



UNIVERSIDADE DO ESTADO DA BAHIA
DEPARTAMENTO DE TECNOLOGIAS E CIÊNCIAS SOCIAIS – CAMPUS III

ARIÂNELY DE ALMEIDA MATOS NASCIMENTO

DIREITOS AUTORAIS NA ERA DIGITAL:
Uma análise sobre a proteção de obras intelectuais utilizadas no
processo de treinamento das inteligências artificiais generativas

JUAZEIRO – BA
2024

ARIÂNELY DE ALMEIDA MATOS NASCIMENTO

DIREITOS AUTORAIS NA ERA DIGITAL:

Uma análise sobre a proteção de obras intelectuais utilizadas no processo de treinamento das inteligências artificiais generativas

Trabalho de Conclusão de Curso – TCC apresentado como requisito parcial para obtenção do título de Bacharel em Direito no Curso de Direito do *Campus* III da Universidade do Estado da Bahia (UNEB).

Orientador(a): Profa. Ma. Mary Monalisa de Carvalho Costa

JUAZEIRO – BA

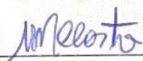
2024

UNIVERSIDADE DO ESTADO DA BAHIA
Autorização Decreto nº 9237/86. DOU 18/07/96. Reconhecimento: Portaria 909/95, DOU 01/08-95
DEPARTAMENTO DE TECNOLOGIA E CIÊNCIAS SOCIAIS- CAMPUS III
COLEGIADO DO CURSO DE DIREITO

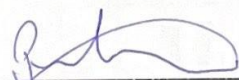


ATA DE DEFESA PÚBLICA DE MONOGRAFIA

Aos onze dias do mês de julho do ano de dois mil e vinte e quatro no Campus II, do DTCS, da UNEB, Juazeiro - BA, reuniram-se sobre a Presidência do(a) Professor(a), orientador(a) Mary Monalisa de Carvalho Costa os professores, Pedro Henrique Matos Souza de Santana, Juliana Cavalcanti Santiago e o(a) Bacharelado(a) **ARIANELY DE ALMEIDA MATOS NASCIMENTO**, que procedeu, em sessão pública, a apresentação de monografia para conclusão de curso, cujo tema versou sobre DIREITOS AUTORAIS NA ERA DIGITAL: UMA ANÁLISE SOBRE A PROTEÇÃO DE OBRAS INTELECTUAIS UTILIZADAS NO PR, sendo a audiência iniciada às 15h (quinze horas), durando a explanação 30 (trinta) minutos, seguindo-se de perguntas elaboradas pelos examinadores, que ao final atribuíram as seguintes notas, respectivamente: 10,0 (dez), 9,0 (nove inteiro) e 10,0 (dez), sendo, assim, obtida a média final 9,7 (nove inteiro e sete décimos). Nada mais havendo foi encerrada a presente Ata, que vai devidamente assinada.



Presidente/orientador



Docente/arguidor



Membro

Dedico este trabalho a minha mãe que sempre me ensinou a trilhar a vida pelo caminho da educação.

AGRADECIMENTOS

Nada mais justo que iniciar essa sessão de agradecimentos pela força maior que move o mundo: Deus. Obrigada por me ter permitido chegar até esse momento.

Agradeço também a minha família, principalmente, a minha mãe, Ariádene, por todo o suporte e incentivo que me deu desde o início dessa trajetória. A meu pai Alex Sandro, pelo apoio que tem me dado ao longo dessa caminhada.

A minha avó Ana Maria, que além de sempre ter sábias palavras para mim, inspira-me diariamente a dar o meu melhor. A meu tio Aristóteles, por toda ajuda. E a minha irmã Aléxia, pelos momentos de distração.

Da mesma forma, sou muito grata aos docentes do curso de Direito da UNEB, do *Campus* III, por todo o conhecimento partilhado. Agradeço ainda, em especial, a minha professora orientadora, Mary Monalisa, por ter aceitado o desafio que foi esse trabalho de pesquisa.

Acrescento ainda minha gratidão aos amigos que fiz durante o percorrer desses anos, especialmente a Anna Thereza, que esteve comigo nos meus momentos mais tristes e nos mais alegres também.

Também sou grata as minhas supervisoras do estágio no MPBA, Dra. Patrícia e Moara por me ensinarem a aplicar o Direito na prática. Jamais irei esquecer de todo o aprendizado que adquiri com vocês.

E, por fim, agradeço a todos os autores que li, artistas que criaram ou performaram obras que vi e musicistas que ouvi. Vocês são essenciais ao mundo, pois o tornam especial, ao inspirar e trazer à tona reflexão para a humanidade.

Sem cultura e sem a relativa liberdade que ela implica, a sociedade, mesmo quando perfeita, não passa de uma selva. É por isso que qualquer criação autêntica é um presente para o futuro.

– Albert Camus

RESUMO

O presente trabalho de conclusão de curso, cuja temática abrange o direito autoral ante o avanço tecnológico proporcionado pela era digital, tem como finalidade discutir a questão relativa à utilização de obras intelectuais, que se encontram sob domínio privado, sem autorização prévia de seus criadores, no processo de treinamento das inteligências artificiais generativas. Para tanto, é abordado, aspectos fundamentais relativos ao conceito, classificação e funcionamento das inteligências artificiais generativas. Posteriormente, também é retratado as particularidades que dizem respeito ao contexto histórico, sistemas de adoção do direito autoral e concepções específicas a respeito da matéria na legislação brasileira. Por fim, é discutido também o cenário jurídico que envolve a problemática que inspira este estudo, trazendo uma análise sobre a questão em âmbito internacional, através de duas das principais ações que tramitam nos Estados Unidos acerca da temática e com a criação do Regulamento da Inteligência Artificial pela União Europeia, bem como na seara nacional, examinando aspectos relativos ao Projeto de Lei nº 2.338/2023, o qual visa a criação de regulamentação específica para as inteligências artificiais no Brasil, tratando, conseqüentemente, sobre o tema que envolve este estudo.

Palavras-chave: Direito autoral; Inteligência artificial; Mineração de dados.

ABSTRACT

This course conclusion work, whose theme covers copyright in the face of technological advances provided by the digital era, aims to discuss the issue relating to the use of intellectual works, which are under private domain, without prior authorization from their creators, in the training process of generative artificial intelligence. To this end, fundamental aspects relating to the concept, classification and functioning of generative artificial intelligence are addressed. Subsequently, particularities relating to the historical context, copyright adoption systems and specific conceptions regarding the matter in Brazilian legislation are also portrayed. Finally, the legal scenario surrounding the issue that inspires this study is also discussed, providing an analysis of the issue at an international level, through two of the main actions processing in the United States that involves this subject and with the creation of the Artificial Intelligence Act by the European Union, as well as in the national field, examining aspects related to Project of Bill No. 2.338/2023, which aims to create specific regulations for artificial intelligence in Brazil, dealing, consequently, with the topic that involves this study.

Key-words: Copyright; Artificial intelligence; Data mining.

LISTA DE ABREVIATURAS E SIGLAS

AGI – *Artificial General Intelligence*;
AIGC – *Artificial Intelligence Generated Content*;
ANI – *Artificial Narrow Intelligence*;
ART – Artigo;
ASI – *Artificial Superintelligence*;
CTIA – Comissão Temporária sobre Inteligência Artificial;
DL – *Deep Learning*;
EUA – Estados Unidos da América;
GANs – *Generative Adversarial Networks*;
GPT – *Generative Pre-trained Transformer*;
IA – Inteligência Artificial;
IAs – Inteligências Artificiais;
KDD – *Knowledge Discovery in Databases*;
LDA – Lei de Direitos Autorais;
LLM – *Large Language Model*;
ML – *Machine Learning*;
NLP – *Natural Language Processing*;
NYTC – New York Times Company;
PL – Projeto de Lei;
RLHF – *Reinforced Learning from Human Feedback*;
SDNY – United States District Court for the Southern District of New York;
SIA – Sistema Nacional de Regulação e Governança de Inteligência Artificial;
UE – União Europeia;
VAEs – *Variational Autoencoders*.

SUMÁRIO

INTRODUÇÃO	11
1. A RESPEITO DA INTELIGÊNCIA ARTIFICIAL GENERATIVA	14
1.1. Conceitos fundamentais e breve histórico das IAs generativas	15
1.2. Funcionamento das IAs generativas	18
1.3. O processo de mineração de dados.....	21
2. SOBRE OS DIREITOS AUTORAIS	24
2.1. Contextualização histórica do direito autoral	25
2.2. Progressão legislativa da matéria no Brasil.....	28
2.3. Direitos morais e patrimoniais do autor conforme a LDA.....	29
3. PERSPECTIVAS E DESAFIOS DOS DIREITOS AUTORAIS DIANTE DOS SISTEMAS DE INTELIGÊNCIAS ARTIFICIAIS GENERATIVAS	35
3.1. Alguns litígios nos EUA envolvendo <i>Copyright</i> e IA generativa	36
3.2. Regulamentação da IA pela União Europeia	39
3.3. Situação no Brasil: o Projeto de Lei nº 2.338/2023 em debate	41
CONSIDERAÇÕES FINAIS	49
REFERÊNCIAS	52

INTRODUÇÃO

É perceptível que no decorrer das últimas décadas, a sociedade tem passado por inúmeras mudanças sendo o elemento central para entender a complexidade dessas transformações a revolução tecnológica. O avanço da tecnologia junto ao processo cada vez mais ascendente da globalização foram essenciais para isso: vive-se a era digital, repleta de praticidade, dinamismo e versatilidade. Nesse contexto, constata-se a busca cada vez mais ávida pela apresentação de inovações tecnológicas ao mundo.

Verifica-se, portanto, que a constante demanda em relação à necessidade de transformações tecnológicas resultaram, nos últimos anos, na propagação das chamadas inteligências artificiais: tecnologia que permite a simulação da inteligência humana por computadores ou máquinas. Assim, a partir do desenvolvimento e consequente disponibilização desses sistemas de IAs para a sociedade, observa-se cada vez mais a expansão sobre a utilização dessas tecnologias em ritmo acelerado pelos mais diferentes setores da sociedade.

Diariamente as pessoas que têm acesso às novas tecnologias vêm adotando os sistemas inteligentes em suas vidas, seja para planejar rotas com o Waze, receber sugestões de filmes pela Netflix, ou recomendações de *playlists* pelo Spotify, ou ainda, para solicitar auxílio a assistentes virtuais que respondem através de comandos de voz como a Alexa, da Amazon. E, como se não bastasse o avanço trazido pelos exemplos mencionados, agora emergem as IAs com capacidade de gerar textos, músicas e imagens que chegam a ser esteticamente idênticas aos de trabalhos realizados por seres humanos: essas são as inteligências artificiais generativas.

As IAs generativas, diferentemente das IAs tradicionais, são capazes de gerar conteúdo com base em comandos de usuários. Mas essa capacidade de “criar” só é possível graças ao treinamento prévio que recebem por meio do acesso a milhões de dados armazenados em bases ou conjunto de dados, contendo os mais variados conteúdos possíveis, incluindo-se, nesse contexto, livros, artigos jornalísticos, obras artísticas e musicais, que, em grande parte, estão sob proteção de direitos autorais.

No entanto, acontece que essas obras protegidas por direitos autorais estão sendo utilizadas de forma desregulada para guiar esse processo de treinamento das IAs generativas, o que tem causado discussão no âmbito jurídico internacional e nacional. Isso porque, em regra, o uso de obras intelectuais, que se encontram em domínio privado, necessita de autorização prévia de seus criadores para serem exploradas. Trata-se de uma forma de defesa do direito patrimonial do autor a quem cabe, primordialmente, o direito exclusivo de fruir economicamente da sua criação.

Sob esse aspecto, inclusive, no que tange à legislação brasileira sobre direitos autorais, a Lei nº 9.610/1998 estabelece no art. 29 que é preciso a autorização prévia e expressa do autor em qualquer modalidade de utilização de sua obra. Dessa forma, entende-se que pela norma brasileira, a utilização indiscriminada de obras protegidas para o processo de treinamento das IAs generativas, viola preceitos basilares do direito patrimonial do autor, causando-lhe prejuízo.

Nesse sentido, debater medidas apropriadas relacionadas a essa problemática, se faz necessário, visto que os direitos autorais devem ser respeitados, uma vez que é através deles que são garantidos aos criadores a valorização e reconhecimento sobre seu trabalho, seja ele de cunho literário, artístico ou científico. Afinal, deve-se admitir que o processo criativo demanda não somente uma enorme quantidade de esforço mental para sua produção, como também de empenho prático para ser realizado. Logo, proteger o produto resultado da criatividade é indispensável, inclusive para incentivar a atividade executada por autores, pintores e musicistas na sociedade.

Diante desse cenário, verifica-se que a temática apresentada é relevante para a atualidade, uma vez que se tem notado paulatinamente o avanço das IAs generativas nos diversos setores da sociedade. Nesse contexto, evidencia-se que o consequente dilema envolvendo o desenvolvimento dessas novas tecnologias e o uso indiscriminado de obras protegidas por direitos autorais para o treinamento delas, representa a justificativa que move o presente trabalho de pesquisa, posto que se faz necessário promover reflexões a respeito dessa questão.

Assim, o presente estudo, embasado na metodologia da pesquisa bibliográfica, a qual permite o levantamento, a seleção e a análise de diferentes conceitos referentes ao tema abordado, apresenta como objetivo geral analisar os direitos autorais dos criadores que têm suas obras intelectuais utilizadas no processo de treinamento dos sistemas de inteligência artificial generativa. Para tanto, ainda se faz

necessário apresentar os objetivos específicos, os quais são essenciais para uma compreensão mais detalhada sobre a problemática e que correspondem a cada um dos capítulos desenvolvidos neste trabalho de pesquisa.

Desse modo, no primeiro capítulo, busca-se investigar como funcionam esses sistemas de IAs generativas, entendendo seus conceitos elementares, breve histórico de criação, funcionamento e processo de treinamento. Já no segundo capítulo, pretende-se examinar aspectos essenciais relativos aos direitos autorais, como seu conceito, seu surgimento, sua evolução no contexto internacional e nacional, bem como sua regulamentação no Brasil.

E, finalmente, no terceiro capítulo objetiva-se debater a questão relativa ao uso de obras intelectuais, sob domínio privado, no processo de treinamento dos sistemas de IAs generativas, estabelecendo relações com as concepções do direito autoral. Assim, trazendo à tona, a problemática que tem motivado discussões jurídicas sobre a proteção dos direitos patrimoniais dos autores não só no Brasil, como também nos Estados Unidos e na União Europeia.

1. A RESPEITO DA INTELIGÊNCIA ARTIFICIAL GENERATIVA

A maior parte dos progressos tecnológicos alcançados no decorrer dos últimos anos é resultado do que cientistas computacionais definem como *machine learning*, método estatístico que permite às máquinas passarem por um processo de "aprendizagem" a partir de dados. Esse mecanismo integra a seara do que se conhece como inteligência artificial, a qual, por sua vez, deve ser compreendida como “[...] a capacidade de um sistema de obter resultados de forma ilimitada e independente, pois é dessa maneira que um sistema digital de processamento de informações pode de fato emular a inteligência humana” (Santos, 2020, p. 28).

Sendo, portanto, “[...] campo da ciência destinado a fornecer máquinas com a capacidade de realizar funções como lógica, raciocínio, planejamento, aprendizagem e percepção [...]” (Gabriel, 2022, p. 56), a inteligência artificial, é classificada em três níveis evolutivos: Limitada (ANI – *Artificial Narrow Intelligence*), Geral (AGI – *Artificial General Intelligence*) e Superinteligência (ASI – *Artificial Superintelligence*). No entanto, ressalta-se que na atualidade apenas a inteligência artificial limitada está disponível na prática, ou seja, por enquanto, a inteligência artificial geral e a superinteligência artificial são apenas previstas no âmbito teórico.

Tratando-se, portanto, da inteligência artificial limitada que está sendo de fato desenvolvida na prática, evidencia-se que até o presente momento, esta se subdivide em duas espécies: a IA tradicional e a IA generativa. Nesse sentido, evidencia-se que a IA tradicional, é assim chamada por já ser amplamente estudada e por ter sido progressivamente adotada por gigantes da tecnologia nas últimas décadas. Assim sendo, as IAs tradicionais caracterizam-se por serem treinadas por um vasto conjunto de dados empregados para instruir determinado modelo a executar tarefas específicas, como reconhecimento e comando por voz: um exemplo clássico seria, portanto, as assistentes virtuais como a Siri da Apple ou a Bixby da Samsung.

Por outro lado, a IA generativa, embora também seja treinada por uma quantidade absurda de dados, ela diferencia-se da tradicional, por ser capaz de criar conteúdo como textos, imagens ou músicas com base nos dados aos quais foi instruída. Este tipo de IA por ser relativamente nova no cenário tecnológico atual, ainda está em constante aprimoramento, mas ganhou grande repercussão após o lançamento do ChatGPT em novembro de 2022 pela OpenAI: a popularidade do

mencionado *chatbot* foi tamanha que “com menos de dois meses de existência [...] ultrapassou a marca dos 100 milhões de usuários” (Caparroz, 2023).

Foi com o sucesso do ChatGPT que se passou a haver uma intensificação do interesse sobre a inteligência artificial generativa. Diante desse contexto, e considerando-se o objetivo geral a ser alcançado neste trabalho, visa-se analisar nos tópicos a seguir, não somente o conceito, mas também o breve histórico relativo a criação e o funcionamento dessa espécie de IA.

1.1. Conceitos fundamentais e breve histórico das IAs generativas

Entende-se inteligência artificial generativa como uma espécie de “[...] modelo de *machine learning* treinado para criar dados novos, em vez de fazer uma previsão sobre um conjunto de dados específicos” (Zewe, 2023, tradução nossa). Desse modo, “um sistema de IA generativo é aquele que aprende a gerar conteúdo similar em conformidade com os dados nos quais foi treinado” (*ibidem*). No entanto, embora o alvoroço sobre as inteligências artificiais generativas tenha surgido com o lançamento do ChatGPT, a tecnologia em si não é totalmente nova.

Em verdade, as IAs generativas possuem estreita ligação com os processos estatísticos desenvolvidos em meados da década de 1950 como os Modelos Ocultos de Markov e os Modelos de Mistura Gaussiana (Cao *et al.*, 2018). Os mencionados modelos foram capazes de gerar dados sequenciais, dando início ao desenvolvimento tecnológico sobre a temática, contudo, destaca-se que foi com o advento do *deep learning* que houve avanço quanto ao desenvolvimento das inteligências artificiais generativas.

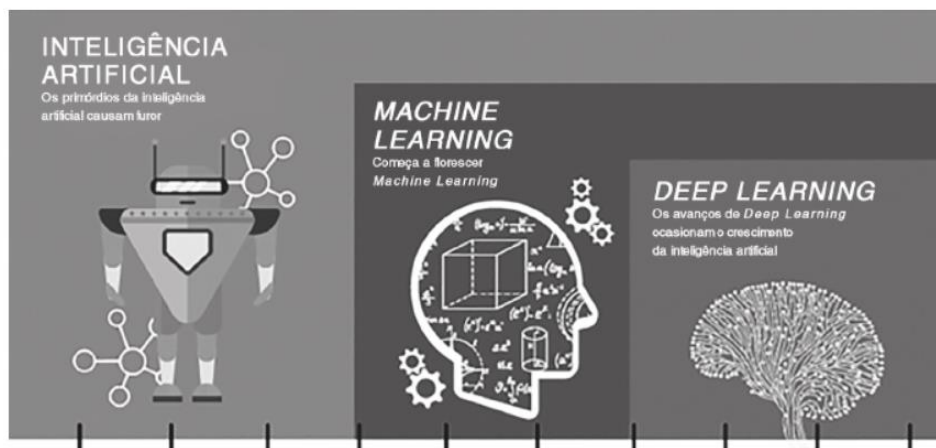
O *deep learning* (DL), é uma subárea do *machine learning* (ML), sendo necessário, portanto, entender este último para compreender o primeiro. Assim, conforme apresentado no início deste capítulo, o *machine learning* que em sua tradução literal significa “aprendizado de máquina”, caracteriza-se por ser um método algorítmico diferenciado por meio do qual “[...] os programadores humanos não precisam especificar um código que determine as ações ou previsões que o programa deva realizar em certa situação” (Gabriel, 2022, p. 73), ou seja, no caso da aplicação do ML, o próprio programa computacional, por meio da análise de dados, “aprenderá

a melhorar seu desempenho após fazer observações sobre o mundo” (Russel; Norving, 2022, p. 590).

Sob esse aspecto, Arthur Samuel (1959, p. 585, tradução nossa), pioneiro na área, destacou que o fato de “programar computadores para aprender com a experiência deverá eventualmente eliminar a necessidade de grande parte deste esforço de programação detalhada”, o que, de fato, é um avanço tecnológico notável na área da computação. Nesse sentido, o ML, como subcampo da inteligência artificial, abrange também uma série de outros métodos de aplicação algorítmica, que por sinal, “[...] são muitos, cada qual com aplicações em domínios específicos de atuação, conforme o contexto e o problema a ser resolvido” (Gabriel, 2022, p. 73).

É nesse cenário, entre os mais variados métodos algorítmicos utilizados pela *machine learning*, que se tem o *deep learning*, ou seja, este último, que em sua tradução literal significa “aprendizado profundo”, é subcampo do ML, este que por sua vez, é subcampo da inteligência artificial (Figura 1). Sabendo disso, evidencia-se que o DL se caracteriza por usar “[...] algoritmos de redes neurais artificiais, aprofundando o processamento em camadas de neurônios artificiais para resolver problemas mais complexos, aproximando-se mais do que entendemos por ‘pensamento’ humano” (*ibidem*, p. 86).

Figura 1 – Evolução da IA desde o seu início como disciplina do conhecimento, passando pelo surgimento de metodologias e algoritmos de ML até o florescimento do *Deep Learning*.



Fonte: Gabriel (2022)

É, portanto, a partir do *deep learning*, que a inteligência artificial generativa passa a ter seu desenvolvimento alavancado. Afinal, conforme é destacado por Cao

et al. (2018), antes do DL, os métodos algorítmicos utilizados como base para os programas computacionais eram limitados, não tendo, por exemplo, uma alta capacidade para gerar imagens ou textos com alto grau de complexidade. Nesse sentido, destaca-se que o grande marco significativo sobre a seara da inteligência artificial generativa veio em 2014 com a criação das Redes Adversariais Generativas (em inglês: *Generative Adversarial Networks* – GANs), as quais apresentaram resultados impressionantes sobre processamento de imagens. Por conseguinte,

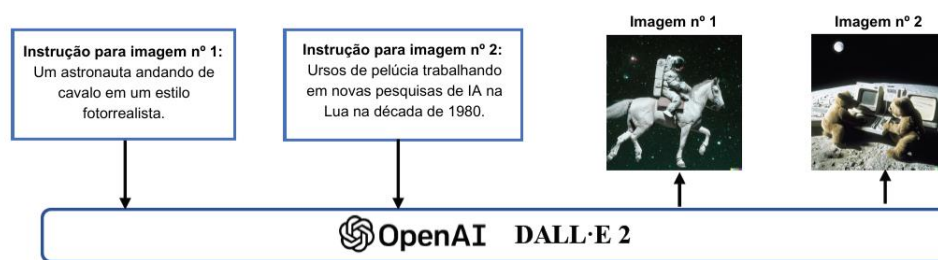
Variational Autoencoders (VAEs) e outros métodos como modelos generativos de difusão também foram desenvolvidos para um controle mais refinado sobre o processo de geração de imagens e para a capacidade de gerar imagens de alta qualidade (Cao *et al.* 2018, p. 4, tradução nossa)

A respeito dos “Modelos de Difusão”, salienta Adam Zewe (2023) que eles foram criados por pesquisadores da Universidade de Stanford e da Universidade da Califórnia, sendo desenvolvidos com a capacidade de aprimorar seus próprios resultados, gerando “[...] novas amostras de dados que se assemelham as de um conjunto de dados em treinamento e têm sido usados para criar imagens de aparência realista” (*ibidem*, tradução nossa). Posteriormente, em 2017, a inteligência artificial generativa subiu mais um nível de avanço tecnológico com a criação do modelo *Transformer*, “[...] uma rede neural que aprende o contexto e, assim, o significado com o monitoramento de relações em dados sequenciais como as palavras desta frase” (Merritt, 2022).

De acordo com Cao *et al.* (2018 p. 4, tradução nossa), o *Transformer* “[...] se tornou a espinha dorsal dominante para muitos modelos generativos em vários domínios”, inclusive no âmbito do processamento de linguagem natural (em inglês: *Natural Language Processing* – NLP), que lida com os desafios relacionados à compreensão e reprodução automática de linguagem humana por computadores. Assim, ao proporcionar um método mais avançado para os programas computacionais interpretarem, compreenderem e manipularem a linguagem humana, observa-se que muitos dos atuais modelos de grandes linguagens (em inglês: *Large Language Model* – LLM), adotam “[...] a arquitetura do *Transformer* como seu principal bloco de construção, oferecendo vantagens sobre os blocos de construção anteriores” (*ibidem*).

A propósito, um exemplo notável para tanto, é o próprio *Generative Pre-trained Transformer*, mais conhecido como GPT da OpenAI. Outro exemplo de inteligência artificial generativa também desenvolvida pela mesma empresa é o DALL-E, que utilizando o modelo base do *Transformer*, é capaz de criar imagens personalizadas de alta qualidade a partir de descrições textuais fornecidas por determinado usuário ao sistema (Figura 2).

Figura 2 – Exemplos de AIGC na geração de imagens. Instruções de texto são fornecidas ao modelo OpenAI DALL-E 2, que gera duas imagens de acordo com as instruções.



Fonte: Cao *et al.* (2018, tradução nossa)

Em suma, todos esses modelos de IAs generativas não seriam possíveis sem o avanço do *deep learning*, que por sua vez, não seria possível sem o desenvolvimento do *machine learning*. Um engloba o outro e juntos são fundamentais para a compreensão do funcionamento dos sistemas de inteligência artificial generativa, o que será apresentado no ponto a seguir.

1.2. Funcionamento das IAs generativas

Conforme exposto anteriormente, as inteligências artificiais generativas são modelos de *machine learning* que, utilizando-se do *deep learning*, são capazes de gerar textos, imagens e outros conteúdos de alta qualidade com base nos dados nos quais foram treinados. Desse modo, nas palavras de Martineau (2023, tradução nossa), as IAs generativas “[...] podem pegar dados brutos – digamos, toda a Wikipédia ou as obras coletadas de Rembrandt¹ – e “aprender” a gerar resultados estatisticamente prováveis quando solicitado”. Mas como de fato esses modelos generativos funcionam?

¹ Rembrandt Harmenszoon van Rijn foi um célebre pintor holandês do movimento Barroco.

Para compreender o funcionamento das IAs generativas precisa-se, previamente “[...] entender como o cérebro pensa [...] só então poderemos perguntar como podemos construir máquinas inteligentes” (Hawkins; Blakeslee, 2004, p. 6, tradução nossa). Nesse sentido, evidencia-se que para Jeff Hawkins e Sandra Blakeslee (2004), o cérebro, por meio da atividade dos neurônios, opera fazendo previsões contínuas sobre eventos futuros e, em seguida, aprende com as discrepâncias entre suas previsões e a realidade subsequente, refinando, cada vez mais, sua capacidade de prever eventos, construindo, assim, uma representação interna do mundo, sendo a inteligência humana resultado desse mecanismo.

Desse modo, Greg Pavlik (2023) afirma que as inteligências artificiais generativas funcionam exatamente da forma descrita pela teoria de Hawkins e Blakeslee. Destaca-se, para tanto, que “todos os modelos de IA generativa começam com uma rede neural artificial codificada em *software*” (*ibidem*), sendo essa rede neural artificial uma espécie de mecanismo computacional inspirado nas redes neurais biológicas existentes nos cérebros humanos. Nesses termos, salienta-se que

[...] uma rede neural artificial é composta de várias unidades de processamento bastante simples, geralmente conectadas por canais de comunicação associados a determinado peso. O comportamento inteligente da rede vem das interações entre suas unidades de processamento. Diferentemente, no entanto, do cérebro biológico – em que qualquer neurônio pode se conectar com qualquer outro neurônio ao redor –, as redes neurais artificiais têm camadas discretas, conexões e direções de propagação de dados (Gabriel, 2022, p. 82).

Ante o exposto, percebe-se que embora as redes neurais artificiais sejam treinadas para funcionar de forma similar ao cérebro humano, elas diferem deste último por ainda serem menos sofisticadas quando comparadas com a eficiência estrutural do processo sináptico. Isso porque, segundo Kaufman (2022), em uma sinapse a comunicação de uma informação de um neurônio para outro pode ser bloqueada, alterada ou até mesmo amplificada. Assim, a noção inicial captada pelo neurônio pré-sináptico pode chegar ao neurônio pós-sináptico de forma distinta daquela que foi originalmente transmitida, demonstrando a extraordinária habilidade do cérebro de transformar informações.

Assim, evidencia-se que “a neuroplasticidade – capacidade do cérebro de mudar, adaptar-se e moldar-se em nível estrutural e funcional, quando sujeito a novas

experiências do ambiente interno e externo – gera uma complexidade que é difícil reproduzir em uma máquina” (Kaufman, 2022, p. 14). Desse modo, enquanto o cérebro humano é capaz de apreender concepções do mundo a partir de um raciocínio intuitivo, “as redes neurais artificiais de hoje precisam ver milhares de exemplos antes de começarem a detectar tais padrões” (Heikkilä; Heaven, 2022).

Sob esse aspecto, explica Martha Gabriel (2022) que a maioria dos modelos de redes neurais segue um processo de treinamento através desses exemplos, os quais acabam permitindo a assimilação de informações a partir da análise de dados reais, sem necessitar de um conjunto de regras rígidas pré-definidas por algoritmos. Para esse treinamento, Pavlik (2023) ensina que os modelos de grandes linguagens, recebem “[...] enormes volumes de texto para processar e são encarregados de fazer previsões simples, como a próxima palavra em uma sequência ou a ordem correta de um conjunto de frases”.

No entanto, conforme ressalta Pavlik (*ibidem*), nos estágios iniciais do treinamento, as previsões de resultado tendem a não ser muito boas. Só que a cada vez que o modelo realiza uma nova previsão de resposta, ele verifica a exatidão dos dados conforme sua base de treinamento, e vai aprimorando seus parâmetros, como se fosse um humano em processo de aprendizagem. Logo,

Após o treinamento inicial, os modelos de IA generativa podem ser ajustados por meio de uma técnica de aprendizagem supervisionada, como a aprendizagem por reforço a partir de *feedback* humano (*Reinforced Learning from Human Feedback*, RLHF). Na RLHF, o resultado do modelo é fornecido a revisores humanos que fazem uma avaliação binária positiva ou negativa – polegar para cima ou para baixo – que é reinserida no modelo. A RLHF foi usada para ajustar o modelo GPT 3.5 da OpenAI para ajudar a criar o *chatbot* ChatGPT que se tornou viral (PAVLIK, 2023).

Em síntese, é assim que se dá o funcionamento das inteligências artificiais generativas. Observa-se, contudo, que para todo o desenvolvimento delas há a essencialidade dos dados para seu treinamento. Em verdade, o funcionamento das IAs generativas dependem de “[...] enormes quantidades de dados para treinar e sustentar seus algoritmos” (Hall, 2023, tradução nossa), ou seja, sem acesso aos dados, os modelos generativos obviamente não funcionariam. Mas, para compreender como essas inteligências artificiais acessam e utilizam esses dados, é necessário

adentrar em outro ponto imprescindível para o desenvolvimento das IAs generativas: a mineração de dados (ou *data mining*).

1.3. O processo de mineração de dados

O avanço contínuo da tecnologia nas últimas décadas resultou em uma ampla disponibilidade de elementos computacionais potentes, inclusive no que diz respeito a coleta de dados e mídias de armazenamento. Nesse sentido, segundo apontam Han, Kamber e Pei (2011), a partir de 1960, a tecnologia da informação passou por uma evolução sistemática, partindo de sistemas primitivos de processamento de arquivos para avançados sistemas de bases de dados em meados de 1980.

Sabendo disso, evidencia-se que “as origens da mineração de dados remontam ao final dos anos 80, quando o termo começou a ser usado, pelo menos na comunidade de pesquisa” (Coenen, 2011, p. 25, tradução nossa). No entanto, nota-se que a partir do desenvolvimento do campo da “[...] IA (especialmente ML), associadas com a explosão da disponibilidade de dados não estruturados no mundo (*big data*²), o campo de *data mining* dá um salto em aplicação e importância” (Gabriel, 2022, p. 84).

Afinal, da década de 80 para os dias atuais, constata-se o crescimento cada vez mais exponencial do volume de dados disponibilizados, o que aconteceu devido ao processo de informatização da sociedade, bem como do avanço acelerado sobre o desenvolvimento de poderosas tecnologias de coleta e armazenamento de dados. Nesse sentido, como ressaltam Han, Kamber e Pei (2011, p. 1, tradução nossa), “vivemos num mundo onde grandes quantidades de dados são recolhidas diariamente. A análise desses dados é uma necessidade importante”.

Nesse cenário, a mineração de dados se torna essencial. Mas o que seria esse procedimento? Antes de conceituá-lo de fato é preciso salientar, conforme indicam Goldschmidt, Passos e Bezerra (2015, p. 2), que “a expressão Mineração de Dados

² “A tradução literal de *big data* significa ‘grandes dados’. Em tecnologia, o conceito é aplicado a uma área que coleta e analisa informações diversas a partir de um grande volume de dados” (ESPM).

(*Data Mining*), mais popular, é, na realidade, uma das etapas da Descoberta de Conhecimento em Bases de Dados³. Assim, para os autores em questão,

A etapa de Mineração de Dados compreende a busca efetiva por conhecimentos úteis no contexto da aplicação de KDD. É a principal etapa do processo de KDD. Tanto que alguns autores se referem à Mineração de Dados e à Descoberta de Conhecimento em Bases de Dados como sinônimos. Esta etapa envolve a aplicação de algoritmos sobre os dados em busca de conhecimento implícito e útil (Goldschmidt; Passos; Bezerra, p. 21).

Ante o exposto, é possível conceituar a mineração de dados, ou *data mining* como “[...] o processo de descoberta de padrões e conhecimentos relevantes a partir de grandes quantidades de dados” (Han; Kamber; Pei, 2011, p. 8, tradução nossa). Isso significa que, assim como no processo de mineração industrial, para a mineração de dados é preciso, obviamente, que haja fontes para efetuação do “garimpo”. Nesse caso, as fontes podem ser os “[...] bancos de dados, armazéns de dados, a *Web*, outros repositórios de informações ou dados que são transmitidos dinamicamente para o sistema” (*ibidem*). Logo, resta-se evidenciado que no processo de mineração de dados “a meta é extrair informações de uma base de dados e transformá-las em uma estrutura compreensível para uso futuro” (Gabriel, 2022, p. 83).

Desse modo, havendo uma fonte para ser explorada, que geralmente são compostas pelas mais variadas quantidades de conteúdos incluindo-se livros, músicas, artigos, matérias jornalísticas, entre outros materiais afins, o processo de mineração de dados vem a ocorrer passando, em regra, por três etapas. Como destaca Kalpana Tyagi (2024) a primeira dessas fases, diz respeito a coleta e acesso dos conteúdos que conseqüentemente compõem algumas das fontes, enquanto na segunda, há a extração especializada dos referidos conteúdos e, por fim, na terceira há a mineração propriamente dita, com a obtenção das informações de qualidade extraídas durante o procedimento.

Sabendo disso, e, como o ponto central de “[...] toda a cadeia de valor da IA generativa é o processo de leitura automática de grande volume de dados para descobrir padrões, gerar novo conhecimento e extrair informações” (TYAGI, 2024, tradução nossa), o processo de mineração de dados é fundamental para ela. Afinal,

³ Do inglês KDD – *Knowledge Discovery in Databases*. Os autores optaram pela tradução do termo, mas continuaram utilizando a sigla em conformidade com a abreviação em língua inglesa.

para funcionarem de forma eficaz, os modelos de IAs generativas precisam ser treinados por dados de qualidade, o que só é possível graças ao *data mining*.

Entretanto, a grande problemática que permeia essa questão sobre a mineração de dados e conseqüentemente inspira o estudo do presente trabalho de pesquisa diz respeito à exploração sobre obras protegidas por direitos autorais utilizadas no processo de treinamento das inteligências artificiais generativas. Afinal, como ressalta Phil Hall (2023, tradução nossa),

[...], muitos desses modelos foram treinados por material em grande parte – e, possivelmente, indiscriminadamente – extraído da *web* (alguns deles de domínio público e outros produtos de empresas privadas, como organizações de notícias, estúdios de cinema e redes sociais).

À vista disso, questiona-se: a ausência de autorização prévia dos autores e de outros titulares de direitos sobre a utilização de suas obras para treinamento desses sistemas generativos configura violação de direito autoral? Para responder essa pergunta é preciso compreender um pouco sobre os direitos autorais, o que será abordado no capítulo a seguir.

2. SOBRE OS DIREITOS AUTORAIS

As intensas mudanças sociais advindas das inovações tecnológicas favorecem um panorama desafiador nas diversas áreas do Direito, inclusive naquela que diz respeito a proteção de obras intelectuais: o Direito Autoral. Assim, é evidente que para o objetivo principal deste trabalho de pesquisa seja alcançado, é imprescindível compreender o âmbito dos direitos autorais, especialmente no que tange a legislação brasileira.

No entanto, antes de abordar a matéria de fato, é preciso que seja compreendido o que é ser autor, questão que inclusive foi debatida por Michel Foucault em uma conferência no Collège de France⁴ em 1969. Para o renomado filósofo (2009, p. 267), “essa noção do autor constitui o momento crucial da individualização na história das ideias, dos conhecimentos, das literaturas, e também na história da filosofia, e das ciências”. No mesmo sentido, entende Roland Barthes (2004, p. 58) que

O autor é uma personagem moderna, produzida sem dúvida por nossa sociedade na medida em que, ao sair da Idade Média, com o empirismo inglês, o racionalismo francês e a fé pessoal da Reforma, ela descobriu o prestígio do indivíduo ou, como se diz mais nobremente, da "pessoa humana".

Com esse entendimento introdutório sobre o autor, pode-se então afirmar que “[...] quando falamos de direito de autor, estamos nos referindo às leis que têm por objetivo garantir ao autor um reconhecimento moral e uma participação financeira em troca da utilização da obra que ele criou” (Afonso, 2009, p. 10). Assim, acrescenta-se ainda, que “[...] o Direito de Autor ou Direito Autoral é o ramo do Direito Privado que regula as relações jurídicas, advindas da criação e da utilização econômica de obras intelectuais estéticas e compreendidas na literatura, nas artes e nas ciências” (Bittar, 2019, p. 41).

Desse modo, resta-se evidente que o bem jurídico passível de proteção, isto é, o objeto do direito autoral é a obra intelectual, resultado da criação artística, literária

⁴ Estabelecimento renomado de ensino e pesquisa na França, fundado em 1530 por François I. A conferência em questão aconteceu na sala nº 6, conforme consta no Relatório de Sessão no volume 3 da coleção “Ditos & Escritos” de Michel Foucault.

ou científica. Assim, destaca-se que no decorrer deste capítulo, serão discutidas questões reflexivas que dizem respeito aos aspectos históricos sobre a evolução dos direitos autorais, a diferença dos sistemas de proteção autoral adotados no mundo, a regulamentação da matéria pela legislação brasileira, abordando, ainda, acepções relativas aos direitos morais e patrimoniais do autor sob a égide da Lei nº 9.610/1998.

2.1. Contextualização histórica do direito autoral

O surgimento da regulamentação sobre direito autoral remonta a Inglaterra do século XVIII. No entanto, para entender o contexto histórico por trás da criação da primeira lei que rege os direitos autorais, é preciso voltar no tempo, mais especificamente para o ano de 1448, quando Johannes Gutenberg criou a primeira tipografia móvel da história, transformando a produção de cópias de escritos, que até então, por mais de vinte séculos, eram feitas à mão (Lemelson-MIT).

Tal inovação possibilitou a impressão de manuscritos em larga escala e sua ampla distribuição ao público. Esse acontecimento foi crucial para a conscientização da necessidade de um sistema de proteção específico, afinal, as criações intelectuais passaram a ser vistas como mercadorias capazes de gerar lucro. Entretanto, os primeiros a se beneficiarem dessa nova forma de comércio foram os impressores, que receberam o direito de reproduzir manuscritos antigos.

Isso porque, nessa época, “era mais visível a atividade técnica tipográfica [...]” (Netto, 2023, p. 46). Ou seja, valorizava-se, então, a atividade dos impressores, considerando-se “[...] a necessidade de proteção do seu investimento e atividades comerciais, para justificar a tutela jurídica no regime da concessão estatal de privilégios e consequentes penalizações a quem os desrespeitasse” (*ibidem*).

Em consequência disso, houve o surgimento do que pode ser considerado como a primeira classe organizada de comerciantes de obras intelectuais no campo literário. Desse modo, “um novo comércio emergia: gráficas e vendedores de livros, na Inglaterra denominados *stationers*, que eram os precursores dos editores modernos” (Stewart, 1983, p. 22 *apud* Netto, 2023, p. 47).

Essa categoria, interessada nos ganhos econômicos advindos dessas publicações, tornou-se a intermediária entre os autores das obras e o público

consumidor. Impulsionava-se assim, significativamente o fortalecimento dos direitos patrimoniais ligados à exploração comercial de obras intelectuais. No entanto, evidencia-se que “[...] tais intermediários reivindicavam para si a titularidade desses direitos. Por isso, os ‘privilégios’ obtidos naquela ocasião devem ser considerados mais propriamente ‘editoriais’ do que ‘autorais’” (Netto, 2023, p. 47).

Apenas em 1710, inspirada pelo *Labor Theory of Property*⁵ de John Locke, sobreveio a *Statute of Anne*, diploma legal que suprimiu o sistema de privilégios até então conferido para os *stationers*. Assim sendo, foi instituído o *Copyright*, sistema de proteção de direito autoral do *common law*⁶, que revogou o favorecimento a guilda dos impressores e livreiros ingleses, garantindo aos autores o monopólio sobre a publicação de suas obras no país. Destaca-se, portanto, nas palavras de Otávio Afonso (2009, p. 4) que “esta foi a primeira lei sobre direito de autor no sentido moderno da expressão, que reconheceu pela primeira vez a existência de um direito individual de proteção sobre uma obra impressa”.

Posteriormente, na Espanha, em 1763, foi abolido por decreto real, o sistema que privilegiava impressores e livreiros, concedendo-se com exclusividade ao autor, o direito de outorga da impressão de suas obras. E, na França, a Assembleia Constituinte da Revolução Francesa, realizada em 1789, assegurou aos autores não somente o direito sobre a representação de suas obras, como também garantiu a eles o direito sobre a reprodução delas, ampliando-se, também, o campo de proteção para além de obras literárias, alcançando-se, assim, obras artísticas e musicais, estabelecendo o sistema do *droit d’auteur*.

Diferentemente do *Copyright*, que levava em conta uma orientação mais mercantil, o *droit d’auteur*, que em sua tradução literal do francês para o português significa “direito do autor”, baseava-se numa orientação mais individualista e humanística, levando em consideração a preservação de garantias morais e patrimoniais para os criadores de obras intelectuais. Assim sendo, o supracitado sistema francês inspirou outros países europeus que também se dedicaram a

⁵ Teoria desenvolvida por John Locke no Capítulo V do livro Segundo Tratado Sobre o Governo Civil (1689), por meio da qual apresenta sua famosa justificativa para a propriedade privada de bens e terras com base no esforço ou trabalho exercido pelos indivíduos.

⁶ Trata-se do sistema jurídico adotado pelos países de língua inglesa, estruturado com base em decisões e costumes.

regulamentar a matéria, passando, então, a ser uma proteção característica no âmbito do *civil law*⁷. Nesse sentido, evidencia-se que

Na Alemanha, o reconhecimento aos direitos de autores e editores, com amplitude legislativa, deu-se pelo Código Civil alemão, de 1794, culminando com a lei especial de 1837. Na Itália, o direito do autor não era reconhecido, garantindo-se privilégios somente ao editor, desde 1603, e apenas tendo em seus territórios sido reconhecido os direitos dos autores já no limiar do século XVIII. Finalmente, a então Rússia czarista, em 1830, edita sua primeira lei autoral, reconhecendo direitos aos autores literários. (Fragoso, 2009, p. 68-69 *apud* Netto, 2023, p. 49)

Já em Portugal, constata-se que “somente com o advento do Decreto Régio de 18 de julho de 1851 inseriu-se o Direito de Autor na ordem jurídica deste país, sob inspiração de Almeida Garret, autor do projeto que resultou no referido Decreto” (Fragoso, 2009, p. 68-69 *apud* Netto, 2023, p. 49). Contudo, observando-se o crescente desenvolvimento das relações internacionais, passou-se a questionar sobre a necessidade de proteger as criações intelectuais em âmbito mundial. Isso porque, com os intercâmbios culturais cada vez mais intensificados pela globalização, muitos autores e artistas reclamavam sobre a ausência de proteção de suas obras no cenário internacional.

Partindo-se do pressuposto de salvaguardar garantias e direitos aos autores e artistas, surge nesse cenário, a primeira proposta do que viria a ser a Convenção de Berna. Elaborada por um congresso internacional realizado em Roma no ano de 1882, promovido pela *Société des gens de lettres*⁸ e pela *Association Littéraire et Artistique Internationale*⁹, a proposta da mencionada Convenção era proporcionar o reconhecimento mútuo dos direitos autorais entre os Estados que viessem a aderir, promovendo assim o desenvolvimento de normas internacionais padronizadas sobre a matéria.

Assim, em 1886 inicialmente contando com França, Alemanha, Itália, Grã-Bretanha, Suíça, Bélgica, Espanha, Haiti, Libéria e Tunísia como países signatários,

⁷ Sistema jurídico proveniente da Europa Continental, estruturado por leis e adotado pela maioria dos países do mundo.

⁸ Sociedade fundada em 1838 para a defesa dos interesses morais e materiais dos escritores membros.

⁹ Associação criada em 1878 que tinha como objetivo primordial instigar a criação de uma Convenção Internacional para a proteção da propriedade literária e artística.

foi consagrada a Convenção de Berna para a Proteção de Obras Literárias e Artísticas, a qual ampliou de modo definitivo os direitos de autor em âmbito mundial.

A partir de então, as legislações internas dos países que passaram a aderir a mencionada Convenção, foram se ajustando em conformidade com a orientação jurídica estabelecida. Nesses termos, segundo a Organização Mundial da Propriedade Intelectual, em agosto de 2023, a Convenção de Berna contava com 181 países signatários, entre eles, o Brasil (WIPO, 2023).

2.2. Progressão legislativa da matéria no Brasil

Verifica-se no decorrer da história brasileira, que a primeira Constituição a trazer uma proteção dos direitos autorais foi a de 1891, mais especificamente no art. 72, § 26. Foi a partir do mencionado dispositivo legal que o autor passou a ter direito exclusivo quanto a reprodução de sua obra, podendo este se utilizar da imprensa para tanto ou de qualquer outro mecanismo (Brasil, 1891). Sete anos após a promulgação da supracitada Constituição, foi criada, segundo Branco e Paranaguá (2009), a primeira lei brasileira a tratar especificamente da proteção autoral no Brasil: a Lei nº 496 de 1898, também conhecida como Lei Medeiros e Albuquerque.

Sob esse aspecto, evidencia-se que a referida Lei foi capaz de criar “[...] um conjunto de regras de defesa da propriedade intelectual num país que, então, contava apenas quinze milhões de habitantes e reduzido número de autores e de casas editoras” (Silva, 1968, p. 56). Contudo, há de se destacar que a Lei nº 496/1898 ficou em vigência por apenas 18 anos, tendo sido revogada após promulgação do Código Civil de 1916. Nesse sentido, evidencia-se que a primeira codificação civil brasileira

[...] classificou o direito de autor como bem móvel, fixou o prazo prescricional da ação civil por ofensa a direitos autorais em cinco anos e regulou alguns aspectos da matéria nos capítulos “Da propriedade literária, artística e científica”, “Da edição” e “Da representação dramática (Paranaguá; Branco, 2009, p. 18).

Observa-se, então, que o supracitado Código, sob a influência de previsões contidas em tratados e convenções internacionais inovou a legislação da matéria no território nacional, bem como, em determinados aspectos, melhorou as previsões contidas na Lei nº 496/1898. Posteriormente, como destaca Netto (2023, p. 53),

Em virtude da acentuada evolução da matéria, refletida no progresso tecnológico, nas convenções internacionais firmadas pelo Brasil, na significativa ampliação inserida na Carta Constitucional de 1967 (e mantida na emenda de 1969), da abrangência do direito exclusivo de titularidade do autor para a acepção genérica “utilização” e não mais apenas “reprodução” de obra intelectual, bem como da profusão de textos legais esparsos promulgados entre o início da vigência do Código Civil de 1916 e os cinquenta anos que se seguiram sob a sua égide, surgiu a ideia de elaboração de um Código de Direito de Autor e Conexos [...]

No entanto, a ideia de uma codificação relativa ao direito autoral não prosperou e apenas em 1973 foi instituída a Lei nº 5.988, apresentando 134 artigos em meio a nove títulos, regulando “[...] a matéria pela primeira vez de maneira completa em nosso país” (Paranaguá; Branco, 2009, p. 19). Em seguida, com a promulgação da Constituição de 1988, o direito autoral foi promovido a cláusula pétrea, constando sua previsão no art. 5º, XXVII e XXVIII da Carta Magna (Brasil, 1988).

Diante desse novo contexto no âmbito constitucional e em decorrência da evolução tecnológica cada vez mais evidente na década de 1990 com a popularização global da rede de computadores, ou seja, da *internet* e do surgimento de novos meios eletrônicos, bem como de mídias digitais, percebeu-se a necessidade de uma nova Lei para tratar sobre a matéria, assim, foi promulgada em 1998 a Lei nº 9.610. Vigente até os dias atuais, a referida Lei não só altera, como atualiza e consolida a legislação relativa aos direitos autorais no Brasil.

2.3. Direitos morais e patrimoniais do autor conforme a LDA

Em consonância com o sistema do *droit d’auteur*, a atual legislação brasileira visa assegurar uma dupla salvaguarda aos direitos do autor: a proteção aos seus direitos morais e patrimoniais. No entanto, antes de abordar os mencionados direitos, é necessário compreender alguns pontos basilares trazidos pela Lei nº 9.610/1998, como as obras a serem protegidas pelos direitos autorais e quem pode ser considerado autor.

Sabendo disso e como exposto inicialmente neste capítulo, o surgimento do direito de autor está estritamente ligado a criação da obra intelectual. Nesse sentido, o art. 7º da Lei de Direitos Autorais (LDA) apresenta um rol exemplificativo de obras intelectuais passíveis de serem protegidas pelo direito autoral:

Art. 7º São obras intelectuais protegidas as criações do espírito, expressas por qualquer meio ou fixadas em qualquer suporte, tangível ou intangível, conhecido ou que se invente no futuro, tais como:

I - os textos de obras literárias, artísticas ou científicas;

II - as conferências, alocações, sermões e outras obras da mesma natureza;

III - as obras dramáticas e dramático-musicais;

IV - as obras coreográficas e pantomímicas, cuja execução cênica se fixe por escrito ou por outra qualquer forma;

V - as composições musicais, tenham ou não letra;

VI - as obras audiovisuais, sonorizadas ou não, inclusive as cinematográficas;

VII - as obras fotográficas e as produzidas por qualquer processo análogo ao da fotografia;

VIII - as obras de desenho, pintura, gravura, escultura, litografia e arte cinética;

IX - as ilustrações, cartas geográficas e outras obras da mesma natureza;

X - os projetos, esboços e obras plásticas concernentes à geografia, engenharia, topografia, arquitetura, paisagismo, cenografia e ciência;

XI - as adaptações, traduções e outras transformações de obras originais, apresentadas como criação intelectual nova;

XII - os programas de computador;

XIII - as coletâneas ou compilações, antologias, enciclopédias, dicionários, bases de dados e outras obras, que, por sua seleção, organização ou disposição de seu conteúdo, constituam uma criação intelectual (Brasil, 1998).

Portanto, se constatada que determinada criação se enquadra nas possibilidades trazidas pelo supracitado dispositivo jurídico, ou que mesmo não se enquadrando nas hipóteses previstas, pertença a seara artística, científica ou literária, a obra poderá ser protegida pelos direitos autorais. No entanto, há de se ressaltar que em obediência a previsão estabelecida pelo *caput* do art. 7º, a obra, passível de proteção autoral, deverá ser exteriorizada. Sobre essa questão enfatiza Bittar (2019, p. 58) que

São protegidas as obras que se exteriorizam pela palavra oral (discurso, conferência, aula, palestra) ou escrita (livro, artigo, verbete), gestos (mímica, pantomima, gesto, coreografia), sinais ou traços (desenho, mapa), sons (melodia, ópera, obra radiofônica), imagens (filme, videofilme, show, novela), figuras (pintura, escultura, arquitetura) e pela combinação de um ou mais meios de expressão (obra teatral, cinematográfica e radiofônica).

Com essa noção sobre as obras intelectuais que podem ser protegidas pelos direitos autorais, passa-se a compreensão relativa à autoria. A respeito desse ponto, destaca-se do art. 11 da LDA que “autor é a pessoa física criadora de obra literária, artística ou científica” (Brasil, 1998), podendo este ser identificado, nos termos da Lei, através de seu nome civil completo ou abreviado ou até mesma por meio de suas iniciais, pseudônimo ou qualquer outro sinal convencional. Assim sendo, concisamente entende-se que

O autor é o sujeito originário do direito de autor e o direito de autor nasce da criação intelectual. Uma vez que esta somente pode ser realizada pelas pessoas físicas, a consequência natural é que a titularidade originária corresponda a pessoa física que cria a obra (Lipzyc, 1996, p. 2 *apud* Netto, 2023, p. 76).

Logo, apenas após reconhecida a autoria de determinada pessoa é que se incide sobre ela os direitos do autor, os quais subdividem-se em morais e patrimoniais. Nesse sentido, seguindo-se a ordem prevista pela LDA, será abordado, primeiramente, as pontuações relativas aos direitos morais do autor, estes que conforme aponta Afonso (2009, p. 35), “são direitos pessoais que têm origem no reconhecimento de que a obra é um prolongamento da personalidade de seu criador”.

Depreende-se, portanto, que os direitos morais do autor dizem respeito a sua vinculação com a obra criada, revelando uma emanção da sua própria personalidade, estando o autor eternamente conectado a sua criação. Por esse motivo, é que Machado de Assis sempre será o autor de *Dom Casmurro*, bem como Clarice Lispector sempre será a autora de *A Hora da Estrela*, devendo ser referenciados a qualquer momento em que suas obras venham a ser utilizadas, seja para estudos, traduções ou adaptações. À vista disso, por meio das palavras de Pontes de Miranda (1974, p. 143, *apud* Netto, 2023, p. 109), resta-se evidenciado que

Essa identificação pessoal, essa ligação do agente à obra, essa relação de autoria, é vínculo psíquico, fático, inabluível (*sic*), portanto indissolúvel, como toda relação causal fática, e entra no mundo jurídico, como criação, como ato-fato jurídico.

Sob esse aspecto, ressalta-se que, pelo art. 27 da Lei nº 9.610/1998, os direitos morais foram classificados como inalienáveis e irrenunciáveis. Ainda, versando sobre essa matéria, é válido destacar que a atual Lei de Direitos Autorais, mais

especificamente no art. 24, traz uma previsão quanto ao que pode ser considerado direitos morais do autor. Nesses termos, tem-se que:

Art. 24. São direitos morais do autor:

I - o de reivindicar, a qualquer tempo, a autoria da obra;

II - o de ter seu nome, pseudônimo ou sinal convencional indicado ou anunciado, como sendo o do autor, na utilização de sua obra;

III - o de conservar a obra inédita;

IV - o de assegurar a integridade da obra, opondo-se a quaisquer modificações ou à prática de atos que, de qualquer forma, possam prejudicá-la ou atingi-lo, como autor, em sua reputação ou honra;

V - o de modificar a obra, antes ou depois de utilizada;

VI - o de retirar de circulação a obra ou de suspender qualquer forma de utilização já autorizada, quando a circulação ou utilização implicarem afronta à sua reputação e imagem;

VII - o de ter acesso a exemplar único e raro da obra, quando se encontre legitimamente em poder de outrem, para o fim de, por meio de processo fotográfico ou assemelhado, ou audiovisual, preservar sua memória, de forma que cause o menor inconveniente possível a seu detentor, que, em todo caso, será indenizado de qualquer dano ou prejuízo que lhe seja causado (Brasil, 1998).

Por conseguinte, quanto aos direitos patrimoniais do autor, afirma Afonso (2009, p. 39) que estes “são os direitos que decorrem da utilização econômica da obra intelectual e da sua comunicação pública, tanto pelo próprio autor como por aqueles por ele autorizados”. Sobre esta espécie de direitos do autor, a Lei nº 9.610/1998 traz determinados aspectos que devem ser observados, entre eles, destacam-se para o desenvolvimento desta trabalho, aqueles relativos à temporariedade e à prévia autorização para utilização da obra.

Em se tratando de temporariedade, é crucial salientar que a LDA estabelece em seu art. 41 que “os direitos patrimoniais do autor perduram por setenta anos contados de 1º de janeiro do ano subsequente ao de seu falecimento, obedecida a ordem sucessória da lei civil” (Brasil, 1998). Desse modo, entende-se que para haver a proteção de uma determinada obra pelos direitos autorais, é necessário que ela esteja dentro do período protetorial estabelecido.

Isso acontece porque, após esse período, a obra entra em domínio público, passando a poder ser economicamente explorada por qualquer pessoa sem precisar de autorização prévia. Sob esse aspecto, nas palavras de Otávio Afonso (2009, p. 50), compreende-se que “uma vez vencido o prazo de proteção, ou seja, o período durante

o qual a exploração econômica da obra é exclusividade do autor ou dos titulares de direito, ocorre a extinção dos direitos exclusivos, no seu aspecto patrimonial”.

Em contrapartida, a obra ainda pertencendo a ordem do domínio privado, passível de proteção autoral, exige a autorização prévia e expressa do autor para qualquer que seja a modalidade de utilização dela por terceiros. Nesses termos, apresenta o art. 29 da Lei nº 9.610/1998, um rol exemplificativo sobre as mais variadas espécies de utilização das obras que precisam ser previamente autorizadas pelos seus criadores:

Art. 29. Depende de autorização prévia e expressa do autor a utilização da obra, por quaisquer modalidades, tais como:

I - a reprodução parcial ou integral;

II - a edição;

III - a adaptação, o arranjo musical e quaisquer outras transformações;

IV - a tradução para qualquer idioma;

V - a inclusão em fonograma ou produção audiovisual;

VI - a distribuição, quando não intrínseca ao contrato firmado pelo autor com terceiros para uso ou exploração da obra;

VII - a distribuição para oferta de obras ou produções mediante cabo, fibra ótica, satélite, ondas ou qualquer outro sistema que permita ao usuário realizar a seleção da obra ou produção para percebê-la em um tempo e lugar previamente determinados por quem formula a demanda, e nos casos em que o acesso às obras ou produções se faça por qualquer sistema que importe em pagamento pelo usuário;

VIII - a utilização, direta ou indireta, da obra literária, artística ou científica, mediante:

a) representação, recitação ou declamação;

b) execução musical;

c) emprego de alto-falante ou de sistemas análogos;

d) radiodifusão sonora ou televisiva;

e) captação de transmissão de radiodifusão em locais de frequência coletiva;

f) sonorização ambiental;

g) a exibição audiovisual, cinematográfica ou por processo assemelhado;

h) emprego de satélites artificiais;

i) emprego de sistemas óticos, fios telefônicos ou não, cabos de qualquer tipo e meios de comunicação similares que venham a ser adotados;

j) exposição de obras de artes plásticas e figurativas;

IX - a inclusão em base de dados, o armazenamento em computador, a microfilmagem e as demais formas de arquivamento do gênero;

X - quaisquer outras modalidades de utilização existentes ou que venham a ser inventadas (Brasil, 1998).

Dessa maneira, depreende-se que a tradução de *Capitães de Areia* para inglês em 1988 precisou de autorização prévia de Jorge Amado, bem como para a adaptação de *Rota 66* para minissérie em 2022, foi necessária anuência expressa de Caco Barcellos. Nesse contexto, resta clarificado, conforme aponta Afonso (2009, p. 44) que “[...] se a obra preexistente se encontra no domínio privado, é obrigatório que o seu autor autorize a realização da obra derivada”.

Sob esse aspecto, nota-se que dentre as diversas modalidades possíveis que exigem autorização prévia e expressa de criadores para utilizações de obras tuteladas pelos direitos autorais, evidencia-se, para a questão de pesquisa em pauta, o disposto no art. 29, X. Afinal, constata-se, por meio da leitura do referido dispositivo legal, que a utilização de obras por quaisquer outras modalidades que venham a ser inventadas, como é o caso do treinamento das IAs generativas, requer a autorização prévia e explícita do autor.

Desse modo, fica evidente que a utilização para qualquer modalidade que seja, de determinada obra, sob domínio privado, sem autorização expressa e prévia de seu criador, configura uma violação dos direitos patrimoniais do autor, conduta inclusive tipificada no Código Penal, mais especificamente no art. 184 do mencionado diploma legal (Brasil, 1941). Nesses termos, em se tratando da legislação brasileira, o ato de utilizar obra sob proteção do direito autoral, sem o devido consentimento do autor, no processo de treinamento das IAs generativas seria considerado uma atitude reprovável.

3. PERSPECTIVAS E DESAFIOS DOS DIREITOS AUTORAIS DIANTE DOS SISTEMAS DE INTELIGÊNCIAS ARTIFICIAIS GENERATIVAS

Ante o exposto no capítulo anterior, fica nítido que pelo direito brasileiro, a utilização de obras protegidas por direitos autorais para o treinamento das IAs generativas, sem autorização prévia e expressa do seu criador, constitui uma violação dos direitos patrimoniais do autor. Por outro lado, sem acesso a essas obras, os modelos atuais de *machine learning* teriam seu desenvolvimento empacado já que é através delas que as IAs generativas “adquirem” conhecimento após o processo de mineração de dados.

Assim, como aponta Tyagi (2024, tradução nossa), “os dados são o alimento que os algoritmos precisam digerir e regurgitar para irradiar alguns padrões e conhecimentos”, sendo, portanto, o processo de mineração de dados, parte fundamental no treinamento das inteligências artificiais generativas. Isso porque, como visto no primeiro capítulo deste trabalho de pesquisa, é através do *data mining* que os dados serão “garimpados” para fornecer uma melhor qualidade para o desempenho dessas espécies de IAs e “embora os dados em si ‘não’ sejam protegidos pelos direitos do autor, a ‘forma criativa’, nomeadamente a ‘obra’, é” (*ibidem*).

Dessa forma, é necessário que haja o devido respeito aos direitos autorais dos criadores, possibilitando o reconhecimento e valoração das obras intelectuais utilizadas para treinamento das IAs generativas. Afinal, sistemas como o GPT da OpenAI ou o BARD do Google, somente conseguem produzir resultados similares aos dos humanos porque foram treinados por meio de dados que constituem criações de artistas, escritores, cientistas e afins, ou seja, eles só são capazes de tamanho feito em razão da

[...] oportunidade de ‘analisar as criações humanas’. Assim, essa ‘usurpação parasitária do mercado de produções literárias e artísticas’ pela IA generativa, que adquiriu a capacidade de criar, implica que o autor humano original deve ser devidamente recompensado por suas obras (*ibidem*).

Diante do cenário apresentado, e sabendo que determinadas obras sob domínio privado passaram pelo processo de mineração de dados e, conseqüentemente, já treinaram esses sistemas de inteligência artificial generativa, como ficam os direitos autorais de seus criadores? A resposta para essa questão inclui um debate complexo que envolve desde processos movidos contra empresas de

tecnologia como a OpenAI e a Microsoft nos Estados Unidos, até a recente aprovação da regulamentação sobre a inteligência artificial pela União Europeia e empenho para normatização da matéria no Brasil, o que será abordado nos tópicos a seguir.

3.1. Alguns litígios nos EUA envolvendo *Copyright* e IA generativa

O sucesso das IAs generativas gerou uma série de ações contra empresas de tecnologia, a maioria delas versando justamente sobre questões relativas à violação de direitos autorais. De acordo com Lyagi (2024, tradução nossa) “em 22 de janeiro de 2024, havia mais de 16 ações judiciais pendentes somente nos EUA. Estas incluem 13 ações coletivas e 3 outras ações judiciais”. Mas, como o intuito neste trabalho não é exaurir a matéria litigiosa estadunidense, será destacado apenas duas dessas ações que são relevantes para o presente estudo.

A primeira delas a ser abordada é a referente ao caso 1:23-cv-08292 da Authors Guild¹⁰ *et al.* contra a OpenAI *et al.*, proposta em 19 de setembro de 2023 perante o United States District Court for the Southern District of New York (SDNY)¹¹. Na peça inicial do caso em questão, alegou a parte autora que a ré se utilizou de obras protegidas por direitos autorais para treinamento de seus modelos de grandes linguagens (em inglês: *Largue Language Models* – LLMs), sem autorização prévia dos escritores. Sob esse aspecto, explica a parte autora que o uso não autorizado e consequentemente não monetizado das obras literárias prejudica os seus criadores, ainda argumentando que

Os réus poderiam ter “treinado” seus LLMs em obras de domínio público. Eles poderiam ter pagado uma taxa de licenciamento razoável para usar obras protegidas por direitos autorais. O que os réus não poderiam fazer era esquivarem-se totalmente do *Copyright Act* para impulsionar seu lucrativo empreendimento comercial, pegando quaisquer conjuntos de dados de livros relativamente recentes que pudessem obter sem autorização. Não há nada de justo nisso. O uso não autorizado pelos réus das obras protegidas por direitos autorais dos requerentes apresenta, portanto, um caso de violação direta aplicando-se ao caso danos sobre direitos autorais conforme lei amplamente estabelecida e reconhecida (SDNY, 2023, p. 2-3, tradução nossa).

¹⁰ Organização profissional criada em 1912 para defesa dos direitos de autores publicados nos EUA.

¹¹ Uma das quatro Cortes Distritais do estado de Nova York. Assim como nos cases, será utilizado a abreviação de seu nome neste trabalho.

Como prova do alegado a parte autora afirmou que a própria OpenAI admitiu ter utilizado para treinar seus modelos de LLMs o “[...] ‘Common Crawl’ e dois ‘*corpora*¹² de livros retirados da Internet’ de ‘alta qualidade’, que se chamam de ‘Books1’ e “Books2”” (*ibidem*, p. 12). Destaca ainda a Authors Guild *et al.* que a Common Crawl é “[...] um *corpus*¹³ vasto e crescente de ‘dados brutos da *web*, extrações de metadados e extrações de texto’ extraídos de bilhões de páginas da *web*” (*ibidem*, p. 13), contendo nele diversos “[...] textos de livros copiados de *sites* piratas” (*ibidem*).

Em resposta a requerente, a OpenAI *et al.* em 26 de janeiro de 2024 se reservou ao direito de não se opor à moção movida pela parte autora, destacando ainda que ante sua escolha de não se manifestar “[...] não adota, concorda ou admite a veracidade de qualquer uma das alegações dos requerentes” (SDNY, 2024, p. 3, tradução nossa). Em junho de 2024, o processo ainda tramitava no SDNY e sem dúvidas o resultado dele criará precedentes de longo alcance tanto para os detentores de direitos autorais quanto para os desenvolvedores de IAs generativas.

Quase acompanhando a Authors Guild *et al.*, a New York Times Company (NYTC) ingressou com uma ação contra a Microsoft Corporation e a OpenAI, em 27 de dezembro de 2023, perante o mesmo Tribunal Distrital, também alegando violação de direitos autorais: trata-se, portanto, do caso 1:23-cv-11195. De acordo com a parte autora, as inteligências artificiais generativas das réis, mais especificamente a Copilot e o ChatGPT, modelos de grandes linguagens,

[...] foram construídos copiando e usando milhões de artigos de notícias protegidos por direitos autorais do The Times, como investigações aprofundadas, artigos de opinião, resenhas, instruções guias e muito mais. Embora os réus tenham se envolvido em cópias em larga escala de muitas fontes, eles deram ênfase especial ao conteúdo do Times ao construir seus LLMs – revelando uma preferência que reconhece o valor dessas obras (SDNY, 2023, p. 2, tradução nossa).

Argumentou ainda a NYTC que ao utilizarem “[...] a valiosa propriedade intelectual de terceiros desta forma, sem pagar por isso, tem sido extremamente

¹² Plural de “*corpus*”.

¹³ Trata-se de uma espécie de coleção de textos sobre determinado tema.

lucrativo para as rés” (*ibidem*, p. 3). Isso porque, conforme apontou a parte autora, o uso de seu conteúdo protegido por direitos autorais no treinamento dessas LLMs possibilitou o aumento da capitalização de mercado da Microsoft Corporation e da OpenAI com o lançamento de suas respectivas inteligências artificiais generativas.

Por meio de memorando, a OpenAI argumentou em 26 de fevereiro de 2024, que o caso em questão, assim como tantos outros, dizem respeito a hipótese de *fair use*, destacando, para tanto, que mesmo se tratando de textos protegidos por direitos autorais, os mesmos se encontram disponíveis para o público, podendo, nesse caso, ser utilizados no processo de treinamento das IAs generativas. Nesse sentido ressaltou a requerida que

Por uma boa razão, há uma longa história de precedentes que sustentam que é perfeitamente legal usar conteúdo protegido por direitos autorais como parte de um processo tecnológico que (como aqui) resulta na criação de produtos novos, diferentes e inovadores (SDNY, 2024, p. 3, tradução nossa).

Da mesma forma, a Microsoft Corporation alegou a necessidade de defesa do *fair use* no que tange os treinamentos dessas novas tecnologias. Sob esse aspecto, a mencionada requerida destacou que pretende “[...] justificar a sua posição de que as tecnologias de IA generativa não constituem violação de direitos de autor” (SDNY, 2024, p. 5), demonstrando que “[...] os modelos generativos de aprendizado de máquina de IA são uma tecnologia legal, assim como muitas outras novas tecnologias introduzidas ao longo dos anos” (*ibidem*).

Percebe-se, mediante o exposto, que o caminho seguido pelas empresas de tecnologia nessa última ação é argumentar em favor da utilização do *fair use*, um “[...] instituto essencialmente de direito norte-americano” (Ascensão, 2022, p. 85) por meio do qual se permite o acesso a material protegido por direitos autorais sem a autorização prévia e expressa do autor. No entanto, assim como o caso do Authors Guild *et al. versus* OpenAI *et al.*, o caso do The New York Times Company *versus* Microsoft Corporation e OpenAI, até o momento do desenvolvimento deste trabalho não foi resolvido.

Nesse sentido, é válido destacar que não há uma conclusão exata sobre a temática. Contudo, compartilhar a essência dos argumentos utilizados nesses processos judiciais em andamento é fundamental para o entendimento a respeito do

conflito que envolve os direitos autorais e as IAs generativas, pois seus resultados influenciarão na abordagem da questão nos EUA e conseqüentemente na conjuntura jurídica mundial.

3.2. Regulamentação da IA pela União Europeia

Diante do avanço cada vez mais notável das inteligências artificiais, países ao redor do mundo passaram a se preocupar sobre a necessidade de regulamentação sobre a matéria. Nesse contexto, destaca-se que a primeira Lei criada para tratar sobre as IAs veio da União Europeia (UE), a qual foi aprovada por unanimidade pelos 27 Estados-Membros da UE em março de 2024, com previsão para entrar em vigência a partir de julho de 2025. Entre os diversos pontos abordados, o Regulamento da Inteligência Artificial, ocupou-se de tratar sobre a questão relativa à proteção dos direitos autorais ante a problemática das IAs generativas.

Assim, destaca-se do texto disponibilizado alguns considerandos¹⁴ para compreensão do posicionamento regulatório. Entre eles, no considerando nº 105 o Parlamento Europeu reforça aquilo que foi abordado anteriormente neste trabalho de pesquisa sobre as IAs generativas, ou seja, que “o desenvolvimento e o treino de tais modelos exigem o acesso a grandes quantidades de texto, imagens, vídeos e outros dados” (União Europeia, 2024, p. 97). Sendo assim, evidencia-se da leitura do supracitado considerando que

Qualquer utilização de conteúdos protegidos por direitos de autor exige a autorização dos titulares dos direitos em causa, a menos que se apliquem exceções e limitações pertinentes em matéria de direitos de autor. A Diretiva (UE) 2019/790 introduziu exceções e limitações que permitem reproduções e extrações de obras ou outro material para efeitos de prospeção¹⁵ (*sic*) de textos e dados, sob determinadas condições. Ao abrigo destas regras, os titulares de direitos podem optar por reservar os seus direitos sobre as suas obras ou outro material para impedir a prospeção (*sic*) de textos e dados, a menos que tal seja feito para fins de investigação científica. Sempre que os direitos de exclusão tenham sido expressamente reservados de forma

¹⁴ Espécie de “parágrafos” introdutórios que podem ser utilizados como forma de justificar o que será abordado em determinada regulamentação.

¹⁵ A versão do Regulamento disponibilizado em língua portuguesa é baseada no português de Portugal. Lá, diferentemente do Brasil, *mining* é traduzido como “prospeção”.

adequada, os prestadores de modelos de IA de finalidade geral¹⁶ têm de obter uma autorização dos titulares de direitos caso pretendam realizar uma prospeção (*sic*) de textos e dados nessas obras (*ibidem*).

Em subsequência, destaca-se do considerando de nº 106 que qualquer que seja o prestador a colocar uma IA no mercado da União, deverá respeitar a proteção do direito autoral, “[...] independentemente da jurisdição em que tem lugar os atos relevantes em matéria de direitos de autor subjacentes ao treino desses modelos de IA de finalidade geral” (União Europeia, 2024, p. 98). Já no considerando nº 107, ressaltou o Parlamento Europeu que, atendendo ao processo de transparência sobre os dados, principalmente sobre aqueles sob proteção de direitos autorais, concernirá aos prestadores das IAs a disponibilização “[...] ao público um resumo suficientemente pormenorizado dos conteúdos utilizados para o treino do modelo [...]” (*ibidem*, p. 99).

Por conseguinte, no considerando nº 108 o Parlamento Europeu (*ibidem*) fez questão de evidenciar que caberá ao Serviço IA¹⁷ o controle sobre o cumprimento das mencionadas obrigações pelos prestadores de serviços das inteligências artificiais na União Europeia. Ante todos os considerandos voltados ao direito autoral e as IAs, constata-se que a legalização de fato da matéria se dá no Capítulo V, Seção 2, art. 53.º, parágrafo 1, c, conforme é possível verificar a seguir:

Artigo 53.º

[...]

1. Os prestadores de modelos de IA de finalidade geral devem:

[...]

c) Aplicar uma política para dar cumprimento ao direito da União em matéria de direitos de autor e direitos conexos e, em especial, identificar e cumprir, nomeadamente através de tecnologias de ponta, uma reserva de direitos expressa nos termos do artigo 4.º, 3.º, da Diretiva (UE) 2019/790 (União Europeia, 2024, p. 292).

Os artigos da Diretiva ao qual o dispositivo legal acima faz referência, diz respeito, respectivamente, a utilização de mineração de dados para fins de investigação científica e as exceções ou limitações que devem ser observadas para

¹⁶ IA de finalidade geral é um termo mais amplo que pode incluir outros tipos de inteligências artificiais além das gerativas. Trata-se de uma expressão de aplicabilidade mais genérica.

¹⁷ O Serviço IA é um gabinete composto por 5 unidades que terá “[...] um papel fundamental na aplicação do Regulamento Inteligência Artificial, especialmente no que diz respeito aos modelos de IA de finalidade geral” (Comissão Europeia, 2024).

tal. Nesses termos, entende-se pelo artigo 3.º da mencionada Diretiva que em se tratando de órgãos de investigação científica ou de instituições responsáveis pelo patrimônio cultural, a mineração de dados sobre material protegido pelos direitos autorais será autorizada (União Europeia, 2019).

Ressalta-se, portanto, do dispositivo legal em pauta, que cabe aos titulares dos direitos em questão “[...] aplicar medidas para assegurar a segurança e a integridade das redes e bases de dados em que as obras ou outro material protegido são acolhidos” (União Europeia, 2019, p. 113), desde que tais medidas não excedam o necessário para que as instituições e órgãos supracitados alcancem seu objetivo. Isto quer dizer que calhará aos detentores de direitos autorais, averiguar em quais locais suas obras estão sendo inseridas, atentando a proteção de seus direitos sem prejudicar o desenvolvimento científico e cultural.

Enquanto isso, o artigo 4.º traz previsão relativa à exceção de mineração de dados sem fins de investigação científica, determinando que os Estados-Membros deverão contemplar uma limitação ou uma exceção aos direitos do autor e alguns direitos conexos no que tanger reproduções e extrações de obras para as quais exista alguma espécie de acesso legalizados das obras para fins de mineração de dados. No entanto, “[...] esta exceção ou limitação, todavia, só se aplicará se o uso das obras não estiver reservado expressamente pelos titulares de direitos” (ROCHA, 2022, p. 130).

Antes o exposto, fica evidente a importância sobre o entendimento das previsões contidas na referida legislação, uma vez que se trata da primeira Lei a versar sobre a temática de regulamentação das IAs no mundo. Nesse aspecto, observa-se que o Regulamento da Inteligência Artificial trouxe uma perspectiva inovadora no que diz respeito à compreensão de como se dará a proteção aos direitos autorais quando se tratar do processo de mineração de dados, fundamental para o treinamento das IAs generativas, inspirando outros países a se dedicarem sobre a regularização da matéria.

3.3. Situação no Brasil: o Projeto de Lei nº 2.338/2023 em debate

No Brasil, o avanço cada vez mais célere quanto à adoção das inteligências artificiais nos mais diversos setores sociais incentivou o Poder Público a criar diretrizes

que visem assegurar a ética, a transparência e a responsabilidade no uso dessas espécies de tecnologia. Desse modo, conforme aponta Ribeiro (2024), a primeira proposta de regulamentação sobre a matéria se deu com o Projeto de Lei (PL) nº 5051/2019, de autoria do Senador Styvenson Valentim.

O referido PL propunha a utilização da inteligência artificial em conformidade com a dignidade humana, a diversidade, a proteção de dados pessoais, a confiabilidade e a auditoria de sistemas, e em favor da valorização do trabalho e do progresso econômico, defendendo que sistemas de decisão baseados em IA devem ser sempre complementares à tomada de decisões humanas. Posteriormente, em 2020, o deputado Eduardo Bismark “[...] apresentou o Projeto de Lei nº 21/2020, com o objetivo de estabelecer um marco legal para o desenvolvimento e uso da IA, tanto pelo setor público quanto por entidades privadas e indivíduos” (Ribeiro, 2024).

A partir de então, as discussões sobre a necessidade de regulamentação da matéria intensificaram-se no Congresso Nacional e, em 2022, o Senado Federal “[...] criou uma Comissão de Juristas responsável por aprofundar o estudo sobre as propostas existentes, incluindo os PL 5051/2019, 21/2020 e outros projetos relacionados” (*ibidem*). A Comissão de Juristas responsável por subsidiar elaboração de substitutivo sobre inteligência artificial no Brasil (CJSUBIA) organizou audiências públicas com a presença de diversos especialistas, autoridades governamentais e membros da sociedade civil para tratar sobre a temática.

Assim, logo após a supracitada Comissão ter finalizado suas atividades, foi entregue um relatório referente à regulamentação do uso das IAs, que resultou no Projeto de Lei nº 2338/2023. Mais detalhado que os seus antecessores, o PL, apresentado por Rodrigo Pacheco, aborda entre vários pontos, a questão que envolve o direito autoral e o processo de mineração de dados, este último essencial para o treinamento das IAs generativas, o que é possível constatar no art. 42, presente na Seção III do Capítulo VIII:

Art. 42. Não constitui ofensa a direitos autorais a utilização automatizada de obras, como extração, reprodução, armazenamento e transformação, em processos de mineração de dados e textos em sistemas de inteligência artificial, nas atividades feitas por organizações e instituições de pesquisa, de jornalismo e por museus, arquivos e bibliotecas, desde que:

I – não tenha como objetivo a simples reprodução, exibição ou disseminação da obra original em si;

II – o uso ocorra na medida necessária para o objetivo a ser alcançado;

III – não prejudique de forma injustificada os interesses econômicos dos titulares; e

IV – não concorra com a exploração normal das obras.

§ 1º Eventuais reproduções de obras para a atividade de mineração de dados serão mantidas em estritas condições de segurança, e apenas pelo tempo necessário para a realização da atividade ou para a finalidade específica de verificação dos resultados da pesquisa científica.

§ 2º Aplica-se o disposto no *caput* à atividade de mineração de dados e textos para outras atividades analíticas em sistemas de inteligência artificial, cumpridas as condições dos incisos do *caput* e do § 1º, desde que as atividades não comuniquem a obra ao público e que o acesso às obras tenha se dado de forma legítima.

§ 3º 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) (Brasil, 2023, p. 27).

Da leitura do dispositivo legal exposto, é possível notar que a posição adotada pelo PL nº 2338/2023 assemelha-se com o estabelecido pelo artigo 4.º e 3.º, da Diretiva nº 2019/790 da União Europeia no que concerne à mineração de dados de obras protegidas por direitos autorais. Em ambos os textos, verifica-se a possibilidade de utilização do procedimento quando envolver inteligências artificiais usadas para organizações e instituições de pesquisa ou cultura. No entanto, ressalta-se que o texto do PL difere da Diretiva com o § 2º, o qual traz previsão sobre a possibilidade de utilização de obras sob domínio privado no processo de mineração de dados para “outras atividades analíticas” relativas à inteligência artificial.

Assim, depreende-se do mencionado dispositivo que sendo cumpridas as condições estabelecidas pelos incisos do *caput* do art. 42, bem como do respectivo § 1º, conquanto que sejam obtidas de “forma legítima” e não disponibilizadas ao público, será possível a mineração de dados de obras que se encontram sob domínio privado sem constituir violação de direitos autorais. Dessa forma, estando incluso entre as “outras atividades analíticas em sistemas de inteligência artificial”, o processo de treinamento das IAs generativas, o § 2º do PL nº 2338/2023, estaria oportunizando a utilização de obras intelectuais sem a autorização prévia e expressa dos criadores, indo de encontro ao previsto pelo art. 29, X da Lei de Direitos Autorais.

Em tramitação no Congresso Nacional, o referido Projeto de Lei tem preocupado o setor artístico, literário e musical brasileiro, tanto que em 08 de abril de 2024, entidades representantes dos mencionados setores enviaram uma carta de

recomendação ao Senado Federal para ser apreciada pela Comissão Temporária sobre Inteligência Artificial (CTIA). A carta em questão “[...] propõe a inclusão de dispositivos específicos [...] que visam assegurar os direitos dos artistas e criadores frente aos desafios impostos pelo avanço da IA, especialmente a IA generativa” (Rolling Stone Brasil, 2024).

Nesse sentido, as entidades representativas frisaram na carta enviada que para a norma ser justa devem ser respeitados os preceitos da Lei de Direitos Autorais, além de ser necessário um controle sobre a qualidade das bases de dados utilizadas para o processo de *data mining* e uma compensação que seja digna para os criadores das obras intelectuais usadas no treinamento das IAs. Assim, visando assegurar os direitos de propriedade intelectual, ressaltaram ainda no documento enviado que o PL 2338/2023 deve ser construído observando os seguintes pontos:

- a) O uso de obras e produções protegidas para mineração de dados e desenvolvimento de ferramentas de IA deverá estar submetido a autorização prévia.
- b) Os conteúdos gerados por IA não poderão ser assemelhados ou protegidos pelas normas da propriedade intelectual.
- c) Apesar dos setores da indústria criativa não encontrarem espaço para novas exceções e limitações aos direitos autorais, caso a legislação venha a inserir alguma previsão, as exceções e limitações da mineração de textos e dados e o desenvolvimento de ferramentas de IA deverão ser restritas e submetidas ao teste dos três passos previsto na lei brasileira e nos tratados internacionais, que regulam os princípios dos direitos autorais, sempre preservada a prerrogativa do titular de direitos autorais autorizar ou proibir o uso de sua obra ou produção. As exceções nunca deverão ser consideradas ou utilizadas para a mineração de dados ou o treinamento comercial, e os usos não comerciais só deverão ser permitidos em determinadas situações estabelecidas por diretrizes claras e jamais admitir prejuízo injustificado aos titulares de direitos autorais.
- d) O treinamento de sistemas de IA e a aplicação dos algoritmos pelos serviços devem ser transparentes e permitir aos titulares de direitos autorais o controle e o acompanhamento sobre os modelos de uso de suas obras e produções.
- e) É importante que as normas estabeleçam a responsabilidade civil objetiva das empresas e dos desenvolvedores de IA como regime aplicável aos danos causados por ferramentas de inteligência artificial.
- f) O ônus da prova deverá recair sempre sobre as empresas e os desenvolvedores de IA. Não há dúvida de que a IA representará uma estratégica ferramenta para o desenvolvimento, mas isso jamais poderá significar a substituição da criação e da centralidade do ser humano diante das expressões culturais, que representam a verdadeira força da inovação e da criatividade (Sanches; Rosa, 2024 *apud* Rolling Stone Brasil, 2024)

Assim, evidencia-se que, em 18 de junho de 2024, foi apresentado Substitutivo ao PL nº 2338/2023, no qual a matéria em questão recebeu tratamento mais especializado, sendo deslocada da Seção III para a Seção IV e recebendo titulação específica de “Direitos de autor e conexos” dentro da proposta regulatória. Ademais, acrescenta-se que deixou de ser regulamentada por um único artigo, passando a ser normatizada por cinco artigos (do art. 60 ao art. 65).

Logo, inspirado pelo considerando nº 107 do Regulamento da Inteligência Artificial da UE, o art. 60 do Substitutivo prevê que em hipótese de utilização de obras protegidas no processo de treinamento dos sistemas de IAs, o desenvolvedor da referida tecnologia deverá informar quais conteúdos foram usados. Por conseguinte, percebe-se que o art. 61 do Substitutivo altera significativamente o previsto no art. 42 do texto inicial do PL nº 2338/2023. Nesses termos, observa-se o seguinte:

Art. 61. 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 por organizações e instituições de pesquisa, jornalismo, museus, arquivos, bibliotecas e educacionais, desde que observadas as seguintes condições:

I - o acesso tenha se dado de forma lícita;

II - não tenha fins comerciais;

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

IV - 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.

§1º Eventuais cópias 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 Este artigo não se aplica a instituições coligadas ou controladas por entidade com fins lucrativos que forneça ou opere sistemas de IA ou que tenham, entre elas, participação acionária.

§ 3º 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) (Brasil, 2024, p. 73-74).

Nota-se, portanto, que o art. 61 do Substitutivo proporciona um avanço significativo ao detalhar mais claramente as condições e limitações para a utilização de conteúdos protegidos por direitos autorais no processo de mineração de dados

voltada à inteligência artificial. Desse modo, constata-se que a nova redação se preocupa em proteger obras intelectuais, sob domínio privado, de serem exploradas por meio de acesso ilícito, para fins comerciais ou de outra forma que venha a causar prejuízo econômico injustificado aos titulares de direitos autorais.

Além disso, reformula por completo o § 2º em comparação com o do art. 42 do PL nº 2338/2023, retirando a hipótese do uso de obras protegidas no processo de mineração de dados para “outras atividades analíticas” relativas à inteligência artificial. Nesses termos, a nova previsão traz a exclusão da possibilidade de utilização automatizada de conteúdos protegidos conforme disposto no *caput* do art. 61 quando se tratar de instituições coligadas a entidades lucrativas que operem com sistemas de IA, evitando potenciais conflitos de interesse e de exploração comercial dos dados minerados.

Já o art. 62 reserva ao titular dos direitos de autor e conexos a possibilidade de “[...] proibir a utilização dos conteúdos de sua titularidade no desenvolvimento de sistemas de inteligência artificial nas hipóteses não contempladas pelo art. 61 desta Lei, na forma do regulamento” (Brasil, 2024, p. 74). Em sequência, observa-se outra inovação trazida pelo art. 63, por meio do qual fica estabelecido que se configura como infração à ordem econômica “[...] a discriminação de titulares de direitos de autor e conexos que exerçam as prerrogativas estabelecidas nesta Lei por agentes de sistemas de IA ou por entidades do mesmo grupo econômico” (*ibidem*). Esse dispositivo ainda é complementado com o seguinte parágrafo único:

Art. 63

[...]

Parágrafo único. A discriminação prevista no *caput* levará em consideração se um agente de sistemas de IA ou integrante do mesmo grupo econômico prejudica o titular de direitos autorais ou conexos em serviços ou aplicações oferecidos que não estejam diretamente ligados ao sistema em que o titular exerceu a prerrogativa (Brasil, 2024).

Ademais, o art. 64 acrescenta a necessidade de que seja estabelecido um *sandbox*, ou seja, um ambiente regulatório experimental pelo Sistema Nacional de

Regulação e Governança de Inteligência Artificial (SIA)¹⁸ para que haja as devidas disposições quanto à transparência e à remuneração adequada pelos desenvolvedores de inteligências artificiais que utilizem conteúdos protegidos por direitos autorais no processo de treinamento dessas tecnologias e que venham a ser disponibilizadas com finalidade comercial. Nesse sentido, destaca-se ainda como complemento do mencionado dispositivo, as seguintes diretrizes:

Art. 64

[...]

Parágrafo único. O ambiente regulatório experimental (*sandbox*) deve assegurar:

I - que os titulares de direitos de autor e conexos tenham plena capacidade de negociar e autorizar direta ou coletivamente a utilização dos conteúdos dos quais são titulares, podendo fazê-lo de forma gratuita ou onerosa.

II - que o cálculo da remuneração a que se refere o caput considere elementos relevantes, tais como o poder econômico do agente de inteligência artificial, o grau de utilização dos conteúdos e os efeitos concorrenciais dos resultados em relação aos conteúdos originais utilizados.

III - a livre negociação na utilização dos conteúdos protegidos, com promoção de um ambiente de pesquisa e experimentação que possibilite o desenvolvimento de práticas inovadoras, e que não restrinjam a liberdade de pactuação entre as partes envolvidas, nos termos dos arts. 156, 157, 421, 422, 478 e 479 do Código Civil Brasileiro - Lei nº 10.406/2002 e o art. 4º. da Lei 9.610/1998.

IV – que a remuneração se aplicará somente nas hipóteses de disponibilização comercial dos sistemas de inteligência artificial, em conformidade com o art. 62 e ressalvadas as hipóteses de usos permitidos previstos no art. 61.

V – que a remuneração a que se refere este artigo diz respeito:

a) aos titulares de direitos de autor e de direitos conexos nacionais ou estrangeiros domiciliados no Brasil;

b) a pessoas domiciliadas em país que assegure a reciprocidade na proteção, em termos equivalentes a este artigo, aos direitos de autor e aos direitos conexos de brasileiros, conforme disposto nos artigos 2º, parágrafo único, e 97, § 4º, da Lei n. 9.610, de 1998, sendo vedada a cobrança nos casos em que a reciprocidade não estiver assegurada.

Por fim, o art. 65 do Substitutivo define que em se tratando da utilização pelos sistemas de inteligência artificial de conteúdos de imagem, áudio, voz ou vídeo relativos às pessoas naturais, deverá ser respeitado os direitos de personalidade, em

¹⁸ De acordo com o próprio art. 4º, X do Substitutivo ao PL nº 2338/2023, trata-se de um ecossistema regulatório coordenado por autoridade competente para a plena implementação e fiscalização do cumprimento da Lei específica sobre Inteligência Artificial em todo o território nacional, promovendo a segurança jurídica.

conformidade com a previsão contida no Código Civil e em legislação pertinente sobre a matéria. Essas mudanças proporcionam um melhor equilíbrio entre os interesses dos titulares de direitos autorais e os desenvolvedores de sistemas de inteligência artificial.

Isso porque, as novas previsões trazidas conciliam o incentivo à inovação enquanto asseguram a devida proteção aos criadores sem contrariar a LDA, versando, portanto, de forma mais apropriada sobre os direitos dos autores no que diz respeito à autorização da utilização de suas obras para o processo de treinamento das IAs, transparência quanto ao uso dos conteúdos, bem como previsão quanto à compensação justa e adequada. No entanto, é válido ressaltar que até o momento do desenvolvimento deste trabalho, a votação sobre o relatório Substitutivo do PL nº 2338/2023 apresentado pelo Senador e membro da Comissão Temporária sobre Inteligência Artificial no Brasil, Eduardo Gomes, não foi realizada.

O fato é que a regulamentação da inteligência artificial no Brasil é essencial para definir limites claros sobre o uso e desenvolvimento dessas novas tecnologias de forma coerente com os preceitos sociais, evitando, assim, abusos. Nesse sentido, no que diz respeito aos direitos autorais, é válido ressaltar que por mais impressionantes que sejam os resultados alcançados pelas IAs generativas, não se deve deixar de lado a necessidade de valoração das obras criadas pelos humanos que venham a servir de base para o treinamento e consequente “aprendizado” dessas ferramentas. Afinal, não se deve confundir apoio e investimento sobre o avanço tecnológico com exploração indiscriminada dos trabalhos intelectuais de artistas e cientistas.

CONSIDERAÇÕES FINAIS

O avanço constante da tecnologia tem proporcionado à sociedade o acesso a diversas inovações, entre elas, a inteligência artificial generativa. Essas espécies de IAs, capazes de gerar conteúdo de alta qualidade de forma prática e eficiente, embora sejam um marco no avanço computacional, têm trazido desafios na seara do direito autoral. Isso porque, para o desenvolvimento dessa automatização de criações, os modelos de IAs generativas como o GPT, o DALL-E e o Midjourney precisaram passar por um processo de treinamento prévio, utilizando uma alta gama de dados repletos de grandes quantidades de obras intelectuais, muitas das quais podem estar protegidas por direitos autorais.

Ante o exposto, um aprofundamento sobre a temática tornou-se crucial. Assim, é possível constatar que foram apresentados no decorrer deste trabalho levantamentos de conceitos e de exemplos fundamentais que contribuíram significativamente para reflexão, compreensão e conhecimento a respeito dos assuntos abordados. Nessa dinâmica, tornou-se imperativo realizar uma investigação sobre o funcionamento e processo de treinamento dos sistemas de inteligência artificial generativa, elucidando ainda, em especial, o procedimento de mineração de dados, o que validou o cumprimento do primeiro objetivo específico.

Por conseguinte, no propósito de examinar a proteção sobre a criação intelectual, este estudo também permitiu considerações importantes no tocante ao conceito de autoria, da história dos direitos autorais, dos diferentes sistemas existentes no contexto mundial, bem como da sua progressão no âmbito legislativo brasileiro até a Lei nº 9.610/1998. Ressaltando, no caso deste último diploma legal, conceitos e previsões relativos aos direitos morais e patrimoniais do autor no contexto nacional e, conseqüentemente, alcançando o segundo objetivo específico.

Após identificar e refletir sobre esses pontos essenciais, foi evidenciado no terceiro objetivo específico, o debate jurídico envolvendo a proteção dos direitos autorais de criadores ante as IAs, examinando-se ações judiciais nos Estados Unidos, bem como marco regulatório na União Europeia e desenvolvimento de regulamentação no Brasil. Sendo assim, foi analisado, no caso dos Estados Unidos, país que segue o sistema de proteção do *Copyright*, o resultado dos julgamentos dos casos *Authors Guild et al. v. Open AI et al.* e *The New York Times Company v. Microsoft*

Corporation & OpenAI, que embora ainda esteja em tramitação no SDNY, foi crucial para a presente pesquisa, uma vez que versa sobre a problemática abordada e seu resultado trará precedentes quanto à questão da utilização de obras protegidas por direitos autorais, sem autorização prévia e expressa de seus criadores, no processo de treinamento de inteligências artificiais generativas.

Quanto ao marco regulatório da União Europeia, verifica-se que a análise realizada no presente estudo a respeito da previsão específica quanto ao art. 53.º, parágrafo 1, c do Regulamento da Inteligência Artificial, foi fundamental para se ter uma noção sobre como a primeira Lei criada para lidar com IAs abordou a problemática. Destacando-se, para tanto, como o diploma legal em questão abarcou a necessidade de respeitar as regras e os critérios estabelecidos pelos arts. 3.º e 4.º da Diretiva 2019/790 da UE quanto às exceções e limitações da possibilidade de uso de obras intelectuais, sob domínio privado, no procedimento de mineração de dados, desde que obtidas de forma lícita.

Ainda sobre essa questão de regulamentação, também foi abordado o desenvolvimento de Projetos de Lei no contexto brasileiro que visam a regulamentação das IAs no país, atentando-se, mais especificamente, ao PL nº 2338/2023 que, em seu texto inicial, trouxe no art. 42 disposições relativas à temática pauta do presente estudo. Sob esse aspecto, foi analisado como a proposta regulatória da matéria levou a comunidade artística, literária e musical brasileira a clamar, por meio de carta direcionada ao Congresso Nacional, o respeito aos preceitos da Lei nº 9.610/1998, reivindicando que sejam acrescentados no referido PL os direitos garantidos aos criadores sobre autorização prévia, controle e compensação pelo uso de suas respectivas obras para treinamento de inteligências artificiais generativas, e, como mais recentemente foi apresentado Substitutivo ao PL nº 2338/2023, o qual dispõe de forma mais especializada e clarificada sobre a matéria.

Diante de tudo o que foi tratado, percebe-se, portanto, que o objetivo geral norteador do presente estudo foi alcançado, já que é possível constatar a realização da análise a respeito dos direitos autorais dos criadores que tem suas obras intelectuais utilizadas no processo de treinamento dos sistemas de inteligência artificial generativa. No entanto, é válido destacar que a dimensão em que se encontra os direitos autorais frente às IAs generativas é desafiadora e exige muita reflexão, pois, a utilização de obras intelectuais que se encontram sob domínio privado, de

maneira indiscriminada para treinar esses modelos de *machine learning* vem provocando incertezas que trazem à baila discussões sobre a questão da necessidade de autorização prévia e de compensação dos autores.

Assim sendo, salienta-se que o intuito do presente trabalho de pesquisa não é desmerecer o desenvolvimento das IAs generativas, afinal, elas representam um importante avanço da era digital. Por outro lado, é crucial que seu desenvolvimento não prejudique os direitos patrimoniais dos criadores que têm suas obras utilizadas no processo de treinamento dessas tecnologias. Em verdade, é preciso estabelecer um consenso em que haja a valorização e o reconhecimento do trabalho criativo dos escritores, artistas e cientistas sem estagnar o processo de avanço tecnológico das IAs generativas.

Por fim, ressalta-se que a problemática que envolve a presente pesquisa ainda está em constante evolução na sociedade atual. Nesse ínterim, espera-se que este trabalho permita novas reflexões para uma melhor compreensão a respeito da temática abordada, promovendo novos interesses voltados a harmonizar os benefícios tecnológicos proporcionados pelas IAs generativas com a proteção dos direitos autorais de obras usadas no seu processo de treinamento. Portanto, é fundamental que se continue a discutir e adaptar políticas e regulamentações à medida que essa tecnologia avança, garantindo assim um desenvolvimento equânime para todos os envolvidos.

REFERÊNCIAS

AFONSO, Otávio. **Direito Autoral: Conceitos Essenciais**. Barueri: Editora Manole, 2009. *E-book*. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788520442791/>. Acesso em: 31 mai. 2024.

ASCENÇÃO, José de Oliveira. **Estudos de José de Oliveira Ascensão sobre direito autoral & sociedade informacional**. Curitiba: IODA, 2022.

BHARTES, Roland. **O rumor da língua**. São Paulo: Martins Fontes, 2004.

BITTAR, Carlos Alberto. **Direito de Autor**. 7 ed. Rio de Janeiro: Forense, 2019.

BRASIL. Câmara dos Deputados. **Projeto de Lei nº 21, de 2020**. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1853928&filename=Tramitacao-PL%2021/2020. Acesso em: 25 jun. 2024.

BRASIL. **Constituição da República dos Estados Unidos do Brasil (de 24 de fevereiro de 1891)**. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao91.htm. Acesso em: 28 mai. 2024.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 28 mai. 2024.

BRASIL. **Decreto-Lei nº 2.848, de 7 de dezembro de 1940**. Código Penal. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del2848compilado.htm. Acesso em 21 jun. 2024.

BRASIL. **Lei nº 9.610, de 19 de fevereiro de 1998**. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l9610.htm. Acesso em: 29 mai. 2024.

BRASIL. Senado Federal. **Projeto de Lei nº 2338, de 2023**. Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=8009064&ts=1718902534445&disposition=inline>. Acesso em: 25 jun. 2024.

BRASIL. Senado Federal. **Substitutivo do Projeto de Lei nº 2338, de 2023**. Dispõe sobre o desenvolvimento, fomento, uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9640105&ts=1720798348647&rendition_principal=S&dispositivo=inline. Acesso em 25 jun. 2024.

BRASIL. Senado Federal. **Projeto de Lei nº 5051, de 2019**. Estabelece os princípios para o uso da Inteligência Artificial no Brasil. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=8009064&ts=1718902534445&disposition=inline>. Acesso em: 25 jun. 2024.

CAO, Yihan *et al.* A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT. **ArXiv, Cornell University**, 07 mar. 2023. Disponível em: <https://arxiv.org/pdf/2303.04226v1>. Acesso em: 13 jun. 2024.

CAPARROZ, Leo. ChatGPT supera TikTok e é plataforma com o crescimento mais rápido de usuários. **Superinteressante**. Disponível em: <https://super.abril.com.br/tecnologia/chatgpt-supera-tiktok-e-e-plataforma-com-o-crescimento-mais-rapido-de-usuarios/>. Acesso em: 20 jun. 2024.

COENEN, Frans. Data mining: past, present and future. **The Knowledge Engineering Review, Cambridge University Press**, Vol. 26, 25th Anniversary Issue, February, 2011. Disponível em: <https://www.cambridge.org/core/journals/knowledge-engineering-review/article/abs/data-mining-past-present-and-future/EE2E494D98BCE76EBE3FE07897540C43>. Acesso em 20 jun. 2024.

COMISSÃO EUROPEIA. Comissão cria o Serviço IA para reforçar a liderança da UE em matéria de inteligência artificial segura e fiável. **Shaping Europe's digital future, European Union**, 29 mai. 2024. Disponível em: <https://digital-strategy.ec.europa.eu/pt/news/commission-establishes-ai-office-strengthen-eu-leadership-safe-and-trustworthy-artificial>. Acesso em: 24 jun. 2024.

ESPM. Entenda o que é o big data. **Blog ESPM**. Disponível em: <https://www.espm.br/blog/entenda-o-que-e-o-big-data/>. Acesso em 20 jun. 2024.

FOUCAULT, Michel. **Estética**: literatura, pintura, música e cinema. Coleção: Ditos & escritos, vol. III. 2 ed. Rio de Janeiro: Forense Universitária, 2009.

GABRIEL, Martha. **Inteligência Artificial: Do Zero ao Metaverso**. Barueri: Grupo GEN, 2022. *E-book*. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9786559773336/>. Acesso em: 12 jun. 2024.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel; BEZERRA, Eduardo. **Data Mining**. Rio de Janeiro: Grupo GEN, 2015. *E-book*. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788595156395/>. Acesso em: 20 jun. 2024.

HALL, Phil. Now That Generative AI Is Here, Where Will All The Data Come From?. **Forbes**, 22 nov. 2023. Disponível em: <https://www.forbes.com/sites/forbestechcouncil/2023/11/22/now-that-generative-ai-is-here-where-will-all-the-data-come-from/>. Acesso em: 15 jun 2024.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data mining: concepts and techniques**. 3. ed. Waltham: Morgan Kaufmann, 2011.

HAWKINS, Jeff; BLAKESLEE, Sandra. **On Intelligence: How a New Understanding of the Brain will Lead to the Creation of Truly Intelligent Machines**. New York: Times Books, 2004.

HEIKKILÄ, Melissa; HEAVEN, Will Douglas. Yann LeCun has a bold new vision for the future of AI. **MIT Technology Review**. Disponível em: <https://www.technologyreview.com/2022/06/24/1054817/yann-lecun-bold-new-vision-future-ai-deep-learning-meta/>. Acesso em: 20 jun. 2024.

KAUFMAN, Dora. **Desmistificando a inteligência artificial**. Belo Horizonte: Grupo Autêntica, 2022. *E-book*. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9786559281596/>. Acesso em: 12 jun. 2024.

LEMELSON-MIT. Johann Gutenberg – Movable Type Printing Press. **Lemelson-MIT**. Disponível em: <https://lemelson.mit.edu/resources/johann-gutenberg>. Acesso em: 30 mai. 2024.

MARTINEAU, Kim. What is generative AI?. **IBM Research**, 20 abr. 2023. Disponível em: <https://research.ibm.com/blog/what-is-generative-AI>. Acesso em: 15 jun. 2024.

MERRITT, Rick. O que é um Modelo Transformer?. **NVIDIA**, 19 abr. 2022. Disponível em: <https://blog.nvidia.com.br/blog/o-que-e-um-modelo-transformer/>. Acesso em: 15 jun. 2024.

NETTO, José Carlos Costa. **Direito autoral no Brasil**. 4 ed. São Paulo: SaraivaJur, 2023. *E-book*. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9786553624634/>. Acesso em: 30 mai. 2024.

PARANAGUÁ, Pedro; BRANCO, Sérgio. **Direitos Autorais**. Rio de Janeiro: Editora FGV, 2009.

PAVLIK, Greg. O que é IA generativa? Como funciona?. **Oracle**, 15 set. 2023. Disponível em: <https://www.oracle.com/br/artificial-intelligence/generative-ai/what-is-generative-ai/>. Acesso em: 15 jun. 2024.

RIBEIRO, Renato Janine. Regulação da Inteligência Artificial no Brasil e no mundo. **Sociedade Brasileira para o Progresso da Ciência – SBPC**, 19 abr. 2024. Disponível em: <https://portal.sbpcnet.org.br/noticias/editorial-regulacao-da-inteligencia-artificial-no-brasil-e-no-mundo/>. Acesso em: 25 jun. 2024.

ROCHA, Maria Victória. Prospecção de texto e dados (text and data mining) na Diretiva relativa aos direitos de autor e direitos conexos no Mercado Único Digital. **RRDDIS – Revista Rede de Direito Digital, Intelectual & Sociedade**, Curitiba, v. 2, n. 4, p. 99-164, 2022.

ROLLING STONE BRASIL. Entidades enviam carta ao Senado pela proteção dos direitos autorais e regulamentação do uso de IA. **Rolling Stone Brasil**, 08 abr. 2024. Disponível em: <https://rollingstone.uol.com.br/musica/entidades-enviam-carta-ao-senado-pela-protecao-dos-direitos-autorais-e-regulamentacao-do-uso-de-ia/>. Acesso em: 25 jun. 2024.

RUSSELL, Stuart Jonathan; NORVIG, Peter. **Inteligência Artificial: Uma Abordagem Moderna**. 4 ed. Grupo GEN, 2022. *E-book*. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788595159495/>. Acesso em: 14 jun. 2024.

SAMUEL, Arthur Lee. Some studies in machine learning using the game of checkers. Originally published in IBM Journal, Vol. 3, No. 3. July, 1959. **MIT Computer Science and Artificial Intelligence Laboratory**. Disponível em: <https://people.csail.mit.edu/brooks/idocs/Samuel.pdf>. Acesso em: 14 jun. 2024.

SANTOS, Manoel Joaquim Pereira dos; JABUS, Wilson Pinheiro; ASCENSÃO, José de Oliveira. **Direito autoral**. São Paulo: SaraivaJur, 2020. *E-book*. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9786555591521/>. Acesso em: 12 jun. 2024.

SILVA, Francisco de Oliveira e. O Direito Autoral no Brasil. **Revista de Direito do Ministério Público do Estado da Guanabara**, Rio de Janeiro, nº 06, p. 56-60, set/dez de 1968.

TYAGI, Kalpana. Copyright, text & data mining and the innovation dimension of generative AI. **Journal of Intellectual Property Law & Practice, Oxford University**, 09 mar. 2024. Disponível em: <https://academic.oup.com/jiplp/article/19/7/557/7624901>. Acesso em: 16 jun. 2024.

UNIÃO EUROPEIA. **Diretiva 2019/790 de 17 de abril de 2019**. Relativa aos direitos de autor e direitos conexos no mercado único digital e que altera as Diretivas 96/9/CE e 2001/29/CE. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32019L0790>. Acesso em 24 jun. 2024.

UNIÃO EUROPEIA. **Regulamento da Inteligência Artificial**. Cria regras harmonizadas em matéria de inteligência artificial e que altera os Regulamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e as Diretivas 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_PT.pdf. Acesso em 24 jun. 2024.

UNITED STATES OF AMERICA. District Court for the Southern District of New York (SDNY). **Case 1:23-cv-08292**. Authors Guild *et al.* v. Open AI *et al.* (Complaint filed on September 19, 2023). Disponível em: https://storage.courtlistener.com/recap/gov.uscourts.nysd.606655/gov.uscourts.nysd.606655.1.0_1.pdf. Acesso em: 22 jun. 2024.

UNITED STATES OF AMERICA. District Court for the Southern District of New York (SDNY). **Case 1:23-cv-08292**. Authors Guild *et al.* v. Open AI *et al.* (Response filed on January 26, 2024). Disponível em: <https://storage.courtlistener.com/recap/gov.uscourts.nysd.606655/gov.uscourts.nysd.606655.63.0.pdf>. Acesso em: 22 jun. 2024.

UNITED STATES OF AMERICA. District Court for the Southern District of New York (SDNY). **Case 1:23-cv-11195**. The New York Times Company v. Microsoft Corporation and OpenAI (Complaint filed on December 27, 2023). Disponível em: https://storage.courtlistener.com/recap/gov.uscourts.nysd.606655/gov.uscourts.nysd.606655.1.0_1.pdf. Acesso em: 23 jun. 2024.

UNITED STATES OF AMERICA. District Court for the Southern District of New York (SDNY). **Case 1:23-cv-11195**. The New York Times Company v. Microsoft Corporation and OpenAI (Microsoft Corporation's Memorandum filed on February 26, 2024). Disponível em:

<https://storage.courtlistener.com/recap/gov.uscourts.nysd.612697/gov.uscourts.nysd.612697.48.0.pdf>. Acesso em 23 jun. 2024.

UNITED STATES OF AMERICA. District Court for the Southern District of New York (SDNY). **Case 1:23-cv-11195**. The New York Times Company v. Microsoft Corporation and OpenAI (OpenAI's Memorandum filed on February 26, 2024).

Disponível em:

<https://storage.courtlistener.com/recap/gov.uscourts.nysd.612697/gov.uscourts.nysd.612697.52.0.pdf>. Acesso em 23 jun. 2024.

WIPO. **Berne Convention for the Protection of Literary and Artistic Works:**

Contracting parties. Disponível em:

<https://www.wipo.int/export/sites/www/treaties/en/docs/pdf/berne.pdf>. Acesso em: 02 jun. 2024.

ZEWE, Adam. Explained: Generative AI. **MIT News**, 09 nov. 2023. Disponível em:

<https://news.mit.edu/2023/explained-generative-ai-1109>. Acesso em: 13 jun. 2024.