado em 14 de agosto de 2026.

Instrumento	Data de vigência / última atualização	Status / função	Relevância para o Art. 22
Políticas publicitárias[82]	Atualizadas em 15 de julho de 2026 / 10 de agosto de 2026	Instrumento de políticas publicitárias.	Proíbe publicidade política e exclui conteúdo político do posicionamento de anúncios.
Termos para Desenvolvedores de Aplicativos[83]	Atualizados em 9 de julho de 2026	Termos condicionais para aplicativos personalizados, Conectores, plug-ins, GPT Actions[84] e funcionalidades relacionadas de desenvolvedores.	Exige conformidade com a legislação e as políticas da OpenAI e exige os avisos e divulgações necessários aos usuários para os aplicativos abrangidos.
Restrições a campanhas políticas[85]	Atualizadas em 15 de julho de 2026	Orientações da Central de Ajuda que explicam as Políticas de Uso da OpenAI no contexto político e de campanhas.	Explica campanhas políticas proibidas, bots de campanha voltados ao público, direcionamento de eleitores/públicos, contato automatizado, mensagens de campanha em escala, falsos apoios, mídia sintética enganosa e prática de astroturfing.
Informações eleitorais e salvaguardas em 2026[86]	Publicadas em 27 de maio de 2026	Descrição pública das medidas de integridade eleitoral e das salvaguardas de produto da OpenAI.	Descreve a abordagem da OpenAI para informações eleitorais confiáveis, transparência de IA, prevenção de uso indevido, aplicação contra interferência eleitoral e desmobilização e monitoramento de viés político.
Transparência e moderação de conteúdo[87]	Atualizadas em 29 de julho de 2026	Descrição pública dos mecanismos de detecção, análise, aplicação de políticas, recurso e melhoria contínua da OpenAI.	Descreve tecnologias automatizadas, análise humana, detecção proativa, denúncias de usuários, medidas de aplicação das políticas, recursos, controles de visibilidade de GPTs, restrições de compartilhamento de conteúdo e melhoria contínua.

[82] Disponível em: https://openai.com/pt-BR/policies/ad-policies/. Acessado em 14 de agosto de 2026.
[83] Disponível em: https://openai.com/pt-BR/policies/developer-apps-terms/. Acessado em 14 de agosto de 2026.
[84] Integrações configuradas no GPT *builder* que permitem a um GPT conectar-se e utilizar APIs externas definidas pelo próprio criador. Cada Ação (Action) utiliza configurações de autenticação e um esquema OpenAPI que descreve as operações disponíveis.
[85] Disponível em: https://help.openai.com/en/articles/20001255-political-campaigning-restrictions. Acessado em 14 de agosto de 2026.
[86] Disponível em: https://openai.com/index/election-safeguards-2026/. Acessado em 14 de agosto de 2026.
[87] Disponível em: https://openai.com/pt-BR/transparency-and-content-moderation/. Acessado em 14 de agosto de 2026.

Instrumento	Data de vigência / última atualização	Status / função	Relevância para o Art. 22
Denúncia de conteúdo no ChatGPT e nas plataformas da OpenAI[88]	Atualizadas em agosto de 2026	Orientações da Central de Ajuda sobre denúncia de conteúdo ou atividade.	Descreve canais para denunciar conversas do ChatGPT, links compartilhados, respostas, GPTs, anúncios, publicações em fóruns e ChatGPT Sites que possam violar os Termos de Uso da OpenAI ou a legislação aplicável.
Como identificamos conteúdo problemático em nossos serviços para pessoas físicas[89]	Atualizadas em agosto de 2026	Orientações da Central de Ajuda sobre detecção e aplicação para ChatGPT, ImageGen e GPTs.	Descreve ferramentas automatizadas, denúncias humanas, equipes treinadas de análise, advertências, bloqueio de respostas do modelo, prevenção de compartilhamento, restrições de distribuição de GPTs, recursos e medidas limitadas de conta.
Entenda suas responsabilidades em relação aos seus ChatGPT Sites[90]	Atualizados em julho/agosto de 2026	Orientações da Central de Ajuda para usuários divulgadores de ChatGPT Sites.	Explica a responsabilidade do usuário divulgador pelos Sites e pelo conteúdo enviado por visitantes, a conformidade com as políticas, a denúncia de Sites e a capacidade da OpenAI de remover ou restringir Sites quando exigido por políticas, legislação ou considerações de segurança da plataforma.

11.3. Conformidade com os requisitos do Art. 22

A tabela abaixo identifica como cada requisito específico do Art. 22 da Portaria é tratado neste Plano.

Dispositivo do Art. 22	Requisito	Tratamento neste Plano
Art. 22, I	Identificar os instrumentos normativos que regem, no âmbito do serviço, o uso da	Tratado na Seção 11.2. A OpenAI identifica os instrumentos contratuais, as políticas, as especificações e as orientações públicas que regem o uso do ChatGPT

[88] Disponível em: https://help.openai.com/pt-br/articles/10245791-como-denunciar-conte%C3%BAdo-no-chatgpt-e-nas-plataformas-da-openai. Acessado em 14 de agosto de 2026.

[89] Disponível em: https://help.openai.com/pt-br/articles/8940831-como-identificamos-conte%C3%BAdo-problem%C3%A1tico-em-nossos-servi%C3%A7os-para-pessoas-f%C3%ADsicas . Acessado em 14 de agosto de 2026.

[90] Disponível em: https://help.openai.com/pt-br/articles/20001337-entenda-suas-responsabilidades-em-rela%C3%A7%C3%A3o-aos-seus-chatgpt-sites. Acessado em 14 de agosto de 2026.

	plataforma e o tratamento de conteúdo político-eleitoral por terceiros, incluindo termos de serviço, termos de uso, padrões comunitários, políticas de uso aceitável ou documentos equivalentes, e demonstrar compatibilidade com a legislação eleitoral.	e de serviços relacionados da OpenAI em contextos relevantes para conteúdo político-eleitoral.
Art. 22, I	Demonstrar compatibilidade entre os instrumentos normativos e o quadro jurídico eleitoral.	Os instrumentos identificados na Seção 11.2 contêm disposições que apoiam objetivos de integridade eleitoral, incluindo restrições a campanhas políticas, interferência eleitoral, desmobilização, conteúdo enganoso gerado por IA, manipulação política direcionada, uso não autorizado de semelhança, conteúdo íntimo não consentido, ameaças/assédio, contorno das salvaguardas e publicidade política.
Art. 22, II	Descrever a arquitetura de moderação de conteúdo do provedor e as medidas disponíveis para responder a violações ou riscos, incluindo remoção, bloqueio, redução de visibilidade, rotulagem, marcação, contextualização, limitação de compartilhamento, notas da comunidade, desmonetização e medidas equivalentes.	Conforme explicado neste Plano, a OpenAI não opera uma arquitetura de moderação semelhante à de uma rede social. Para o ChatGPT, controles análogos incluem recusas do modelo, respostas seguras, bloqueio de prompts ou resultados, advertências, restrições de compartilhamento de conteúdo, restrições de visibilidade ou da Store de GPTs, canais de denúncia, restrições de conta ou de acesso, recursos e outras medidas de aplicação das políticas específicas do produto. Medidas como notas da comunidade, redução de visibilidade no feed e desmonetização não são apresentadas como medidas de moderação do ChatGPT porque não correspondem à funcionalidade do ChatGPT tratada neste Plano.
Art. 22, II	Descrever os critérios técnicos, legais e de proporcionalidade que orientam a escolha da medida aplicável e as salvaguardas contra intervenção indevida em conteúdo lícito.	Os materiais públicos da OpenAI descrevem fatores e mecanismos de alto nível para aplicação, incluindo requisitos legais, gravidade, violações reiteradas, denúncias de usuários, análise humana quando apropriado, recursos e reavaliação. O Plano também se baseia na distinção entre usos políticos proibidos e trabalho cívico ou de campanha responsável, dirigido por humanos e permitido, conforme descrito nas Políticas de Uso e nas Restrições a campanhas políticas da OpenAI.
Art. 22, III	Descrever procedimentos, fluxos de decisão, estruturas de governança e mecanismos de supervisão para aplicação efetiva das regras de uso.	Os materiais públicos da OpenAI descrevem denúncias, análise, aplicação, recursos e melhoria contínua em nível funcional. Podem ser apresentadas denúncias relativas a conversas do ChatGPT, links compartilhados, respostas, GPTs, anúncios, fóruns e ChatGPT Sites. Determinado conteúdo ou atividade denunciado pode

		ser analisado por sistemas automatizados e equipes treinadas de análise, e a aplicação pode incluir advertências, bloqueio, restrições de compartilhamento, restrições de visibilidade de GPTs, limitações de acesso, suspensão, desativação e recursos, quando aplicável.
Art. 22, IV	Identificar alterações materiais aos termos de uso ou às políticas de conteúdo adotadas para conformidade com a Resolução e suas alterações, incluindo redação modificada, data de vigência e método de notificação aos usuários; atualizações editoriais ou não relacionadas são excluídas.	Os instrumentos existentes já incluem restrições relevantes para eleições, incluindo restrições a campanhas políticas, interferência eleitoral, desmobilização, enganosidade, publicidade política, manipulação política direcionada e contorno das salvaguardas.
Art. 22, parágrafo único	Para a proibição de propaganda eleitoral por mensagens em massa por meio de tecnologias não fornecidas pelo provedor e em violação aos termos de uso do provedor, apresentar mecanismos disponíveis para identificar e registrar padrões anômalos ou automatizados de atividade que possam indicar o uso de ferramentas externas de mensagens em massa, dentro dos limites técnicos do serviço.	O ChatGPT não é um serviço de mensagens em massa, e os casos de uso da API e empresariais estão fora do escopo principal deste Plano. Contudo, as Políticas de Uso e as Restrições a campanhas políticas da OpenAI proíbem o uso dos serviços da OpenAI para mensagens de campanha em escala, contato automatizado de campanha, determinação de quais indivíduos devem receber determinadas mensagens, chatbots de campanha voltados ao público e conexão de conteúdo de mensagens a ferramentas de distribuição de terceiros. Na medida em que uso indevido relevante seja denunciado ou detectado em serviços controlados pela OpenAI, ele poderá ser tratado por meio dos processos aplicáveis de denúncia, análise, investigação e aplicação de políticas.
Art. 22, parágrafo único	Apresentar procedimentos de investigação e resposta aplicáveis a notificações recebidas ou sinais identificados, incluindo medidas de contenção e preservação de evidências.	Como o ChatGPT não é, por si só, um serviço de mensagens em massa, não há um fluxo específico do ChatGPT para disseminação de mensagens em massa. Denúncias ou sinais de uso indevido podem ser analisados e escalados por meio dos processos aplicáveis de denúncia, investigação e aplicação de políticas.

12. MEDIDAS DE INICIATIVA PRÓPRIA (ART. 23)

12.1. Metodologia para o planejamento de ações corretivas e preventivas

O Art. 23 da Portaria dispõe sobre as medidas de iniciativa própria pelo provedor. Para os deveres de conteúdo aberto previstos no Art. 9º-D, III, IV e VI, da Resolução, o Art. 23 exige que o Plano descreva a metodologia para planejar ações corretivas e preventivas, os

mecanismos de aprimoramento dos sistemas de recomendação e as medidas de transparência dos resultados, incluindo os critérios próprios do provedor para aferir a suficiência dessas ações. O Art. 23 também dispõe sobre a remoção de perfis falsos, automatizados ou robôs nos termos do Art. 38-A, § 2º, da Resolução, exigindo que o provedor informe se opta por exercer essa faculdade e, em caso afirmativo, descreva os critérios e procedimentos aplicáveis. Caso o provedor não exerça essa faculdade, o Art. 23 exige uma justificativa fundamentada que demonstre a compatibilidade dessa opção com a avaliação do provedor quanto à probabilidade, gravidade e potencial alcance dos riscos relevantes, bem como medidas preventivas ou de mitigação alternativas, quando aplicável.

Nos termos do Art. 4º, § 2º, da Portaria, as obrigações previstas nos Arts. 11 a 23 aplicam-se apenas quando materialmente incidentes sobre as funcionalidades efetivamente disponibilizadas pelo provedor. Com base nisso, o Art. 23 não é materialmente incidente sobre o ChatGPT nem sobre os serviços e funcionalidades relacionados tratados neste Plano como categoria autônoma de medidas de iniciativa própria relativas a deveres de conteúdo aberto, aprimoramento de sistemas de recomendação ou remoção facultativa de perfis.

Conforme demonstrado neste Plano, o ChatGPT não é operado como uma rede social de uso geral, feed público, plataforma de conteúdo aberto, sistema de recomendação de conteúdo político-eleitoral ou ambiente de disseminação pública baseado em perfis. As interfaces públicas ou de compartilhamento limitadas tratadas neste Plano, como links compartilhados, páginas públicas de GPTs, GPT Store ou interfaces públicas de descoberta de GPTs e ChatGPT Sites, não transformam o ChatGPT em uma plataforma de conteúdo aberto ou sistema de recomendação de conteúdo político-eleitoral. Essas interfaces são tratadas em outras partes deste Plano na medida em que criem questões relativas a denúncias, restrições, remoções, controle de acesso, controle de visibilidade ou aplicação de políticas.

Diante disso, a OpenAI não apresenta uma metodologia específica do Art. 23 para o planejamento de ações corretivas e preventivas relativas aos deveres de conteúdo aberto, mecanismos de aprimoramento de sistemas de recomendação, transparência dos sistemas de recomendação de conteúdo político-eleitoral ou remoção facultativa de perfis públicos falsos, automatizados ou robôs. As salvaguardas e os controles relevantes do ChatGPT são tratados no âmbito das obrigações materialmente incidentes sobre os serviços e funcionalidades efetivamente disponibilizados pela OpenAI, incluindo as salvaguardas de IA generativa nos termos do Art. 15, os canais de denúncia e interação nos termos do Art. 17, a avaliação e a mitigação de riscos nos termos do Art. 20, as Políticas de Uso e a prevenção de abuso por terceiros nos termos do Art. 22 e as obrigações relativas à publicidade nos termos dos Arts. 18 e 19.

12.2. Consequências para os requisitos do Art. 7º

Considerando que o Art. 23 não é materialmente incidente sobre os serviços e funcionalidades tratados neste Plano, nenhum quadro específico de medidas de iniciativa própria nos termos do Art. 23 é apresentado para fins do Art. 7º da Portaria. Os requisitos do Art. 7º relativos a prazo, resultado esperado e trajetória de alcance aplicam-se às medidas de conformidade adotadas para cumprir uma obrigação aplicável. Quando a obrigação subjacente do Art. 23 não é materialmente incidente sobre a funcionalidade do serviço relevante, como neste caso, não há medidas específicas para as quais um prazo de implementação, resultado esperado ou trajetória de alcance separados seriam significativos.

13. REQUISITOS PROCEDIMENTAIS E CRONOGRAMA

13.1. Obrigações de apresentação e atualização

Obrigação	Prazo/Gatilho	Referência	Status
Apresentação inicial do Plano	16 de agosto do ano eleitoral	Art. 24	Apresentado
Atualização após alteração substancial	No prazo de 48 horas após alteração substancial de risco ou de funcionalidades	Art. 24, parágrafo único	Conforme necessário
Relatório consolidado	19 de dezembro do ano eleitoral	Art. 28	Pendente
Comunicação de risco extraordinário	No prazo de 3 dias após a identificação	Art. 28, § 1º	Conforme necessário

13.2. Controle de versões

Versão	Data	Descrição das alterações	Motivo da atualização
1.0	16 de agosto de 2026	Apresentação inicial	Art. 24 - prazo inicial para apresentação

14. MAPA DE REFERÊNCIAS CRUZADAS À RESOLUÇÃO (ART. 6º)

A tabela abaixo identifica, para cada dispositivo da Resolução referido na Portaria, a seção deste Plano em que a respectiva obrigação é tratada, seja por meio da descrição das medidas adotadas para assegurar o cumprimento da Resolução, seja por meio de justificativa

fundamentada quanto à sua não aplicabilidade. A análise detalhada de cada obrigação encontra-se na seção correspondente.

Dispositivo da Resolução	Seção do Plano
Art. 9º-B (Rotulagem de conteúdo sintético)	Seção 6.5
Art. 9º-C (Proibição de conteúdo fabricado ou manipulado para desinformação)	Seções 6.2, 6.3 e 6.4
Art. 9º-D (Deveres de diligência do provedor)	Seções 3, 6, 9, 10 e 11
Art. 9º-E (Responsabilidade solidária pela indisponibilização imediata)	Seções 3, 6.2 e 6.3
Art. 27-A (Repositório de anúncios e transparência)	Seção 8
Art. 28 (Propaganda na Internet - quadro geral)	Seções 6, 8 e 11
Art. 29 (Propaganda eleitoral paga - regras de impulsionamento)	Seção 8
Art. 30 (Liberdade de expressão, proibição do anonimato e direito de resposta)	Seções 3 e 11
Art. 32 (Responsabilidade do provedor após notificação judicial)	Seção 3
Art. 33 (Mensagens eletrônicas - identificação e descadastramento)	Seção 11 (Art. 22)
Art. 33-A (Informações aos usuários sobre o tratamento de dados para propaganda eleitoral)	Seção 10

15. SUBMISSÃO

A OpenAI submete este Plano de boa-fé e confia na exatidão e completude das informações apresentadas.

Brasília, 16 de agosto de 2026

OpenAI OpCo, LLC

OpenAI

Versão Pública

I. ANEXOS

Anexo	Título	Descrição	Classificação
Doc. 1	Model Spec	Página da OpenAI sobre o Model Spec, que descreve o comportamento pretendido para os modelos que sustentam os produtos da OpenAI, incluindo a plataforma de API.	Público
Doc. 2	Informações eleitorais e salvaguardas em 2026	Tradução livre da página da OpenAI sobre medidas para ajudar a proteger eleições em países e territórios ao redor do mundo (título original "Election information and Safeguard in 2026").	Público
Doc. 3	Restrições a campanhas políticas	Página da OpenAI que explica o que "campanha política" significa segundo as Políticas de Uso da OpenAI e fornece exemplos não exaustivos de usos relacionados a campanhas que são permitidos ou proibidos	Público

II. PÁGINAS PÚBLICAS CITADAS PELA OPENAI NO PLANO

ChatGPT	https://chatgpt.com/
Termos de Uso	https://openai.com/pt-BR/policies/row-terms-of-use/
Termos de serviço	https://openai.com/pt-BR/policies/service-terms/
Apresentamos a busca do ChatGPT	https://openai.com/pt-BR/index/introducing-chatgpt-search/
Apps no ChatGPT	https://help.openai.com/pt-br/articles/11487775-apps-in-chatgpt
Perguntas frequentes sobre links compartilhados do ChatGPT	https://help.openai.com/pt-br/articles/7925741-perguntas-frequentes-sobre-links-compartilhados-do-chatgpt
GPTs no ChatGPT	https://help.open com/pt-br/articles/8554407-gpts-in-chatgpt
ChatGPT Sites	https://openai.com/academy/chatgpt-sites/
Transparência e moderação de conteúdo	https://openai.com/pt-BR/transparency-and-content-moderation/
Compartilhar e publicar GPTs	https://help.openai.com/pt-br/articles/8798878-sharing-and-publishing-gpts
O que você precisa saber sobre o fim da série Sora	https://help.openai.com/pt-br/articles/20001152-what-to-know-about-the-sora-discontinuation

Políticas publicitárias	https://openai.com/pt-BR/policies/ad-policies/
Contrato de Prestação de Serviços da OpenAI	https://openai.com/pt-BR/policies/services-agreement/
Kodex	https://app.kodexglobal.com/signin
Como denunciar conteúdo no ChatGPT e nas plataformas OpenAI	https://help.openai.com/pt-br/articles/10245791-como-denunciar-conte%C3%BAdo-no-chatgpt-e-nas-plataformas-da-openai
Denunciar conteúdo	https://openai.com/pt-BR/form/report-content/
Partidos políticos registrados no TSE	https://www.tse.jus.br/partidos/partidos-registrados-no-tse
Novo módulo SGIP3 Consulta – Tribunal Superior Eleitoral	https://www.tse.jus.br/partidos/partidos-registrados-no-tse/informacoes-partidarias/modulo-consulta-sgip3
OpenAI Model Spec	https://model-spec.openai.com/2025-12-18.html
Políticas de uso	https://openai.com/policies/usage-policies/
Restrições a campanhas políticas	https://help.openai.com/pt-br/articles/20001255-restri%C3%A7%C3%B5es-a-campanhas-pol%C3%ADticas
Election information and safeguards in 2026	https://openai.com/index/election-safeguards-2026/
AP Elections: Information & Services	https://www.ap.org/elections/
Apresentamos o ChatGPT Imagens 2.0	https://openai.com/pt-BR/index/introducing-chatgpt-images-2-0/
Como identificamos conteúdo problemático em nossos serviços para pessoas físicas	https://help.openai.com/pt-br/articles/8940831-como-identificamos-conte%C3%BAdo-problem%C3%A1tico-em-nossos-servi%C3%A7os-para-pessoas-f%C3%ADsicas
Entenda suas responsabilidades em relação aos seus ChatGPT Sites	https://help.openai.com/pt-br/articles/20001337-entenda-suas-responsabilidades-em-rela%C3%A7%C3%A3o-aos-seus-chatgpt-sites
Entendendo as injeções imediatas: um desafio de segurança de vanguarda	https://openai.com/pt-BR/index/prompt-injections/
GPT-5.6: inteligência de fronteira que acompanha a sua ambição	https://openai.com/pt-BR/index/gpt-5-6/
Construindo um sandbox seguro e eficaz para viabilizar o Codex no Windows	https://openai.com/pt-BR/index/building-codex-windows-sandbox/
Projetando agentes de IA para resistir à injeção de prompt	https://openai.com/pt-BR/index/designing-agents-to-resist-prompt-injection/
Sinais de proveniência (Content Credentials, SynthID) no conteúdo gerado pela OpenAI	https://help.openai.com/pt-br/articles/8912793-sinais-de-proveni%C3%AAncia-credenciais-de-conte%C3%BAdo-synthid-em-conte%C3%BAdo-gerado-pela-openai
Verifique conteúdo gerado pela OpenAI	https://openai.com/pt-BR/research/verify/
Cartão do sistema do Sora 2	https://openai.com/index/sora-2-system-card/
Understanding the source of what we see and hear online	https://openai.com/index/understanding-the-source-of-what-we-see-and-hear-online/
Defining and evaluating political bias in LLMs	https://openai.com/index/de ing-and-evaluating-political-bias-in-llms/
Política de compartilhamento e publicação	https://openai.com/pt-BR/policies/sharing-publication-policy/
Termos do desenvolvedor de aplicativos	https://openai.com/pt-BR/policies/developer-apps-terms/