

A Portaria 463 mandou testar as salvaguardas de IA e mandou guardar o resultado

O art. 15, V mantém o resultado dos testes de resistência fora da camada pública, e a detecção de conteúdo sintético alheio depende da marca que o próprio gerador quis inserir

20 set. 2026 · v. 1.0 · Infralegal setorial · Ato normativo infralegal · Brasil · Em vigor · art. 19 · art. 20

Fábio Cabral · ORCID 0009-0005-6669-0424 · CC-BY-4.0 · DOI 10.5281/zenodo.22864362

O QUE MUDOU

Em 27 de julho de 2026 o Tribunal Superior Eleitoral editou a Portaria nº 463, que regulamenta o art. 125-B da Resolução TSE nº 23.610/2019 e obriga provedores de aplicação de internet a apresentar plano de conformidade para as Eleições Gerais de 2026, com prazo fixado pelo art. 24 em 16 de agosto do ano da eleição.

A portaria alcança expressamente a inteligência artificial generativa. O art. 3º, III, obriga a apresentar plano o provedor de aplicação de internet que possua funcionalidade de geração, por usuários, de conteúdo sintético por meio de sistema de inteligência artificial generativa, e o § 1º fixa o corte em cinco milhões de usuários ativos mensais no Brasil, com possibilidade de o Tribunal exigir plano simplificado de quem não alcance esse número.

O art. 15 é o dispositivo que lista o que esses provedores devem apontar no plano. São cinco itens, as salvaguardas técnicas contra a geração de conteúdo que favoreça candidatura ou contenha indicação de voto, as salvaguardas contra conteúdo sexual não consentido envolvendo pessoas candidatas e contra violência política de gênero, a descrição funcional dos mecanismos destinados a prevenir, detectar, bloquear e mitigar tentativas de contornar as salvaguardas, os mecanismos de identificação e rotulagem de conteúdo sintético em publicações orgânicas, e a existência e a periodicidade de testes de resistência às salvaguardas e às tentativas de contorno, a serem realizados antes do início e durante o período eleitoral.

O inciso V, que trata desses testes, termina com uma ressalva, *ipsis litteris*, cujos resultados específicos serão mantidos exclusivamente na camada de acesso restrito prevista no art. 8º, § 1º, II, desta portaria, sem necessidade de inclusão no relatório previsto no art. 28 desta portaria.

Essa ressalva convive com um dever de sinal contrário. O art. 8º, § 1º, I, define a camada de acesso público, o § 2º obriga o provedor a torná-la pública de forma acessível aos eleitores e à sociedade civil organizada, e o § 3º veda que a classificação como restrita abranja a integralidade de uma medida ou categoria de medidas, determinando que permaneçam disponíveis ao público, no mínimo, a existência da medida, sua descrição funcional, a obrigação a que se vincula e o indicador agregado de resultado. O mesmo § 3º define esse indicador com uma cláusula antievasão expressa, a métrica que demonstre, em período determinado, o grau de alcance do resultado esperado, vedada a mera descrição de processo ou a simples declaração de existência de mecanismo.

Há ainda um terceiro dispositivo, e é ele que faz a articulação. O art. 7º, § 4º autoriza que, quando a medida depender de dados gerados durante o processo eleitoral, a trajetória de alcance seja substituída por compromisso de monitoramento, com indicação da metodologia de coleta, da periodicidade de aferição e dos critérios gerais de cálculo, sendo admitida a complementação posterior dos valores efetivamente apurados. Esses valores vão ao relatório consolidado do art. 28,

que deve ser encaminhado ao Tribunal até 19 de dezembro do ano eleitoral, e cuja parte não restrita deve ser divulgada na camada pública. O § 5º do art. 28 declara a finalidade de tudo isso, *ipsis litteris*, para fins de controle social e supervisão regulatória. O primeiro turno é em 4 de outubro de 2026 e o segundo em 25 de outubro.

Em 14 de setembro de 2026, respondendo ao pedido de acesso à informação nº 78812109191933, a Assessoria Especial de Enfrentamento à Desinformação do Tribunal informou que os planos estão publicados no Portal da Justiça Eleitoral. São nove, hospedados pelo próprio Tribunal, de Google, Kwai, LinkedIn, Mercado Livre, Meta, OpenAI, Telegram, TikTok e X. A mesma resposta trouxe quatro negativas. Até aquela data nenhum plano havia sido declarado apto na forma do art. 27, § 1º, não havia sido instaurado regime de supervisão regulatória direta do art. 26 em relação a qualquer provedor, não houve declaração de recusa de fornecimento de informação do art. 9º, § 2º, e não foi determinada auditoria técnica do art. 10, III. O Tribunal acrescentou que todos os planos seguem em fase de análise e de diligências, razão pela qual não concluiu a avaliação quanto ao cumprimento integral das obrigações e dos prazos, inclusive no que se refere à eventual omissão de provedores e à adequada divulgação das informações classificadas na camada de acesso público.

Lidos os nove planos, o art. 7º, § 4º foi invocado pelo nome. O plano da Meta registra, sobre as restrições eleitorais dos sistemas de recomendação, que visto que a eficácia dessas restrições específicas para as eleições depende de dados gerados durante o período eleitoral, a Meta assume um compromisso de monitoramento nos termos do Art. 7º, § 4º, da Portaria nº 463/2026/TSE, e que os valores efetivamente mensurados serão informados no relatório consolidado. O plano da Kwai registra que, quando a métrica depende da operação durante o período eleitoral, o Plano adota compromisso de monitoramento, com metodologia e periodicidade, sem antecipar valores que somente poderão ser apurados durante a execução. O plano do X registra que os controles eleitorais serão ativados progressivamente, com previsão de que todos estejam plenamente operacionais até 4 de outubro de 2026, e que os indicadores serão monitorados durante o período eleitoral e reportados no relatório consolidado do artigo 28 da Portaria, nos termos do artigo 7º, § 4º.

O campo do indicador agregado de resultado vem, portanto, preenchido com nome de métrica e não com valor de métrica, e isso não decorre de omissão dos provedores. O plano do Telegram, único dos nove que adota literalmente o gabarito da portaria medida por medida, declara como indicadores a taxa de conformidade no cumprimento de ordens, o número de contas, bots e canais rotulados ou removidos por indício de falsidade ou automação, o volume de denúncias e o número de solicitações recebidas. São nomes de métricas sem valor e sem período, exatamente o que o art. 7º, § 4º permite.

Sobre a rotulagem de conteúdo sintético exigida pelo art. 15, IV, os planos descrevem soluções convergentes e descrevem também o seu limite. O plano da Meta divide o tratamento entre conteúdo gerado pelas próprias ferramentas, ao qual são aplicados marca d'água invisível, marcadores visíveis e metadados C2PA e IPTC, e conteúdo gerado por ferramentas de terceiros, e sobre este registra, *ipsis litteris*, que a detecção depende da identificação de indicadores padrão do setor de imagens de IA, particularmente credenciais de conteúdo C2PA e metadados IPTC, que outras empresas incluem em seus conteúdos gerados por IA, acrescentando que este é um esforço contínuo e em evolução enquanto a indústria trabalha coletivamente para desenvolver padrões comuns para proveniência de conteúdo. O plano do Google descreve o SynthID como tecnologia que utiliza marcas d'água invisíveis para determinar se imagens, vídeos e áudios foram gerados ou editados especificamente pelos modelos de IA do Google. O plano da OpenAI descreve a identificação de conteúdo sintético por metadados C2PA e marca d'água SynthID e registra que, no conteúdo gerado pela OpenAI, a detecção do SynthID fornece um indicador de origem da OpenAI, e que uma correspondência positiva não indica precisão, titularidade, contexto ou que o conteúdo não tenha sido alterado em relação à

sua forma original.

O mesmo plano da OpenAI é o que vai mais longe na descrição dos próprios limites, e três passagens merecem transcrição. Sobre os metadados criptográficos, *ipsis litteris*, os metadados C2PA podem ser removidos ou perdidos por meio de upload, download, edição, conversão, compartilhamento ou remoção deliberada, e 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. Sobre a marca d'água, *ipsis litteris*, o SynthID pode se degradar após transformações significativas, e a ausência de uma marca d'água detectável não é conclusiva. E sobre o conjunto, *ipsis litteris*, no conteúdo gerado pela OpenAI, nenhuma solução individual de proveniência proporciona identificação absoluta de conteúdo gerado por IA.

Há ainda, no mesmo plano, uma declaração de escopo que pesa mais do que todas em um regime destinado a enfrentar desinformação. Sobre texto e código, *ipsis litteris*, o texto e os dados estruturados gerados pelo ChatGPT atualmente não carregam marcas d'água de proveniência legíveis por máquina, e a empresa acrescenta que 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.

Registre-se, por fim, o efeito jurídico da apresentação. O art. 27, caput, assegura em caráter provisório o credenciamento do art. 27-A, § 4º, e o cadastro do art. 29, § 9º, ambos da Resolução nº 23.610/2019, ao provedor que apresentar o plano no prazo do art. 24, independentemente da conclusão da análise de seu mérito. Como nenhum plano foi declarado apto, os nove provedores atravessam o período eleitoral sob credenciamento provisório obtido pelo ato de entregar.

Dois defeitos da camada pública, verificáveis por qualquer pessoa que baixe os arquivos, completam o quadro. O plano do LinkedIn tem catorze páginas, porém o rodapé de todas as páginas anuncia quinze páginas e o arquivo publicado tem catorze, com o texto terminando na conclusão. Não é possível dizer se o rodapé está errado ou se falta uma página. Na versão pública da Meta, com noventa e uma páginas, há trinta e sete marcações de abertura de conteúdo confidencial, e a página 34 é inteiramente vazia, contendo apenas o par de marcadores e o rodapé, logo depois de o texto anunciar que apresentaria os resultados consolidados em uma matriz de riscos e a metodologia de mensuração das medidas de mitigação, em conformidade com o art. 20 da portaria.

APRECIÇÃO

O que segue é apreciação do autor, e pode ser dispensada por quem queira citar apenas os fatos acima.

A Portaria nº 463/2026 fez algo que o direito brasileiro de transparência algorítmica raramente faz. Ela nomeou o vício antes de ele acontecer. A fórmula vedada a mera descrição de processo ou a simples declaração de existência de mecanismo é a definição técnica exata do que os relatórios de transparência das plataformas vinham entregando havia uma década, e o art. 8º, § 3º a proíbe com todas as letras. É boa norma, e merece ser dito antes de qualquer crítica.

O que a mesma portaria fez em seguida foi retirar, em dois lugares distintos, aquilo que daria cumprimento à sua própria exigência.

O primeiro lugar é o art. 7º, § 4º. Se o indicador precisa demonstrar o grau de alcance em período determinado, ele precisa de um número medido, e um número medido sobre o período eleitoral só existe depois do período eleitoral. A portaria resolveu isso autorizando substituir a trajetória de alcance por compromisso de monitoramento e admitindo a complementação

posterior dos valores. O resultado é que a camada de acesso público nasce sem o elemento que o art. 8º, § 3º elegeu como seu mínimo indispensável, e permanece assim durante todo o tempo em que a eleição acontece. Não houve evasão de ninguém. Meta, Kwai e X invocaram o dispositivo pelo nome, porque ele estava lá para ser invocado. Quando nove documentos redigidos por nove equipes jurídicas distintas convergem na mesma lacuna, a lacuna não está neles.

O segundo lugar é o art. 15, V, e é o mais grave, porque ali não há adiamento, há exclusão. De todos os deveres de IA do art. 15, o inciso V é o único que produz evidência de eficácia. Salvaguarda contra recomendação de voto, salvaguarda contra conteúdo sexual não consentido, mecanismo contra contorno, rótulo de conteúdo sintético, todos esses são medidas que se descrevem. Só o teste de resistência mede se elas aguentam. E é exatamente o resultado desse teste que a portaria manda manter exclusivamente na camada restrita, dispensando-o até do relatório consolidado. O número que diria se as salvaguardas funcionam não chega em dezembro. Não chega nunca.

Some-se a isso o que os próprios planos dizem sobre detecção. A rotulagem de conteúdo sintético do art. 15, IV funciona em dois regimes. Sobre o conteúdo que a plataforma mesma gerou, ela rotula com alta confiança, porque conhece a própria marca. Sobre o conteúdo gerado por terceiro, a Meta declara que a detecção depende de credenciais C2PA e metadados IPTC que outras empresas incluem, o Google descreve o SynthID como capaz de determinar se o conteúdo foi gerado especificamente pelos modelos de IA do Google, e a OpenAI descreve a correspondência positiva do SynthID como indicador de origem da OpenAI. Cada laboratório reconhece o próprio fabricante. Ninguém detecta o conteúdo de ninguém, salvo quando o gerador quis ser encontrado.

E a OpenAI, que é quem descreve com mais franqueza os próprios limites, fecha os dois lados da inferência. A presença da marca não prova, porque uma correspondência positiva não indica precisão, titularidade nem contexto. A ausência da marca também não prova, porque os metadados podem ser removidos ou perdidos em upload, edição ou conversão, e porque a marca d'água pode se degradar após transformações significativas. É o obrigado dizendo ao tribunal, por escrito e antes de qualquer litígio, que o rótulo não decide a questão que o rótulo existe para decidir.

Isso desloca o ponto de apoio do regime inteiro. A obrigação de rotular é dirigida ao provedor de aplicação que distribui o conteúdo, mas a capacidade de cumpri-la depende de um ato do provedor que gerou o conteúdo, que pode ser outro, pode estar fora do Brasil, pode não ter cinco milhões de usuários mensais e, portanto, pode não estar sujeito à portaria. Quem não assina nada com o TSE é justamente quem decide se a marca existe. O deepfake produzido em modelo aberto, executado em máquina local e publicado sem metadado não é alcançado por nenhuma dessas medidas, e nenhum dos nove planos afirma o contrário.

E há um vão anterior a esse, declarado no mesmo plano. O texto e os dados estruturados gerados pelo ChatGPT, segundo a OpenAI, atualmente não carregam marcas d'água de proveniência legíveis por máquina. Num regime cuja finalidade declarada é enfrentar a desinformação eleitoral, a modalidade que mais circula em conversa, em grupo de mensagem e em publicação orgânica é justamente aquela para a qual a proveniência ainda não existe. O art. 15, IV exige descrever o mecanismo de rotulagem de conteúdo sintético, e a descrição fiel, nessa modalidade, é a de que não há o que rotular.

O desenho brasileiro repete isso no marco geral. O art. 19 do PL nº 2.338/2023 põe o dever de

incluir identificador sobre quem gera o conteúdo sintético, e o seu § 2º promete que a autoridade competente, com o Conselho Permanente de Cooperação Regulatória de Inteligência Artificial, disponibilizará biblioteca de softwares para facilitar a sinalização. O art. 20 promete que o poder público, com a iniciativa privada e a sociedade civil, promoverá capacidades para identificar e rotular conteúdo sintético e estabelecer a autenticidade e a proveniência do conteúdo digital. São promessas programáticas, na forma de regulamentação. A marca é dever exigível hoje, a capacidade de verificar a marca é projeto para depois. A portaria eleitoral não inventou a assimetria, ela apenas foi a primeira a aplicá-la a uma eleição.

Este é o terceiro caso que o Observatório registra com a mesma estrutura. Na nota 4, criou-se o dever de marcar conteúdo sintético e não se criou a obrigação de deixar verificar a marca. Na nota 5, a autoridade publicou o quanto da multa e não publicou o porquê, porque a motivação estava em relatório de instrução não público. Agora, exigiu-se testar as salvaguardas de IA e determinou-se que o resultado do teste não saia da camada restrita. Três regimes distintos, três autoridades distintas, e a mesma operação, a obrigação é criada e a verificabilidade por terceiro não vem junto. O que este caso acrescenta é o refinamento, porque aqui a norma não esqueceu a verificabilidade. Ela a escreveu com precisão incomum no art. 8º, § 3º, e depois a desfez em dois outros artigos seus, um adiando e outro reservando.

O que decorre disso para o desenho dos arts. 19 e 20 do PL nº 2.338/2023 fica para outro lugar. Esta nota registra o mecanismo e para aqui.

MÉTODO

Leitura direta das fontes primárias, todas congeladas antes da redação.

A Portaria TSE nº 463/2026 foi lida no texto compilado do sítio do Tribunal, capturado em 2 de setembro de 2026, de onde foram transcritos os arts. 3º, 7º, 8º, 9º, 10, 15, 24, 25, 26, 27 e 28. Nenhuma afirmação sobre o conteúdo da portaria se apoia em cobertura secundária.

A resposta do Tribunal veio do pedido de acesso à informação nº 78812109191933, protocolado no Serviço de Informação ao Cidadão do TSE em 9 de setembro de 2026 com sete itens, todos ancorados em dispositivo da portaria e nenhum alcançando a camada de acesso restrito do art. 8º, § 1º, II. A resposta foi recebida por mensagem eletrônica da Ouvidoria em 14 de setembro de 2026, dentro do prazo do art. 11 da Lei nº 12.527/2011, sem indeferimento, sem invocação de sigilo e sem prorrogação.

Os nove planos foram baixados do endereço indicado pelo Tribunal em 20 de setembro de 2026. Todos os trechos de plano citados nesta nota foram transcritos dos arquivos assim obtidos, e as passagens sobre compromisso de monitoramento e sobre detecção de conteúdo sintético foram conferidas uma a uma no documento de origem.

Os arts. 19 e 20 do PL nº 2.338/2023 foram lidos na versão do autógrafo do Senado.

Nenhuma afirmação sobre entrega fora do prazo do art. 24 é feita, porque o Tribunal não informou as datas de protocolo.

PONTAS SOLTAS

- Não se afirma que qualquer provedor tenha descumprido a portaria. Essa avaliação compete ao

Tribunal, que declarou em 14 de setembro de 2026 não tê-la concluído, inclusive quanto à adequada divulgação das informações da camada de acesso público.

- A lista de nove é a lista de quem apresentou plano, e não a lista de obrigados. Quem é alcançado pelo art. 3º, III, e pelo corte de cinco milhões de usuários do § 1º não foi apurado. A Anthropic não consta da lista, e isso não autoriza afirmar omissão.
- Não foi verificado se cada provedor tornou pública a camada de acesso público no próprio sítio, como determina o art. 8º, § 2º, ou se a publicação existe apenas no portal do Tribunal.
- Não se sabe desde quando a seção de planos está no ar no Portal da Justiça Eleitoral. Os arquivos respondem cabeçalho de última modificação de 3 de setembro de 2026, o Internet Archive tem captura da página de 10 de agosto sem a seção e de 17 de setembro com a seção, e não há captura do dia 3 de setembro.
- Não foi verificado se algum plano foi atualizado na forma do art. 24, parágrafo único, que obriga comunicação em até quarenta e oito horas quando há alteração substancial de risco ou de funcionalidades.

FONTES DOCUMENTAIS ARQUIVADAS

Portaria TSE nº 463, de 27 de julho de 2026, texto compilado, que regulamenta o art. 125-B da Resolução TSE nº 23.610/2019 (arquivado em 3 set. 2026)

sha256 ba336cdf42d84b074413465a32a0f6b34000d3180184a8cddad089dbe23215d8

Resposta do Tribunal Superior Eleitoral ao pedido de acesso à informação nº 78812109191933, de 14 de setembro de 2026, transcrição literal (arquivado em 20 set. 2026)

sha256 d59778059cb8b5cb9a8b45e18f977a7fd536315236306f498bfa80b97027828a

Plano de conformidade do Telegram para as Eleições Gerais de 2026, versão pública (arquivado em 20 set. 2026)

sha256 ab75f84ec08e8bcd28ff0f72c27d9b8bc8af217472c177c3b19887b6ebe26eb6

Plano de conformidade da Meta para as Eleições Gerais de 2026, versão pública (arquivado em 20 set. 2026)

sha256 caac059a25ba5594842d24e7e7b012cb794e0cd4d600afab50182ca79b9afaf5

Plano de conformidade da Kwai para as Eleições Gerais de 2026, versão pública (arquivado em 20 set. 2026)

sha256 69691370a92b914496d27442d7b07f3699df020f8dcf2a44628dc5d01b09d61e

Plano de conformidade do X e do Grok para as Eleições Gerais de 2026, versão pública (arquivado em 20 set. 2026)

sha256 8cc7806c32d241a113051896c977dd7da419813c1382beff577161873ee4c3b2

Plano de conformidade da OpenAI para as Eleições Gerais de 2026, versão pública (arquivado em 20 set. 2026)

sha256 453940e57bf21369ba323ac11cdf611fa4fda35cb57490eafd378b46c7f90086

Plano de conformidade do Google para as Eleições Gerais de 2026, versão pública (arquivado em 20 set. 2026)

sha256 fa4b8f55948e0d52e3735b11f690ce708b5afff5086a5c0b45896f3fe887ad6a

Plano de conformidade do LinkedIn para as Eleições Gerais de 2026, versão pública (arquivado em 20 set. 2026)

sha256 f988741f2defc7decdd3ff69596e364be0723069a574d54af8b1f614a1b3a30f

Fonte oficial: <https://www.tse.jus.br/legislacao/compilada/prt/2026/portaria-no-463-de-27-de-julho-de-2026>

COMO CITAR

CABRAL, Fábio. A Portaria 463 mandou testar as salvaguardas de IA e mandou guardar o resultado. *Observatório IA + Direito*, nota nº 10, 20 set. 2026. Disponível em: <https://fabiocabral.pro/observatorio/2026-09-salvaguardas-de-ia-com-teste-reservado>. Acesso em: [data]. DOI: 10.5281/zenodo.22864362.