

A marca d'água chegou ao Brasil sem passar pelo Congresso

O texto do Claude passou a ser marcado no mundo inteiro por força do art. 50 do AI Act, e ninguém fora do fornecedor consegue ler a marca

18 ago. 2026 · v. 1.1 · Autorregulação · Ato privado de conformidade · Estados Unidos · Em vigor · art. 19 · art. 20 · art. 4

Fábio Cabral · ORCID 0009-0005-6669-0424 · CC-BY-4.0 · DOI 10.5281/zenodo.22002225

O QUE MUDOU

Em 14 de agosto de 2026 a Anthropic publicou como funciona a marca d'água no texto do Claude. O método é descrito como uma versão da abordagem SynthID-Text, do Google DeepMind, publicada na Nature em 2024, e não acrescenta nada ao texto. Ele troca a fonte da aleatoriedade na escolha entre palavras igualmente boas, de modo que a sequência produzida fique consistente com uma chave. Quem tem a chave calcula a probabilidade de o modelo ter participado do texto. Quem não tem, não vê nada.

O fundamento declarado é o cumprimento do art. 50 do Regulamento (UE) 2024/1689, o AI Act, e do Code of Practice on Transparency of AI-generated Content, código voluntário assinado em julho de 2026 por cerca de 190 organizações, entre elas Anthropic, Google, Meta, Microsoft, Mistral e OpenAI. O art. 50, n. 2, obriga os prestadores a assegurar que os resultados sintéticos sejam marcados em formato legível por máquina e detectáveis como artificialmente gerados ou manipulados, com soluções eficazes, interoperáveis, sólidas e fiáveis, sem nomear tecnologia. As obrigações de transparência do art. 50 são aplicáveis desde 2 de agosto de 2026, e o adiamento promovido pelo Regulamento (UE) 2026/1744 alcançou o regime de alto risco, não este.

A empresa declara que aplica a marcação globalmente porque ainda não dispõe de forma durável de restringi-la por região, e que a API de detecção será oferecida em breve, sem data. Declara ainda os limites do instrumento. A marca indica apenas que o modelo provavelmente esteve envolvido, não distingue texto escrito pelo modelo de texto humano por ele pesadamente editado, não confirma autoria humana quando não é encontrada, não carrega informação sobre o usuário e nada diz sobre titularidade ou responsabilidade. E é mais esparsa em passagens factuais, onde não existem escolhas equivalentes entre palavras, quase nula em código e possivelmente inexistente quando o modelo apenas revisa gramática e pontuação.

Nos dias seguintes ao anúncio surgiram repositórios e serviços web prometendo remover a marca. A imprensa especializada apurou, em 13 de agosto, que essas ferramentas removem caracteres invisíveis e metadados de arquivo, e que o autor da mais visível delas registra que a remoção efetiva da marca estatística não está disponível. Nenhuma alegação de remoção pode ser conferida hoje, porque não existe detector público contra o qual testar o resultado.

O quadro normativo em volta tem três outras peças. O art. 50, n. 7, acrescentado ao AI Act pelo Regulamento (UE) 2026/1744, incumbiu a Comissão de incentivar códigos de práticas da União para detecção, marcação e rotulagem, e a autorizou a adotar ato de execução com regras comuns caso considere o código inadequado. O California AI Transparency Act, operativo desde 2 de agosto de 2026, exige de prestadores com mais de um milhão de usuários mensais a incorporação de disclosure latente, a opção de disclosure visível e uma ferramenta pública e gratuita de detecção, e exclui do seu âmbito o conteúdo em texto, alcançando apenas imagem, vídeo e áudio. Na China, as

medidas de rotulagem de conteúdo sintético e a norma nacional obrigatória GB 45438-2025 vigoram desde 1º de setembro de 2025, com rótulo explícito e rótulo implícito, este predominantemente construído como metadado padronizado no arquivo.

No Brasil não há obrigação geral de marcação. A única regra vigente de identificação de conteúdo sintético é o art. 9º-B da Resolução TSE nº 23.610/2019, na redação da Resolução nº 23.755/2026, que impõe ao responsável por propaganda eleitoral informar de modo explícito, destacado e acessível que o conteúdo foi fabricado ou manipulado e qual tecnologia foi utilizada. É rótulo visível, não marca invisível. O art. 19 do PL 2338/2023, que impõe identificador de proveniência ao conteúdo sintético e promete, no § 2º, biblioteca de softwares com padrão internacional amplamente reconhecido, segue aguardando parecer do relator na Comissão Especial da Câmara.

APRECIÇÃO

Três consequências, e a ordem importa.

A primeira é de fonte. Um dever europeu passou a produzir efeito no Brasil sem norma brasileira, sem tratado e sem ato da autoridade nacional. Não houve extraterritorialidade jurídica, houve decisão de engenharia. A empresa marca no mundo inteiro porque separar por região sairia caro ou frágil, e o resultado prático é que o parâmetro que hoje alcança o texto produzido por um advogado brasileiro foi escrito em Bruxelas. Enquanto o art. 19 do PL 2338 espera parecer, o assunto que ele regula já está decidido na prática, e quem decidiu não foi o Congresso.

A segunda é de verificação, e é a que ninguém está discutindo. Marcar virou dever, verificar não. O AI Act exige solução eficaz, interoperável, sólida e fiável, e não exige detector acessível a terceiro. O código europeu é voluntário, ainda que o novo art. 50, n. 7, permita à Comissão substituí-lo por ato de execução. A única jurisdição que impôs ferramenta pública e gratuita de detecção foi a Califórnia, e ela deixou o texto de fora. Sobra um arranjo desconfortável. Existe marcação global de texto e não existe, em lugar nenhum, obrigação de que alguém consiga lê-la. Quem coloca a marca é quem informa se ela está lá, e a essa altura o § 2º do art. 19 do PL, que promete biblioteca de softwares e padrão internacional, deixa de ser dispositivo decorativo e passa a ser o único ponto do texto brasileiro que enfrenta o problema certo. Ele precisa de detector público, com taxa de erro publicada e continuidade declarada, e não apenas de ferramenta para marcar.

A terceira é de prova, e vem por declaração do próprio fornecedor, o que a torna rara e valiosa. A marca não distingue quem escreveu de quem editou, não prova autoria humana pela ausência e enfraquece nas passagens factuais, porque ali as escolhas entre palavras não são equivalentes. Ou seja, o instrumento é mais fraco justamente no trecho que carrega a citação de julgado, o número do dispositivo e o dado, que é onde mora a alucinação que já rendeu condenação por litigância de má-fé no Brasil. Um resultado de detector, portanto, não sustenta sozinho juízo de má-fé, de plágio ou de fraude acadêmica, e a ausência de marca não prova trabalho humano. Quem for usar isso em processo disciplinar ou judicial precisa de contexto, de cadeia de custódia, de versionamento e de contraditório, e não de um número.

Uma correção final, dirigida à leitura que circulou nas redes. Não é verdade que a marca tenha sido removida horas depois de anunciada. O que apareceu em horas foi o mercado que promete removê-la, e o que ele entrega, por ora, é limpeza de metadados. O caminho conhecido para desfazer a marca é a reescrita integral por outro modelo, admitida pela própria empresa e testada em literatura ainda preliminar. Isso não enfraquece a crítica, fortalece. O problema da marca d'água não é ter caído, é ter nascido inverificável, e é isso que a torna imprópria como

prova e adequada apenas como indício de proveniência entre agentes que cooperam. Regulação que aposta em marcação sem construir a capacidade pública de verificação escolhe o lado fácil do problema e produz um efeito perverso conhecido, o descuidado fica marcado e quem tem motivo para esconder consegue esconder.

MÉTODO

Leitura integral do anúncio institucional da Anthropic de 14/08/2026, capturado e arquivado com sha256 no acervo. Leitura direta dos arts. 50 e 113 no texto consolidado do AI Act em português, 02024R1689 PT 27.07.2026 001.001, conferindo as marcas B e M1 que separam a redação de 2024 da alteração do Regulamento (UE) 2026/1744. Leitura dos arts. 19 e 20 do PL 2338 na versão do autógrafo do Senado. Conferência da natureza e da lista de signatários do código europeu no portal da Comissão, do art. 9º-B no repositório do TSE, do California AI Transparency Act com a alteração da AB 853, das medidas chinesas de rotulagem e da ficha de tramitação da Câmara. O fenômeno das ferramentas de remoção foi conferido em imprensa especializada, e não nos repositórios. Dossiê de verificação interno com 22 itens, 11 confirmados, 3 corrigidos, 3 não verificados e 5 omissões, sobre relatório derivado que não é citável.

PONTAS SOLTAS

- A API de detecção anunciada não tem data, preço, política de acesso nem taxa de erro publicada.
- Não foi apurado se OpenAI, Google e Meta já implantaram marcação de texto em produção, nem com quais chaves e detectores.
- Não se apurou se a marcação alcança os modelos servidos no Brasil por intermediários de nuvem, nem como interage com integrações por API de terceiros.
- O texto integral do artigo do SynthID-Text na Nature não foi obtido, apenas a referência bibliográfica. Os números do experimento em produção não estão verificados aqui.
- As contagens de estrelas dos repositórios de remoção divergem entre coberturas e não foram conferidas na fonte.
- Não foi levantada manifestação da ANPD, do CNJ ou da OAB sobre marcação de conteúdo sintético.

FONTES DOCUMENTAIS ARQUIVADAS

How Claude's text watermark works, Anthropic, 14/08/2026 (arquivado em 18 ago. 2026)

sha256 020bcfe3e923df67229e81af9ad9d5a9497b74c2b977f657cca4cdedc7ca2a43

Obra de terceiro: preservada no acervo com hash, não republicada.

Fonte oficial: <https://www.anthropic.com/news/claude-text-watermark>

COMO CITAR

CABRAL, Fábio. A marca d'água chegou ao Brasil sem passar pelo Congresso. *Observatório IA + Direito*, n. 4, 18 ago. 2026. Disponível em: <https://fabiocabral.pro/observatorio/2026-08-marca-dagua-claude>. Acesso em: [data].