



AUTORIA E DIREITOS AUTORAIS DE OBRAS CRIADAS POR INTELIGÊNCIA ARTIFICIAL GENERATIVA

Lylian Nascimento¹

Editora Responsável: Ana Luiza Neder

Revisão textual: Agnes Meneghin e Tailine Hijaz

RESUMO

O uso crescente da inteligência artificial (IA) na produção de textos criativos tem intensificado o debate jurídico e acadêmico acerca da noção de autoria e dos direitos autorais. O presente estudo analisa a autoria de obras geradas com o auxílio de chatbots de IA, como o ChatGPT, à luz da doutrina contemporânea e da legislação brasileira. O objetivo consiste em discutir as diferentes possibilidades de atribuição de autoria e de titularidade dos direitos sobre o conteúdo criativo, distinguindo tais categorias, que não necessariamente coincidem no direito autoral, seja ao usuário, à empresa desenvolvedora ou aos detentores dos dados de treinamento, com base nos regimes de apropriação já debatidos na literatura jurídica. A metodologia adotada é de natureza bibliográfica e documental, com ênfase na análise da Lei de Direitos Autorais (Lei nº 9.610/98) (BRASIL, 1998). Os resultados indicam que a tese mais compatível com os princípios atuais do ordenamento jurídico brasileiro é a da autoria do usuário, em razão de sua intencionalidade,

¹ Mestranda em Linguagens, Usos e Tecnologias pela Universidade Federal de Sergipe (UFS). Possui pós-graduação em Direito do Trabalho e Processo do Trabalho. Atuou como bolsista no Programa Institucional de Bolsas de Iniciação à Docência (PIBID – 2021-2022) e no Programa Institucional de Bolsas de Iniciação Científica (PIBIC-CNPq – 2021 e 2022). ORCID: <https://orcid.org/0009-0003-4329-7115>. E-mail: lylianvan@gmail.com

Recebido: 15/08/2025. Aprovado: 19/06/2026.

criatividade e atuação curatorial, compreendendo a IA como instrumento. Conclui-se pela necessidade de avanço regulatório no Brasil, bem como pela relevância da articulação entre direito, sociolinguística e ciência da computação para a construção de soluções juridicamente seguras e socialmente responsáveis.

Palavras-chave: Inteligência Artificial Generativa; Direitos Autorais; Autoria; Sociolinguística; Legislação Brasileira.

AUTHORSHIP AND COPYRIGHT OF WORKS CREATED BY GENERATIVE ARTIFICIAL INTELLIGENCE

ABSTRACT

The increasing use of artificial intelligence (AI) in the production of creative texts has intensified legal and academic debates concerning the notions of authorship and copyright. This study analyzes the authorship of works generated with the assistance of AI chatbots, such as ChatGPT, in light of contemporary legal scholarship and Brazilian legislation. The objective is to examine the different possibilities for attributing authorship and ownership of rights over creative content, distinguishing these categories, which do not necessarily coincide under copyright law, whether to the user, the developing company, or the creators of the training data, based on appropriation regimes discussed in legal literature. The methodology adopted is bibliographic and documentary research, with a focus on the analysis of Brazilian Copyright Law (Law No. 9,610/98). The results indicate that the thesis most consistent with the current principles of the Brazilian legal system is that of user authorship, due to the user's intentionality, creativity, and curatorial role, considering AI as a tool. The study concludes by emphasizing the need for regulatory development in Brazil, as well as the importance of collaboration among law, sociolinguistics, and computer science to ensure legal certainty and promote the development of ethical and socially responsible AI systems.

Keywords: Generative Artificial Intelligence; Copyright; Authorship; Sociolinguistics; Brazilian Legislation.

INTRODUÇÃO

Em um mundo onde máquinas escrevem romances, quem detém os direitos de sua criação? A crescente capacidade da inteligência artificial generativa, como o *ChatGPT*, de produzir conteúdo criativo, incluindo romances, roteiros e outros gêneros textuais, tem gerado debates significativos no campo jurídico e acadêmico. O artigo intitulado *Como escrever um romance com ChatGPT*, publicado no site Mundo Escrito (2023), ilustra a viabilidade dessa colaboração entre humano e máquina, destacando exemplos de como *prompts* bem elaborados podem gerar narrativas completas com o suporte da inteligência artificial. A literatura especializada destaca a importância de distinguir diferentes formas de utilização da inteligência artificial no processo criativo. Em alguns casos, a IA atua como ferramenta auxiliar, como em correções gramaticais ou sugestões estilísticas; em outros, desempenha papel central na geração dos elementos criativos da obra. Essa distinção é fundamental para a análise jurídica da autoria, uma vez que o grau de intervenção humana influencia diretamente a configuração da proteção autoral.

O debate sobre autoria em obras geradas por sistemas computacionais não é recente, remontando a discussões anteriores ao desenvolvimento da inteligência artificial generativa contemporânea. Já na década de 1970, relatórios técnicos como o da National Commission on New Technological Uses of Copyrighted Works (Fuld, 1979) analisaram os desafios jurídicos relacionados à criação por computadores, especialmente no contexto de programas capazes de gerar resultados de forma automatizada. A partir da década de 1980, esses debates foram aprofundados pela doutrina autoral, especialmente em relação a obras criadas por computadores (Lee, 2024, p. 1452-1453). No entanto, a sofisticação atual dos modelos generativos, como os sistemas baseados em *prompt engineering*, impulsiona uma reavaliação urgente dos paradigmas legais vigentes. Globalmente, diferentes jurisdições e órgãos reguladores têm se posicionado, refletindo a complexidade do tema. Nesse cenário, a perspectiva de Lee (2024) é relevante ao defender que obras criadas com

o auxílio da inteligência artificial devem ser elegíveis para proteção autoral, desde que haja uma intervenção criativa humana mínima.

No Brasil, a situação é igualmente desafiadora: a Lei de Direitos Autorais (Lei nº 9.610/98) (BRASIL, 1998) foi elaborada em um contexto anterior ao surgimento da inteligência artificial generativa tal como hoje a conhecemos. Cabe ressaltar que a inteligência artificial já existia à época de sua elaboração, mas não os modelos generativos contemporâneos. A legislação exige uma figura física como autor, o que gera lacunas e incertezas diante da nova realidade de coautoria ou autoria assistida por inteligência artificial generativa. Diante dessa exigência legal, Schirru (2019, p. 2) levanta a hipótese de que as obras geradas através de sistemas de inteligência artificial, ou por quaisquer agentes não-humanos, não fariam jus à proteção autoral por não atender aos requisitos legais necessários, estando, portanto, imediatamente em domínio público. Isso reforça a necessidade urgente de reavaliar os paradigmas legais vigentes.

No ordenamento jurídico brasileiro, a expressão direitos autorais possui amplo sentido, englobando tanto os direitos de autor quanto os direitos conexos. Os direitos de autor referem-se diretamente à criação intelectual e ao vínculo entre o autor e a obra, conforme o Art. 11 da referida lei, que estabelece que "autor é a pessoa física criadora de obra literária, artística ou científica". Esses direitos compreendem tanto os direitos morais, que são inalienáveis e irrenunciáveis, relacionados à paternidade e à integridade da obra (Art. 24), quanto os direitos patrimoniais, que dizem respeito à exploração econômica da obra e podem ser transferidos ou licenciados a terceiros (Arts. 28 a 45). Por sua vez, os direitos conexos são atribuídos a sujeitos que, embora não sejam autores, participam da difusão ou interpretação da obra. Assim, embora toda obra deva ter um autor, nem sempre esse autor será o titular dos direitos patrimoniais dela decorrentes.

Além da Lei de Direitos Autorais, as discussões sobre a governança e proteção de dados emergem neste debate na medida em que a construção de Modelos de Linguagem de Grande Escala (LLMs) depende da utilização de vastos bancos de dados informacionais. Contudo, embora a conformidade com as bases legais de tratamento de dados seja um requisito ético indissociável da estruturação de inteligências artificiais, as implicações desses ecossistemas tecnológicos operam de forma proeminente sobre a seara da

propriedade intelectual e a tutela das criações intelectuais preexistentes utilizadas na fase de aprendizado das máquinas.

Nesse contexto, a problematização central deste ensaio reside na incerteza jurídica quanto à autoria de um romance concebido com o auxílio de um *chatbot* de IA: a quem, de fato, pertence o resultado criativo?

O debate acerca da titularidade de direitos sobre obras produzidas com auxílio de inteligência artificial já vem sendo discutido na doutrina jurídica contemporânea. Esses estudos demonstram que a problemática da autoria em contextos de produção assistida por IA envolve diferentes possibilidades interpretativas dentro do direito autoral. A partir dessa indagação, formulam-se três perguntas norteadoras para esta análise:

1. A autoria do conteúdo gerado pertence a quem escreveu o prompt?
2. Seria titular de direitos a empresa que fornece e mantém o serviço de *chatbot*?
3. Ou seriam os autores dos dados utilizados para treinar o modelo de IA os legítimos detentores de algum direito sobre a obra gerada?

Este estudo adota uma abordagem de pesquisa bibliográfica e documental, focando na análise da legislação brasileira. Para responder às perguntas de pesquisa, serão examinados os princípios da Lei de Direitos Autorais (Lei nº 9.610/98) (BRASIL, 1998), que define a autoria e os critérios para proteção da obra intelectual. A partir dessa base legal e doutrinária, serão construídas as argumentações para cada uma das teses de autoria, visando oferecer uma compreensão das implicações jurídicas.

A segurança jurídica desse processo é vital, especialmente no Brasil. A boa notícia é que o cenário tem ganhado atenção legislativa com a tramitação do Projeto de Lei nº 2338/2023, que, conforme sua ementa, "Dispõe sobre o uso da Inteligência Artificial" (BRASIL, 2023). Mais especificamente, o Art. 1º do Projeto visa estabelecer

[...] normas gerais de caráter nacional para o desenvolvimento, implementação e uso responsável de sistemas de inteligência artificial (IA) no Brasil, com o objetivo de proteger os direitos fundamentais e garantir a implementação de sistemas seguros e confiáveis, em benefício da pessoa

humana, do regime democrático e do desenvolvimento científico e tecnológico (BRASIL, 2023).

No entanto, apesar desses avanços, o texto legal não enfrenta de forma direta a questão da autoria de obras geradas por inteligência artificial, mantendo em aberto um dos principais desafios jurídicos no campo dos direitos autorais.

Cumprе assinalar que o Projeto de Lei avança significativamente ao incorporar uma seção dedicada às exceções de direitos autorais. O seu artigo 42 regulamenta que as atividades de mineração de textos e dados por sistemas automatizados não constituem ofensa a direitos autorais quando realizadas por organizações de pesquisa, de jornalismo, museus, arquivos e bibliotecas, desde que o uso ocorra na medida necessária e não prejudique a exploração normal das obras originais. Ademais, o diploma normativo traz disposições fundamentais sobre a transparência no ambiente digital: o artigo 7º garante o direito à informação prévia quanto às interações com sistemas de inteligência artificial, estabelecendo que o caráter automatizado deve ser informado, quando couber, por meio de ícones ou símbolos facilmente reconhecíveis. Complementarmente, o artigo 11 exige o envolvimento humano significativo e a determinação humana final em decisões algorítmicas que tenham impactos irreversíveis ou que envolvam riscos à vida ou à integridade física.

É justamente para instituir essas garantias e ferramentas de governança que o projeto se apresenta, buscando criar condições de previsibilidade. Afinal, enquanto não houver um marco legal consolidado no ordenamento jurídico, as controvérsias relacionadas ao uso da inteligência artificial continuarão sendo levantadas por criadores, desenvolvedores e usuários dessas tecnologias, sem que haja o respaldo normativo necessário para garantir segurança jurídica no Brasil.

1. DISCUSSÃO SOBRE A TITULARIDADE DE DIREITOS EM OBRAS GERADAS POR IA

1.1 A autoria de quem deu o comando no chatbot

A tese que atribui a autoria ao indivíduo que formula os comandos para o *chatbot*, um programa de inteligência artificial projetado para interagir com usuários por meio de conversas, simulando um diálogo humano, é a que mais se aproxima da concepção tradicional de autoria, além de estar em maior consonância com os princípios da Lei de Direitos Autorais brasileira. O argumento central sustenta que a inteligência artificial atua como uma ferramenta, um meio pelo qual a criatividade humana é articulada e concretizada. Assim como o pintor utiliza pincéis e tela para materializar sua visão artística, o *prompt*, instrução textual específica fornecida pelo usuário para orientar a geração de conteúdo pela IA, serve como instrumento para estruturar e expressar suas ideias. A autoria, portanto, configura-se como uma extensão da criatividade e da intenção comunicativa do indivíduo.

Essa perspectiva é reforçada, inclusive, pelo próprio site que ensina a escrever um romance com o auxílio do ChatGPT. Nele, o processo é descrito como bastante simples, permitindo que o autor conclua a primeira versão da obra em questão de horas ou até minutos. No entanto, ressalta-se que será necessário um período adicional, de alguns dias ou até semanas, para realizar as edições e a revisão do texto final.

Tal observação corrobora a noção de que a criação exige a atuação humana não apenas na formulação dos comandos, mas também na curadoria e no refinamento do texto gerado. A intervenção humana é constante e decisiva. O processo de criação de um romance com o ChatGPT, por exemplo, não se resume a um único comando. Pelo contrário, envolve uma série de interações: refinamento de *prompts*, seleção de trechos gerados, edição, reescrita e uma curadoria ativa que molda a narrativa, os personagens e o estilo, conforme indicado pelo próprio site Mundo Escrito. A intencionalidade artística, a escolha do enredo, o desenvolvimento dos arcos dramáticos e a coerência temática são decisões que

partem do humano. O texto produzido pela IA surge, assim, como uma resposta ao direcionamento criativo do *prompt*, e não como uma criação autônoma da máquina.

Corroborando essa visão, Lee (2024, p. 1525-1532, tradução nossa) resgata o histórico debate jurídico sobre a autoria fotográfica no final do século XIX, destacando que, à época, também se questionava se a fotografia poderia ser considerada uma obra de cunho autoral ou apenas uma reprodução mecânica da realidade. Críticos como Baudelaire chegaram a rejeitar seu valor artístico, alegando que o fotógrafo apenas apertava um botão, sem exercer criatividade. No entanto, juristas como Eugène Pouillet argumentaram que o autor da fotografia faz escolhas conscientes sobre enquadramento, iluminação, distância e perspectiva, decisões que envolvem julgamento estético e intenção comunicativa. Essa concepção prevaleceu, e o entendimento jurídico se consolidou no sentido de que a autoria está presente quando há concepção intelectual por parte do fotógrafo.

Nas palavras de Auroux (1998, p. 320), "nenhum instrumento (mesmo uma calculadora eletrônica ou um computador), por definição, é autônomo". Ainda que essas ferramentas possam formalizar e externalizar capacidades humanas, elas não substituem a competência humana de forma absoluta, mas a prolongam e transformam.

Não obstante os argumentos que equiparam o *prompt* a uma ferramenta de expressão criativa, é imperioso reconhecer a existência de uma sólida divergência doutrinária e pretoriana em nível internacional. Parte considerável dos estudiosos e das autoridades de propriedade intelectual sustenta que a contribuição do usuário por meio de comandos textuais não atinge o patamar mínimo de originalidade e controle sobre o resultado final necessário para configurar a autoria jurídica.

O principal expoente desse entendimento contraposto é o United States Copyright Office (USCO). Em suas diretrizes oficiais e decisões administrativas recentes, como no célebre caso da obra "Zarya of the Dawn", o órgão norte-americano firmou posicionamento no sentido de que elementos gerados de

forma autônoma por sistemas de IA generativa a partir de *prompts* não são passíveis de proteção por direitos autorais. O fundamento central reside na ausência de controle preditivo e factual sobre a expressão final da obra, uma vez que o usuário não dita um resultado específico e a máquina atua de forma imprevisível, impedindo que o humano seja considerado a 'mente criadora' (*master mind*) do resultado. No caso específico de “Zarya of the Dawn”, essa regra resultou na exclusão da proteção autoral sobre as imagens geradas pela IA Midjourney, embora o USCO tenha mantido os direitos autorais sobre os elementos textuais e a diagramação, que foram criados integralmente pelo esforço criativo da autora humana. Analisando o caso, o *Harvard Journal of Law & Technology Digest* sintetiza o papel do usuário diante do mecanismo de inteligência artificial:

O Copyright Office concluiu que as imagens geradas pela Midjourney não eram produto de autoria humana porque o usuário não exerce "controle preditivo" suficiente sobre os resultados do sistema. Em vez de atuar como uma ferramenta auxiliar tradicional conduzida por um artista humano, o sistema Midjourney opera de maneira mais próxima a um patrono ou cliente que comissiona um artista, fornecendo instruções gerais sobre o que a obra deve retratar, mas deixando a execução criativa real e os detalhes expressivos específicos a cargo da máquina. (HARVARD JOURNAL OF LAW & TECHNOLOGY DIGEST, 2023, tradução nossa).

Diante desse cenário, e sopesando as divergências apontadas, pode-se sustentar que a originalidade da obra, requisito essencial para sua proteção autoral, não reside propriamente no algoritmo, mas na complexidade, intencionalidade e inventividade dos comandos elaborados, bem como na intervenção humana que os acompanha. Contudo, permanece a indagação: até que ponto essa intervenção é suficiente para configurar uma autoria juridicamente protegida?

Sob a ótica legal, a Lei nº 9.610/98, em seu Art. 7º, define obras intelectuais como "criações do espírito, expressas por qualquer meio ou fixadas em qualquer suporte, tangível ou intangível" (BRASIL, 1998). A exigência de um autor pessoa natural para a titularidade de direitos autorais é central no direito brasileiro.

Pode-se então argumentar que o “espírito” da criação reside na atuação do *prompt*, que concebe a ideia original, orienta o sistema por meio de comandos específicos e imprime sua marca autoral no resultado por meio da curadoria, seleção, organização e edição dos conteúdos gerados. Essa “marca autoral” se manifesta na escolha de nuances narrativas, no desenvolvimento de personagens complexos ou na estruturação de enredos inovadores, que demonstram a intenção criativa do usuário.

Contudo, a linha entre uma intervenção humana substancial e um comando meramente funcional pode ser tênue. Obras que resultem de *prompts* excessivamente genéricos, com pouca ou nenhuma curadoria humana, poderiam ter sua originalidade questionada, dificultando a atribuição de autoria exclusiva ao “prompt”. Por exemplo, um *prompt* como “escreva um romance sobre amor” irá gerar um texto genérico, enquanto “crie um romance distópico onde a IA controla emoções humanas, com um protagonista que não aceita essa realidade e busca restaurar a individualidade, com um final feliz” demonstra uma intervenção e intencionalidade muito mais claras do usuário, se aproximando da expressão criativa individual.

1.2 A autoria da empresa que fornece o serviço de chatbot

A ideia de que a autoria de um romance gerado por inteligência artificial pertence à empresa que desenvolve e oferece o serviço de *chatbot*, como a *OpenAI* com o *ChatGPT*, encontra pouco respaldo legal. Um exemplo emblemático dessa controvérsia foi a ação judicial movida pela empresa de *mídia Raw Story Media* contra a *OpenAI*, julgada pela Corte Federal de Nova York. A juíza Colleen McMahon rejeitou a acusação de violação de direitos autorais, destacando que, dada a imensidão de dados utilizados no treinamento da IA, seria praticamente impossível comprovar a reprodução de conteúdos protegidos específicos, além da ausência de dano concreto à parte autora. Isso sublinha a complexidade da questão, pois os modelos de IA não replicam cópias exatas, mas sim aprendem com os dados para gerar novas informações, o que levanta dúvidas sobre a responsabilização de uma tecnologia que “imita” de forma indireta (laderoza, 2024).

A capacidade de gerar conteúdo é uma função inerente ao programa, criada e aprimorada pela própria empresa. No entanto, a IA carece de intenção criativa ou “espírito”,

elementos que são requisitos para a atribuição de autoria conforme a Lei de Direitos Autorais brasileira. A máquina, por mais sofisticada que seja, é incapaz de expressar subjetividade autoral. Essa limitação é reforçada pelo artigo 11 da Lei nº 9.610/98, que atribui a condição de autor exclusivamente a uma pessoa física (BRASIL, 1998).

Desse modo, a obra literária gerada com o auxílio da IA constitui, essencialmente, um resultado da interação entre a ferramenta e o discernimento humano. A IA, apesar de simular uma criação original, não é autônoma; ela processa informações pré-programadas, moldadas por dados de vivências humanas, mas sem qualquer consciência ou criatividade genuína. “Se quisermos pensar os instrumentos linguísticos nos termos modernos da IA, então devemos concebê-los como *extensões artificiais da inteligência humana*” (Auroux, 1998, p. 320).

Nesse contexto, os termos de uso de plataformas como o ChatGPT frequentemente estabelecem que a titularidade do conteúdo gerado permanece com os usuários. Contudo, tais disposições devem ser interpretadas com cautela, uma vez que refletem opções contratuais adotadas pelas empresas, não constituindo, por si sós, fundamento suficiente para a definição do enquadramento jurídico da autoria à luz da legislação vigente.

Assim, embora possam indicar uma tendência prática de atribuição de direitos ao usuário, esses instrumentos não substituem a análise jurídica propriamente dita, que deve considerar os critérios estabelecidos pelo ordenamento jurídico, especialmente no que se refere à exigência de autoria humana e aos requisitos de originalidade.

A posição de Luca Schirru (2019) oferece uma possível abordagem ao problema ao direcionar a análise para a natureza do produto final. Segundo o autor, por não atenderem aos requisitos legais de “criação do espírito” por uma pessoa física, os conteúdos desenvolvidos por agentes não humanos não fariam jus à proteção autoral, podendo ser compreendidos como integrantes do domínio público.

Nessa perspectiva, o domínio público não deve ser interpretado como um “cemitério de obras que perderam interesse”, mas como a “situação normal” da obra intelectual e um espaço de livre circulação e diálogo social. A ausência de proteção autoral sobre conteúdos gerados por IA, portanto, não configuraria um vácuo jurídico, mas um mecanismo de preservação do interesse público. Sob essa ótica, um domínio público robusto funcionaria

como insumo essencial para a criatividade humana, favorecendo a contínua produção de novas obras.

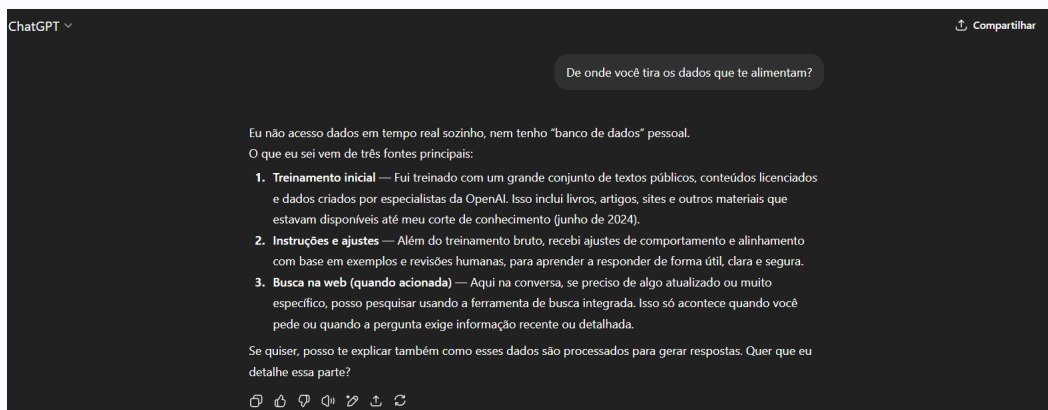
Schirru (2019) também alerta para o risco de desvirtuamento do direito autoral quando este passa a ser mobilizado prioritariamente como instrumento de proteção de investimentos econômicos das empresas desenvolvedoras. Ao sustentar a não incidência de proteção sobre os resultados da IA, preserva-se o paradigma antropocêntrico do direito brasileiro, que exige a contribuição humana como fundamento para a exclusividade. Nesse contexto, a compreensão do domínio público como espaço de circulação livre torna ainda mais relevante a discussão sobre a origem, a qualidade e a diversidade dos dados utilizados no treinamento desses sistemas.

A inteligência artificial (IA), especialmente os grandes modelos de linguagem, enfrenta desafios significativos em relação à transparência, principalmente no que diz respeito às fontes de dados e aos processos de tomada de decisão. A falta de clareza sobre como esses sistemas são treinados levanta preocupações éticas e de privacidade.

De acordo com a Política de Privacidade da *OpenAI*, a empresa coleta dados pessoais desde o momento da criação da conta para a utilização de seus serviços. A evolução dos modelos de linguagem demonstra um avanço notável nas capacidades da ferramenta, o que também aprofunda as questões sobre o uso de dados. Inicialmente, o ChatGPT não tinha a capacidade de acessar a internet em tempo real. No entanto, essa limitação foi superada com o lançamento do modelo GPT-4, que passou a permitir a navegação na web para fornecer informações mais atualizadas. Recentemente, a *OpenAI* anunciou o GPT-5, que promete um raciocínio ainda mais profundo para problemas complexos, respostas mais rápidas e uma utilidade maior para consultas do "mundo real". A empresa o descreve como "o colaborador de escrita mais competente até o momento", capaz de transformar ideias simples em textos convincentes e com "profundidade e ritmo literários".

Para entender melhor as origens de seu conhecimento, foi feita uma pergunta direta ao ChatGPT: "De onde você tira os dados que te alimentam?". A resposta da ferramenta oferece uma visão sobre suas fontes, embora mantenha uma certa generalidade:

Figura 1 – Resposta do ChatGPT sobre as fontes de dados de treinamento.



Fonte: Elaboração própria a partir de interação com ChatGPT (2025).

A imagem mostra a resposta do *chatbot*, que afirma ser treinado com "um grande conjunto de textos públicos, conteúdos licenciados e dados criados por especialistas da OpenAI", além de receber "instruções e ajustes" e poder realizar "busca na web (quando acionada)". No entanto, a resposta também destaca a falta de transparência sobre a origem e a seleção desses dados. Essa opacidade é crucial para o debate, pois os sistemas de IA operam recombinação de dados preexistentes de vastos bancos de dados. Mesmo com a afirmação do ChatGPT sobre a utilização de conteúdo licenciado, a ausência de detalhes sobre quais conteúdos foram usados e sob quais condições levanta sérias questões. Isso nos leva diretamente a um dos principais dilemas éticos e legais da IA generativa: o uso de material protegido por direitos autorais.

A falta de clareza sobre as fontes de dados alimenta a discussão sobre se esses modelos foram treinados ilegalmente com obras de artistas, escritores e criadores, violando seus direitos de propriedade intelectual. A autoria de uma obra gerada por IA, tema central deste trabalho, não pode ser analisada de forma isolada, pois está profundamente vinculada à origem dos dados que alimentaram o modelo. O questionamento "quem é o autor?" desdobra-se inevitavelmente em "de quem são os dados que possibilitaram essa autoria?"

1.3 A autoria de quem gerou os dados que alimentaram o modelo

Determinar a autoria aos criadores dos dados que serviram de base para o treinamento de modelos de inteligência artificial generativa é, provavelmente, o aspecto mais complexo e controverso no debate sobre direitos autorais de obras geradas por IA. O

ChatGPT foi treinado por um grande conjunto de dados que incluem textos, imagens, códigos e outras mídias, muitos dos quais protegidos por direitos autorais. A capacidade da IA de produzir linguagem natural, simular estilos e gerar narrativas coerentes deriva diretamente do processamento e da identificação de padrões presentes nesses dados preexistentes.

A IA não inventa ou cria algo do nada; ela sintetiza, recombina e reinterpreta informações e padrões extraídos de seu treinamento. Um romance gerado por um *chatbot*, portanto, pode apresentar fragmentos que guardam semelhança com as obras originais que compuseram o conjunto de dados que a treinou. Esse fenômeno aproxima-se do que a literatura especializada denomina “memorização”, isto é, situações em que modelos de inteligência artificial reproduzem, de forma parcial ou aproximada, conteúdos presentes nos dados de treinamento. A controvérsia reside em determinar se essa “recombinação” constitui uma nova obra original, uma obra derivada, ou se representa uma violação dos direitos autorais das fontes primárias.

A Lei nº 9.610/98 protege as obras intelectuais e, de forma específica, suas derivações (adaptações, traduções, arranjos musicais, etc.) (BRASIL, 1998). O Art. 29 da referida lei estabelece que o uso de obras preexistentes para a criação de obras derivadas exige, via de regra, a autorização prévia e expressa do autor da obra original (BRASIL, 1998). Nesse sentido, o processo de treinamento de uma IA, com a consequente geração de uma nova obra, pode ser equiparado a uma “cópia” ou “transformação” não autorizada, violando os direitos dos autores originais dos dados utilizados.

A simples agregação de dados para treinamento, mesmo que não resulte em cópias diretas, poderia ser interpretada como uma exploração econômica de conteúdo protegido sem a devida remuneração ou autorização. O principal desafio reside na dificuldade prática de rastrear e quantificar a contribuição de milhões de obras em um conjunto de dados para uma única obra gerada. Imputar a autoria a todos os criadores dos dados originais seria inviável.

Nesse sentido, destaca-se a importância de dados linguísticos estruturados e coletados dentro dos parâmetros do ordenamento de ética em pesquisa e da Lei Geral de

Proteção aos Dados, o que releva o trabalho de documentação linguística e sua importância para uma IA generativa ética e responsável (cf. Freitag et al., 2025).

A sociolinguística, com sua expertise em reconhecer a variação linguística e a heterogeneidade dos usos da língua em diferentes contextos sociais, tem um papel crucial na curadoria e na construção de amostras linguísticas mais diversas para o treinamento de LLMs. A aplicação de seus fundamentos pode mitigar os vieses existentes em bases de dados homogêneas, contribuindo para o desenvolvimento de sistemas de IA que sejam mais inclusivos, precisos e éticos, evitando a reprodução de preconceitos relacionados a gênero, raça, classe e regiões.

Além disso, a sociolinguística contribui diretamente para o cumprimento de preceitos legais e éticos, como a LGPD, ao propor metodologias para a coleta e o processamento de dados que preservam a privacidade dos falantes, e ao garantir que as amostras linguísticas sejam representativas da diversidade social e cultural, e não apenas de variedades linguísticas hegemônicas. A curadoria de dados sob uma perspectiva sociolinguística assegura que os LLMs não apenas reproduzam, mas também valorizem a complexidade e a riqueza da linguagem humana.

1.4 O papel crucial da sociolinguística na construção de LLMs éticos e seguros

Pode-se compreender a Inteligência Artificial como uma tentativa de reproduzir, por meio de programas computacionais, a mente humana e sua capacidade de processar informações. Enquanto nós, humanos, adquirimos a linguagem através de um processo complexo, que vai desde a exposição inicial a sons e palavras até a interação social e a prática constante, a IA funciona como um “cérebro digital”, aprendendo e tomando decisões a partir de dados. Nos casos de IA generativa que operam a partir de linguagem natural, esse processamento é possibilitado pelos LLMs (*Large Language Models*, ou grandes modelos de linguagem), responsáveis por executar tarefas de NLP (Processamento de Linguagem Natural).

Com os avanços dos *chatbots*, cresce a expectativa de que as respostas sejam cada vez mais humanizadas, capazes de construir narrativas completas e envolventes, como um romance convincente. No entanto, como discutido anteriormente, a IA não cria a

partir do nada: ela sintetiza, recombina e reinterpreta informações e padrões extraídos de um vasto banco de dados produzido por seres humanos.

Nesse cenário, a sociolinguística torna-se não apenas uma aliada técnica, mas também um pilar ético na construção de LLMs. Afinal, compreender a variação e o uso da língua em diferentes comunidades não é suficiente se o processo de coleta e seleção dos dados não for transparente, inclusivo e metodologicamente rigoroso. Ao integrar conhecimentos sociolinguísticos à curadoria de dados, torna-se possível desenvolver modelos que não apenas reproduzam interações mais próximas da comunicação humana real, mas que também respeitem a diversidade linguística e os princípios éticos, ao processo de coleta e curadoria de dados.

A sociolinguística, em seu conceito mais difundido, é a área que estuda a relação entre língua e sociedade. No Brasil, a vertente variacionista da sociolinguística é, por sua vez, uma área muito produtiva e abrangente no cenário brasileiro da pesquisa linguística. Contudo, essa abordagem apresenta diferenças em relação à proposta original de seu fundador, William Labov. A sociolinguística brasileira busca documentar a diversidade linguística por meio da construção e análise de bancos de dados linguísticos, coleções organizadas e sistematizadas de amostras de língua coletadas em comunidades de fala. Essas comunidades são formadas por grupos de pessoas que interagem regularmente e compartilham normas, padrões e expectativas sobre o uso da língua, utilizando uma variedade linguística comum e seguindo regras implícitas para diferentes contextos comunicativos.

Dessa forma, a sociolinguística brasileira se caracteriza pela sua operacionalização por meio desses bancos de dados linguísticos, que consistem em amostras sistematizadas coletadas em um universo delimitado como comunidade de fala (Freitag, 2016, p. 452). Na prática, isso significa que, para construir bases de treinamento que efetivamente representem a diversidade linguística, é necessário adotar procedimentos planejados de seleção de informantes e constituição de amostras que considerem variáveis como sexo/gênero, faixa etária, escolaridade, zona de residência e contexto sociocultural.

Para alcançar esse rigor metodológico e ético, a sociolinguística variacionista baseia-se em uma abordagem adaptada dos preceitos de William Labov, focando na

entrevista sociolinguística como principal instrumento para a geração de dados. Conforme detalhado no livro *Documentação sociolinguística: coleta de dados e ética em pesquisa* (Freitag, 2017, p. 13):

Dimensionada a amostra e identificados os informantes, o procedimento seguinte no processo de constituição do banco de dados é a realização da coleta. A estratégia adotada é a realização da entrevista sociolinguística, um protocolo desenvolvido e aprimorado com o objetivo de fazer com que o informante se desligue da situação de interação semidirigida diante de um gravador e monitore o mínimo sua fala, numa tentativa de fazer com que o seu vernáculo emergja (LABOV, 1972), minimizando os efeitos do “paradoxo do observador”: o pesquisador precisa se inserir na comunidade para observar o uso linguístico vernacular, mas a sua inserção na comunidade interfere no uso linguístico vernacular que vai ser observado.

Para que o pesquisador obtenha uma coleta eficaz, a pesquisa sociolinguística adota uma série de estratégias éticas e metodológicas. Freitag (2017) detalha protocolos que incluem a abordagem cuidadosa ao informante, a seleção criteriosa dos participantes e o estabelecimento de um código de conduta para o documentador, com o objetivo de garantir a representatividade e a autenticidade da amostra linguística.

Além disso, a proteção ética dos participantes é uma preocupação central, alinhada diretamente aos preceitos da LGPD. As diretrizes de pesquisa exigem que o consentimento livre e esclarecido dos informantes seja obtido e que os projetos passem por comitês de ética, conforme indica Freitag (2017). Essa abordagem rigorosa garante que o processo de coleta de dados não apenas seja metodologicamente sólido, mas também respeite a privacidade e a autonomia dos indivíduos, estabelecendo uma base segura e legal para a construção de qualquer banco de dados linguístico, incluindo aqueles usados para o treinamento de LLMs.

A conexão fundamental entre LLMs e dados linguísticos está justamente nesse processo de aprendizado a partir de grandes bases textuais, que refletem a diversidade e as variações da língua em uso. Quanto mais representativos e variados forem os dados linguísticos utilizados, maior será a capacidade do modelo em lidar com diferentes formas de linguagem e contextos comunicativos. Desse modo, as contribuições dos linguistas podem tornar o treinamento de uma IA mais eficiente, contemplando diferentes variedades

linguísticas por meio de amostras representativas. Em última análise, a sociolinguística não apenas reforça a importância da diversidade, mas oferece os métodos e critérios necessários para incorporá-la de forma concreta às bases de uma tecnologia que está redefinindo a comunicação humana.

CONCLUSÃO

A inteligência artificial generativa tem remodelado, de forma decisiva, as fronteiras da criação intelectual, desafiando os pressupostos clássicos de autoria e os marcos regulatórios dos direitos autorais. No contexto jurídico brasileiro, a atribuição de autoria a obras produzidas por sistemas de IA evidencia um campo normativo ainda incipiente, marcado por lacunas e incertezas decorrentes da ausência de regulamentação específica.

Este trabalho analisou as principais correntes acerca da titularidade de direitos em obras geradas com o auxílio de inteligência artificial, demonstrando que a Lei de Direitos Autorais (Lei nº 9.610/98) (BRASIL, 1998), ao exigir a figura de uma pessoa física como autora, encontra limitações para abarcar a complexidade da criação assistida por IA. A partir da análise desenvolvida, conclui-se que a tese mais compatível com o ordenamento jurídico brasileiro é a da autoria do usuário, em razão de sua intencionalidade, intervenção criativa e atuação curatorial no processo de produção do conteúdo.

A relevância desse debate ultrapassa a proteção de obras literárias, alcançando toda a cadeia produtiva da economia criativa. A ausência de diretrizes normativas claras compromete a segurança jurídica e impacta diretamente criadores, desenvolvedores e usuários dessas tecnologias. Nesse cenário, a tramitação do Projeto de Lei nº 2338/2023 representa um avanço importante, ainda que insuficiente diante da complexidade do tema.

Diante disso, torna-se imprescindível a atualização das políticas de propriedade intelectual, de modo a garantir a proteção dos criadores, a justa remuneração e o estímulo à inovação, sem prejuízo do desenvolvimento tecnológico. A construção de um marco regulatório eficaz demanda soluções que conciliem os princípios tradicionais dos direitos autorais com as especificidades da inteligência artificial.

Por fim, destaca-se o papel fundamental da sociolinguística na construção de inteligências artificiais mais éticas e representativas, especialmente no que se refere à

curadoria de dados linguísticos diversos e socialmente situados. A articulação entre direito, sociolinguística e ciência da computação mostra-se, portanto, essencial para o desenvolvimento de sistemas de IA que sejam não apenas tecnicamente avançados, mas também juridicamente seguros e socialmente responsáveis.

REFERÊNCIAS

AUROUX, Sylvain. **A filosofia da linguagem**. Campinas: Editora da Unicamp, 1998.

BONI, Ana Paula. Direitos autorais: Inteligência Artificial altera criação de conteúdo. **KLA Law**, São Paulo, 9 set. 2024. Disponível em:
<https://klalaw.com.br/direitos-autorais-inteligencia-artificial-criacao-conteudo/>. Acesso em: 21 jun. 2025.

BRASIL. **Lei nº 9.610, de 19 de fevereiro de 1998**. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. Brasília, DF: Presidência da República. Disponível em:
https://www.planalto.gov.br/ccivil_03/leis/l9610.htm. Acesso em: 21 jun. 2025.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em:
https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 9 ago. 2025.

BRASIL. **Projeto de Lei n. 2.338, de 2023**. Dispõe sobre o uso da Inteligência Artificial. Brasília, DF: Senado Federal, 2023. Disponível em:
<https://legis.senado.leg.br/sdleg-getter/documento?dm=9347622>. Acesso em: 21 jun. 2025.

FREITAG, Raquel Meister Ko. **Sociolinguística no/do Brasil**. Cadernos de Estudos Linguísticos, v. 58, n. 3, p. 445-460, 2016.

FREITAG, Raquel Meister Ko. **Documentação sociolinguística**: coleta de dados e ética em pesquisa. São Cristóvão: Editora UFS, 2017.

FREITAG, Raquel Meister Ko et al. **Plataforma da Diversidade Linguística Brasileira**: dados linguísticos para uma IA brasileira. Cadernos de Linguística, v. 6, n. 4, p. e863-e863, 29 dez. 2025.

FULD, Stanley H. **Final Report of the July 31, 1978**. National Commission on New Technological Uses of Copyrighted Works. National Commission on New Technological Uses of Copyrighted Works, Washington, DC., 1979. Disponível em:
<https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/PB85225621.xhtml>. Acesso em: 18 mar. 2026.

HARVARD JOURNAL OF LAW & TECHNOLOGY DIGEST. **Zarya of the Dawn**: How AI is Changing the Landscape of Copyright Protection. Harvard Journal of Law & Technology (JOLT), Cambridge, 1 mar. 2023. Disponível em:
<https://jolt.law.harvard.edu/digest/zarya-of-the-dawn-how-ai-is-changing-the-landscape-of-copyright-protection>. Acesso em: 17 jun. 2026.

IADEROZA, Thiago Chiavegatto. Decisão judicial rejeita ação de violação de direitos autorais contra OpenAI. **JOTA**, Brasília, 23 nov. 2024. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/decisao-judicial-rejeita-acao-de-violacao-de-direitos-autorais-contr-openai>. Acesso em: 21 jun. 2025.

LABOV, William. **Sociolinguistic patterns**. Philadelphia: University of Pennsylvania Press, 1972.

LEE, Edward. Prompting progress: authorship in the age of AI. **Florida Law Review**, Gainesville, v. 76, n. 5, p. 1445–1549, 2024. Disponível em: <https://scholarship.law.ufl.edu/flr/vol76/iss5/5>. Acesso em: 21 jun. 2025.

MUNDO ESCRITO. **Como escrever um romance com o ChatGPT?** Veja, passo a passo. Mundo Escrito, [S. l.], 20 set. 2023. Disponível em: <https://mundoescrito.com.br/como-escrever-um-romance-com-chatgpt/>. Acesso em: 21 jun. 2025.

OPENAI. ChatGPT. **Resposta sobre a transparência de dados de treinamento**. Conversa com a Inteligência Artificial. Data da conversa: 13 ago. 2025.
OPENAI. Introducing GPT-5. Disponível em: <https://openai.com/index/introducing-gpt-5/>. Acesso em: 13 ago. 2025.

OPENAI. **Política de privacidade**. Disponível em: <https://openai.com/pt-BR/policies/row-privacy-policy/>. Acesso em: 13 ago. 2025.

OPENAI. **Termos de uso**. Disponível em: <https://openai.com/pt-BR/policies/row-terms-of-use/>. Acesso em: 13 ago. 2025.

PRASS, Rochele Moura; MÜGGE, Ernani; BERNASIUK, Helen Lentz Ribeiro. Autoria em tempos de inteligência artificial generativa: um olhar para a produção ficcional contemporânea no Brasil. **Texto Digital**, Florianópolis, v. 19, n. 2, p. 75–106, 20 dez. 2023.

SCHIRRU, Luca. **Inteligência Artificial e o Direito Autoral: O Domínio Público em Perspectiva**, 22 mar. 2019. Disponível em: <https://itsrio.org/wp-content/uploads/2019/04/Luca-Schirru-rev2-1.pdf>. Acesso em: 30 jun. 2026.

UNITED STATES COPYRIGHT OFFICE. **Re: Zarya of the Dawn** (Registration # VAu001480196). Decisão administrativa em sede de procedimento de cancelamento de registro. Washington, DC: Library of Congress, 21 fev. 2023. Disponível em: <https://www.copyright.gov/docs/zarya-of-the-dawn.pdf>. Acesso em: 17 jun. 2026.