

## CRIMINALIZAÇÃO DO USO INDEVIDO DA INTELIGÊNCIA ARTIFICIAL: DESAFIOS LEGAIS E IMPACTOS NA RESPONSABILIDADE PENAL

CRIMINALIZATION OF THE MISUSE OF ARTIFICIAL INTELLIGENCE: LEGAL  
CHALLENGES AND IMPACTS ON CRIMINAL RESPONSIBILITY

CRIMINALISATION DE L'USAGE ABUSIF DE L'INTELLIGENCE ARTIFICIELLE : DÉFIS  
JURIDIQUES ET IMPACTS SUR LA RESPONSABILITÉ PÉNALE

Aline da Silva Araújo<sup>1</sup>  
Camila de Sousa da Silva<sup>2</sup>  
Delner do Carmo Azevedo<sup>3</sup>

**RESUMO:** Este artigo propõe uma análise sobre como a inteligência artificial pode replicar discriminações e outras violações de direitos no contexto da persecução penal. A pesquisa baseia-se na abordagem dialética, utilizando como técnica de coleta a documentação indireta, com ênfase em fontes bibliográficas. A principal conclusão destaca que, ao contrário do que alguns autores defendem que decisões automatizadas seriam mais objetivas, consistentes e neutras, a realidade mostra que decisões orientadas por sistemas que simulam o comportamento humano podem, na verdade, amplificar violações de direitos. A principal contribuição deste estudo é ressaltar a importância de reconhecer que o desenvolvimento, a aplicação e o uso da inteligência artificial ou de qualquer algoritmo devem ser conduzidos em um ambiente comprometido com o respeito e a promoção dos direitos humanos.

**Palavras-chave:** Inteligência artificial. Crimes. Responsabilidade penal.

**ABSTRACT:** This article analyzes how artificial intelligence can replicate discrimination and other rights violations in the context of criminal prosecution. The research adopts a dialectical approach, utilizing indirect documentation as a collection technique, with an emphasis on bibliographic sources. The main conclusion highlights that, contrary to claims that automated decisions are more objective, consistent, and neutral, reality shows that decisions guided by systems simulating human behavior can amplify rights violations. The study's primary contribution lies in emphasizing the importance of ensuring that the development, application, and use of artificial intelligence or any algorithm take place in an environment committed to respecting and promoting human rights.

147

**Keywords:** Artificial Intelligence. Crimes. Criminal responsibility.

**RESUMEN:** Cet article propose une analyse de la manière dont l'intelligence artificielle peut reproduire des discriminations et d'autres violations des droits dans le cadre de la poursuite pénale. La recherche repose sur une approche dialectique, utilisant la documentation indirecte comme technique de collecte, avec une attention particulière portée aux sources bibliographiques. La principale conclusion souligne que, contrairement à l'idée selon laquelle les décisions automatisées seraient plus objectives, cohérentes et neutres, la réalité montre que les décisions guidées par des systèmes simulant le comportement humain peuvent, en fait, amplifier les violations des droits. La principale contribution de cette étude est de mettre en avant l'importance de garantir que le développement, l'application et l'utilisation de l'intelligence artificielle ou de tout algorithme se déroulent dans un environnement engagé envers le respect et la promotion des droits humains.

**Mots-clés :** Intelligence artificielle. Crimes. Responsabilité pénale.

<sup>1</sup> Graduanda no Curso de Direito; Centro Universitário São Lucas.

<sup>2</sup> Graduanda no Curso de Direito; Centro Universitário São Lucas.

<sup>3</sup> Orientador no curso de direito. Centro Universitário São Lucas.

## I. INTRODUÇÃO

Este trabalho aborda a questão da responsabilização penal por atos ilícitos praticados por sistemas de inteligência artificial. Seu objetivo é analisar, com base em um estudo detalhado da legislação brasileira e em teorias nacionais e internacionais, se é viável responsabilizar penalmente esses sistemas e, a partir dessa análise, identificar quem responderia juridicamente por tais atos. A motivação para este estudo vem da rápida evolução dos sistemas de IA, que têm influenciado significativamente a forma como as pessoas agem, pensam e se comunicam. Muitos sistemas de inteligência artificial possuem autonomia suficiente para se autor reprogramarem ou tomarem decisões de forma independente, sem intervenção humana.

A justificativa do estudo se dá sobre as violações de direitos tendem a ser mais evidentes quando são causadas pela ação ou omissão de pessoas do que quando resultam de processos e decisões geradas por algoritmos. Isso ocorre porque, no contexto computacional, as regras e definições que estruturam a lógica dos programas, e que podem influenciar os resultados obtidos, muitas vezes são ocultas pela falta de transparência ou pela imprevisibilidade e autonomia de certos sistemas complexos.

No entanto, nem sempre esses sistemas seguem as instruções originalmente programadas, pois são suscetíveis a falhas, mesmo sem interferência humana. Isso reforça a necessidade de mais investigações, visto que as decisões autônomas desses sistemas podem resultar em crimes e ainda existem muitas incertezas sobre como o direito penal brasileiro lida com essa questão.

148

A pesquisa é fundamentada principalmente em fontes bibliográficas e documentais, utilizando métodos como leitura, seleção, fichamento, organização e síntese de informações. A abordagem é qualitativa e exploratória, e a coleta de dados segue o procedimento de pesquisa bibliográfica.

Para facilitar a compreensão do texto para os leitores e manter um estilo coerente com o uso de outros autores do Direito, este trabalho adota o termo "algoritmo" de forma mais abrangente, sem o rigor técnico da palavra, incluindo também o que na Computação é chamado de "código". Essa escolha de intercambiar os termos é feita especificamente pelos motivos mencionados.

## 2. CONCEITO E NOÇÕES DE INTELIGÊNCIA ARTIFICIAL

Conforme Kaplan e Haenlein (2019) definem inteligência artificial como a capacidade de um sistema de analisar corretamente dados externos, aprender a partir deles e empregar esse aprendizado para atingir objetivos específicos, adaptando-se de forma flexível, a inteligência artificial é um campo que explora a criação e análise de agentes computacionais que atuam de forma inteligente. Para Russell e Norvig (2020), o conceito envolve o estudo de agentes que recebem estímulos do ambiente e respondem a eles. Assim, o conceito de IA, apesar de amplo, baseia-se na capacidade de um sistema para interpretar dados, aprender com eles e aplicá-los para cumprir tarefas e objetivos de forma adaptativa.

Nas últimas duas décadas, novas técnicas computacionais foram desenvolvidas. Com o aumento da capacidade do hardware e o uso de grandes conjuntos de dados para treinamento em aprendizado de máquina, foi possível aplicar algoritmos matemáticos ao campo da inteligência artificial.

Assim, passou a ser viável que tarefas complexas antes realizadas por humanos fossem desenvolvidas e executadas por meio de algoritmos. O avanço das técnicas computacionais possibilitou que algoritmos pudessem, com base em dados iniciais, estabelecer relações, realizar funções e fazer previsões de forma autônoma, sem intervenção humana direta durante o processo de aprendizado.

O termo “inteligência artificial” é frequentemente usado de forma intercambiável com “aprendizado de máquina”, que pode ser supervisionado (onde algoritmos aprendem com dados de treinamento fornecidos) ou não supervisionado (identificando novos padrões nos dados sem conhecimento prévio). Dessa forma, o aprendizado de máquina permite reconhecer e inferir padrões e executar tarefas específicas sem instruções explícitas de um operador. (STAUB, 2018).

O termo agente refere-se a algo ou alguém que executa ações. Porém, nem todo agente é inteligente, um agente é considerado inteligente quando realiza ações adequadas para sua situação e objetivos, adapta-se a mudanças no ambiente e nas metas, aprende com a experiência e faz escolhas apropriadas de acordo com suas limitações perceptivas e computacionais. (BARTNECK et al., 2020, p.34).

A inteligência artificial pode ser classificada em níveis, como a IA forte (STAUB, 2018) e a IA fraca (MÁNTARAS, 2019). A IA forte é aquela em que o agente consegue replicar o comportamento humano a ponto de ser quase indistinguível de uma pessoa. A IA fraca, por outro lado, refere-se a agentes que demonstram alguns traços de inteligência humana, mas

apenas em uma tarefa ou campo específico, como no caso de chatbots, que respondem a perguntas em temas para os quais foram treinados.

## 2.1. SISTEMA DE INTELIGÊNCIA ARTIFICIAL

A área de Inteligência Artificial (IA), um ramo da ciência da computação, está em constante avanço e impacta várias esferas, como a Medicina, o mercado financeiro, as engenharias e, em particular, o Direito, foco deste estudo. Para Coppin (2013, p. 4), embora não haja uma definição exata, a IA pode ser compreendida de forma simplificada como sistemas que aparecem ser inteligentes, executando tarefas de maneira autônoma. Tais técnicas são aplicadas em problemas que variam de simples a complexos e repetitivos.

Brookshead (2013, p. 414) complementa, definindo a IA como um estudo voltado para a criação de máquinas autônomas, capazes de realizar atividades sem a intervenção humana. Dada a influência significativa desses sistemas na vida em sociedade, esta pesquisa abordará as formas de aprendizado das máquinas, o processo de tomada de decisões e as características de seu desenvolvimento.

A inteligência, tradicionalmente entendida como a capacidade humana de compreender, raciocinar e resolver problemas, já era debatida quanto à possibilidade de ser replicada por máquinas. Os otimistas sugerem que, à medida que se comprehendem os processos físicos humanos, poderia ser viável reproduzir tais sistemas em máquinas, ainda que com variações materiais. Por outro lado, céticos argumentam que a mente humana, sendo imaterial, não pode ser reproduzida em robôs puramente materiais.

150

A presença da Inteligência Artificial (IA) em nosso cotidiano tem crescido rapidamente. Em muitas situações, utilizamos modelos baseados em algoritmos de IA sem sequer perceber (FACELI, 2021), desde atividades rotineiras, como ler e-mails, lavar roupas, dirigir veículos autônomos ou semiautônomos, até escolher filmes ou séries em plataformas de streaming. Em outros casos, o uso da IA é explicitamente mencionado, criando a expectativa de algo mais avançado. Nos últimos anos, o número de produtos e serviços que promovem o uso de IA em seu marketing aumentou consideravelmente, o que passou a ser visto como um selo de qualidade superior em relação aos concorrentes, embora nem sempre isso seja verdadeiro, tanto em termos de uso real de IA quanto de superioridade.

Mas o que impulsionou tanto esse crescimento? O fator principal foi o desenvolvimento acelerado de tecnologias para extração, armazenamento, transmissão e processamento de dados, que são os quatro pilares do avanço tecnológico que suportaram a expansão da IA.

No primeiro pilar, a extração de dados avançou com o desenvolvimento de sensores mais rápidos, acessíveis e sofisticados, incluindo câmeras de alta tecnologia. No segundo, o armazenamento de dados evoluiu devido ao uso de novos materiais que aumentaram a capacidade de armazenagem, melhorando a confiabilidade e reduzindo os custos. Apesar de parecer um tema recente, a ideia de uma "máquina pensante" remonta a tempos antigos. Em 1950, o matemático Alan Turing propôs um teste para verificar se uma máquina poderia simular o pensamento humano. Esse teste, conhecido como Teste de Turing, consiste em uma pessoa fazer perguntas a um humano e a um sistema de IA, sem saber qual é qual. (FACELI, 2021). Se a máquina consegue "enganar" o questionador, ela passa no teste.

### 3. RESPONSABILIDADE JURÍDICA E ÉTICA NA INTELIGÊNCIA ARTIFICIAL

Em 2012, a União Europeia começou a discutir o aspecto jurídico das ações realizadas por robôs no contexto da inteligência artificial. Em 2017, o Parlamento Europeu aprovou a Resolução 2015/2013, que regulamenta os impactos dessa revolução tecnológica em todos os Estados-membros, com foco especial na IA. Entre as diretrizes estabelecidas, inclui-se a criação de um código de ética para engenheiros, programadores e desenvolvedores, visando garantir o respeito à dignidade humana, à privacidade e à segurança (PEREIRA; TEIXEIRA, 2019, p. 128).

151

A resolução também abordou a necessidade de instituir um estatuto jurídico para robôs, com o objetivo de definir a responsabilidade por danos causados ou por decisões tomadas de maneira autônoma (PEREIRA; TEIXEIRA, 2019). Para que essa responsabilidade seja atribuída, o parágrafo 1º dos “Princípios gerais relativos ao desenvolvimento da robótica e da inteligência artificial para uso civil” determina que o robô deve ter: aquisição de autonomia por meio de sensores e/ou troca de dados com o ambiente (interconectividade) e a análise desses dados; capacidade de autoaprendizado com a experiência e interação (critério opcional); um suporte físico mínimo; adaptação de comportamento e ações ao ambiente; e ausência de vida no sentido biológico.

A Resolução também prevê a possibilidade de responsabilizar quem programou ou instruiu a máquina, de acordo com o nível de autonomia dado à IA. Assim, os atos prejudiciais

cometidos pelos robôs podem, em algumas situações, acarretar responsabilidade para o criador humano (PEREIRA; TEIXEIRA, 2019).

Nesse contexto, o pesquisador Gabriel Hallevy (2010) definiu três modelos no sistema jurídico dos EUA para tratar da responsabilidade penal em casos que envolvem IA: o modelo *perpetration-by-another*, o modelo *natural-probable-consequence* e o modelo *direct liability*. O primeiro, *perpetration-by-another*, atribui a culpa ao programador. O segundo, *natural-probable-consequence*, responsabiliza o programador ou o usuário, considerando o risco assumido e a falta de medidas preventivas. No modelo *direct liability*, há possibilidade de responsabilização direta da IA, com cinco sanções penais possíveis: (a) desativação temporária da IA; (b) restrição de suas áreas de atuação; (c) direcionamento para uso social; (d) trabalho compulsório em tarefa específica; ou, até mesmo, (e) desligamento completo da tecnologia (PAULA; CORNWALL; CABRAL, 2019, p. 112).

### 3.1. O USOS INDEVIDO DA INTELIGÊNCIA ARTIFICIAL E A RESPONSABILIDADE PENAL

O uso indevido da Inteligência Artificial (IA) tem se tornado uma preocupação crescente, à medida que a tecnologia avança em diversas áreas, incluindo a justiça, a medicina, a educação e o setor financeiro. No entanto, a IA também tem sido empregada para a prática de crimes, levando a questionamentos sobre a responsabilidade penal de indivíduos ou empresas envolvidas em atividades ilícitas. A capacidade da IA de processar grandes volumes de dados, fazer previsões e tomar decisões autônomas a torna uma ferramenta poderosa, mas também suscetível ao abuso.

Um dos principais riscos do uso indevido da IA está relacionado a crimes cibernéticos, como o roubo de dados pessoais e financeiros. Algoritmos de IA podem ser programados para realizar ataques de phishing em larga escala, automatizando a criação de e-mails fraudulentos e aumentando a eficiência de estelionatários. A IA também pode ser utilizada para a criação de *deepfakes*, vídeos falsificados que podem ser usados para difamação, extorsão e manipulação de informações. Nesse contexto, surge a questão da autoria desses crimes: seria a responsabilidade atribuída ao programador, ao usuário da tecnologia ou à própria máquina, uma vez que ela pode agir de maneira autônoma?

Além disso, a IA tem sido empregada em práticas discriminatórias, como o uso de algoritmos de seleção de candidatos a empregos ou concessão de crédito que, ao serem mal

calibrados, podem resultar em discriminação racial, de gênero ou socioeconômica. Tais discriminações podem ser tratadas como crimes de violação de direitos humanos ou de igualdade, gerando uma responsabilidade penal tanto para as empresas que implementam essas tecnologias quanto para os desenvolvedores dos sistemas de IA.

Outro crime potencialmente associado à IA é o uso de algoritmos para manipulação de mercado, como a prática de *Insider trading*, onde algoritmos são empregados para analisar informações privilegiadas e realizar transações financeiras antes que o mercado tenha acesso aos dados. Esse tipo de prática pode ser enquadrado em crimes financeiros, com implicações graves para a integridade dos mercados financeiros.

A responsabilidade penal diante do uso indevido da IA levanta um desafio complexo. Tradicionalmente, o Direito Penal exige a presença de dolo ou culpa por parte de um agente humano. No entanto, com a autonomia crescente das máquinas, surge a questão de como atribuir responsabilidade penal quando a IA age de forma inesperada ou em conflito com a intenção do seu programador. Uma das abordagens em discussão é a de considerar a responsabilidade penal objetiva para casos em que, embora não haja intenção criminosa explícita, o impacto da ação da IA seja prejudicial e cause danos significativos. Outra proposta seria responsabilizar os desenvolvedores ou empresas que falharem em implementar medidas de segurança, não testando adequadamente seus sistemas de IA para evitar abusos e danos.

Além disso, a dificuldade de traçar uma linha clara entre a ação humana e a ação da IA exige uma revisão dos conceitos tradicionais de imputação penal. Alguns defendem a criação de uma figura jurídica intermediária, como o "agente automatizado", que poderia ser responsabilizado de maneira compartilhada com os humanos que programam, mantêm ou operam essas tecnologias.

Em resposta ao aumento de crimes associados à IA, alguns países têm começado a desenvolver legislações específicas para lidar com as implicações legais da Inteligência Artificial. A União Europeia, por exemplo, propôs um regulamento abrangente sobre IA buscando estabelecer um quadro jurídico que defina claramente as responsabilidades das empresas e a proteção dos direitos dos indivíduos, enquanto busca garantir que as inovações tecnológicas não sejam usadas de forma prejudicial.

O desafio maior será, portanto, equilibrar a inovação tecnológica com a segurança jurídica, assegurando que os direitos dos cidadãos sejam protegidos sem sufocar o progresso tecnológico. A legislação sobre crimes envolvendo IA e a responsabilidade penal é um campo

em constante evolução, que exigirá uma abordagem multidisciplinar, envolvendo juristas, especialistas em ética, cientistas da computação e outros profissionais para garantir uma regulação justa e eficiente.

Como apontado por Morais (2023, p. 12) e Hulsman (2012, p. 36), a resposta penal do Estado, conforme é aplicada atualmente, não pode ser considerada uma verdadeira punição, ou, pelo menos, não se alinham adequadamente com os objetivos do Direito Penal relacionados à prevenção e repressão de crimes. Na prática, trata-se de um processo comunicativo disfarçado de uma ação eficaz, repleto de rituais aceitos pela sociedade, mas que, na verdade, oculta a verdadeira natureza da criminalidade, suas causas e os efeitos da criminalização. Essa sanção estatal, simbólica por natureza, poderia ser um argumento contra a aplicação da resposta penal a máquinas, por várias razões.

Primeiramente, se a privação de liberdade de seres humanos já é questionada sob diversas perspectivas criminológicas, seria ainda mais difícil justificar a expansão do Direito Penal para além dos limites da pessoa humana, especialmente considerando que a responsabilização penal das pessoas jurídicas já enfrenta inúmeras críticas.

Além disso, sob o ponto de vista político, considerando que o Direito Penal é uma das ferramentas comunicativas mais poderosas que o Estado controla, como destacado por Hulsman (2012) citado por Morais (2023, p. 13), não há interesse em controlar corpos não humanos. A pena privativa de liberdade tem um valor simbólico significativo no processo de construção da imagem do Estado perante a sociedade. A simples desativação ou modificação de códigos-fonte, no caso de uma punição direcionada à inteligência artificial, não teria o mesmo impacto narrativo que a prisão de indivíduos na luta contra o crime.

154

Por outro lado, a criação irrestrita de qualquer tipo de Inteligência Artificial, sem a devida imposição de limites ou regulamentações, gera uma série de preocupações e riscos importantes, especialmente no campo do Direito Penal. Esses riscos envolvem questões éticas, sociais e jurídicas que demandam atenção cuidadosa para assegurar a proteção dos direitos individuais e a segurança da sociedade como um todo. (PEREIRA; TEIXEIRA, 2019).

Esse cenário cria um complexo imbróglio judicial, multifacetado e desafiador. A imprevisibilidade do comportamento das máquinas dificulta a determinação de responsabilidades pelos atos ilícitos, uma vez que o desenvolvimento, a implementação e a operação de sistemas de IA envolvem várias partes, como programadores, fabricantes de hardware e proprietários dos sistemas.

De acordo com Bartneck (2020) são necessárias reflexões no campo do Direito Penal, pois, atualmente, é possível criar programas que, mesmo gerando consequências desastrosas por se tornarem imprevisíveis para seus criadores, não sofreram qualquer resposta estatal. “É fundamental estabelecer restrições nas capacidades de tomada de decisões das IAs, embora isso contrarie a ideia essencial de aprendizado de máquina”. Caso essas limitações não sejam impostas, ao menos uma avaliação humana final deve ser realizada (MORAIS, 2023, p. 14).

### **3.2. O CÓDIGO PENA E A OBSERVÂNCIA DO USO DA INTELIGÊNCIA ARTIFICIAL**

No entanto, ao analisar o conceito legal de crime estabelecido no artigo 1º da Lei de Introdução ao Código Penal, observa-se que não há uma especificação que restrinja a imputação do delito exclusivamente à pessoa natural. Isso sugere a possibilidade de atribuir culpabilidade tanto à inteligência artificial quanto à pessoa jurídica.

Quanto ao critério analítico, a questão se torna mais complexa ao considerar os diferentes sistemas que desenvolveram teorias para definir o crime. Como já foi mencionado, o sistema finalista de Hans Welzel, adotado pelo Código Penal brasileiro, define o crime como uma conduta humana direcionada a um fim, com dolo ou culpa voltados para a vontade, incluindo negligência, imperícia ou imprudência (PAULA; CORNWALL; CABRAL, 2019). É nesse ponto que surge um dos principais desafios para a responsabilização penal da inteligência artificial no Brasil: a questão da vontade.

155

No que diz respeito à definição de culpabilidade no sistema finalista, há também uma limitação atual para que seja possível atribuir essa característica à IA. A imputabilidade, conforme definida por Estefam e Gonçalves (2019, p. 386), "consiste na capacidade mental de compreender o caráter ilícito do fato", o que só poderia ser atribuído à IA se fosse comprovado o desenvolvimento tecnológico necessário para possibilitar essa capacidade.

A responsabilidade penal no sistema jurídico brasileiro está relacionada à possibilidade de impor sanções a indivíduos que realizem atos considerados criminosos de acordo com as normas estabelecidas na legislação. Isso implica na aplicação de consequências legais, como penas privativas de liberdade, restrições de direitos, multas, entre outras, aos agentes que cometem infrações. (MORAIS, 2023).

A aplicação dessa responsabilidade está diretamente ligada à observância dos elementos do crime, conforme definidos pelo Código Penal Brasileiro. Estes elementos incluem a

tipicidade (quando a ação se encaixa na descrição legal de um crime), a antijuridicidade (quando a ação é contrária à lei), a culpabilidade (a capacidade do agente de compreender a ilicitude de sua conduta e agir em conformidade) e a punibilidade (a possibilidade de impor uma pena) (REALE, 2020. p. 42).

A responsabilidade penal pode ser atribuída a qualquer pessoa capaz de entender e querer, desde que realize uma conduta que se enquadre nos tipos penais estabelecidos pela legislação. Isso abrange tanto indivíduos quanto pessoas jurídicas, conforme as disposições legais. (STAUB, 2018). No caso das pessoas jurídicas, a responsabilidade penal é tratada por normas específicas e pode resultar em sanções como multas e a proibição de atividades, entre outras.

No contexto jurídico brasileiro, a responsabilidade penal é aplicada exclusivamente a indivíduos com plena capacidade mental e maioridade, bem como a pessoas jurídicas, especialmente em crimes ambientais, conforme disposto pela Constituição Federal. (BRASIL, 1988). Não há previsão para outras categorias de agentes infratores, muito menos para entidades abstratas, como programas de inteligência artificial.

Neste cenário, a crescente utilização de sistemas autônomos e algoritmos inteligentes levanta questões significativas sobre a responsabilidade penal de tais entidades não humanas. No direito penal brasileiro, compreender e aplicar as normas sobre as Inteligências Artificiais surge como um desafio complexo, aprofundar a análise sobre a possível punição das Inteligências Artificiais no contexto jurídico brasileiro, investigando se essas entidades podem ser responsabilizadas penalmente e, caso afirmativo, como esse processo poderia ocorrer. Para tanto, será adotado o método de revisão bibliográfica integrativa, com uma análise dos textos existentes sobre o tema. (SOUZA, 2020).

156

Ao explorar essa questão em constante evolução, o artigo pretende esclarecer os desafios éticos e legais que surgem à medida que as IAS se tornam mais sofisticadas, autônomas e inseridas na sociedade e no sistema jurídico brasileiro.

Para determinar a potencial consciência da ilicitude na IA, seria preciso demonstrar que os robôs têm a capacidade de compreender que a conduta praticada é criminosa, assim como a exigibilidade de conduta diversa, uma vez que essa característica pressupõe que o agente poderia agir de outra maneira, mas optou por não o fazer (PAULA; CORNWALL; CABRAL, 2019).

Embora, atualmente, não seja possível afirmar que a inteligência artificial tenha vontade própria ou a capacidade de compreender o caráter ilícito de suas ações, as discussões apresentadas anteriormente indicam que essa questão permanece em aberto.

Segundo o autor,

A responsabilidade direta é justificada, pois as máquinas têm a capacidade de captar informações sensoriais e interpretar dados. Elas podem ser programadas para cumprir um propósito específico e agir conforme esse objetivo. Porém, um grande desafio é que as máquinas não possuem sentimentos humanos, o que as impede de entender plenamente as motivações emocionais. Isso, no entanto, não impede que elas cometam crimes, já que nem todas as infrações exigem emoções humanas. Assim, é possível ajustar a responsabilidade direta das máquinas, considerando essa limitação emocional (HALLEVY, 2019, p. 8-15).

Por outro lado, há um argumento contrário baseado na ausência de componente subjetivo, especialmente sob a ótica da teoria do crime. O sistema de IA não possui vontade, uma característica exclusiva dos seres humanos. Além disso, carece de consciência e moral, aspectos fundamentais para avaliar a reprovação de suas ações, o que o torna incapaz de se alinhar às normas legais. (DIVINO, 2020). Isso enfraquece as bases éticas das infrações e limita sua eficácia em termos de censura e prevenção.

Somente os seres humanos têm a capacidade de agir com intenção e consciência, direcionando suas ações a um propósito específico, algo que os sistemas de IA não podem fazer. Esse sistemas não conseguem praticar atos voluntários guiados psiquicamente a um objetivo, tampouco agir com culpa. Para que um robô inteligente seja considerado responsável por um crime, seriam necessárias revisões profundas na estrutura da teoria criminal. (SOUZA, 2020).

Ainda, há o entendimento de que é improvável atribuir subjetivamente danos causados por programas altamente complexos. Além disso, os elementos subjetivos das leis penais, projetados para atribuir responsabilidade, podem dificultar a imputação de culpa quando o ato é cometido por um sistema autônomo (SOUSA, 2020, p. 78).

As críticas às abordagens de Hallevy são baseadas em argumentos filosóficos e metajurídicas, questionando a atribuição de responsabilidade aos sistemas de IA. Para esses críticos, as ações realizadas pela IA, como um atropelamento, são apenas execuções de algoritmos, sem significado intrínseco, pois a IA não possui vontade própria. Além disso, a comparação da IA a um inimputável é questionada, pois, mesmo com a falta de capacidade, a IA é uma entidade diferente dos seres humanos, levantando dúvidas sobre a validade dessa analogia (DIVINO, 2020, p. 178).

Na literatura brasileira, existe uma quantidade considerável de estudos que exploram a possibilidade de responsabilização em casos civis envolvendo sistemas autônomos. No entanto, no âmbito criminal, essa discussão ainda é incipiente. Observa-se que a doutrina civil enfrenta menos desafios e apoia de forma mais clara a ideia de responsabilização civil para sistemas

autônomos, com foco na reparação dos danos materiais sofridos pela vítima, priorizando os prejuízos causados, em vez da análise da conduta do agente responsável.

### 3.3. ALGORITMOS ENVIESADOS NA DISCRIMINAÇÃO

Uma outra questão observada nas pesquisas envolve a análise do uso inadequado da inteligência artificial, especialmente no que se refere às tendências discriminatórias presentes em algoritmos enviesados.

Um exemplo disso está nos estudos sobre o uso de IA em decisões judiciais, onde se discute os vieses algorítmicos em tecnologias de aprendizado de máquina. Esses algoritmos são projetados para que o sistema aprenda automaticamente a partir de inferências feitas em grandes volumes de dados, identificando padrões e realizando generalizações, sem necessidade de uma programação específica e explícita (CHAVES JUNIOR et al., 2023, p.9).

Assim, a IA fornece resultados baseados nos objetivos estabelecidos previamente e nos dados que foram coletados, o que pode impactar consideravelmente o ambiente em que atua. A interação da IA com o meio externo é primeiramente moldada pelos dados que recebe e, em seguida, pela influência daqueles que a programam.

É relevante lembrar que os vieses refletem preconceitos humanos e influências ideológicas, que acabam gerando decisões que não seguem uma lógica fundamentada. De acordo com os estudos, como os algoritmos são desenvolvidos por pessoas, os chamados "algoritmos enviesados" tendem a reproduzir os mesmos preconceitos e vieses dos programadores ou dos dados que utilizam, espelhando estruturas sociais predominantes (CABRAL et al., 2023; CHAVES JUNIOR et al., 2023, p.8).

Essas pesquisas indicam que algoritmos de IA, quando aplicados no sistema judicial, podem operar de forma a reforçar desvantagens já enfrentadas por grupos socialmente marginalizados, amplificando preconceitos e segregando ainda mais esses grupos (CABRAL et al., 2023; CHAVES JUNIOR et al., 2023).

Pesquisadores apontam que, embora tais comportamentos discriminatórios sejam mais evidentes em processos criminais, eles também aparecem em questões civis, como nas limitações de crédito concedidas por empresas de cartão de crédito, onde há viés sexista; no tratamento preconceituoso contra mulheres negras, e na seleção de currículos masculinos em detrimento de femininos por algumas empresas. Também há discriminação racial em

algoritmos de reconhecimento facial, cujas taxas de precisão podem variar de acordo com as características de raça e gênero dos criadores desses modelos (CHAVES JUNIOR et al., 2023).

No campo penal, como discutem Chaves Junior et al. (2023, p.11-14), algoritmos enviesados podem criar lacunas nos modelos desenvolvidos, onde a falta de representação adequada de determinados grupos compromete sua inclusão estatística, gerando consequências sociais graves, intensificando racismo e preconceitos dentro do Sistema Penal, e violando direitos fundamentais garantidos pela Constituição. Tudo isso é agravado pela aparente "neutralidade matemática" que essas IAs parecem exibir.

## CONSIDERAÇÕES FINAIS

O trabalho demonstrou que o avanço das inteligências artificiais apresenta complexos desafios éticos e jurídicos, especialmente quando se trata da tomada de decisão autônoma por esses sistemas. A ciência ainda precisa desenvolver meios mais claros para explicar e ilustrar como esses processos decisórios realmente ocorrem, pois essa é uma área extremamente complexa. Observa-se que o atual estilo de vida é influenciado pela presença de riscos, os quais não surgem apenas da falta de conhecimento científico, mas são também uma consequência do próprio progresso das áreas produtivas da sociedade. Isso configura o que chamamos de sociedade de riscos.

No Direito, esses avanços trazem dilemas éticos significativos, particularmente no que concerne à responsabilidade penal em casos de uso indevido da IA. A atribuição de responsabilidade torna-se fundamental para garantir que, mesmo em situações em que o risco não se materializa, a norma jurídica de proteção seja respeitada. No entanto, a responsabilização penal depende do enquadramento do ato como crime típico, seja ele intencional (doloso) ou negligente (culposo), além de exigir a comprovação do nexo causal e da tipicidade.

Diante desse contexto, percebe-se uma lacuna preocupante na doutrina, especialmente quanto ao poder-dever de garante daqueles que concebem e capacitam as IAS. A responsabilidade penal deve, prioritariamente, recair sobre os agentes humanos que, por imprudência ou omissão, tornam essas entidades tecnológicas autônomas sem considerar suas possíveis consequências lesivas.

Ainda, o uso indevido da IA, como a substituição massiva de funcionários por chatbots, evidencia práticas problemáticas que podem transferir a responsabilidade criminal para os programadores ou até mesmo para os próprios sistemas, um raciocínio falho e perigoso. Logo,

antes de se questionar a possibilidade de punição direta das inteligências artificiais, é essencial definir a responsabilidade daqueles que negligenciam seu papel de supervisão e controle ético sobre essas tecnologias.

Conclui-se, portanto, que a implementação de valores éticos e sociais no desenvolvimento da IA é primordial, pois o ser humano deve ser o filtro final e determinante para as consequências das decisões das máquinas. Dessa forma, o adequado controle e a responsabilização dos agentes humanos são cruciais para assegurar o uso ético e seguro dessas tecnologias.

## REFERÊNCIAS

BARTNECK, Christoph; LÜTGE, Christoph; WAGNER, Alan; WELSH, Sean. What is AI? In: BARTNECK, Christoph; LÜTGE, Christoph; WAGNER, Alan; WELSH, Sean (coords.). *An Introduction to Ethics in Robotics and AI*. Cham: Springer, 2020. p. 5-16. Disponível em: [https://link.springer.com/chapter/10.1007/978-3-030-51110-4\\_2](https://link.springer.com/chapter/10.1007/978-3-030-51110-4_2). Acesso em: out. 2024.

BROOKSHEAR, JG. *Ciência da computação*. Disponível em: Minha Biblioteca, (II<sup>a</sup> edição). Grupo A, 2013. Disponível em: [https://www.academia.edu/35087288/Livro\\_CIENCIA\\_DA\\_COMPUTACAO\\_UMA\\_VISAO\\_ABRANGENTE](https://www.academia.edu/35087288/Livro_CIENCIA_DA_COMPUTACAO_UMA_VISAO_ABRANGENTE) acessado em: out. 2024.

160

BRASIL. Constituição (1988). *Constituição da República Federativa do Brasil*. Brasília: Senado. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/constituicao/constituicao.htm](http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm). Acessado out. 2024.

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.animaeducacao.com.br/server/api/core/bitstreams/845dd7a4-2a0d-43d7-87b2-0dbd456fdb6a/content>. Acesso em: out. 2024.

CHAVES JÚNIOR, 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/revistadodireito/article/view/4768>. Acesso em: out. 2024. doi:<https://doi.org/10.18256/2238-0604.2023.v19i2.4768>.

COPPIN, Ben. **Inteligência artificial**. Tradução: Jorge Duarte Pires Valério. Rio de Janeiro: LTC, 2013. E-book. Disponível em: <http://periodicos.ufc.br/nomos/article/view/20493> acesso em: out. 2024.

DIVINO, S. **Responsabilidade penal de Inteligência Artificial:** uma análise sob a ótica do naturalismo biológico de John Searle. Disponível em:

[www.researchgate.net/publication/344549566\\_Responsabilidade\\_penal\\_de\\_Inteligencia\\_Artificial\\_uma\\_analise\\_sob\\_a\\_otica\\_do\\_naturalismo\\_biológico\\_de\\_John\\_Searle](http://www.researchgate.net/publication/344549566_Responsabilidade_penal_de_Inteligencia_Artificial_uma_analise_sob_a_otica_do_naturalismo_biológico_de_John_Searle). Acesso em: 10 de out. 2024.

ESTEFAM, André; GONCALVES, Victor Eduardo Rios. **Direito penal esquematizado: parte geral.** 8. ed. São Paulo: Saraiva, 2019.

FACELI, Katti; LORENA, Ana C.; GAMA, João; et al. **Inteligência Artificial -Uma Abordagem de Aprendizado de Máquina.** Rio de Janeiro: Grupo GEN, 2021. Ebook. ISBN 9788521637509. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788521637509/>. Acesso em: out. 2024.

GONÇALVES, Carlos Roberto. **Direito civil: parte geral.** 25. ed. São Paulo: Saraiva, 2018.

HALLEVY, P. G. **The basic models of criminal liability of AI systems and outer circles,** 2019. Disponível em: <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3402527](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3402527)>. Acesso em: 18 out. 2024.

HALLEVY, Gabriel. **The Criminal Liability of Artificial Intelligence Entities.** Ono Academic College, Faculty of Law. 2020. Disponível em: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1564096](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1564096). Acesso em: 04 out. 2024.

KAPLAN, Andreas; HAENLEIN, Michael. **A brief history of artificial intelligence:** on the past, present, and future of artificial intelligence, California Management Review, California, v. 61, n. 4, p. 5-14, ago. 2019. Disponível em: <https://journals.sagepub.com/doi/10.1177/0008125619864925>. Acesso em: out. 2024.

MÁNTARAS, Ramon. **Towards artificial intelligence:** advances, challenges, and risks. Metode Science Studies Journal, [s.l.], v. 1, n. 9, p. 119-125, mar. 2019. Disponível em: <https://ojs.uv.es/index.php/Metode/article/view/11145>. Acesso em: out. 2024.

MORAIS, Renato Watanabe de. **Programações podem ser punidas?** Responsabilidade penal em decisões tomadas pela inteligência artificial. Rio de Janeiro: Revista Científica do CPJM, 2023. Disponível em: <https://rcpj.mnuvens.com.br/revista/article/download/217/187>. Acesso em: 18 ou. 2024.

PAULA, Alice Lima; CORNWALL, Bruno Meirelles de M.; CABRAL, Dalila M. **Breves**

**reflexões sobre a inteligência artificial e seus impactos no campo do Direito Penal.** In: CHAVES, Natália Cristina (org.). **Direito, tecnologia e globalização.** [online]. Porto Alegre, 2019. p. 98-117. Disponível em: [https://www.direito.ufmg.br/wpcontent/uploads/2019/12/direito\\_tecnologia\\_globalizacao.pdf](https://www.direito.ufmg.br/wpcontent/uploads/2019/12/direito_tecnologia_globalizacao.pdf). Acesso em 15 set. 2024.

PEREIRA, Uiara Vendrame; TEIXEIRA, Tarcísio. **Inteligência artificial: a quem atribuir responsabilidade?** Revista de Direitos e Garantias Fundamentais. 2019, v.20 n.2, p. 119-142. Disponível em: [https://sisbib.mnuvens.com.br/direitos\\_garantias/article/view/1523](https://sisbib.mnuvens.com.br/direitos_garantias/article/view/1523). Acesso em: out. 2024.

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

**REGULAMENTO INTELIGÊNCIA ARTIFICIAL. Parlamento aprova legislação histórica.**  
Disponível em: <https://www.europarl.europa.eu/news/pt/press-room/20240308IPR19015/regulamento-inteligencia-artificial-parlamento-aprova-legislacao-historica> acessado em: out. 2024.

RUSSELL, Stuart; NORVIG, Peter. **Artificial intelligence: a modern approach**. New Jersey: Pearson. 2020.

SOUZA, Susana Aires de. “**Não fui eu, foi a máquina**”: Teoria do crime, responsabilidade e inteligência artificial. Inteligência artificial no Direito Penal. Coimbra: Almedina, 2020, p. 68-III.

STRAUB, Stefan. **From big data to deep learning: a leap towards strong AI or ‘intelligentia obscura’?** Big Data and Cognitive Computing, [s. l.], v. 2, n. 3, p. 1-19, Jul. 2018. Disponível em: <https://www.mdpi.com/2504-2289/2/3/16/pdf>. Acesso em: out. 2024.