

RESPONSABILIZAÇÃO PENAL DA INTELIGÊNCIA ARTIFICIAL: UMA REVISÃO INTEGRATIVA SOBRE A POSSIBILIDADE DE ENTIDADES TECNOLÓGICAS SEREM CRIMINALMENTE PUNIDAS

*Letícia Andrade dos Santos**

*Luana Minikel de Oliveira***

*Renata Pedrolli Renz****

*Tarsila Helena Bastiani Kretzer*****

Resumo: Este artigo buscou entender se as Inteligências Artificiais (IAs) podem ser penalmente responsabilizadas no âmbito do direito penal brasileiro e, se sim, como essa responsabilização funcionaria. Trata-se de revisão integrativa em que foram identificados 7 arquivos que compuseram o espelho do estudo, a partir da base de dados Google Acadêmico, no idioma português. Os resultados apontam que, atualmente, não há uma correta regularização do uso e criação das IAs. Desse modo, quando surgem problemas ocasionados pela sua utilização, cria-se uma complexa situação jurídica. Além disso, surgem Inteligências Artificiais baseadas em machine learning cujos algoritmos se tornam enviesados, mimetizando comportamentos sociais discriminatórios. Portanto, é possível extrair dos estudos analisados a necessidade de regularizar o desenvolvimento das IAs, levando em conta os valores éticos e sociais juridicamente relevantes, bem como delimitar a responsabilização dos agentes que dão autonomia a essas entidades tecnológicas.

Palavras-chave: inteligência artificial; responsabilização; direito penal; Brasil; crimes digitais.

*Graduanda do Curso de Graduação em Direito da Universidade Federal de Santa Catarina – UFSC

**Graduanda do Curso de Graduação em Direito da Universidade Federal de Santa Catarina – UFSC

***Graduanda do Curso de Graduação em Direito da Universidade Federal de Santa Catarina – UFSC

****Graduanda do Curso de Graduação em Direito da Universidade Federal de Santa Catarina – UFSC

1. INTRODUÇÃO

A rápida evolução das tecnologias da informação e o advento da Inteligência Artificial (IA) têm desencadeado transformações significativas em diversas esferas da sociedade contemporânea, sendo que as discussões sobre a temática vêm angariando notável relevância nos últimos anos.

A Inteligência Artificial (IA) é um sistema de computador projetado para realizar tarefas que normalmente requerem inteligência humana. Esse sistema é capaz de aprender, raciocinar, tomar decisões e executar ações com base em dados e algoritmos. A Inteligência Artificial pode ser classificada em várias categorias, dependendo de sua capacidade e funcionalidade, sendo um campo multidisciplinar que envolve contribuições de diversas áreas do conhecimento, como a matemática, a lógica, a psicologia e a filosofia (Russell Stuart J., Norvig, Peter, 2003).

Essa é uma tecnologia em rápida evolução e tem o potencial de transformar a vida das pessoas de muitas maneiras e, à medida que a IA continua a se desenvolver, é provável que tenha um impacto ainda maior no cotidiano da população.

Desse modo, percebe-se que a crescente presença de IA em diversos aspectos da sociedade moderna desencadeou uma série de desafios no campo do direito, em particular no âmbito do direito penal.

Em 2015, em Baunatal, próxima a Frankfurt, Alemanha, um trágico incidente ocorreu na planta industrial da Volkswagen. Durante a montagem de um robô estacionário, um jovem de 22 anos foi agarrado pela máquina e esmagado contra uma placa de metal. As primeiras especulações apontavam para um erro humano como causa, levantando questões sobre responsabilidades legais e possíveis ações judiciais, enquanto se avaliava se a máquina em si estava programada corretamente (Phillip, 2015, apud Moraes, 2023, p. 11).

Três anos mais tarde, um veículo autônomo operado pela empresa de transporte por aplicativo Uber, com um motorista de reserva ao volante como medida de emergência, atropelou fatalmente uma mulher em uma rua de Tempe, no estado do Arizona, Estados Unidos. Uma investigação preliminar revelou que o veículo estava se deslocando a aproximadamente 65 quilômetros por hora no momento em que colidiu com uma pedestre que caminhava com sua bicicleta na via. Surpreendentemente, o carro não diminuiu a velocidade antes do impacto, e o motorista de segurança da Uber não apresentou sinais de distração. O clima estava claro e seco (Wakabayashi, 2018, apud Moraes, 2023, p. 11).

Em 5 de janeiro de 2020, teve início uma investigação militar conduzida pelo exército dos Estados Unidos para apurar um ataque ocorrido no dia anterior em um colégio militar em Trípoli, a capital da Líbia. O incidente resultou na trágica perda de 32 estudantes e deixou vários feridos. O relatório finalizado pelo Conselho de Segurança das Nações Unidas indicou que o ataque foi executado por um drone não tripulado, sem intervenção de comando humano. A máquina disparou um míssil, originalmente destinado a atingir veículos blindados, guiado por um sistema semi-automático a laser. O projétil atravessou duas paredes e explodiu no pátio da escola (Conselho de Segurança da ONU, 2021, apud Moraes, 2023, p. 12).

Assim, à medida que as Inteligências Artificiais (IAs) desempenham papéis significativos em processos de tomada de decisão, monitoramento, e até mesmo na aplicação da lei, estando cada vez mais autônomas, surge a necessidade premente de estabelecer um quadro legal que defina se essas entidades tecnológicas serão responsabilizadas por suas ações e como essa responsabilização funcionaria.

A responsabilidade penal no direito brasileiro refere-se à possibilidade de aplicação de sanções penais a indivíduos que tenham praticado condutas consideradas criminosas de acordo com as normas estabelecidas pelo ordenamento jurídico. Assim, implica a atribuição de consequências legais, como penas privativas de liberdade, penas restritivas de direitos, multas, entre outras, ao agente que cometeu um delito.

A aplicação da responsabilidade penal no direito brasileiro está vinculada à observância dos elementos do crime, conforme definidos pelo Código Penal Brasileiro. Esses elementos incluem a tipicidade (conduta que se enquadra na descrição legal do crime), a antijuridicidade (contrariedade da conduta ao ordenamento jurídico), a culpabilidade (capacidade do agente de entender a ilicitude de sua conduta e agir de acordo) e a punibilidade (possibilidade de aplicação de uma pena) (Reale Júnior, Miguel, p. 42).

A responsabilidade penal pode ser aplicada a qualquer pessoa que tenha capacidade de entendimento e vontade, desde que cometa uma conduta que se enquadre nos tipos penais previstos na legislação. Isso inclui não apenas indivíduos, mas também pessoas jurídicas, de acordo com as disposições legais. No caso das pessoas jurídicas, a responsabilidade penal está prevista em normas específicas e pode envolver sanções como multas, proibição de atividades, entre outras medidas.

Dessa forma, no contexto jurídico brasileiro, a atribuição de responsabilidade ocorre exclusivamente em relação a indivíduos mentalmente sãos e maiores de idade, bem como a pessoas jurídicas, especificamente no âmbito dos crimes ambientais, conforme estabelecido pela própria Constituição Federal. Não existe qualquer disposição relativa a outras categorias de agentes perpetradores de crimes, muito menos em relação a entidades abstratas, como programas de inteligência artificial.

Levando isso em conta, nota-se que no contexto jurídico, a crescente presença de sistemas autônomos e algoritmos inteligentes suscita questionamentos cruciais sobre a responsabilidade penal atribuída a essas entidades não humanas. No âmbito do direito penal brasileiro, a compreensão e a aplicação das normas em relação às Inteligências Artificiais emergem como um desafio complexo.

Nesse contexto, este artigo visa aprofundar a análise sobre a possível penalização das Inteligências Artificiais no âmbito jurídico brasileiro, investigando se tais entidades podem ser submetidas à responsabilidade penal e, em caso afirmativo, como esse processo de responsabilização se desenvolve. Para isso, utilizaremos como método a revisão bibliográfica integrativa, fazendo uma análise dos textos publicados sobre a temática elencada.

Ao investigar essa questão complexa e em rápida evolução, espera-se lançar luz sobre os desafios éticos e legais que surgem à medida que as IAs se tornam cada vez mais desenvolvidas, autônomas e integradas à sociedade e ao sistema jurídico brasileiro.

2. PROCEDIMENTOS METODOLÓGICOS

Trata-se de uma Revisão Bibliográfica Integrativa (RBI). Este é um método de pesquisa que busca sintetizar e integrar resultados de estudos relevantes sobre um determinado tema. É uma forma de revisão de literatura que vai além de uma simples sumarização de resultados, pois busca identificar padrões e tendências, além de explorar as implicações teóricas e práticas dos resultados encontrados. (Mendes, et. al., 2008, p. 760). O percurso metodológico seguiu as seguintes etapas: seleção do tema, coleta dos dados, avaliação dos dados coletados, análise e interpretação dos mesmos, refinamento e apresentação dos resultados.

Para direcionar esta revisão, foi elaborada a seguinte pergunta norteadora: “As Inteligências Artificiais podem ser responsabilizadas no âmbito do direito pe-

nal brasileiro? Caso afirmativo, como essa responsabilização se desenvolve?”. O levantamento dos arquivos foi realizado no segundo semestre de 2023, por meio da internet, através da base de dados do Google Acadêmico, com resultados em português (Brasil).

Inicialmente, com o objetivo de restringir a busca à pergunta orientadora, foram escolhidas as seguintes palavras-chave: inteligência artificial, direito, crimes digitais e responsabilização. Com base nessas definições, foram obtidos aproximadamente 16.100 resultados. Dentre os analisados, verificou-se que 43,75% eram artigos, 25% eram livros, 18,75% eram monografias e 12,75% eram dissertações.

A partir dos dados coletados, refinou-se a pesquisa para alcançar maior precisão em relação ao tema proposto. Como critérios objetivos de filtragem, substituiu-se a palavra-chave "responsabilização" por "responsabilidade penal" e incluiu-se a palavra "Brasil" para delimitar a região estudada. Essa modificação resultou em, aproximadamente, 4.470 documentos.

Além disso, com o intuito de restringir ainda mais os resultados, optou-se por aplicar um filtro cronológico, limitando a pesquisa aos estudos do ano corrente, através da seleção da opção de período específico (2023). Nessa busca, foram encontrados 492 documentos. Entre os resultados analisados, 50% eram artigos, 27,5% eram monografias, 13,75% eram livros ou capítulos de livros, 7,5% eram dissertações e apenas 1,25% eram teses.

Considerando esses aspectos, foram examinados os documentos que melhor se enquadraram à delimitação do tema. Para isso, realizou-se a leitura de 90 documentos, analisando não apenas os títulos, mas também os resumos. A análise dos arquivos baseou-se nos critérios de observância do título do arquivo e do resumo, bem como na identificação de palavras-chave no corpo do texto. Como resultado desse processo, foram selecionados 35 trabalhos, distribuídos entre artigos (19), monografias (10), dissertações (4) e teses (2).

3. RESULTADOS

Os trabalhos foram avaliados com base nos arquivos alinhados à proposta de pesquisa, utilizando a categorização temática. Nesta fase, os 35 documentos iniciais foram submetidos a uma análise subjetiva, que incluiu a leitura de seus resumos e introduções, resultando na exclusão daqueles que não abordaram o tema proposto.

Assim, a pesquisa foi limitada a 20 arquivos e, após a leitura na íntegra desses documentos, foram selecionados 7. Essa escolha foi feita por meio de uma nova aplicação do filtro subjetivo, visando uma melhor adequação dos documentos à temática escolhida. Dentre os resultados escolhidos, 87,5% eram artigos e 12,5% eram monografias.

A leitura completa dos resultados selecionados permitiu identificar subtemas comuns nas discussões, os quais serão elucidados no próximo tópico. Os critérios para extrair esses dados dos arquivos incluíram: relevância para o tema da pesquisa, qualidade metodológica e originalidade das contribuições. Os arquivos selecionados estão dispostos na tabela abaixo.

Autoria	Título do documento selecionado	Objetivo(s)
Renato Watanabe de Moraes	Programações podem ser punidas? Responsabilidade penal em decisões tomadas pela inteligência artificial.	Investigar se no Direito Penal tradicional, com uma visão centrada no ser humano, existe a possibilidade de robôs cometerem crimes e serem responsabilizados por eles, observando a fronteira desafiadora para o nexo de causalidade
Cristiane Helena de Paula Lima Cabral; Galvão Rabello; Gustavo Henrique de Medina Esteves; Victor Henrique Matos Batista.	Discriminação Algorítmica: os riscos do uso da inteligência artificial nos julgamentos criminais pelo poder judiciário brasileiro.	Analisar o uso da inteligência artificial nos julgamentos brasileiros, especialmente os criminais, observando a ideia de que “robôs” acabam por reproduzir determinadas condutas da sociedade e tendem a ser extremamente discriminatórios.

<p>Raquel Bellini Salles; Thais Silva Costa</p>	<p>A securitização dos danos causados por inteligência artificial.</p>	<p>Analisar os impactos da inteligência artificial na responsabilidade civil considerando a dificuldade de caracterização dos pressupostos da obrigação de indenizar, os limites das ações indenizatórias individuais e o potencial lesivo das novas tecnologias.</p>
<p>Zéu Palmeira Sobrinho.</p>	<p>Inteligência Artificial (IA) e Assédio Moral nas Relações de Trabalho.</p>	<p>Investigar o impacto dos sistemas de decisões, fundamentados na inteligência artificial, debruçando-se sobre os efeitos de tais modalidades decisórias, tanto na esfera de aplicação do trabalho, quanto na esfera da chamada política criminal atuarial.</p>
<p>Airto Chaves Junior; Bárbara Guasque; Thiago Santos Aguiar de Pádua.</p>	<p>Segregação racial e vieses algorítmicos: máquinas racistas no âmbito do controle penal.</p>	<p>Verificar se o emprego de Sistemas de Inteligência Artificial poderia abrandar a altíssima carga de seletividade do Controle Penal, notadamente aquelas fundadas no racismo e na discriminação.</p>
<p>Sergio Marcos Carvalho de Ávila Negri; Joana de Souza Machado; Carolina Fiorini Ramos Giovanini; Nathan Pascoalini Ribeiro Batista.</p>	<p>Sistemas de Inteligência Artificial e Avaliações de Impacto para Direitos Humanos.</p>	<p>Analisar estratégias de regulação de sistemas de inteligência artificial, com foco em modelos mais recentes baseados na classificação de riscos</p>

Giordano Bruno Alves Fernandes	Responsabilidade Civil e Inteligência Artificial: da Necessidade de Legislação Específica sobre Inteligência Artificial no Brasil.	Responsabilidade Civil e Inteligência Artificial: da Necessidade de Legislação Específica sobre Inteligência Artificial no Brasil.
--------------------------------	--	--

Tabela 1: Descrição dos estudos utilizados neste trabalho, incluindo autores, título e objetivo geral de cada um.

4. DISCUSSÃO DOS RESULTADOS

Inicialmente, é importante elucidar que, conforme se observa a partir do tópico anterior, poucos trabalhos selecionados delimitavam-se à temática proposta por este artigo. Apenas um dos documentos abordava a questão da responsabilização penal da Inteligência Artificial, este sendo este escrito por Renato Watanabe de Moraes e intitulado “Programações podem ser punidas? Responsabilidade penal em decisões tomadas pela inteligência artificial”.

Os demais documentos perpassaram pela temática abordando-a de forma secundária, focados em outros objetivos, como, por exemplo, a análise do uso de IAs no judiciário brasileiro, a necessidade de uma legislação que regule o desenvolvimento e os comportamentos da Inteligência Artificial e as questões éticas envolvidas com o uso dessa tecnologia.

Isso demonstra que o assunto ainda é pouco discutido no Brasil, apesar da relevância do debate e necessidade premente de entender como o direito penal brasileiro está reagindo ou como reagirá à criação de entidades abstratas autônomas e capazes de cometer crimes.

Contudo, a partir dos documentos elencados, constatou-se a presença de alguns subtemas que se demonstraram importantes nos debates realizados, os quais serão abordados a seguir.

Conforme já comentado, os dados foram extraídos, analisados e discutidos por meio da categorização temática, buscando elencar, através do filtro subjetivo, os pontos mais relevantes para o tema proposto. Os critérios utilizados para a seleção das categorias foram: repetição do debate entre os resultados selecionados

e relevância da discussão para a temática deste artigo.

4.1 RAZÃO PARA PUNIR INTELIGÊNCIAS ARTIFICIAIS

A primeira categoria elencada nesta revisão integrativa é um questionamento que surgiu em alguns dos arquivos selecionados e envolve a problemática da razão para punir inteligências artificiais.

Conforme elencado por Morais (2023, p. 12) e Hulsman (2012, p. 36), a reação penal do Estado, conforme empregada atualmente, não constitui verdadeiramente uma forma de punição, ou pelo menos não se alinha de maneira adequada com os objetivos proclamados pelo Direito Penal em relação à prevenção e repressão. Na realidade, trata-se de um processo comunicativo disfarçado de uma resposta eficaz, repleto de rituais conhecidos e aceitos pelo público, mas que acabam por encobrir a verdadeira natureza da criminalidade, suas causas e os impactos da criminalização. Esse símbolo representado pela sanção estatal poderia ser um argumento relevante contra a resposta penal direcionada a máquinas, por diversas razões.

Primeiramente, se a privação de liberdade de seres humanos já é questionável sob algumas perspectivas criminológicas, menos justificativa haveria para ampliar o alcance do Direito Penal além dos limites que envolvem a pessoa humana, sendo que a responsabilização penal da pessoa jurídica, por exemplo, já enfrenta diversas críticas.

Além disso, do ponto de vista político, considerando que o Direito Penal é uma das ferramentas comunicativas mais poderosas controladas pelo Estado, conforme destacado por Hulsman (2012) apud Morais (2023, p. 13), não há interesse no controle de corpos não humanos. A pena privativa de liberdade representa um capital estatal significativo no processo de construção de imagens perante a sociedade. A mera desativação ou modificação dos códigos-fonte, no caso de uma suposta punição de IA, não geraria o mesmo impacto que o aprisionamento de indivíduos na narrativa da luta contra o crime.

Por outro lado, a autorização irrestrita para a criação de qualquer tipo de Inteligência Artificial, sem a imposição de limites ou controles adequados, levanta uma série de preocupações e riscos significativos, principalmente para o campo do Direito Penal. Esses riscos abrangem questões éticas, sociais e legais, que exigem uma atenção cuidadosa para garantir a proteção dos direitos individuais e a segu-

rança da sociedade como um todo.

Exemplos dos riscos decorrentes da falta de regulação incluem a possibilidade de as IAs, ao ganharem autonomia na tomada de decisões e estabelecerem suas próprias etapas de processamento de dados, cometerem atos ilícitos de forma independente. Ou seja, a partir de suas redes neurais, as IAs passarem a violar direitos e a privacidade das pessoas.

Cria-se, a partir desse cenário, um imbróglio judicial complexo e multifacetado. A falta de previsibilidade do comportamento das máquinas torna desafiador determinar a responsabilidade pelos atos ilícitos, dado que o desenvolvimento, implementação e operação de sistemas de IA envolvem diversas partes, como programadores, fabricantes de hardware e proprietários do sistema.

Contudo, reflexões são necessárias no campo do Direito Penal, posto que, atualmente, é possível criar programas com a garantia de que, mesmo que resultem em consequências desastrosas ao se tornarem imprevisíveis para seus criadores, não haverá resposta estatal. Assim, é crucial incorporar restrições nas capacidades de tomada de decisões das IAs, muito embora isso vá contra a ideia fundamental de aprendizado de máquina. Se não houver limitação, ao menos uma avaliação final humana deve ser realizada (Morais, 2023, p. 14).

4.2 ALGORITMOS ENVIESADOS COMO PERPETUADORES DE DISCRIMINAÇÃO

Outra problemática percebida nos estudos encontrados diz respeito à análise do mal uso da inteligência artificial, decorrente de tendências discriminatórias dos algoritmos, os chamados algoritmos enviesados.

Exemplo disto, são os estudos que abordam o uso de IAs nos sistemas de decisões judiciais, envolvendo, em especial, discussões acerca dos vieses algorítmicos na estruturação de máquinas como o machine learning, as quais possibilitam ao sistema o aprendizado automático (por conta própria) a partir de inferências em bancos de dados, reconhecendo padrões e generalizações, sem a necessidade de uma programação precisa e explícita (Chaves Junior et al., 2023, p.9).

Desta forma, de acordo com o conjunto de objetivos previamente estabelecidos e baseado nos bancos de dados coletados, o algoritmo fornecerá um resultado que pode influenciar todo o ambiente. Isso porque, a forma como a IA se relaciona com o mundo exterior é determinada, primeiramente, pelos dados coletados e, depois, por quem a programou.

Importante ressaltar que vieses referem-se a pré-concepções humanas de aspectos não racionais, dotados de influência político ideológica, as quais resultam em tomadas de decisões de forma irrefletida, sem embasamento lógico. No contexto abordado nos estudos, considerando que os algoritmos são desenvolvidos por indivíduos, os algoritmos enviesados seriam aqueles que reproduzem preconceitos e vieses dos seus programadores ou dos dados com os quais interagem, mimetizando dinâmicas sociais predominantes (Cabral et al., 2023; Chaves Junior, et al., 2023, p.8).

Esses estudos apontam para o desenvolvimento de programação discriminatória de algoritmos de IAs utilizadas no âmbito do Poder Judiciário, os quais acabam operando de forma que uma parcela já marginalizada socialmente seja novamente prejudicada pelo resultado da ferramenta, catalisando preconceitos e segregando setores sociais (Cabral et al., 2023; Chaves Junior, et al., 2023).

Segundo os pesquisadores, esses comportamentos, embora bastante incisivos no âmbito criminal, também abrangem discriminação de gênero e racial em questões civis, como o viés sexista na concessão de limite de crédito por operadoras de cartão de crédito; misógino na estereotipação de mulheres negras como prostitutas (Cabral et al., 2023, p.); sexista e misógino na seleção de currículos masculinos por empresas e racista, quando a precisão de algoritmos de reconhecimento facial está, em geral, diretamente relacionada com as características de gênero e raça dos criadores dos modelos utilizados (Chaves Junior, et al., 2023).

Conforme apontam Chaves Junior et al. (2003, p.11-14), no âmbito penal, algoritmos enviesados podem acarretar em pontos cegos nos modelos criados, posto que se determinados grupos não forem bem representados, não terão uma representatividade estatística, acarretando consequências sociais danosas, catalisando injustiças sociais e intensificando o racismo e preconceito que permeiam no Sistema Penal e, conseqüentemente, lesando Direitos Fundamentais constitucionalmente assegurados. Isto tudo com a falsa aparência de neutralidade matemática apresentada pelas IAs.

4.3 A NECESSIDADE DE INCORPORAÇÃO DE VALORES ÉTICOS E SOCIAIS NO DESENVOLVIMENTO DE IA'S

Através da análise dos textos selecionados, notou-se que a incorporação de valores éticos e sociais no desenvolvimento de Inteligências Artificiais é essencial para assegurar que essas tecnologias beneficiem a sociedade de maneira ética,

justa e responsável.

A integração de valores éticos não apenas implica transparência nas decisões das IAs, favorecendo a compreensão e confiança dos usuários e da sociedade em geral, mas também assegura que essas tecnologias sejam desenvolvidas com respeito à privacidade dos indivíduos. Tal prática visa garantir a segurança e confidencialidade dos dados pessoais, fatores que desempenham um papel crucial no desenvolvimento de IAs seguras e confiáveis, resultando em uma maior aceitação dessas tecnologias.

Ademais, a incorporação de valores éticos assegura que as IAs estejam alinhadas com princípios morais fundamentais da sociedade, promovendo o desenvolvimento de tecnologias que respeitem a dignidade e os direitos humanos. Dessa forma, é possível que as IAs compreendam melhor o contexto humano, levando em consideração aspectos culturais, sociais e emocionais nas interações com as pessoas. Essa abordagem é essencial para garantir que elas contribuam positivamente para o avanço da sociedade como um todo.

Somando-se a isso, a garantia de que as IAs sejam desenvolvidas considerando o impacto social é crucial para a promoção de tecnologias responsáveis e benéficas para a sociedade. Essa abordagem contribui para evitar a reprodução de preconceitos e discriminações presentes nos dados de treinamento desses algoritmos, promovendo equidade e justiça nas decisões automatizadas.

Além disso, a inserção de valores sociais também contribui para garantir que as IAs estejam em conformidade com normas reconhecidas internacionalmente, promovendo consistência e aceitação global e, ainda, para prevenir o uso indevido dessas tecnologias para fins prejudiciais, garantindo uma aplicação ética e responsável.

Assim, é imperioso lançar uma atenção rigorosa sobre o desenvolvimento dos sistemas de IA, preocupando-se com a base de dados utilizada quando do treinamento e validação dos modelos, buscando assegurar que padrões éticos e responsáveis sejam incorporados nas entidades (Chaves Junior, et al, 2023, p. 22).

Uma maneira de assegurar que as inteligências artificiais sejam desenvolvidas com responsabilidade e incorporem valores éticos e sociais é por meio do método ART - accountability, responsibility, and transparency. Em resumo, a accountability implica que a máquina seja capaz de explicar e justificar suas decisões (prestação de contas). Isso requer a capacidade de demonstrar o processo de tomada de decisão utilizado. É importante destacar que, neste ponto, surgem desafios, espe-

cialmente no aprendizado de máquina, onde é difícil demonstrar precisamente os critérios considerados pela tecnologia antes de tomar uma decisão, especialmente em modelos de várias camadas (Dignum, 2019 apud Fernandes, 2023, p. 30).

A responsibility refere-se ao papel das pessoas em relação à inteligência artificial. É crucial estabelecer a capacidade de vincular as decisões da IA com as entradas recebidas e as ações dos interessados envolvidos na tomada de decisão da máquina. Por exemplo, o uso da IA para fins ilegais, como no caso do deepfake, e na vigilância e obtenção de dados pessoais de terceiros destaca situações em que pessoas ou instituições se beneficiam das ações da máquina, mesmo que estas não tenham sido explicitamente programadas para tal (Dignum, 2019 Apud Fernandes, 2023, p. 30).

Já a transparency diz respeito à capacidade de descrever, inspecionar e reproduzir os mecanismos pelos quais a IA toma decisões e aprende a se adaptar ao ambiente, incluindo a origem dos dados usados e criados pela IA. Contudo, a transparência enfrenta desafios, uma vez que quase sempre implica na divulgação do código do sistema, e a maioria dos sistemas é desenvolvida pelo setor privado.

No entanto, esse debate revela-se ainda bastante incipiente. Acredita-se que, com a aceleração constante na evolução dos sistemas de IA, as discussões éticas assumirão um papel cada vez mais crucial nas próximas décadas. Isso poderá resultar em regulamentações mais adequadas para o desenvolvimento da tecnologia, contribuindo, assim, para uma maior precisão nas implicações legais para os indivíduos envolvidos no sistema sócio-tecnológico da inteligência artificial.

5. CONSIDERAÇÕES FINAIS

A presente revisão integrativa evidenciou que o debate sobre a responsabilização penal de Inteligências Artificiais é algo complexo e desafiador. Por meio dos estudos analisados, foi possível perceber que, atualmente, não existem previsões legais para a punição criminal de entidades não humanas no âmbito do direito penal brasileiro. Contudo, ficou demonstrado que reflexões sobre a temática são necessárias, considerando todo o atual cenário de desregulação e insegurança jurídica.

Com base nos resultados selecionados, percebe-se que, no Brasil, a discussão sobre a temática ainda é inexpressiva, sendo que, no ano de 2023, no respectivo banco de dados utilizado, foi encontrado apenas 1 artigo que abordava o assunto proposto.

Analisando esse cenário, nota-se que a situação é alarmante. A temática da responsabilização das Inteligências Artificiais (IAs) é uma questão difícil e em rápida evolução, na medida em que as IAs se tornam cada vez mais desenvolvidas, autônomas e integradas à sociedade e ao sistema jurídico brasileiro, e, apesar disso, as academias pouco estão produzindo sobre o assunto.

Estamos realizando, em paralelo, a mesma revisão integrativa com o filtro para língua inglesa e, com base nisso, nos parece que os estudos nessa língua estão consideravelmente mais avançados quanto ao tema elencado que no Brasil.

Sob uma análise mais crítica, percebe-se que há uma indiferença por parte da doutrina com relação a um elemento fundamental, que é o poder-dever de garante, essencial à teoria do crime. Falou-se sobre a suposta possibilidade de responsabilização penal de IAs e robôs, sem se debruçar sobre as ações potencialmente dolosas, ou no mínimo imprudentes, dos agentes que capacitam e autonomizam essas entidades tecnológicas, fazendo-se indiferentes às enormes possibilidades lesivas das mesmas.

Ouvida-se que, a título de exemplo, um modelo de substituição massiva e estrutural de funcionários de atendimento por chatbot é algo, por si só, problemático por parte de gestores de empresas. A transferência da imputação penal dos sujeitos que consentem as possibilidades lesivas para os próprios programas de computador ou, quando muito, para seus programadores, se mostra como uma premissa oculta de todo o raciocínio apresentado pela literatura estudada.

Assim, acredita-se que, antes de se debater sobre a possibilidade de Inteligências Artificiais serem punidas, o primordial seria delimitar a responsabilidade daqueles que dão autonomia para esses sistemas tecnológicos, em casos em que são omissos e imprudentes ao contratarem e utilizarem tais entidades.

Conforme elencado ao longo das discussões acerca dos resultados encontrados, acredita-se ser crucial incorporar restrições nas capacidades de tomada de decisões das IAs, adotando valores éticos e sociais no seu desenvolvimento. Contudo, mostra-se primordial estabelecer a responsabilidade no uso dessas tecnologias, sendo que o crivo final, determinante para o aprendizado das máquinas e suas consequências, deve ser o humano.

REFERÊNCIAS

BATISTA, N.P.R.; GIOVANINI, C.F.R.; MACHADO, J.S.; NEGRI, S.M.C.A. Sistemas de Inteligência Artificial e Avaliações de Impacto para Direitos Humanos. Culturas Jurídicas. 13/07/2023. Disponível em: <https://periodicos.uff.br/culturasjuridicas/article/view/56809>.

CABRAL, C.H.P.L.; RABELLO, G.; ESTEVES, G.H.M.; BATISTA, V.H.M. Discriminação algorítmica: os riscos do uso da inteligência artificial nos julgamentos criminais pelo poder judiciário brasileiro. Disponível em: <https://repositorio-api.ani-maeducacao.com.br/server/api/core/bitstreams/845dd7a4-2a0d-43d7-87b2-0db-d456fdb6a/content>. Acesso em: 25/10/2023.

CHAVES JUNIOR, A.; GUASQUE, B.; PÁDUA, T.S.A. Segregação racial e vieses algorítmicos: máquinas racistas no âmbito do controle penal. Revista Brasileira de Direito, Passo Fundo, v. 19, n. 2, p. e4768, set. 2023. ISSN 2238-0604. Disponível em: <https://seer.atitus.edu.br/index.php/revistadedireito/article/view/4768>. Acesso em: 16 nov. 2023. doi:<https://doi.org/10.18256/2238-0604.2023.v19i2.4768>.

CONSELHO DE SEGURANÇA DA ONU. Resolução 1970 (2011). Index: S/2021/229, 08 mar. 2021. Disponível em: <https://undocs.org/S/2021/229>.

FERNANDES, G. Responsabilidade Civil e Inteligência Artificial: da Necessidade de Legislação Específica sobre Inteligência Artificial no Brasil. Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/28173>.

MENDES, K.D.S.; SILVEIRA, M.F.; GALVÃO, C.M.; SANTOS, L.M. (2008). Revisão integrativa: método de pesquisa para a incorporação de evidências na saúde. Revista da Escola de Enfermagem da USP, 42(4), 758-764.

MORAIS, R.W. Programações podem ser punidas? Responsabilidade Penal em Decisões Tomadas pela Inteligencia Artificial. Revista Científica do CPJM, 2(07), 130-148. Disponível em: <https://rcpjm.emnuvens.com.br/revista/article/view/217>

RUSSELL, S.J.; NORVIG, P., "Artificial Intelligence: A Modern Approach," (3rd ed.), Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13- 790395-2. Disponível em: https://people.engr.tamu.edu/guni/csce421/files/AI_Russell_Norvig.pdf

REALE JUNIOR, M. Fundamentos de Direito Penal, 5ª ed. São Paulo: Grupo GEN, 2020. p. 42

SOBRINHO, Z.P. A INTELIGÊNCIA ARTIFICIAL (IA) E ASSÉDIO MORAL NAS RELAÇÕES DE TRABALHO. Revista FIDES, v. 14, n. 1, p. 45-57, 14 jun. 2023. Disponível em: <http://www.revistafides.ufrn.br/index.php/br/article/view/658>

PAULA, C.H.; CABRAL, L.; RABELLO, G.; ESTEVES, Gustavo Henrique de Medina; BATISTA, Victor Henrique Matos. Discriminação Algorítmica: os riscos do uso da inteligência artificial nos julgamentos criminais pelo poder judiciário brasileiro. Revista nima, [s.l.], 16 jun. 2023. Disponível em: <https://repositorio-api.animaeducacao.com.br/server/api/core/bitstreams/845dd7a4-2a0d-43d7-87b2-0dbd456fdb6a/content>.

PHILLIP, Abby. Robot grabs man, killshim in German car factory. Washington Post. 2015. Disponível em: <https://www.washingtonpost.com/news/morning-mix/wp/2015/07/02/robot-grabs-man-kills-him-in-german-car-factory/>.

SALLES, Raquel Bellini; COSTA, Thais Silva. A securitização dos danos causados por inteligência artificial. civilistica.com, v. 12, n. 1, p. 1-32, 31 maio 2023. Disponível em: <https://civilistica.emnuvens.com.br/redc/article/view/870>

WAKABAYASHI, Daisuke. Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam. The New York Times. 2018. Disponível em <https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html>.