DO ZERO AO JULGAMENTO: DESAFIOS ÉTICOS E JURÍDICOS DA INTELIGÊNCIA ARTIFICIAL NO DIREITO PENAL

Beatriz Viana dos Santos[1]

Resumo: Este trabalho analisa os impactos jurídicos, sociais e éticos da Inteligência Artificial (IA), com foco nos algoritmos enviesados no Direito Penal e outras áreas. Embora a IA traga avanços operacionais em diversos setores, levanta preocupações sobre transparência, imparcialidade e direitos fundamentais. O estudo destaca que algoritmos, ao serem treinados com dados históricos enviesados, podem reforçar discriminações. Exemplo disso são erros em reconhecimento facial e a exclusão de mulheres nos dados. Também são discutidas iniciativas legislativas, como a LGPD, o Marco Civil da Internet e o PL nº 2.338/2023. Conclui-se que é necessário regulamentar a IA com base em princípios constitucionais e promover uma governança ética, transparente e inclusiva das tecnologias.

Palavras-chaves: Inteligência. Artificial. Algoritmos. Dados. IA.

Abstract: This paper analyzes the legal, social, and ethical impacts of Artificial Intelligence (AI), focusing on biased algorithms in Criminal Law and other fields. Although AI brings operational advancements across various sectors, it raises concerns about transparency, fairness, and fundamental rights. The study highlights that algorithms trained on biased historical data can reinforce discrimination. Examples include facial recognition errors and the exclusion of women from datasets. Legislative initiatives such as the General Data Protection Law (LGPD), the Brazilian Civil Rights Framework for the Internet, and draft law 2.338/2023 are also discussed. The conclusion emphasizes the need to regulate AI based on constitutional principles and to promote ethical, transparent, and inclusive governance of these technologies.

Keywords: Intelligence. Artificial. Algorithms. Data. AI.

1. INTRODUÇÃO

A Inteligência Artificial (IA), antes vista como algo distante e futurista, hoje já faz parte do cotidiano e está presente em diferentes áreas da sociedade. Atualmente, é um dos principais agentes de transformação nas dinâmicas sociais, econômicas e profissionais, influenciando significativamente a forma como indivíduos vivem, trabalham e interagem. No campo jurídico,

[1] Acadêmico do curso Direito da Instituição de Ensino Superior (Anhembi Morumbi) da rede Ânima Educação. E-mail: beatrizssviana223@gmail.com do autor do artigo. Artigo apresentado como requisito parcial para a conclusão do curso de Graduação em Direito da Instituição de Ensino Superior (Anhembi Morumbi) da rede Ânima Educação. 2025. Orientador: Prof. Larissa Carbonari de Almeida Miranda.

é possível observar um crescimento expressivo da presença da IA, especialmente na otimização de procedimentos processuais, na gestão e fluxos trabalho e tomada de decisões legais.

Tal expansão é diretamente impulsionada por avanços tecnológicos recentes, como aumento exponencial da capacidade de processamento computacional, a consolidação do armazenamento em nuvem e a disponibilidade massiva de bancos de dados. Esses fatores têm favorecido a implementação de soluções baseadas em IA em áreas como saúde, segurança pública, mobilidade urbana, finanças e, com intensidade crescente, no sistema de justiça. No âmbito jurídico, já é possível identificar sua aplicação em ferramentas de automação de petições, sistemas de análise preditiva de jurisprudência e mecanismos de triagem documental, os quais proporcionam maior celeridade e eficiência às atividades de advogados, magistrados e servidores públicos.

Apesar de sua adoção prática, a regulação normativa da IA ainda se encontra em estágio inicial no ordenamento jurídico brasileiro. Embora não exista, até o presente momento, uma legislação específica sobre o tema, instrumentos normativos já consolidados exercem influência direta sobre a utilização dessas tecnologias. É o caso do Marco Civil da Internet (Lei nº 12.965/2014), que estabelece princípios para o uso da internet no país e garante direitos fundamentais relacionados à privacidade e à proteção de dados, e da Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018), a qual disciplina o tratamento de dados pessoais e impõe limites ao uso de informações sensíveis, elemento central no funcionamento de sistemas baseados em IA.

É importante destacar, contudo, que os algoritmos que integram a IA não são, como muitas vezes se presume, isentos de vieses. Na prática, eles aprendem com os dados históricos utilizados em seu treinamento, os quais podem estar impregnados de preconceitos raciais, sociais e de gênero. Quando esses dados não são previamente tratados de forma crítica e responsável, os sistemas tendem a reproduzir — e, por vezes, intensificar — desigualdades estruturais já existentes. A esse respeito, a autora Cathy O'Neil (2016) adverte que os sistemas algorítmicos, quando não submetidos a auditoria e transparência adequadas, podem comprometer direitos fundamentais, como o acesso equitativo à informação e o princípio da igualdade material.

A jurisprudência pátria tem refletido essa preocupação. O Tribunal de Justiça do Estado de Minas Gerais, por exemplo, já reconheceu que decisões automatizadas adotadas por plataformas digitais, como o bloqueio de contas, devem observar os princípios do contraditório e da ampla defesa. Mesmo quando intermediadas por sistemas autônomos, tais decisões não

podem prescindir de fundamentação adequada e da possibilidade de contestação por parte do usuário, conforme disposto no Marco Civil da Internet e no Código de Defesa do Consumidor.

Outro campo particularmente sensível diz respeito à aplicação da IA na segurança pública, sobretudo no reconhecimento facial. Pesquisas como a do G1 (2024) apontam que esses sistemas, quando treinados com bases de dados incompletas ou enviesadas, apresentam taxas de erro significativamente maiores em relação a pessoas negras ou pertencentes a grupos minoritários “Nos EUA, por exemplo, uma pesquisa revelou que grandes algoritmos de reconhecimento facial erravam 34% a mais no caso de mulheres negras, em comparação com homens brancos.”. Tal distorção compromete a equidade do sistema de justiça e acarreta riscos concretos de criminalização indevida, reforçando estigmas e desigualdades historicamente consolidadas.

Diante desse cenário, o presente trabalho propõe uma análise crítica acerca dos efeitos jurídicos e sociais da IA, com ênfase nas implicações geradas por algoritmos enviesados. Serão abordados casos emblemáticos relacionados a processos seletivos automatizados, nos quais práticas discriminatórias persistem sob a roupagem da neutralidade tecnológica, bem como crimes digitais cada vez mais sofisticados, como fraudes bancárias mediante troca de identidade visual e a produção de conteúdos íntimos manipulados digitalmente, envolvendo inclusive menores de idade.

O objetivo central consiste em compreender, sob uma perspectiva técnico-jurídica, os riscos e desafios que decorrem da aplicação indiscriminada da IA e propor medidas regulatórias capazes de assegurar a compatibilidade dessas tecnologias com os princípios constitucionais da dignidade da pessoa humana, da igualdade substancial e da proteção integral de direitos fundamentais. Pretende-se, assim, não apenas diagnosticar as fragilidades do atual modelo de governança algorítmica, mas contribuir de forma propositiva para a construção de um marco normativo sólido, ético e transparente, que esteja à altura das transformações promovidas pela era digital

2. O QUE É INTELIGÊNCIA ARTIFICIAL?

Atualmente, a Inteligência Artificial vem se consolidando como uma das principais inovações tecnológicas, com impacto direto em diversos setores — inclusive no campo do Direito. Seu potencial transformador está na capacidade de processar grandes quantidades de dados, identificar padrões e realizar tarefas complexas de forma automatizada. Para isso, a IA se vale de técnicas como o aprendizado de máquina (*machine learning*) e as redes neurais

artificiais, que permitem aos sistemas aprenderem com experiências anteriores e tomar decisões com diferentes níveis de autonomia.

A popularização do uso da IA está diretamente relacionada ao avanço da capacidade de processamento computacional, à ampliação do armazenamento em nuvem e à abundância de dados disponíveis (*big data*). Essa conjunção tecnológica tornou viável a incorporação da IA em diversas áreas, como saúde, transporte, segurança, finanças e, mais recentemente, no campo jurídico como mostra o site da PRODEST (Companhia de Processamento de Dados do Espírito Santo). Nessa última esfera, a IA tem sido empregada em análises preditivas, automatização de trâmites processuais e na formulação de pareceres com base em grandes bases jurisprudenciais.

Do ponto de vista jurídico, no Brasil, ainda não existe uma legislação específica dedicada exclusiva sobre Inteligência Artificial. Um passo importante nessa direção é o Projeto de Lei nº 2.338/2023, que pretende criar um Marco Legal para a Inteligência Artificial no Brasil. A proposta busca estabelecer regras claras para garantir que essas tecnologias sejam desenvolvidas e utilizadas de forma ética, segura e responsável. Entre as preocupações centrais estão a proteção dos direitos fundamentais, a transparência no funcionamento dos algoritmos, a prevenção de discriminações e a identificação de riscos que possam afetar a sociedade. Além desta, há algumas normas importantes que orientam seu uso e trazem impactos significativos. É o caso do Marco Civil da Internet (Lei nº 12.965/2014), que define princípios para o uso ético da internet e para a proteção da privacidade, e da Lei Geral de Proteção de Dados Pessoais – LGPD (Lei nº 13.709/2018), que regula o tratamento de dados pessoais e assegura direitos fundamentais aos cidadãos.

A produção acadêmica tem se debruçado sobre um ponto crucial: a ideia equivocada de que algoritmos são neutros. Como alerta Cathy O'Neil (2016), os sistemas de IA tendem a reproduzir os padrões e preconceitos presentes nos dados utilizados em seu treinamento. Esse problema se torna ainda mais grave quando os algoritmos são aplicados a decisões que afetam diretamente a vida das pessoas, como em processos seletivos, análises de crédito ou até mesmo sentenças judiciais. Nessas situações, o uso indiscriminado da IA pode reforçar desigualdades sociais históricas e ameaçar direitos fundamentais, como a igualdade, a dignidade da pessoa humana e o direito ao contraditório.

> A particularidade desses algoritmos é que eles aprendem por conta própria, fazendo inferências a partir dos dados. Portanto, o aprendizado de máquina é a capacidade destas de aprender com os dados, identificando tendências e padrões em eventos aparentemente aleatórios. (MORAIS, 2023, p. 4)

A particularidade desses algoritmos é que eles aprendem por conta própria, fazendo inferências a partir dos dados. Portanto, o aprendizado de máquina é a capacidade destas de

aprender com os dados, identificando tendências e padrões em eventos aparentemente aleatórios.

A jurisprudência brasileira já começa a se debruçar sobre os desafios impostos pelas novas tecnologias aplicadas ao campo jurídico. Tribunais de Justiça como o de Minas Gerais, por exemplo, têm adotado soluções baseadas em inteligência artificial para otimizar sua atuação. Segundo Karina Farias, o TJMG implantou o robô Radar, que auxilia magistrados na identificação de processos repetitivos, permitindo sua classificação e agrupamento, de modo a promover mais celeridade e coerência decisória no tratamento de demandas semelhantes. Como destaca a autora, a adoção dessas ferramentas, embora organizativas, requer limites regulatórios e éticos para que não se sobreponham aos direitos fundamentais dos jurisdicionados (FARIAS, 2022, p. 100).

Além disso, a aplicação da IA em contextos como segurança pública e reconhecimento facial tem gerado debates na doutrina penal e processual penal. Os elevados índices de erro em sistemas aplicados a indivíduos de grupos minoritários indicam o risco de reforço de práticas discriminatórias e violação de direitos humanos. A complexidade e a opacidade desses sistemas dificultam o controle institucional e a responsabilização jurídica, exigindo, por parte do Estado, um esforço regulatório urgente.

Deste modo, busca-se investigar os efeitos dos vieses algorítmicos na produção e aplicação da IA, com foco especial nos impactos sociais, jurídicos e éticos. Serão abordadas temáticas como racismo estrutural, assédio digital, manipulação de dados, uso indevido de imagens íntimas por meio de *deepfakes* e fraudes digitais que se utilizam de recursos de IA. A proposta é apresentar uma análise crítica dos riscos associados à ausência de uma governança algorítmica efetiva, sugerindo mecanismos regulatórios e políticas públicas que garantam a proteção dos direitos fundamentais diante do avanço tecnológico.

Os aplicativos matemáticos que alimentam a economia de dados são baseados em escolhas feitas por seres humanos falíveis e muitos desses modelos codificam preconceitos, falta de compreensão e vieses humanos em sistemas de software que cada vez mais gerenciam vidas. (ALBINO, 2023, p. 15)

2.1. COMO FUNCIONA A IA?

A IA opera com base na capacidade dos sistemas computacionais de processarem grandes volumes de dados, reconhecendo padrões, inferindo relações e, a partir disso, realizando tarefas que tradicionalmente exigiriam cognição humana. O funcionamento da IA está estruturado em algoritmos matemáticos e estatísticos que, ao interagirem com dados

provenientes de diferentes fontes, conseguem interpretar, categorizar e tomar decisões com base no histórico de aprendizado acumulado. Como explica Nunes e Marques (2018) no artigo Inteligência Artificial e Direito Processual: Vieses Algorítmicos e os Riscos de Atribuição de Função Decisória às Máquinas, esses sistemas vêm sendo cada vez mais incorporados ao cotidiano, da saúde à educação, exigindo análise crítica sobre sua lógica e impacto social.

> De acordo com uma pesquisa realizada pela CBRE, cerca de 48% dos escritórios advocatícios de Londres já utilizam sistemas de inteligência artificial e 41% pretendem implantá-los. Segundo a pesquisa, a IA é utilizada, principalmente, para gerar e revisar documentos e para a eletronic discovery,mas também tem aplicação relevante na realização de pesquisas jurídicas e na due diligence – investigação prévia de companhias antes da realização de negócios. (NUNES e MARQUES, 2018, p. 2)

Entre os modelos mais populares de Inteligência Artificial, um dos que mais chamam a atenção são as redes neurais artificiais. Inspiradas no funcionamento do cérebro humano, elas operam por meio de neurônios artificiais, organizados em camadas interligadas. Quando recebem uma informação, ela percorre essas camadas em uma sequência definida: começa pela entrada, segue por camadas intermediárias — conhecidas como camadas ocultas —, onde os dados são analisados, e termina na camada de saída, que fornece a resposta final. Ao longo desse processo, os caminhos entre os neurônios vão sendo ajustados conforme a rede aprende com novos exemplos e experiências. (FLECK, 2016, p. 52).

Esse processo de aprendizagem ocorre, em geral, por repetição. Ou seja, a rede é treinada com exemplos variados até que consiga gerar respostas coerentes mesmo diante de dados que nunca foram apresentados antes. Essa forma de aprendizado é o que caracteriza o chamado aprendizado de máquina (*machine learning*), uma das bases mais importantes da IA.

No interior das redes neurais, cada camada é responsável por realizar uma transformação matemática específica sobre os dados. A eficácia da transmissão dessas informações é determinada por funções de ativação, que operam como filtros, decidindo se determinada informação deve ser repassada adiante ou descartada. Simon Haykin, explica, em Redes neurais: princípios e práticas, que essas funções são fundamentais para que o sistema modele relações não lineares entre os dados, permitindo respostas complexas mesmo com entradas não estruturadas.

Uma vertente mais avançada desse processo é o aprendizado profundo (*deep learning*), que utiliza redes neurais com múltiplas camadas ocultas. Essas redes profundas são capazes de identificar padrões altamente sofisticados e abstratos, sendo especialmente eficazes em contextos como o reconhecimento facial, a tradução automática, os diagnósticos clínicos por imagem e a análise de sentimentos expressos em linguagem natural. Getúlio Akabane, em “Gestao Estratégica Das Tecnologias Cognitivas” descreve o que é *deep learning*.

> Caracteriza-se pela manipulação de grande volume de dados para ensinar aos computadores como desenvolver processos que apenas os humanos são capazes de desenvolver e como solucionar muitos problemas complexos. (AKABANE, 2025, p. 108)

Para que a IA funcione de maneira confiável, três pilares são fundamentais: a qualidade dos dados utilizados para seu treinamento, a robustez da estrutura algorítmica empregada e a responsabilidade na interpretação e aplicação dos resultados. A ausência de qualquer um desses elementos pode comprometer seriamente a eficácia, a equidade e a confiabilidade dos sistemas automatizados.

2.2. APRENDIZADO DE MÁQUINA E AS REDES NEURAIS ARTIFICIAIS

Para entender como a IA aprende é necessário conhecer dois de seus principais fundamentos: o aprendizado de máquina (*machine learning*) e as redes neurais artificiais. Juntos, esses conceitos explicam como sistemas computacionais conseguem realizar tarefas complexas sem depender de instruções fixas, como ocorria nos modelos tradicionais de programação.

O aprendizado de máquina pode ser entendido como uma forma de fazer com que os computadores aprendam com os dados. Ou seja, em vez de programar linha por linha o que o sistema deve fazer, alimenta-se o algoritmo com uma grande quantidade de informações para que ele descubra, por conta própria, como agir diante de determinadas situações. Esse processo é semelhante ao modo como os seres humanos aprendem com a experiência: a partir da repetição e da observação de padrões.

> Em AM [aprendizado de máquina], dispositivos computacionais são programados para aprender a partir de experiências passadas. Para tal, frequentemente empregam um princípio de inferência denominado indução, que permite extrair conclusões genéricas a partir de um conjunto particular de exemplos. (FACELI, 2021, p.4)

De modo simplificado, o funcionamento do *machine learning* envolve três etapas principais: primeiro, o sistema recebe os dados de entrada, como imagens, números ou textos. Em seguida, um algoritmo constrói um modelo matemático com base nesses dados. Por fim, uma função de avaliação verifica se o modelo está acertando suas previsões e o ajusta, se necessário. Com o tempo e com a exposição a novos dados, o sistema torna-se mais preciso e eficiente. Faceli explica:

Existem três formas principais de aprendizado de máquina:

I. Aprendizado supervisionado: o sistema é treinado com dados que já têm uma resposta conhecida, como uma imagem de um animal com a etiqueta "gato". Assim, ele aprende a associar certas características à resposta correta;
II. Aprendizado não supervisionado: nesse caso, o sistema não recebe respostas prontas. Ele mesmo precisa identificar agrupamentos ou padrões entre os dados fornecidos;
III. Aprendizado por reforço: inspirado no comportamento humano, o sistema aprende por tentativa e erro. Ele recebe recompensas quando acerta e penalizações quando erra, ajustando suas ações ao longo do tempo.

(FACELI, 2021, p.4) lassifica essas modalidades de forma clara ao afirmar que os principais tipos de aprendizado de máquina são: supervisionado, que utiliza dados rotulados; não supervisionado, onde o sistema identifica padrões por conta própria; e o aprendizado por reforço, em que o agente aprende por tentativa e erro.

As redes neurais artificiais, segundo (FLECK, 2016, p. 2), são modelos computacionais que se inspiram no funcionamento do cérebro humano, sendo compostas por neurônios organizados em camadas interligadas, os quais interagem entre si para resolver problemas complexos. Na prática, o funcionamento dessas redes pode ser dividido em três fases: na entrada, os dados são recebidos; nas camadas ocultas, ocorrem os cálculos e ajustes de pesos; por fim, na saída, o sistema fornece uma resposta com base no que aprendeu. Podem variar de simples a bastante complexa, ajudam sistemas de Inteligência Artificial a resolver tarefas difíceis, como identificar rostos ou interpretar exames médicos. À medida que são expostas a mais dados, tornam-se mais eficientes em reconhecer padrões e lidar com novas situações.

No entanto, essas redes também têm suas limitações: exigem alto poder de processamento e podem acabar decorando os dados de treinamento, sem conseguir se adaptar bem a informações diferentes — um problema conhecido como *overfitting*. “Isso significa que o modelo memorizou ou se especializou nos dados de treinamento.” (FACELI, 2021, p.4)

Nesse sentido, autores como Marques e Neto, alertam que a opacidade algorítmica pode inviabilizar o controle democrático sobre decisões que impactam direitos fundamentais, especialmente quando essas decisões se dão de maneira automática, sem justificativa compreensível pelos jurisdicionados. Defende-se, assim, a necessidade de adoção de mecanismos de transparência algorítmica e de auditoria, que permitam a rastreabilidade das decisões automatizadas e assegurem sua conformidade com o ordenamento jurídico.

> A transparência, além de ser dever anexo da boa-fé obje tiva que permeia as relações de trabalho, é fundamento da Lei nº 13.709/18 (Lei Geral de Proteção de Dados Pessoais - LGPD), consoante se extrai do princípio da autodeterminação informa tiva. No mesmo sentido, o art. 6º, inciso IV, do PL 21/2020, prevê o dever de transparência

> sobre o uso e funcionamento dos sistemas de inteligência artificial. (NETO, MARQUES, 2022, p. 7)

Assim, embora o aprendizado de máquina e as redes neurais artificiais representem alguns dos maiores avanços da tecnologia atual, seu uso exige cuidado, responsabilidade e regulamentação adequada. A aplicação ética desses sistemas é indispensável para que a IA contribua de maneira justa e segura para o desenvolvimento social e institucional.

Do ponto de vista do Direito, a ausência de transparência nas decisões automatizadas desafia diretamente princípios constitucionais fundamentais, como a igualdade, o devido processo legal e o direito à informação. Quando não é possível compreender ou auditar como uma decisão foi tomada por um sistema de inteligência artificial, compromete-se a possibilidade de impugnação e o exercício pleno do contraditório. Essa situação é especialmente grave em contextos sensíveis como concessão de crédito, processos seletivos, policiamento ou decisões judiciais automatizadas.

3. OS IMPACTOS REAIS DA IA NA VIDA COTIDIANA

A IA, enquanto campo de desenvolvimento científico e tecnológico, tem se consolidado como um dos principais vetores de transformação da sociedade contemporânea. Antes associada ao campo da ficção ou restrita a laboratórios de pesquisa, passou a integrar de maneira ampla e silenciosa diversos aspectos da vida cotidiana. Seu funcionamento é baseado na coleta e análise de grandes volumes de dados, processados por algoritmos que aprendem padrões e produzem decisões automatizadas.

Nesse contexto, conforme ressalta SANCTIS (2020) “Por isso, está a se defender uma ordenação clara e sistemática de regras existentes sobre a Inteligência Artificial aplicada à justiça.”. Tal constatação justifica a necessidade de observar o avanço da IA à luz dos princípios constitucionais da dignidade da pessoa humana, da igualdade e da liberdade, que passam a ser tensionados frente à expansão de tecnologias que moldam o comportamento humano.

De acordo com levantamento realizado pela MindMiners e divulgado pela Exame (2024), “56% dos brasileiros já percebem os efeitos da IA em sua vida cotidiana”. Este dado reforça que a IA não se limita a setores altamente tecnificados, mas impacta diretamente a população em geral. A constatação exige a observância do princípio da transparência, previsto no art. 6º do Código de Defesa do Consumidor, e implica o reconhecimento do dever informacional como instrumento de efetivação de direitos fundamentais em um contexto digitalizado.

A presença da Inteligência Artificial no cotidiano também é abordada por Lee (2019), que identifica o momento atual como a 'quarta onda da IA', marcada pela capacidade dos sistemas de interagir em tempo real com o mundo físico. Essa integração tecnológica, embora prometa mais conforto e praticidade, também levanta preocupações jurídicas importantes, como as que envolvem privacidade, autonomia e consentimento. Nesse contexto, ganha destaque o princípio da autodeterminação informativa, previsto no Marco Civil da Internet, que se torna essencial para proteger a democracia diante da atuação de sistemas algorítmicos pouco transparentes. Outro ponto de atenção são os vieses algorítmicos, que podem gerar consequências discriminatórias e aprofundar desigualdades já existentes.

Pedro Costa (2019) observa que “Os papéis de gênero ou estereótipos são reforçados neste processo, e os chatbots e assistentes digitais de certa forma simulam atributos, papéis e estereótipos femininos.”, o que evidencia que os sistemas de IA podem replicar e acentuar desigualdades sociais preexistentes. Tais práticas confrontam diretamente o princípio da igualdade substancial e a vedação à discriminação, princípios consagrados no art. 3º, IV, da Constituição Federal, e que devem nortear o desenvolvimento ético e inclusivo da tecnologia.

Casos de manipulação de imagens e vídeos por IA também se tornaram objeto de preocupação. A fronteira entre o real e o simulado está cada vez mais tênue, fato que afeta especialmente adolescentes, expostos a conteúdos forjados sem que consigam discernir sua veracidade. Diante disso, destaca-se a necessidade de proteção de dados sensíveis e da imagem, especialmente quando envolvem menores de idade, aplicando-se os princípios da proteção integral (art. 227 da CF) e da veracidade da informação como base do Estado Democrático de Direito.

O avanço da inteligência artificial tem provocado transformações significativas no mercado de trabalho, especialmente pela substituição de tarefas repetitivas por sistemas automatizados, o que tende a acentuar desigualdades estruturais já existentes. Brynjolfsson e McAfee (2014) destacam que esse processo, embora aumente a eficiência, pode agravar disparidades sociais ao eliminar funções que antes garantiam a subsistência de grandes contingentes populacionais. Diante desse panorama, reforça-se a importância da função social do trabalho, prevista no art. 170 da Constituição Federal, e a necessidade de políticas públicas eficazes voltadas à requalificação profissional.

Nesse contexto de transformação tecnológica acelerada, Rahman (2021) observa que indivíduos com maior nível de escolaridade têm maiores chances de adaptação às novas exigências do mercado. Essa constatação reforça a urgência de universalizar a educação digital, não apenas como meio de inclusão social, mas como estratégia essencial para preparar a

população para os desafios impostos pela IA, em consonância com o disposto no art. 205 da Constituição.

Na discussão sobre a regulação ética da Inteligência Artificial, Floridi e Cowls (2019) apresentam cinco princípios essenciais: justiça, beneficência, explicabilidade, privacidade e responsabilização. A falta de diretrizes claras nesse campo pode colocar em risco direitos constitucionais importantes, como o devido processo legal e o contraditório (art. 5º, incisos LIV e LV), especialmente quando decisões automatizadas impactam diretamente a vida das pessoas. Diante disso, o Projeto de Lei nº 2338/2023 surge como uma iniciativa relevante para criar um marco regulatório que oriente o uso responsável no Brasil.

Marcos regulatórios robustos e vinculantes, como o PL 2.338/2023, são essenciais para garantir o uso ético e seguro da inteligência artificial no Brasil. O projeto estabelece princípios como a centralidade da pessoa humana, não discriminação, transparência e responsabilização. Ele classifica sistemas por grau de risco, exige governança adequada, avaliações de impacto e supervisão humana para sistemas de alto risco. Além disso, prevê mecanismos de fiscalização, responsabilidade civil e proteção de direitos fundamentais. Isso fortalece a confiança pública, estimula a inovação e assegura segurança jurídica para o desenvolvimento tecnológico.

Embora iniciativas privadas tenham papel importante na construção de soluções tecnológicas voltadas ao bem comum — como exemplifica o desafio WatsonX promovido pela IBM em 2024, plataforma de IA e dados projetada para construir, escalar e governar soluções de inteligência artificial de forma confiável. —, a regulação dessas ferramentas não pode se restringir à autodisciplina empresarial. Cabe ao Estado garantir que os sistemas de IA operem em conformidade com os princípios da legalidade, da transparência e da supremacia do interesse público.

3.1. IMPACTOS NO DIREITO PENAL

A incorporação da Inteligência Artificial no Direito Penal tem promovido significativas transformações na persecução penal, na produção de provas e na compreensão das condutas típicas. O uso de algoritmos, redes neurais e sistemas de análise preditiva amplia a capacidade de investigação do Estado, mas também levanta sérias questões jurídicas e éticas.

Mais do que simples automação de tarefas, a IA afeta diretamente princípios estruturantes do processo penal, como o contraditório, a ampla defesa, a imparcialidade do julgador e a dignidade da pessoa humana. Conforme observa Karina da Hora Farias (2022, p. 69):

> Mas, é importante compreender que do ponto de vista técnico e de máquina tal assertiva alega-se ser plenamente possível e conceda decisões, não obstante ser questionável a qualidade desse resultado para o efetivo senso de justiça, a partir da reflexão judicial pautada na isonomia e equidade social. Karina Farias (2022, p. 99):

Na área da segurança pública, ferramentas como o reconhecimento facial, a definição de perfis de risco e a análise de comportamento têm sido cada vez mais utilizadas. No entanto, quando baseadas em dados enviesados, essas tecnologias correm o risco de reforçar estereótipos e ampliar desigualdades sociais. Isso pode afetar diretamente a imparcialidade das decisões judiciais, que passam a refletir padrões ocultos e difíceis de serem percebidos por quem é julgado, sem uma justificativa clara. Além disso, como muitos desses sistemas operam de forma pouco transparente, 'caixas-pretas', termo utilizado por Sofiya Noble em “Algorithms of Oppression”, torna-se difícil entender como as decisões são tomadas, o que pode ferir princípios fundamentais como a legalidade, o direito à motivação das decisões e o devido processo legal.

Segundo O'Neil (2016), os algoritmos de decisão têm o potencial de perpetuar injustiças ao converter preconceitos históricos em previsões que aparentam ser neutras. A crítica, portanto, não está voltada à tecnologia em si, mas à sua introdução sem regulamentação clara, critérios normativos ou garantias mínimas de governança algorítmica. O uso da IA no campo penal exige mecanismos legais que assegurem tanto a eficácia da investigação quanto a proteção dos direitos fundamentais.

3.2. A AUTOMATIZAÇÃO DA JUSTIÇA PENAL: AVANÇOS E LIMITES

O uso da Inteligência Artificial no Direito Penal tem promovido mudanças significativas nos procedimentos judiciais, principalmente no que diz respeito à celeridade processual e à gestão de grandes volumes de dados. Sistemas automatizados vêm sendo testados para auxiliar juízes na dosimetria da pena, como descrito por Sara Pereira e Tarsis Oliveira, 2024, p. 4, e prevê reincidência de réus e até sugerir decisões com base em padrões jurisprudenciais. No entanto, essa aparente eficiência traz consigo riscos que não podem ser ignorados, sobretudo quando se trata da garantia dos direitos fundamentais do acusado.

Além disso, é preciso destacar que os sistemas de IA não operam de forma neutra, pois refletem os dados com os quais foram treinados. Quando alimentados por históricos judiciais enviesados, esses sistemas correm o risco de perpetuar padrões discriminatórios, tornando-se uma ameaça à imparcialidade da justiça. A presunção de inocência e o devido processo legal pode ser comprometido se decisões forem tomadas com base em estatísticas e não em uma análise individualizada do caso concreto.

> O grande risco dessa opacidade dos algoritmos é que os sistemas de Inteligência Artificial podem produzir "arbitrariedade de critérios e de conclusões, associada à discricionariedade, à discrepância com direitos fundamentais e outros princípios jurídicos, associando o sistema ao aprofundamento da desigualdade e imprevisibilidade[...]", a partir de perfis e compreensões automatizadas. (HARTMANN PEIXOTO, 2020; p.28)

Assim, fica evidente que, embora a IA possa contribuir com eficiência, ela jamais deve substituir o julgamento humano, cuja sensibilidade é essencial para garantir justiça em sua plenitude.

3.3. RECONHECIMENTO FACIAL E DISCRIMINAÇÃO ALGORÍTMICA

O uso de tecnologias de reconhecimento facial em segurança pública é um dos temas mais sensíveis quando se trata da aplicação da inteligência artificial no direito penal. Embora prometam maior eficiência na identificação de suspeitos, esses sistemas têm demonstrado falhas significativas, principalmente no reconhecimento de indivíduos negros e de outras minorias. A base de dados utilizada para o treinamento desses algoritmos muitas vezes reflete desequilíbrios históricos, resultando em erros que podem comprometer a liberdade de pessoas inocentes.

Segundo a Revista Brasileira de Direito, (2022, p. 83), "[..] os resultados estão carregados de subjetividade e vieses racistas tendentes a catalisar setores sociais bem delimitados para a marginalização e a discriminação.". Tal risco exige cautela por parte das autoridades públicas, sobretudo diante de um histórico de seletividade penal que já afeta desproporcionalmente determinados grupos sociais.

A utilização desses recursos tecnológicos sem regulamentação específica e sem controle social adequado pode comprometer os princípios da ampla defesa e do contraditório. Em uma prisão baseada exclusivamente em reconhecimento facial equivocado, como garantir que o cidadão tenha acesso à lógica do algoritmo utilizado? A ausência de explicabilidade e transparência torna essas tecnologias verdadeiras "caixas-pretas", imunes ao escrutínio judicial e democrático.

Floridi e Cowls (2019) alertam que nenhum sistema automatizado deve operar sem que suas decisões possam ser justificadas perante a sociedade. Isso demonstra que a aplicação de IA no reconhecimento de suspeitos precisa estar ancorada em diretrizes éticas claras, com salvaguardas institucionais que impeçam abusos e garantam a reparação em casos de erro.

3.4. ALGORITMOS E PRISÕES PREVENTIVAS: UMA NOVA FACE DA CRIMINALIZAÇÃO

Nos últimos anos, ferramentas de análise preditiva têm sido cada vez mais usadas para prever a chance de uma pessoa voltar a cometer crimes e apoiar decisões sobre prisão preventiva. Esses sistemas se baseiam em informações passadas, como idade, onde a pessoa mora, antecedentes criminais e até sua condição socioeconômica. A promessa é que se consiga prever comportamentos futuros com base em padrões estatísticos. No entanto, essa prática tem gerado críticas quanto ao reforço de desigualdades sociais e à criminalização da pobreza.

Como apontado por Gustavo Pedrina1 (2019, p.2) "para o processo penal, importa esclarecer que não há no seu uso a solução singular para o estabelecimento de procedimentos judiciais que levem à aferição da culpa de modo automático e equânime". Ao prever riscos com base em dados históricos, o sistema penal passa a agir não sobre atos efetivamente praticados, mas sobre probabilidades, o que pode configurar uma inversão da lógica do processo penal, que exige materialidade e autoria.

Tais práticas tornam ainda mais evidente o risco de automatizar decisões sensíveis sem considerar os contextos sociais, psicológicos e individuais. A prisão preventiva, por exemplo, já é aplicada de forma abusiva em diversos casos; com a IA, existe a possibilidade de essa distorção ser sistematizada e legitimada pela aparência de neutralidade tecnológica. Decidir sobre a liberdade de alguém com base em um cálculo algorítmico é um retrocesso do ponto de vista dos direitos fundamentais.

> Abordagens policiais são, com muita frequência, realizadas por critérios raciais, geográficos, e por estigmas fundados até mesmo na forma de vestimenta do "suspeito". Ou seja, o gradiente de discricionariedade na ação de investigar é tão elástico que essa discricionariedade se transmuda em arbitrariedade seletiva. (JUNIOR, GUASQUE e PÁDUA, 2023, p. 7)

A fala reforça a importância de reavaliar o papel dos algoritmos na tomada de decisões que afetam diretamente a liberdade das pessoas, sob pena de aprofundar a crise de legitimidade da justiça penal.

> A estes problemas, outros se acrescentam. Alguns atingem mesmo a essência da justiça penal, com a sua possível transformação numa justiça preditiva, orientada pela previsão do acontecimento criminal e pelo fim maior de construção de uma sociedade inócua e segura. Do outro lado da balança estarão, como custos, o sacrifício de direitos fundamentais (por exemplo, da liberdade daquele que é avaliado pelo sistema como sendo potencialmente perigoso) e dos princípios que estruturam um direito penal do facto e do resultado, assente na tutela subsidiária dos bens jurídicos. (SOUSA, 2020, p.33)

3.5. DEEPFAKES E A PRODUÇÃO DE PROVAS NO PROCESSO PENAL

À medida que a Inteligência Artificial avança, surgem também novas formas de manipular provas digitais. Um exemplo preocupante são os *deepfakes* — vídeos e áudios extremamente realistas criados por algoritmos, que conseguem imitar com perfeição a fala e os gestos de qualquer pessoa. Esse tipo de tecnologia coloca em xeque a confiança nas provas no processo penal, já que pode ser usado tanto para acusar alguém injustamente quanto para criar álibis que nunca existiram. “Com a reconstrução digital de imagens e as deepfakes, tornou-se possível, a partir de sistemas de inteligência artificial, criar vídeos de pessoas com base em imagens e vídeos antigos, produzindo-se cenas inéditas” (MEDON, 2021, p. 2)

A manipulação digital da imagem e da voz compromete a veracidade da prova e exige novos critérios de validação jurídica no processo penal. O sistema de justiça precisa adaptar-se rapidamente para distinguir o verdadeiro do falso em um cenário onde os sentidos humanos não são mais suficientes para aferir a autenticidade de um conteúdo audiovisual.

A questão se torna ainda mais preocupante quando tais provas são apresentadas em inquéritos ou ações penais sem a devida perícia técnica. A dificuldade em identificar uma montagem pode comprometer todo o curso do processo e gerar condenações indevidas. Nesse sentido, é imprescindível que o judiciário desenvolva protocolos técnicos para verificação de conteúdo digital e capacite seus agentes para lidar com esses novos desafios.

> A atual tecnologia de prova pericial não é capaz de verificar, por exemplo, uma montagem fotográfica se ela for bem feita, “pixel por pixel”, “bit a bit”. Ainda que o art. 159 do CPP preveja a contratação de assistentes técnicos para auxiliar o perito, isso pode refletir em famílias que não possuem condições de contratar tal profissional para verificar a autenticidade da prova. (NAKANISHI, 2023, p. 29)

A proliferação dos deepfakes exige a criação de perícias especializadas e a atualização constante dos instrumentos processuais, sob pena de fragilizar o direito de defesa e o princípio da verdade real. Assim, é evidente que a aplicação da IA no campo das provas não pode ocorrer sem respaldo científico, jurídico e ético.

3.6. A RESPONSABILIZAÇÃO PENAL DAS EMPRESAS DE TECNOLOGIA

Diante do crescimento acelerado das tecnologias digitais, as grandes empresas de tecnologia, conhecidas como ‘big techs’, passaram a ocupar um papel central não apenas na economia, mas também em decisões que impactam diretamente a vida das pessoas. Essa influência, somada ao amplo acesso a dados sensíveis e à utilização de algoritmos cada vez mais autônomos, levanta um debate essencial: como responsabilizar penalmente essas corporações quando seus sistemas causam danos à sociedade?

O que se observa é uma lacuna preocupante. Muitas dessas empresas operam em um espaço ainda pouco regulamentado, o que pode abrir brechas para abusos e práticas ilícitas, como a facilitação de fraudes financeiras, a lavagem de dinheiro ou mesmo a perpetuação de discriminações algorítmicas. Para Bechara, Tasinaffo e Castilho (2023), essa ausência de regulação pode tornar essas organizações verdadeiros "territórios de irresponsabilidade penal", especialmente quando há omissão na prevenção de riscos conhecidos.

A discussão sobre a responsabilização penal das pessoas jurídicas no Brasil ainda é recente, concentrando-se principalmente em crimes ambientais. No entanto, como argumentam Santos (2024), é urgente ampliar esse escopo para incluir os crimes digitais e econômicos, já que essas empresas lidam com informações valiosas e impactam diretamente os direitos fundamentais de milhões de usuários.

Reconhecer a responsabilidade penal das 'big techs' não significa impedir a inovação, mas garantir que ela ocorra dentro de parâmetros éticos e legais. Exigir medidas de compliance, transparência nos algoritmos e mecanismos eficazes de controle pode contribuir para uma atuação mais justa e responsável dessas corporações. Afinal, diante de seu enorme poder econômico e informacional, é fundamental que também respondam por eventuais danos que venham a causar, não apenas civilmente, mas também na esfera penal.

4. RACISMO ALGORÍTMICO E RECONHECIMENTO FACIAL

Os sistemas de reconhecimento facial têm sido apontados como particularmente problemáticos no que se refere ao racismo algorítmico. Estudos demonstram que esses sistemas apresentam altas taxas de erro no reconhecimento de pessoas negras e de outros grupos racializados.

> A IBM aproveitou a ocasião para declarar que cancelou, de maneira definitiva, o desenvolvimento de software de reconhecimento facial. Traduz-se em forte indicativo de que as ferramentas de reconhecimento facial são ineficientes, além de produzirem danos sociais seletivos e exclusão (JUNIOR, 2023, p.18)

Essa falha não é meramente técnica. Ela possui implicações sérias no campo penal, onde o reconhecimento equivocado de uma pessoa pode levar à sua prisão indevida. A repetição estatística de práticas discriminatórias nos dados históricos utilizados para o treinamento dos algoritmos contribui para a perpetuação dessas desigualdades. Os modelos não avaliam criticamente os dados; apenas os replicam. Assim, os erros de identificação não decorrem de falhas pontuais, mas de uma lógica sistêmica profundamente enraizada.

Diante disso, torna-se urgente repensar os critérios utilizados no treinamento dos sistemas de IA. A correção de vieses raciais não se limita à eliminação de atributos explícitos de raça nos dados, mas requer a reconstrução crítica do modo como esses sistemas são concebidos e implementados.

4.1. O VIÉS RACIAL NOS SISTEMAS DE INTELIGÊNCIA ARTIFICIAL

A aplicação de sistemas de IA no campo jurídico tem revelado um cenário preocupante no que se refere à reprodução de preconceitos raciais. Isso ocorre, sobretudo, quando os algoritmos são treinados a partir de bancos de dados historicamente enviesados, oriundos de contextos sociais marcados pelo racismo estrutural. Nesses casos, os sistemas aprendem com base em informações que, ainda que não intencionalmente, reforçam estereótipos e práticas discriminatórias. Como resultado, observa-se um índice elevado de erros em tecnologias como o reconhecimento facial, afetando desproporcionalmente pessoas negras.

> Hoje, os softwares operam em estados como Bahia, Rio de Janeiro, Ceará, São Paulo, Santa Catarina e Paraíba. Tecnologias que, sabidamente, possuem as maiores falhas operacionais relativas ao falso reconhecimento positivo de pessoas negras. Veja-se que, após cerca de um ano de utilização, as estatísticas relativas ao enviesamento do modelo já começaram a aparecer. Levantamento realizado pela Rede de Observatórios da Segurança constatou que, das 151 prisões pela via do sistema de reconhecimento facial que aconteceram no país, 90% eram de pessoas negras. (JUNIOR, 2023, P.19)

Estudos revelam que os algoritmos aplicados à segurança pública e ao controle penal apresentam significativa taxa de falsos positivos quando se trata de indivíduos negros. Esse padrão se perpetua porque os dados que alimentam os sistemas refletem decisões humanas passadas já contaminadas por discriminação racial. De acordo com Karina da Hora Farias (2022), os sistemas inteligentes, ao se basearem em dados históricos, inevitavelmente reproduzem os vieses neles contidos, colocando em risco a neutralidade esperada de suas decisões.

Essa constatação evidencia a necessidade de supervisão e auditoria permanentes nos sistemas de IA utilizados no Direito Penal. Não se pode admitir que decisões tão sensíveis, como uma prisão preventiva ou a indicação de periculosidade, sejam baseadas em tecnologias que não oferecem garantias de imparcialidade. A ausência de transparência nas decisões automatizadas dificulta, inclusive, o exercício pleno do contraditório e da ampla defesa, pilares do processo penal brasileiro.

> É necessário avaliar o impacto desse novo Judiciário em modo remoto sobretudo a partir dos valores constitucionais em vigor, de maneira a continuar garantindo

> plenamente a dignidade da pessoa, os valores sociais do trabalho, a ampla defesa e contraditório e, sobretudo, o real acesso de todos às novas tecnologias para atuar nesse Judiciário 4.0 de maneira plena, equilibrada com valores humanos e igualitária. (TAVARES, 2022, p. 11)

Portanto, combater o viés racial na IA não é apenas uma demanda técnica, mas sobretudo uma exigência jurídica e ética. O Estado, ao optar pelo uso dessas ferramentas, tem o dever de assegurar que sua utilização não contribua para a manutenção ou agravamento das desigualdades raciais. A regulação da IA precisa dialogar com os princípios constitucionais da igualdade e da dignidade, sob pena de legitimar a seletividade penal por vias automatizadas.

4.2. A ILUSÃO DA NEUTRALIDADE ALGORÍTMICA

A inteligência artificial é frequentemente vista como uma tecnologia objetiva, baseada em dados e livre de preconceitos humanos. No entanto, essa percepção ignora o fato de que os algoritmos são criados por pessoas inseridas em contextos culturais e ideológicos específicos. Assim, os preconceitos estruturais da sociedade acabam refletidos nos sistemas tecnológicos.

> Os dados disponíveis normalmente são os dados não rotulados. Como os algoritmos de machine learning são mais simples de serem desenvolvidos, os próprios programadores ou clientes rotulam os dados a serem minerados pelo programa. A rotulação agrega subjetividade ao processo de análise de dados. É nesse momento que o viés humano pode comprometer a imparcialidade almejada nos processos matemáticos de análise de dados. (MARQUES, 2022, p. 6)

A IA é moldada por escolhas humanas e institucionais. Os mecanismos de busca, como o Google e o Bing, não devem ser compreendidos apenas como ferramentas técnicas que organizam informações de maneira neutra. Eles são, na verdade, construções sociais e políticas, pois refletem as escolhas, valores e ideologias tanto dos engenheiros que os desenvolvem quanto das instituições que os financiam e operam.

Cada algoritmo é programado com critérios específicos que determinam quais resultados aparecem primeiro, quais conteúdos são priorizados e quais são excluídos — e essas decisões não são isentas de influências culturais, econômicas ou ideológicas. Assim, os buscadores podem reforçar visões de mundo dominantes, silenciar vozes minoritárias e até reproduzir preconceitos estruturais. Essa constatação exige uma análise crítica sobre a suposta imparcialidade da tecnologia, evidenciando que até mesmo as ferramentas mais automatizadas carregam intencionalidades humanas.

Essa suposta neutralidade é particularmente perigosa quando aplicada em contextos sociais sensíveis, como segurança pública, crédito, educação e saúde. Quando se acredita que a tecnologia é imparcial, falhas discriminatórias em seus resultados não são questionadas. Como consequência, decisões automatizadas tendem a reproduzir e até intensificar desigualdades

históricas. O problema não é a existência da tecnologia em si, mas o uso acrítico que se faz dela.

Safya Noble, em “Algorithms of Oppression”, mostra como os mecanismos de busca podem perpetuar estereótipos de gênero de forma sutil, mas devastadora. Em testes realizados com termos como “*black girls*” ou “*Latina women*”, os resultados priorizavam conteúdo hipersexualizado, pornográfico ou degradante. Isso indica que os algoritmos “aprendem” com padrões sociais discriminatórios e os replicam em larga escala. Segundo Noble, as sugestões de autocompletar do Google refletem os vieses incorporados em seus algoritmos e conjuntos de dados. (NOBLE, 2018).

Dessa forma, os sistemas de IA se tornam amplificadores de preconceitos. Quando não há uma crítica estruturada sobre os valores embutidos nesses sistemas, corre-se o risco de institucionalizar práticas discriminatórias sob o manto da eficiência tecnológica. Como bem afirma Noble, precisamos começar a enxergar os algoritmos como objetos políticos e culturais que exigem regulação e reforma. (NOBLE, 2018, p. 13), demonstrando a urgência de uma abordagem ética e regulatória na aplicação da inteligência artificial.

4.3. A INVISIBILIDADE E A REPRESENTAÇÃO DAS MULHERES NOS SISTEMAS DIGITAIS

A escassez de dados sobre mulheres na formulação de políticas públicas e no desenvolvimento tecnológico é um dos fatores mais alarmantes no debate sobre machismo e inteligência artificial. Hoje, majoritariamente, o mundo é masculino (tamanho, corpo, estética, preferências etc.) é tratado como universal, e o feminino como nicho; uma abordagem de pesquisa do tipo masculino, a menos que se indique o contrário (Perez, 2019, p. 30). Isso significa que o corpo e as necessidades masculinas são tratados como padrão, enquanto tudo o que se refere ao feminino é visto como exceção ou desvio.

Esse viés de gênero tem impactos concretos e perigosos. Perez revela casos em que sistemas de reconhecimento de voz funcionavam muito melhor para vozes masculinas, carros que não eram testados com manequins femininos e até medicamentos com dosagens calculadas com base apenas em corpos masculinos. Quando a inteligência artificial é treinada com dados incompletos, seus resultados podem ser não apenas imprecisos, mas letais. (Perez, 2019).

Além da ausência de dados, a forma como as mulheres são representadas nos sistemas digitais também revela estruturas de poder que moldam a tecnologia. Ao buscar por termos relacionados a mulheres negras, por exemplo, os resultados tendem a apresentar conteúdos

pornográficos ou depreciativos. O algoritmo é um sistema de tomada de decisão que influencia o conhecimento público e perpetua a desigualdade (Noble, 2018). A hipersexualização das mulheres, especialmente negras e latinas, demonstra como os sistemas de IA reforçam estereótipos raciais e sexuais de forma automatizada, afetando autoestima, segurança e empregabilidade.

Segundo Perez, o desenho de sistemas de transporte, por exemplo, é frequentemente baseado na rotina de deslocamento dos homens, de casa para o trabalho e vice-versa, ignorando o fato de que mulheres realizam múltiplas paradas diárias. As normas e os tipos são desvios das mulheres, e não o contrário (Perez, 2019). Isso reforça a ideia de que a tecnologia continua a ser pensada e produzida por e para homens. Perez também adverte que mesmo os países mais desenvolvidos não estão imunes a esse viés: “sua abordagem comparativa e interseccional mostra que até as partes mais 'desenvolvidas' do mundo [...] não estão imunes aos preconceitos mantidos pelas convenções da lacuna de dados de gênero. “ (Perez, 2019).

Noble complementa essa análise ao afirmar que os algoritmos de busca refletem e reforçam os preconceitos históricos: os algoritmos não são neutros; eles refletem os contextos sociais e históricos em que foram criados (Noble, 2018). Ela alerta que os sistemas automatizados podem amplificar práticas discriminatórias em escala, criando novos meios de exclusão. As convenções dominadas por homens na geração e disseminação de dados perpetuam uma lógica excludente, em que a ausência de mulheres nas equipes de desenvolvimento tecnológico compromete diretamente a representatividade nos produtos criados.

4.4. CAMINHOS PARA A JUSTIÇA ALGORÍTMICA DE GÊNERO

A superação do machismo na inteligência artificial passa pela construção de sistemas mais inclusivos e transparentes. Um dos caminhos é garantir a diversidade nas equipes que desenvolvem tecnologia. Como afirma Noble (2018), equipes diversas têm mais chances de antecipar uma gama mais ampla de danos e desenvolver sistemas mais equitativos. Isso significa que a presença de mulheres, pessoas negras e outras minorias na criação de IA é fundamental para a justiça algorítmica.

> Tratar da Inteligência artificial e os vieses de gênero é cuidar também das mulheres vítimas de enorme discriminação, começando pela diferença salarial entre homens e mulheres. A discriminação tende aumentar com o uso das novas ferramentas, o algoritmo leva em consideração as crenças de quem o elabora. (FORMIGOSA, 2023)

Outro ponto crucial é a revisão dos conjuntos de dados utilizados para treinar os sistemas. Dados enviesados geram decisões enviesadas. O corpo feminino precisa ser incluído como uma variável real, e não como um erro estatístico. Essa mudança exige não apenas ajustes técnicos, mas uma revisão ética e metodológica profunda sobre o que se considera como "dado válido".

Além disso, é necessário que as decisões algorítmicas sejam transparentes e auditáveis. O funcionamento de sistemas que afetam diretamente a vida das pessoas – como concessão de crédito, contratação de seguros ou seleção de currículos – não pode ser uma "caixa preta". A regulação estatal e a vigilância da sociedade civil são ferramentas essenciais para garantir que esses sistemas estejam alinhados aos princípios democráticos. Noble reforça que a regulação e reforma são necessárias para combater os danos causados pelos sistemas automatizados de tomada de decisão. (NOBLE, 2018).

Por fim, a luta por representatividade e justiça não é apenas uma agenda feminista, mas uma questão de humanidade, promover uma IA justa é um dever coletivo, que exige consciência crítica, engajamento político e compromisso ético com a igualdade.

5. CONSIDERAÇÕES FINAIS

Cada vez mais a Inteligência Artificial tem ganhado espaço como uma aliada na modernização de instituições públicas e privadas. Mas, à medida que essa tecnologia avança e passa a ser usada em áreas como a justiça, a segurança, o sistema financeiro, a saúde e até a mídia, fica mais evidente um problema preocupante: o risco de repetir desigualdades antigas, baseadas em raça, gênero e classe. Como foi mostrado ao longo deste trabalho, os algoritmos não são neutros. Eles refletem a sociedade que os criou e, se não forem monitorados com cuidado, podem acabar reforçando essas desigualdades, disfarçadas de decisões técnicas e objetivas. (O'NEIL, 2020; NOBLE, 2018; PEREZ, 2019).

A ausência de representatividade nos dados, o apagamento das vivências femininas e o racismo estrutural nos conjuntos de treinamento revelam uma fragilidade ética e epistêmica nos sistemas algorítmicos atuais. Como demonstrado por Perez (2019), o "gender data gap" compromete desde políticas públicas até decisões automatizadas que afetam a vida de milhões de mulheres.

Dessa forma, torna-se cada vez mais urgente construir uma regulamentação algorítmica comprometida com a justiça social. Isso passa, primeiro, por criar regras claras e eficazes, como

propõe o Projeto de Lei nº 2338/2023 — que assegurem que as decisões automatizadas sejam transparentes, possam ser auditadas e explicadas de forma acessível. Também é fundamental contar com comissões formadas por especialistas de diferentes áreas, como juristas, engenheiros, sociólogos e representantes da sociedade civil, para avaliar o uso ético dessas tecnologias.

Além disso, é preciso garantir que as equipes responsáveis pelo desenvolvimento da inteligência artificial sejam diversas, trazendo olhares e vivências distintas para dentro dos processos de criação. Outro ponto essencial é considerar recortes sociais e interseccionais na hora de construir os bancos de dados que alimentam os algoritmos, para evitar que padrões excludentes sejam repetidos.

Diante desse cenário, torna-se cada vez mais necessário que a sociedade repense a forma como as grandes empresas de tecnologia devem ser responsabilizadas pelos impactos causados pelos sistemas que desenvolvem. As big techs, com seu imenso alcance e domínio sobre dados, frequentemente operam à margem dos mecanismos jurídicos tradicionais. Quando os algoritmos criados por essas corporações contribuem para violações de direitos, reforçam desigualdades ou facilitam práticas criminosas, é essencial que haja uma resposta penal clara e eficaz.

É imprescindível que as instituições públicas brasileiras, sobretudo o Judiciário e o Legislativo, assumam protagonismo no controle democrático das tecnologias de decisão automatizada. A responsabilização civil e penal por danos causados por sistemas de IA precisa ser normatizada, garantindo reparação e prevenção. Os algoritmos devem servir à justiça, e não a substituir.

Mais do que aspectos técnicos e regulatórios, combater os vieses algorítmicos exige também uma educação digital crítica. É essencial formar profissionais conscientes, com base em discussões sobre ética na computação, justiça algorítmica, racismo estrutural e desigualdade de gênero. Afinal, as tecnologias não apenas reproduzem a realidade, elas ajudam a moldá-la. Da mesma forma, a transparência dos órgãos públicos que utilizam inteligência artificial deve ser prioridade. A sociedade precisa saber quais algoritmos estão em uso, quais dados os alimentam e com base em que critérios tomam decisões. Esse é um passo indispensável para garantir o controle social e proteger direitos fundamentais.

> Por isso que deve haver total transparência das fórmulas algorítmicas usadas porque não protegidas por propriedade intelectual. Somente dessa forma, a justiça preditiva deve permitir compreender melhor a maneira pela qual uma questão específica é tratada pelo juiz judicial. (SANCTIS, 2020 p. 54)

Por fim, é preciso compreender que a resolução dos vieses e preconceitos da inteligência artificial não ocorrerá de forma espontânea, nem por mera evolução técnica. É necessário um compromisso ativo, intersetorial e contínuo, que articule Direito, tecnologia, sociedade civil e movimentos sociais. A construção de uma IA mais justa depende de escolhas políticas, decisões institucionais e, acima de tudo, de uma visão de mundo que reconheça a diversidade como valor fundamental. Como reforça Caroline Perez (2019), “incluir mulheres nos dados não é apenas uma questão de justiça; é uma questão de sobrevivência”. Portanto, só haverá verdadeira inovação quando a inteligência artificial for também instrumento de equidade, justiça e dignidade para todas as pessoas. A tecnologia deve estar a serviço da justiça — e não acima dela.

REFERÊNCIAS

AKABANE, Getúlio K. *Gestão estratégica das tecnologias cognitivas*. Érica, 2018. Disponível em: https://app.minhabiblioteca.com.br/reader/books/9788536530000/.

ALBINO, João Pedro; VALENTE, Vânia Cristina Pires Nogueira (Org.). *Inteligência artificial e suas aplicações interdisciplinares*. Editora e-Publicar, 2023. Disponível em: https://dx.doi.org/10.47402/ed.ep.b202320930201.

BECHARA, Fábio Ramazzini; TASINAFFO, Fernanda Lima Venciguerra; CASTILHO, Alexandre Affonso. Análise crítica da responsabilidade penal das pessoas jurídicas frente ao poder econômico das big techs. *Revista Diálogos Possíveis*, São Paulo, 2023.

BRASIL. Lei nº 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. *Presidência da República*, 2014. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). *Diário Oficial da União: seção 1*, Brasília, DF, ano 155, n. 157, p. 59-64, 15 ago. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm.

BRASIL. *Lei nº 8.078, de 11 de setembro de 1990.* Dispõe sobre a proteção do consumidor e dá outras providências. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8078compilado.htm. Acesso em: 21 maio 2025.

BRASIL. Projeto de Lei nº 2.338, de 2023. Dispõe sobre o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana. *Câmara dos Deputados*, 17 mar. 2025. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2868197&filename=PL%202338/2023.

CONSELHO NACIONAL DE JUSTIÇA. *Justiça do Trabalho da 8ª Região debate inteligência artificial na discriminação de gênero.* Agência CNJ de Notícias, 30 mar. 2023. Disponível em: https://www.cnj.jus.br/justica-do-trabalho-da-8a-regiao-debate-inteligencia-artificial-na-discriminacao-de-genero/. Acesso em: 21 maio 2025.

COSTA, Pedro Carvalho Ferreira da. *Conversations with ELIZA: sobre género e inteligência artificial*. Universidade de Lisboa, 2018.

EXAME. *Pesquisa revela que 56% dos brasileiros já percebem impacto da inteligência artificial na sociedade*. s.d. Disponível em: https://exame.com/inteligencia-artificial/pesquisa-revela-que-56-dos-brasileiros-ja-percebem-impacto-da-inteligencia-artificial-na-sociedade/amp/.

FACELI, Katti; LORENA, Ana C.; GAMA, João; AL, et. *Inteligência Artificial: uma abordagem de aprendizado de máquina*. LTC, 2021. Disponível em: Acesso restrito via Minha Biblioteca.

FLECK, Leandro; TAVARES, Maria Hermínia Ferreira; EYNG, Eduardo; HELMANN, Andrieli Cristina; ANDRADE, Minéia Aparecida de Moares. *Redes neurais artificiais: princípios básicos*. *Revista Eletrônica Científica Inovação e Tecnologia*, Medianeira, PR, v. 1, n. 13, p. 47–57, 2016. Disponível em: https://periodicos.utfpr.edu.br/recit/article/viewFile/4330/Leandro.

FLORIDI, Luciano; COWLS, Josh. *A unified framework of five principles for AI in society*. *PhilArchive*, s.d. Disponível em: https://philarchive.org/archive/FLOHTD.

HARTMANN PEIXOTO, Fabiano. *Direito e Inteligência Artificial*. Universidade de Brasília, 2020. Disponível em: https://livros.unb.br/index.php/portal/catalog/book/200.

IBM. *WatsonX: Inteligência Artificial explicável e segura*. 2024. Disponível em: https://www.ibm.com/docs/pt-br/watsonx/w-and-w/2.1.0?topic=cases-watsonxai-use-case.

LEE, Kai-Fu. *A quarta revolução industrial da IA*. Globo Livros, 2019. Disponível em: https://www.kufunda.net/publicdocs/Intelig%C3%AAncia%20artificial%20(Kai-Fu%20Lee).pdf. Acesso em: 19 maio 2025.

MARQUES, Fabíola; MARTINEZ NETO, Aldo Augusto. *Vieses algorítmicos, direitos fundamentais e os sindicatos*. Ano 8 (2022), nº 6, p. 707-729, 2022. Disponível em: https://www.cidp.pt/revistas/rjlb/2022/6/2022_06_0707_0729.pdf.

NAKANISHI, Maria Fernanda Mugnaini. *A problemática jurídica dos deepfakes: uma análise do uso da inteligência artificial na produção de provas e suas repercussões penais.* 2023.

Trabalho de Conclusão de Curso (Bacharelado em Direito) – Centro Universitário de Brasília – UniCEUB, Faculdade de Ciências Jurídicas e Sociais – FAJS, Brasília, 2023.

Disponível em: https://repositorio.uniceub.br/jspui/bitstream/prefix/17157/1/22001667.pdf

NOBLE, Safiya Umoja. *"Algorithms of Oppression": How Search Engines Reinforce Racism*. NYU Press, 2018. Disponível em: Acesso restrito via Minha Biblioteca.

O'NEIL, Cathy. *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown Publishers, 2016.

PEREZ, Caroline Criado. *Mulheres invisíveis: o viés dos dados em um mundo projetado para homens*. Tradução de GUERRA, Renata. Intrínseca, 2022. Acesso restrito via Minha Biblioteca.

RAHMAN, Tarik. *Artificial Intelligence and Ethics: Contemporary Reflections*. Oxford University Press, 2021. Disponível em: https://www.researchgate.net/publication/381574628_Artificial_Intelligence_and_Ethics.

ROSA, Alexandre Morais da; GUASQUE, Bárbara. *Inteligência artificial, vieses algorítmicos e racismo: o lado desconhecido da justiça algorítmica*. *Open Journal of the University of Medellín*, v. 23, n. 50, 2023. Disponível em: https://revistas.udem.edu.co/index.php/opinion/article/download/4046/3797.

SANCTIS, Fausto Martin de. *Inteligência artificial e direito*. Almedina Brasil, 2020. Disponível em: Acesso restrito via Minha Biblioteca.

SANCTIS, Fausto Martin de. *Inteligência Artificial e Direito*. Almedina Brasil, 2020. Disponível em: https://app.minhabiblioteca.com.br/reader/books/9786556270890/.

SANTOS, Letícia Andrade dos et al. Responsabilização penal da inteligência artificial: uma revisão integrativa sobre a possibilidade de entidades tecnológicas serem criminalmente punidas. *Revista de Iniciação Científica*, Florianópolis, 2024.

SOUSA, Susana Aires de. *Um direito penal desafiado pelo desenvolvimento tecnológico: alguns exemplos a partir das neurociências e da inteligência artificial*. s.d.

PEREIRA, Sara Matias Ferrari; OLIVEIRA, Tarsis Barreto. *The use of artificial intelligence in criminal law and its impact on the fundamental rights of non-discrimination and privacy.*

Revista do Instituto de Direito Constitucional e Cidadania – IDCC, Londrina, v. 9, n. 1, e099, jan./jun. 2024. DOI: 10.48159/revistadoidcc.v9n1.e099.

WHEELER, J. Craig. *The Second Machine Age by Erik Brynjolfsson and Andrew McAfee - Commentary*. 2015. Disponível em: https://www.as.utexas.edu/astronomy/education/fall15/wheeler/secure/ExponentialGrowth.pdf .