



SENADO FEDERAL
Gabinete do Senador IZALCI LUCAS

EMENDA Nº
(ao PL 2338/2023)

Suprimam-se os arts. 64 e 65.

Dê-se aos artigos 62 e 63 a seguinte redação:

Art. 62 - Não constitui ofensa aos direitos autorais a utilização automatizada de conteúdos protegidos em processos de mineração de textos e dados para os fins de pesquisa ou desenvolvimento de sistemas de inteligência artificial, desde que observadas as seguintes condições:

I - o acesso tenha se dado de forma legítima;

II - a atividade não tenha como objetivo principal a fruição, reprodução, exibição ou disseminação da obra original em si;

III - a utilização de conteúdos protegidos por direitos de autor e conexos seja feita na medida necessária para o objetivo a ser alcançado, não prejudique injustificadamente os interesses econômicos dos titulares e não concorra com a exploração normal das obras, e;

IV - não exista sinalização expressa junto a obra da oposição a sua análise para o desenvolvimento de sistemas de inteligência artificial, devendo tal oposição ser feita por meio legível por máquina e padronizado quando a obra for disponibilizada na rede mundial de computadores, nos termos desta lei e da regulação.

§1º Eventuais cópias transitórias de conteúdos protegidos por direitos de autor e conexos utilizadas no treinamento de sistemas de inteligência artificial deverão ser mantidas em estritas condições de segurança, e unicamente pelo



tempo necessário para a realização da atividade ou para a finalidade específica de verificação dos resultados.

§ 2º A atividade de mineração de textos e dados que envolva dados pessoais estará sujeita às disposições da Lei no 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais).

Art. 63 A limitação aos direitos de autor prevista no caput somente será aplicável quando o titular de direitos autorais não houver indicado expressamente sua reserva por meios adequados e, no caso de obras disponíveis publicamente na rede mundial de computadores, por meio legível por máquina que siga os padrões de mercado e as normas internacionais, na forma de regulamento.

JUSTIFICAÇÃO

O desenvolvimento responsável de aplicações de IA requer acesso a uma gama diversificada de dados de treinamento para aprender. Para treinar um modelo, são necessários bilhões de pontos de dados para fornecer informações culturalmente relevantes, perspectivas variadas e diferentes formas e uso da linguagem. Isso permite que os modelos de IA identifiquem recursos, relacionamentos e padrões entre pontos de dados.

A lei de direitos autorais deve apoiar a dar suporte ao desenvolvimento da IA. As exceções (e em determinados países, regras de *fair use*) que permitem a análise computacional e a mineração de texto e dados, de forma equilibrada em relação aos direitos autorais, permitem que pesquisadores e inovadores usem material protegido por direitos autorais em certas circunstâncias sem a permissão do titular dos direitos autorais. Tais regras são necessárias para facilitar novas formas de criatividade, expressão e o desenvolvimento de novos produtos que podem ampliar a produtividade e estimular a inovação.

A IA depende do das exceções TDM (*text and data mining*) viáveis porque o desenvolvimento responsável de aplicações de IA geralmente requer acesso a uma gama diversificada de dados de treinamento de alta qualidade. Não é coincidência que os principais países inovadores do mundo tenham regras de



fair use ou uma exceção específica para a análise computacional e a mineração de texto e dados, incluindo para fins e usos comerciais. Sem proteções legais claras para esse tipo de uso, podemos desperdiçar o potencial da IA de fornecer soluções inovadoras para importantes desafios sociais.

Atualmente, há uma discussão significativa sobre obrigar os desenvolvedores de IA a divulgar os conjuntos de dados nos quais tais sistemas foram treinados. A maioria dos LLMs é treinada em uma ampla variedade de dados online disponíveis publicamente, incluindo dados disponíveis na web, em vez de conjuntos de dados "offline". Diante disso, uma regra determinando a publicação do conjunto de dados usados em treinamento seria uma política insustentável por algumas razões:

Primeiro, a fonte de grande parte do treinamento, validação, teste e dados de entrada é o enorme volume de conteúdo disponível em toda a World Wide Web aberta — em contraste com modelos que usam um número limitado de fontes bem definidas e prontamente identificáveis. Segundo, identificar os conjuntos de dados usados para treinar sistemas específicos exporia informações competitivamente sensíveis (e potencialmente protegidas por segredo comercial). E terceiro, os desenvolvedores de IA não têm acesso a informações detalhadas e precisas sobre o status de direitos autorais, propriedade ou termos de licenciamento para todo o conteúdo disponível na web pública. Na verdade, não existe tal fonte de verdade em nenhum lugar do mundo. Por esse motivo, desenvolvedores de IA têm caminhado para anunciar meios técnicos e legíveis por máquina para que detentores de direitos possam exercer o controle sobre quando seus conteúdos disponíveis na web podem ser usados para treinamento de modelos de IA generativos.

Esses controles (que ainda estão em fase de aprimoramento e desenvolvimento) tornam os requisitos de divulgação desnecessários porque permitem que os detentores de direitos saibam e controlem *ex ante* se seu conteúdo online poderá ser usado para treinamento de modelos futuros. Além disso, é importante destacar que esse tipo de controle e a capacidade de escolher se um conteúdo pode ou não ser usado para treinamento pode, também, facilitar novas soluções baseadas no mercado.



A proposta dos artigos 62 e 63 vai exatamente nesse sentido e busca estabelecer um equilíbrio importante entre salvaguardar os interesses legítimos dos titulares de direitos autorais e fomentar o avanço da IA no Brasil. Este artigo permite o uso automatizado de conteúdo protegido para fins de mineração de textos e dados em pesquisa e desenvolvimento de IA, desde que certas condições sejam atendidas.

Em primeiro lugar, a proposta garante que o acesso ao conteúdo seja obtido de forma legítima, evitando mau uso. Em segundo lugar, ela especifica que o objetivo principal não pode ser a exploração direta da obra original em si, mas sim seu uso como parte do processo de desenvolvimento de IA. Isso protege os interesses econômicos dos titulares de direitos, impedindo que o conteúdo seja usado como substituto da obra original.

Importante mencionar que a proposta limita o uso de conteúdo protegido ao estritamente necessário para alcançar os objetivos pretendidos de desenvolvimento de IA, sem prejudicar indevidamente os interesses econômicos legítimos dos titulares de direitos ou concorrer com a exploração normal das obras. Isso está alinhado com os princípios de uso justo e o teste de três etapas estabelecido em tratados internacionais de direitos autorais.

A exigência de que os titulares de direitos indiquem expressamente sua oposição ao uso de suas obras para o desenvolvimento de IA, por meio de sinais legíveis por máquina, garante ainda mais uma abordagem equilibrada. Isso dá aos titulares de direitos a capacidade de controlar o uso de seu conteúdo, ao mesmo tempo que fornece diretrizes claras para os desenvolvedores de IA sobre o escopo permitido de uso.

Em comparação com as abordagens mais permissivas observadas em países como Cingapura e Japão, que não permitem a titulares optarem por proibir a análise computacional e a mineração de dados e de texto, a proposta brasileira mantém salvaguardas mais fortes para os titulares de direitos. As exceções de direitos autorais de Cingapura para mineração de textos e dados são mais amplas, enquanto as recentes alterações do Japão permitem o uso de obras protegidas por direitos autorais sem permissão para fins de desenvolvimento de IA.



Em contraste, a proposta brasileira está alinhada com a abordagem da União Europeia, conforme descrito na Diretiva de Direitos Autorais Digitais da UE, mantendo o País alinhado com legislações internacionais. Ambas as estruturas reconhecem a necessidade de equilibrar os interesses dos titulares de direitos e o desenvolvimento de tecnologias inovadoras, como a IA, que podem impulsionar o progresso econômico e social.

Ao estabelecer esse equilíbrio, tais artigos visam criar um ambiente que fomente o crescimento do setor de IA no Brasil, ampliar a produtividade por meio de IAs capazes de realizar tarefas complexas em português, ao mesmo tempo em que respeita os princípios fundamentais da proteção dos direitos autorais. Essa abordagem ponderada garante que o desenvolvimento da IA possa continuar a prosperar, respeitando também os direitos e interesses daqueles que criam o conteúdo que alimenta esses avanços.

Diante do exposto, solicito o apoio de meus nobres pares nesta Casa para a aprovação desta Emenda.

Sala da comissão, de de .

Senador Izalci Lucas
(PL - DF)

