

# ENVIESAMENTO NA INTELIGÊNCIA ARTIFICIAL: UM ESTUDO DOS SEUS IMPACTOS

RUAN TIAGO SANTANA DE OLIVEIRA<sup>1</sup>  
JEAN CLEVERSON SIMÕES MUTTI AFONSO RÊGO<sup>2</sup>

UNIVERSIDADE DO ESTADO DA BAHIA (UNEB)  
2024

## RESUMO

Este trabalho analisa o fenômeno do enviesamento algorítmico na inteligência artificial (IA), destacando como algoritmos, projetados para decisões imparciais, podem reproduzir e amplificar preconceitos presentes nos dados de treinamento. A pesquisa aborda os fundamentos da IA, incluindo seus paradigmas de aprendizado (supervisionado, não supervisionado, por reforço, por transferência e *Deep Learning*), suas aplicações em nosso dia a dia, desde consultas ao clima pela manhã até a recomendação de um filme a noite, e os impactos práticos de algoritmos enviesados em áreas como recrutamento e justiça criminal. Por meio de casos reais e hipotéticos, demonstra-se como a perpetuação de vieses marginaliza grupos minoritários e reforça desigualdades sociais. Com base em uma revisão da literatura e estudos de caso, o artigo propõe medidas para mitigar esses problemas, como sistemas “*human-in-the-loop*” e maior representatividade nos dados de treinamento, conforme técnicas sugeridas por Munro e Rajkomar et al. Os resultados confirmam a hipótese inicial e reforçam a importância de estratégias éticas e inclusivas no desenvolvimento e aplicação da IA.

**Palavras-chave:** Inteligência Artificial (IA). Enviesamento Algorítmico. Mitigação de Vieses.

## ABSTRACT

This paper analyzes the consequences of algorithmic bias in artificial intelligence (AI), highlighting how algorithms designed for impartial decisions can reproduce and amplify biases present in training data. The research addresses the fundamentals of AI, including its learning paradigms (supervised, unsupervised, reinforcement, transfer, and deep learning), its applications in our daily lives, from checking the weather in the morning to recommending a movie at night, and the practical impacts of envied algorithms in areas such as recruitment and criminal justice. Through real and hypothetical cases, it demonstrates how the perpetuation of lives marginalizes minority groups and reinforces social inequalities. Based on a literature review and case studies, the paper proposes measures to mitigate these problems, such as 'human-in-the-loop' systems and greater representation in training data, according to techniques suggested by Munro and Rajkomar et al. The results confirm the initial hypothesis and reinforce the importance of ethical and inclusive strategies in the development and application of AI.

**Keywords:** Artificial Intelligence (AI). Algorithmic Bias. Bias Mitigation

---

<sup>1</sup> Graduando em Sistema de Informação – ruantsdo@gmail.com

<sup>2</sup> Bacharelado Em Análise De Sistemas, Especialista em Gestão de Pessoas com ênfase em Consultoria, Mestre em Administração de Empresas e Professor da UNEB - jrego@uneb.br

# 1 INTRODUÇÃO

Apesar dos avanços promissores, a Inteligência Artificial (IA) carrega consigo o risco de perpetuar e amplificar desigualdades sociais através de algoritmos enviesados. Além disso, esse avanço acelerado também expõe uma face preocupante: a presença de vieses em algoritmos de IA, o que levanta sérias preocupações éticas e sociais. Com o seu crescente uso em áreas como: medicina, processamento de dados e geração de conteúdo, essas preocupações se tornam ainda mais relevantes. Portanto, embora os sistemas de IA sejam projetados para tomar decisões imparciais, eles podem ser influenciados por vieses presentes nos dados de treinamento, resultando em discriminações contra certos grupos, como mulheres e minorias étnicas. Dessa forma, o trabalho propõe-se a explorar esses vieses, suas causas, impactos e possíveis soluções.

O problema central identificado é que algoritmos de IA, quando treinados com dados tendenciosos ou imprecisos, podem reproduzir ou amplificar esses vieses em suas decisões, causando resultados discriminatórios. Isso é especialmente preocupante em áreas onde a IA atua de forma decisiva, como recrutamento, crédito e reconhecimento facial, levantando sérias preocupações sobre seus impactos negativos na sociedade. Diante desse cenário, a pesquisa busca investigar como esses vieses afetam a sociedade e como podem ser mitigados.

Nesse contexto, a pesquisa é motivada pela necessidade de entender e amenizar falhas nos sistemas de IA que podem perpetuar desigualdades sociais. Nesse sentido, o objetivo geral do estudo é identificar e expor os impactos dos vieses algorítmicos na IA. Além disso, pretende-se identificar métodos de mitigação desses vieses e obter informações sobre as melhores formas de minimizar seus impactos negativos.

A crescente onipresença da IA na vida cotidiana exige a garantia de que tais sistemas operem de forma justa e ética, tornando urgente a investigação e combate aos vieses algorítmicos e suas profundas implicações sociais. Portanto, compreender esses mecanismos é essencial para promover o desenvolvimento responsável da IA.

Este trabalho parte da hipótese de que os algoritmos de inteligência artificial, ao serem treinados com dados tendenciosos, perpetuam e amplificam preconceitos sociais, impactando desproporcionalmente grupos minoritários e vulneráveis. Esses vieses representam desafios éticos e sociais significativos, mas podem ser mitigados por meio de práticas que garantam maior representatividade nos dados e o uso de abordagens éticas no desenvolvimento de sistemas de IA.

Compreender a essência e o funcionamento da IA é essencial para explorarmos seus benefícios e enfrentarmos desafios éticos, como o viés algorítmico, que é o foco central desta análise. Ao adentrar neste universo, buscamos não apenas entender sua transformação em nossa rotina, mas também conscientizar sobre os riscos associados ao seu uso. Esse entendimento servirá de base para analisar os impactos dos vieses algorítmicos, entender e identificar possíveis soluções para mitigar seus efeitos negativos, promovendo um uso mais ético e inclusivo da tecnologia.

Visando entender melhor o que é IA, este artigo está estruturado para abordar de maneira sistemática os aspectos fundamentais, impactos e soluções relacionados ao enviesamento algorítmico na inteligência artificial. E para isso, é importante

primeiramente observar sua evolução histórica e identificar alguns dos principais marcos que moldaram esse campo tão dinâmico. Após essa contextualização, serão exploradas as diversas definições de IA propostas por diferentes especialistas, com o objetivo de esclarecer sua essência. Em seguida, abordaremos de forma simplificada seu funcionamento, as formas de aprendizado e suas aplicações no cotidiano.

## **2 DESCOBRINDO O QUE É INTELIGÊNCIA ARTIFICIAL**

A crescente presença da Inteligência Artificial (IA) em nosso cotidiano, desde *smartphones* a veículos autônomos, exige uma compreensão profunda de seu funcionamento, potencial e, crucialmente, seus impactos sociais e éticos. Esta seção visa desmistificar a IA apresentando seus conceitos-chave, características, aplicações e impactos sociais, com foco especial no desafio do viés algorítmico. Com isso em mente, vamos observar sua evolução histórica e identificar alguns dos principais marcos que moldaram esse campo tão dinâmico.

### **2.1 CONTEXTUALIZAÇÃO HISTÓRICA**

A Inteligência Artificial como a conhecemos hoje, foi construída sobre uma base histórica rica e interdisciplinar, com contribuições de visionários que estabeleceram os alicerces da computação e da IA. Este capítulo revisita momentos-chave dessa evolução, destacando como pioneiros como Alan Turing e iniciativas marcantes nas décadas de 1950 e 1960 impulsionaram o desenvolvimento de conceitos e tecnologias que continuam a moldar o campo da IA.

Alan Turing, pioneiro da computação e considerado o pai da ciência da computação, é reconhecido por ter decifrado o código Enigma, um dos principais códigos utilizados pelos nazistas durante a Segunda Guerra Mundial. Essa conquista, crucial para o sucesso dos aliados, foi possibilitada pela invenção da Máquina de Turing, um modelo teórico de computador capaz de realizar qualquer tipo de cálculo que possa ser expresso por um algoritmo (Brandão, 2017).

Além de sua contribuição para a computação, Turing também deixou um legado significativo em áreas como matemática, lógica, biologia e filosofia. Sua abordagem estruturada da IA, pioneira para a época, serve até hoje como fundamento para os avanços da área (Brandão, 2017).

Entre as décadas de 1950 e 1960, a IA e o conexionismo (abordagem que busca simular a estrutura do cérebro humano por meio de redes neurais artificiais) experimentaram avanços notáveis. Newell e Simon desenvolveram programas capazes de realizar provas geométricas e lógicas. Simultaneamente, Frank Rosenblatt e outros pesquisadores exploraram as redes neurais, encontrando maneiras de treiná-las para realizar diversas tarefas de reconhecimento de padrões (Chrisley; Begeer, 2000, p.361). Ver figura 1.

Fig.1. Algoritmo de bolso para dados de treinamento ilimitados. Exemplos de treinamento não são armazenados.

**GALLANT: PERCEPTRON-BASED LEARNING ALGORITHMS**

**INPUT:** Training Examples  $\{ E^k, C^k \}$ .  $E^k$  is a vector with  $E_0^k \equiv 1$  and other components,  $E_1^k, \dots, E_p^k$ , assuming values in  $\{+1, -1, 0\}$ .  $C^k = \{+1, -1\}$  is the desired response. Examples do not reside in memory but are produced one-by-one (sampling with replacement) whenever needed.

**OUTPUT:**  $W = \langle w_{p+1,0}, w_{p+1,1}, \dots, w_{p+1,p} \rangle$  is a vector of integral 'pocket' weights where  $w_{p+1,0}$  is the bias.

**TEMPORARY DATA:**

$\pi$  = vector of integral perceptron weights,  $\langle \pi_0, \pi_1, \dots, \pi_p \rangle$ .  
 $run_\pi$  = number of consecutive correct classifications using perceptron weights  $\pi$ .  
 $run_W$  = number of consecutive correct classifications using pocket weights  $W$ .

1. Set  $\pi = \langle 0, 0, \dots, 0 \rangle$  and  $run_\pi = run_W = num\_ok_\pi = num\_ok_W = 0$ .
2. Obtain a training example  $E^k$  (with corresponding classification  $C^k$ ).
3. If  $\pi$  correctly classifies  $E^k$ , i.e.
  - $\{ \pi \cdot E^k > 0 \text{ and } C^k = +1 \}$  or
  - $\{ \pi \cdot E^k < 0 \text{ and } C^k = -1 \}$
 Then:
  - 3a.  $run_\pi = run_\pi + 1$ .
  - 3b. If  $run_\pi > run_W$ 
    - Then:
      - 3ba. Set  $W = \pi$
      - 3bb. Set  $run_W = run_\pi$
 Otherwise:
  - 3A. (Change step:) Form a new vector of perceptron weights
    - $\pi = \pi + C^k E^k$
  - 3B. Set  $run_W = 0$
4. End of this iteration. If the specified number of iterations has not been taken then go to 2.

Fonte: Gallant, S. (1990). Perceptron-based learning algorithms

Desenvolvido na década de 1960 por Joseph Weizenbaum, o ELIZA é considerado de acordo com Berry (2023), um dos programas de computador mais influentes já escritos. O ELIZA, e sua persona mais famosa, o DOCTOR, simulava conversação humana, o que atraiu atenção de programadores, gerando discussões e inspirando novas pesquisas na área até os dias de hoje.

## 2.2 O QUE SERIA INTELIGÊNCIA ARTIFICIAL?

Após revisitarmos os marcos históricos que moldaram o desenvolvimento da inteligência artificial, é essencial compreender o que define essa área em termos conceituais. A Inteligência Artificial, embora amplamente reconhecida, apresenta diferentes interpretações entre pesquisadores e organizações. Neste capítulo, exploraremos essas definições, destacando as perspectivas complementares que ajudam a entender a essência, o propósito e as capacidades dessa tecnologia transformadora.

A Inteligência Artificial é um campo da ciência da computação que se dedica a criar sistemas capazes de realizar tarefas complexas que, até então, eram consideradas domínio exclusivo dos humanos.

A IA, segundo Alura (2024), é um ramo da ciência da computação que se dedica à criação de sistemas e algoritmos capazes de realizar tarefas que normalmente demandam inteligência humana. A TOTVS (2024) corrobora com essa definição, afirmando que a inteligência artificial é um campo científico que busca desenvolver e implementar máquinas para executarem atividades humanas de forma autônoma. Complementando, a IBM (2024, p. 1) define IA como: “a utilização de computadores e máquinas para imitar a capacidade de resolução de problemas e tomada de decisão da mente humana.”

Diferentes autores e pesquisadores têm oferecido perspectivas complementares sobre o que define IA. Alguns a veem como um esforço para criar máquinas pensantes, “como mentes, no sentido total e literal” (Haugeland, 1985). Outros enfatizam a capacidade da IA de executar funções que exigem inteligência humana (Kurzweil, 1990). Há ainda quem defina IA como o estudo das faculdades mentais por meio de modelos computacionais (Charniak; Mcdermott, 1985) ou como o projeto de agentes inteligentes (Poole et al., 1998).

Em essência, a inteligência artificial se caracteriza pelo aprendizado, adaptabilidade e autonomia de máquinas e algoritmos, imitando, em diversos graus, as capacidades cognitivas humanas. Mas como, de fato, esses sistemas complexos operam? Quais são os mecanismos que permitem que máquinas aprendam, se adaptem e tomem decisões? Para compreender melhor o funcionamento da IA, exploraremos brevemente seus princípios subjacentes a seguir.

### 2.3 COMO FUNCIONA UMA IA: UMA VISÃO GERAL DOS PRINCÍPIOS

A inteligência artificial opera como um sistema complexo que combina algoritmos, dados e poder computacional para simular aspectos da inteligência humana. Essa integração permite que as máquinas não apenas aprendam e se adaptem, mas também tomem decisões com base em padrões extraídos dos dados processados. Como destaca a TOTVS (2024, p.1), “a IA funciona por meio de uma combinação de algoritmos, dados e computação, um complexo nó de processos que possibilita às máquinas aprender, se adaptar e tomar decisões baseadas em dados”.

A capacidade da IA de “aprender” com dados é um de seus atributos mais impressionantes. Seja utilizando exemplos pré-definidos ou explorando de forma autônoma um vasto conjunto de informações, esses sistemas identificam padrões, fazem previsões e tomam decisões que podem otimizar tarefas e resolver problemas de forma eficiente.

O funcionamento de uma IA pode variar conforme a técnica de aprendizado adotada. De acordo com a Alura (2024), há modelos que correlacionam entradas específicas com saídas desejadas, enquanto outros, sem qualquer orientação prévia, buscam identificar padrões ocultos nos dados. Além disso, algumas IA's aprendem interagindo com o ambiente, ajustando suas ações conforme a experiência adquirida.

Essas diferentes formas de aprendizado são essenciais para moldar a forma como a IA resolve problemas e se aprimora ao longo do tempo. A seguir, exploraremos cada um desses paradigmas em maiores detalhes, desvendando como eles sustentam a evolução e a aplicação prática da IA.

### 3 DESVENDANDO A APRENDIZAGEM EM IA: UMA JORNADA PELOS PARADIGMAS

Na inteligência artificial, o aprendizado de máquina se destaca como uma área essencial e em constante evolução, moldando muitas das inovações que utilizamos no dia a dia. Nesta seção, exploraremos os principais métodos de aprendizado que sustentam esse campo, destacando seus conceitos fundamentais.

*Machine Learning: Essência e Algoritmos de Aprendizagem* – Segundo a IBM (2024), “*Machine learning* é uma área da inteligência artificial e da ciência da computação que se concentra no uso de dados e algoritmos para imitar a maneira como os humanos aprendem, melhorando gradualmente sua precisão”. O aprendizado de máquina permite a identificação automática de padrões e relações em grandes volumes de dados (Bishop, 2006).

*Aprendizagem Supervisionada: Aprendendo com Dados Rotulados* – Esse tipo de aprendizado é o mais comum e envolve duas etapas: treinamento e teste. Durante o treinamento, o modelo recebe dados rotulados, nos quais cada entrada está associada a um rótulo (saída desejada). Assim, ele aprende a relação entre entrada e saída, tornando-se capaz de fazer previsões para novos dados. Exemplos de algoritmos usados aqui incluem árvores de decisão, regressão linear e *Naive Bayes* (Nasteski, 2017).

*Aprendizagem Não Supervisionada: Descobrir Padrões Ocultos* – Ao contrário do método supervisionado, a aprendizagem não supervisionada trabalha com dados não rotulados. O objetivo é identificar padrões e estruturas ocultas nos dados, preparando-os para análises futuras. Essa técnica é especialmente útil para agrupamentos e detecção de anomalias (Ongsulee, 2017).

*Aprendizagem por Reforço: A Aprendizagem pela Experimentação* – “A aprendizagem por reforço é aprender o que fazer – como mapear situações em ações – de modo a maximizar um sinal numérico de recompensa” (Sutton; Barto, 2018). Neste método, o modelo não recebe instruções explícitas, mas descobre quais ações geram melhores recompensas por meio de tentativa e erro. Além da recompensa imediata, ele também aprende a considerar o impacto das ações nas recompensas futuras.

*Aprendizagem Transferida: Reaproveitando Conhecimento para Novas Tarefas* – Essa abordagem permite que o conhecimento adquirido em uma tarefa seja utilizado para resolver problemas semelhantes em um novo contexto. Em vez de começar do zero, o modelo aproveita padrões aprendidos anteriormente para acelerar o processo de aprendizado em um domínio relacionado (LU et al., 2015).

*Deep Learning: Aprendizado de Máquina para Percepção e Análise* – O aprendizado profundo (*deep learning*) é uma técnica poderosa que utiliza redes neurais compostas por várias camadas de processamento. Essas camadas extraem automaticamente características relevantes dos dados, eliminando a necessidade de pré-processamento manual. Aplicações comuns incluem reconhecimento de imagens e voz. À medida que cada camada processa o dado, conceitos mais abstratos são identificados nas camadas superiores (Rusk; Nicole, 2016).

Esses paradigmas de aprendizado moldam a forma como a IA opera e se aprimora. No próximo capítulo, exploraremos com mais detalhes como esses métodos são aplicados e quais desafios e oportunidades surgem ao combiná-los em sistemas inteligentes.

## 4 APLICAÇÕES DA IA E SUA UTILIDADE NOS DIAS ATUAIS

A inteligência artificial está cada vez mais presente em nosso cotidiano de maneira tão sutil que, muitas vezes, nem percebemos sua influência. Desde a consulta ao clima em nossos *smartphones* até a recomendação de séries à noite, algoritmos inteligentes trabalham silenciosamente para tornar nossas atividades diárias mais eficientes.

Na medicina, a IA tem revolucionado o diagnóstico e o tratamento. De acordo com a Anhanguera (2024), algoritmos de aprendizado profundo analisam imagens médicas com alta precisão, identificando doenças como câncer de mama de forma mais rápida e eficiente. Na medicina de precisão, modelos de IA, treinados com dados clínicos e genéticos, ajudam a prever a probabilidade de um paciente desenvolver doenças. Além disso, a automação de tarefas repetitivas, como a triagem de pacientes, libera os profissionais de saúde para se concentrarem em atividades mais complexas e humanizadas.

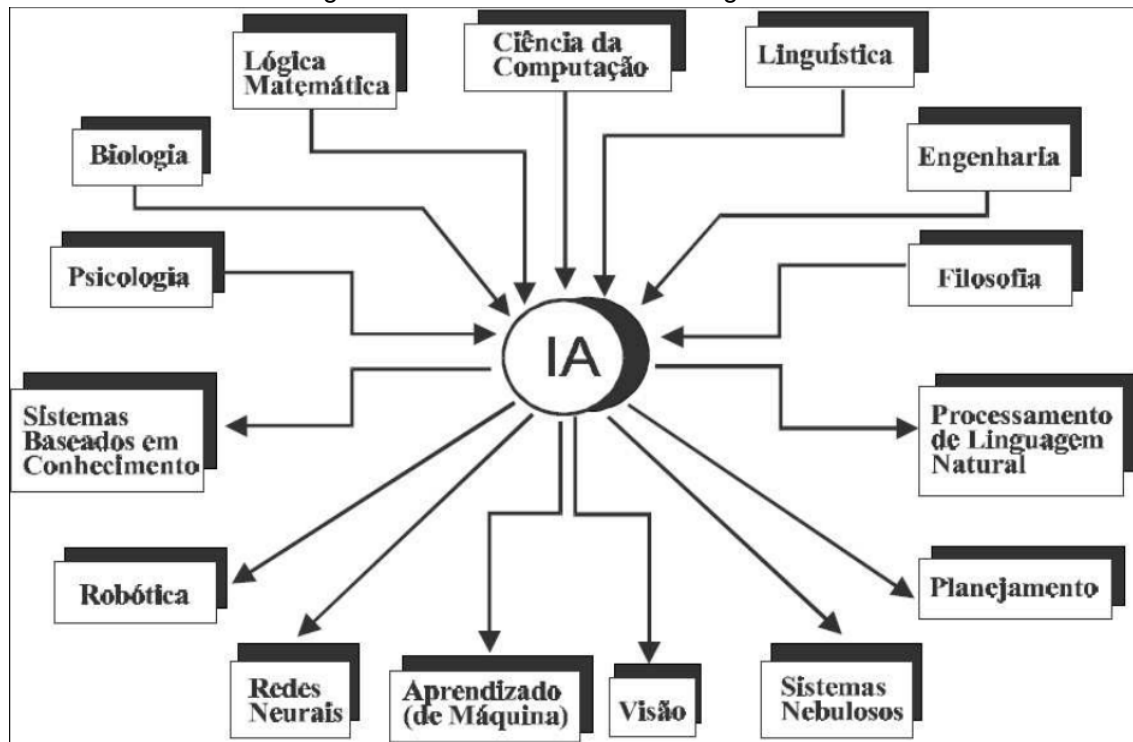
A aplicação da IA não se limita ao campo da saúde. Segundo a TOTVS (2024), a IA permite processar grandes volumes de dados, gerando *insights* estratégicos para empresas. Ela é capaz de identificar padrões, antecipar a demanda por produtos e entender o comportamento do consumidor, aprimorando a oferta de serviços e produtos.

No setor de entretenimento, conforme a Alura (2024), a IA cria oponentes artificiais em jogos de vídeo e desenvolve conteúdo procedural para cenários e personagens em jogos e filmes, proporcionando experiências interativas e imersivas.

A crescente integração da inteligência artificial em diferentes aspectos da sociedade nos convida a refletir sobre sua dualidade: enquanto simplifica atividades diárias e promove avanços significativos, como na medicina e no combate a fraudes, também desafia nossa capacidade de compreender e gerenciar seu impacto. A influência da IA se estende ainda à educação, onde otimiza o aprendizado por meio da personalização de conteúdos; à indústria, onde automatiza processos e reduz custos operacionais; e às finanças, ajudando a combater fraudes e oferecendo produtos personalizados. A presença da IA em diversos setores impulsiona a inovação e permite resolver desafios que antes pareciam intransponíveis. Ver figura 2.

É fascinante perceber como a IA torna tarefas complexas mais acessíveis e otimizadas, seja personalizando conteúdos educativos ou aprimorando experiências de entretenimento. No entanto, sua influência silenciosa também exige atenção redobrada para os efeitos éticos e sociais que podem emergir de sua aplicação, especialmente em setores como saúde e finanças, onde decisões automatizadas têm consequências diretas na vida das pessoas. Isso reforça a importância de um desenvolvimento consciente, que alinhe o progresso técnico às necessidades humanas.

Fig.2. Áreas relacionadas com inteligência artificial



Fonte: Monard; Baranaukas (2000). p.2

E este é apenas o começo. Com o avanço da tecnologia, espera-se que a IA continue abrindo novos horizontes, desde veículos autônomos até descobertas científicas inovadoras, moldando o futuro de formas que ainda estamos começando a imaginar.

## 5 ENVIESAMENTO ALGORITMO E IMPACTOS ADVERSOS DA IA

Embora a inteligência artificial carregue um enorme potencial para transformar positivamente nossas vidas, ela também apresenta desafios éticos que exigem atenção e ações proativas. Entre esses desafios, o viés algorítmico se destaca como uma preocupação central. Pois, se não forem mitigados, esses vieses podem reforçar desigualdades sociais e perpetuar injustiças existentes na sociedade.

De acordo com Awan (2023), o viés algorítmico surge quando os algoritmos são treinados com dados tendenciosos, incompletos ou injustos, ou quando o processo de desenvolvimento exclui grupos sociais específicos. Isso pode ocorrer mesmo de forma não intencional, à medida que sistemas reproduzem padrões discriminatórios presentes nos dados com os quais foram treinados.

O conceito do “efeito Mateus”, derivado da passagem bíblica de Mateus 25:29 que diz: “Pois a todo aquele que tem será dado, e terá em abundância; mas ao que não tem, até o que tem lhe será tirado”, é central na análise de Herzog (2021) sobre o viés algorítmico. Este efeito ilustra a dinâmica de retroalimentação que amplifica as desigualdades existentes, concentrando benefícios e oportunidades em torno daqueles que já os possuem, perpetuando a desvantagem dos grupos marginalizados. No contexto da “justiça estrutural”, o efeito Mateus demonstra como algoritmos de IA, treinados com dados históricos que refletem desigualdades preexistentes, podem perpetuar ou exacerbar essas mesmas desigualdades. A

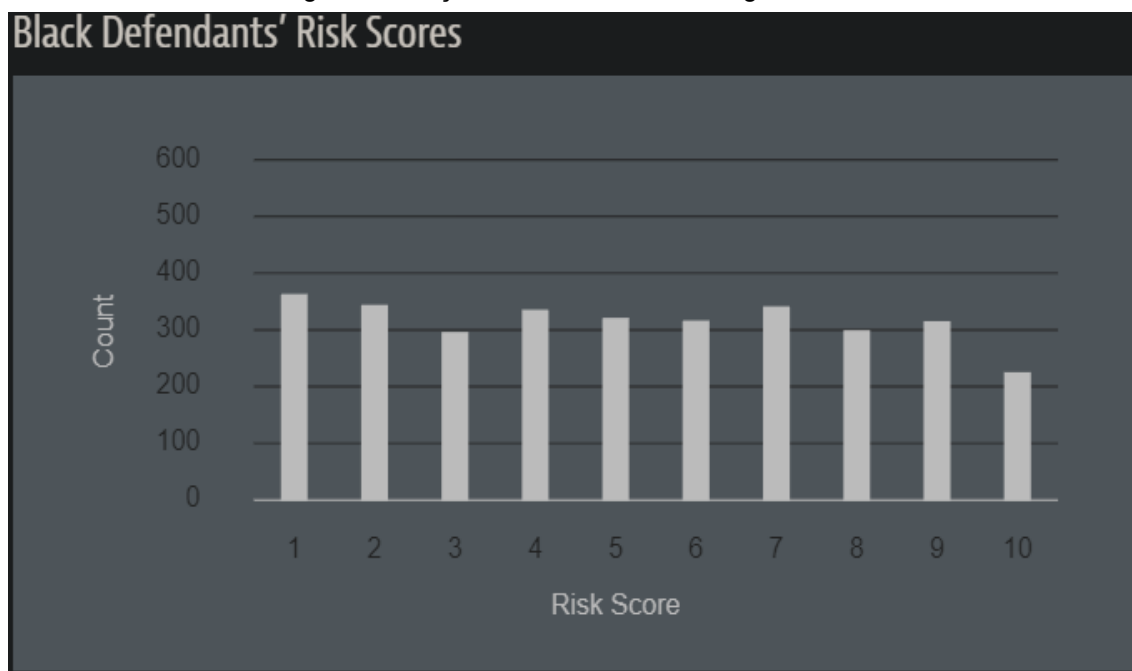
dependência em dados históricos enviesados, frequentemente contendo correlações entre variáveis aparentemente neutras e características protegidas, leva a algoritmos que reproduzem e até mesmo amplificam preconceitos existentes em áreas como acesso ao emprego, crédito, sistema de justiça criminal, entre muitas outras. Conseqüentemente, a utilização de IA sem uma cuidadosa consideração de seus vieses pode aprofundar ainda mais as injustiças sociais.

Pesquisas conduzidas por Hagerty e Rubinov indicam que o impacto social da IA varia entre diferentes regiões do mundo, refletindo as particularidades culturais e sociais locais. Nos Estados Unidos, por exemplo, estudos mostram que tecnologias de IA podem intensificar divisões sociais e agravar desigualdades, especialmente entre grupos marginalizados. Esses padrões de desigualdade tendem a se repetir globalmente, afetando ainda mais os países de baixa e média renda, que podem se beneficiar menos dos avanços tecnológicos e são mais vulneráveis a seus impactos negativos.

O impacto do viés algorítmico se torna especialmente preocupante em áreas sensíveis, como o sistema de justiça criminal. Um exemplo alarmante foi revelado pela ProPublica, que analisou o sistema COMPAS, desenvolvido pela Northpointe (hoje Equivant). Esse sistema, utilizado para avaliar o risco de reincidência de réus nos Estados Unidos, mostrou-se enviesado: pessoas negras eram frequentemente classificadas como alto risco, enquanto pessoas brancas, mesmos reincidentes, eram categorizadas como baixo risco (Vieira, 2019).

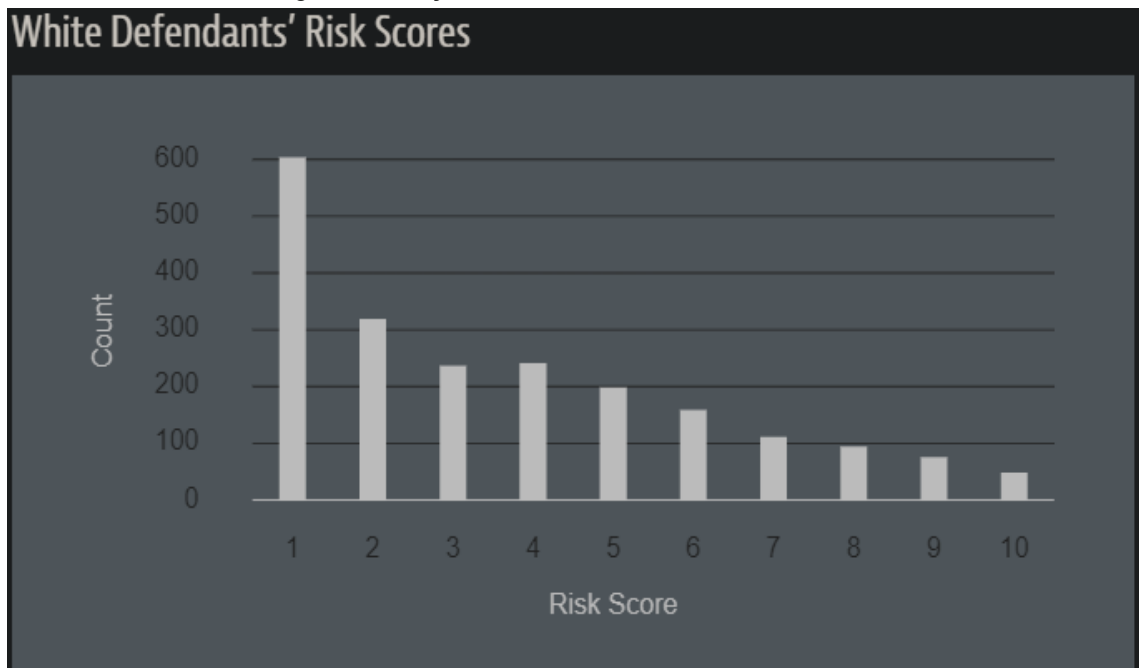
Dieterich, Mendoza e Brennan (2016) destacam que o sistema avalia diversos critérios, como histórico criminal, uso de drogas, nível de educação e meio em que moram. Embora a cor da pele não seja um critério explícito, variáveis indiretas, como bairro e círculo social, podem influenciar as previsões. Segundo a ProPublica, o COMPAS prevê corretamente 61% dos casos, mas negros são classificados como alto risco duas vezes mais que brancos, evidenciando a disparidade racial nas decisões do algoritmo. Ver figuras 3 e 4.

Fig.3. Pontuações de risco dos réus negros



Fonte: Machine Bias; There's software used across the country to predict future criminals. And it's biased against blacks.; Angwin, et al., (2016).

Fig.4. Pontuações de risco dos réus brancos



Fonte: Machine Bias; There's software used across the country to predict future criminals. And it's biased against blacks.; Angwin, et al., (2016).

As Figuras 3 e 4 destacam as discrepâncias nas pontuações de risco atribuídas a réus brancos e negros. Aproximadamente 350 réus negros foram classificados como baixo risco, em comparação com mais de 600 réus brancos na mesma categoria. Por outro lado, ao observar os réus classificados como alto risco, percebe-se uma diferença ainda mais acentuada: cerca de 200 réus negros foram incluídos nesta categoria, enquanto menos de 100 réus brancos receberam a mesma avaliação.

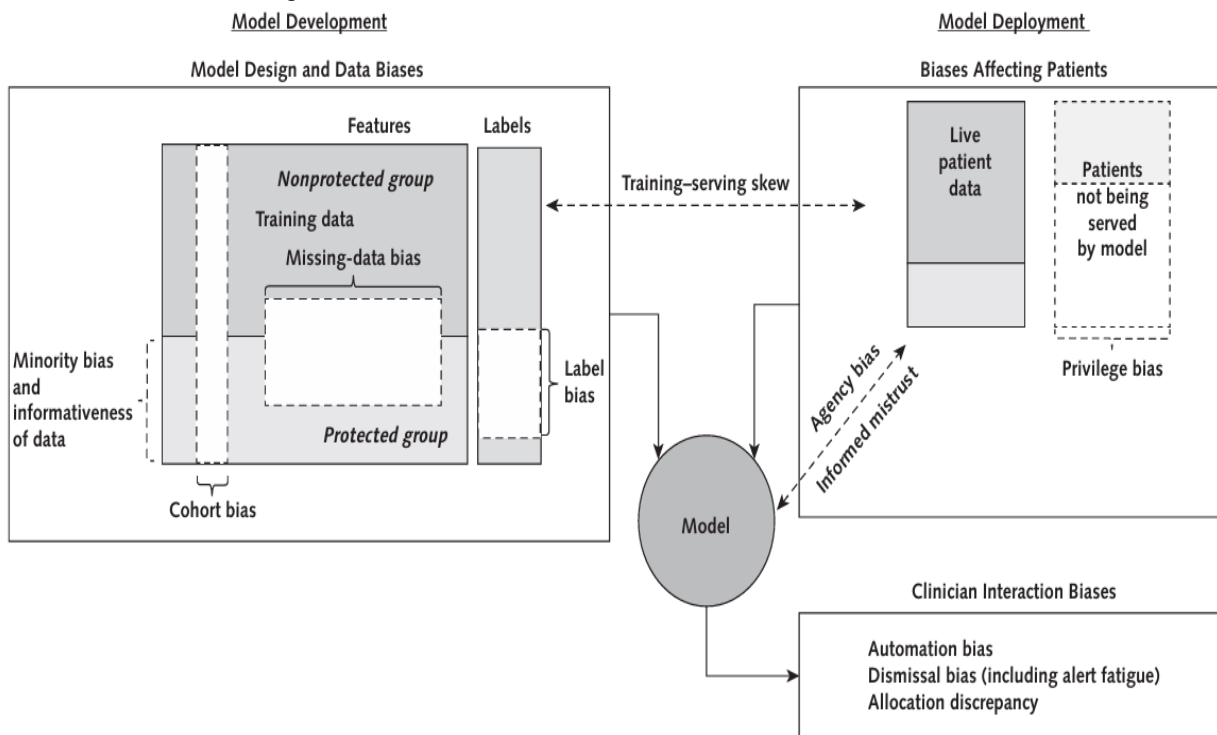
Esses dados evidenciam uma disparidade notável entre as classificações de risco atribuídas aos diferentes grupos raciais, sugerindo um viés algorítmico subjacente no sistema.

Além da justiça, os vieses algorítmicos também afetam a saúde. Panch et al. destacam em seu artigo os vieses algorítmicos e suas causas, de acordo com os autores é importante compreender e mitigar o viés algorítmico em sistemas de saúde. Pois, a falta de clareza na definição de "equidade", a insuficiência de dados representativos e a natureza dos algoritmos de aprendizado profundo são desafios cruciais que precisam ser enfrentados para garantir a justiça e a precisão dos sistemas de IA em saúde.

Rajkomar et al. ilustram o risco de vieses algorítmicos na saúde com um exemplo hipotético: um software de monitoramento para alertar equipes médicas sobre pacientes internados com risco de piora, que requereriam transferência para a UTI em 6 horas. Considerando afro-americanos como grupo protegido neste exemplo, os autores demonstram como o viés pode surgir. Se os dados de treinamento do modelo incluírem poucos pacientes afro-americanos, sua precisão para esse grupo pode ser comprometida. Uma menor sensibilidade pode levar à subdetecção de pacientes com risco, prejudicando-os caso as equipes médicas confiem excessivamente nos alertas, um exemplo de viés de automação. Além disso, um menor valor preditivo positivo para afro-americanos pode resultar em viés de rejeição, uma forma de "fadiga de alerta" onde os médicos passam a desconsiderar alertas para esse grupo por serem mais propensos a falsos positivos. Isso prejudicaria,

desproporcionalmente, os pacientes que realmente necessitam de intervenção. Ver figura 5.

Fig. 5 Estrutura conceitual de como vários vieses se relacionam entre si.



Fonte: Rajkomar et al. Ensuring Fairness in Machine Learning to Advance Health Equity (2018). P.869

A Figura 5 ilustra como os vieses podem surgir e se interrelacionar dentro de um sistema de inteligência artificial. No lado direito da imagem (*Biases Affecting Patients*), observamos os dados atuais dos pacientes, que apresentam uma lacuna representando a ausência de informações sobre aqueles que não pertencem ao grupo favorecido. Quando esses dados são processados pelo modelo (*Model*), o sistema consulta suas instruções de treino e dados de treinamento (*Model Design and Data Biases*) para auxiliar na tomada de decisões.

Contudo, nota-se a presença de uma lacuna adicional (*Missing Data Bias*), resultante da falta de informações provenientes da ausência de dados específicos para aquele grupo de pacientes atualmente internados. Esse déficit compromete a precisão das decisões, levando a potenciais falsos positivos ou negativos. Como mencionado anteriormente, esse tipo de viés pode contribuir para a “fadiga de alerta”, uma situação em que os profissionais de saúde começam a desconsiderar os avisos do sistema devido à imprecisão recorrente. Essa desconfiança pode levar à rejeição de alertas realmente corretos, colocando em risco pacientes que, de fato, necessitam de tratamento especial.

Além das questões sociais e éticas, a IA pode gerar riscos econômicos. Segundo Szczepanski, a concentração de conhecimento e recursos em grandes corporações pode criar “superempresas”, prejudicando a economia global e aumentando a disparidade entre países desenvolvidos e em desenvolvimento. A demanda por profissionais com habilidades específicas deve crescer, enquanto outros trabalhadores podem ser deixados para trás, aumentando a desigualdade e reduzindo salários e arrecadação tributária.

Esses exemplos evidenciam que os vieses algorítmicos representam um desafio significativo para o uso responsável da IA. Se não forem enfrentados com seriedade, esses vieses podem consolidar e ampliar desigualdades já existentes. No entanto, ao abordarmos esses desafios de forma proativa e responsável, podemos garantir que a IA se torne uma força para o progresso social, promovendo um futuro mais justo e igualitário. O desenvolvimento ético da IA não é apenas uma escolha, mas uma responsabilidade coletiva que devemos assumir para moldar um futuro mais inclusivo para todos.

## 6 FORMAS DE AMENIZAR OS VIESES ALGORITMICOS

Os sistemas de aprendizado de máquina atuais, frequentemente dependem de feedback humano para aprender, embora a maioria dos cursos sobre aprendizado de máquina foque principalmente em algoritmos, negligenciando a interação humano-computador.

Em seu livro Munro, apresenta o conceito de "*Human-in-the-Loop Machine Learning*" propondo uma abordagem prática para otimizar o processo completo de aprendizado de máquina, incluindo técnicas como anotação, aprendizado ativo e transfer learning. Métodos como sistemas "*human-in-the-loop*" também ajudam a manter o controle humano sobre decisões sensíveis.

De acordo com Rajkomar et al., para mitigar os riscos do viés algorítmico em sistemas de saúde, é essencial considerar a diversidade e as desigualdades sociais presentes nos dados que alimentam os algoritmos. Nesse sentido, equipes multidisciplinares devem desenvolver algoritmos adaptados ao contexto específico do sistema de saúde em que serão aplicados, assegurando que diferentes necessidades populacionais sejam atendidas. Além disso, a inclusão de profissionais com experiências diversas e a colaboração com especialistas clínicos ajudam a alinhar as soluções tecnológicas às realidades práticas, reduzindo desigualdades preexistentes.

Embora o desafio ainda presente de compreender como algoritmos, especialmente os baseados em aprendizado profundo (*deep learning*), chegam às suas decisões ou previsões em algoritmos complexos, avanços regulatórios, técnicos e educacionais podem construir sistemas de IA mais inclusivos e confiáveis.

Por fim, a capacitação contínua de cientistas de dados sobre questões éticas e a criação de um "Juramento do Cientista de Dados" fortalecem o compromisso com a equidade.

## 7 METODOLOGIA

A metodologia é o conjunto de regras, normas e procedimentos que orientam a condução da pesquisa, fornecendo ferramentas e conhecimentos necessários para coletar, analisar e interpretar dados de forma rigorosa e confiável.

Este projeto se configura como uma pesquisa de natureza básica, cujo objetivo é ampliar o conhecimento sobre os vieses na inteligência artificial e conscientizar sobre seus impactos. Trata-se de uma pesquisa descritiva, uma vez que busca identificar, descrever e apresentar fenômenos já categorizados, analisando como os vieses emergem e se manifestam nos algoritmos de IA.

A abordagem utilizada será qualitativa, combinando a riqueza da análise qualitativa que permitindo compreender percepções e características subjetivas dos

dados, o que facilita a mensuração e identificação de padrões e relações estatísticas.

A pesquisa foi conduzida sob o enfoque hipotético-dedutivo, partindo de uma hipótese inicial sobre os impactos dos vieses em algoritmos de IA e buscando testá-la por meio de dados disponíveis na literatura. A partir de uma análise geral dos estudos e dados, será possível inferir sobre casos mais específicos, confirmando ou refutando a hipótese proposta. Essa análise baseou-se em uma revisão sistemática, estruturada para selecionar, avaliar e sintetizar estudos relevantes sobre o tema.

O procedimento adotado será bibliográfico, com a análise de livros, artigos científicos e publicações relevantes, que constitui a base teórica da investigação. Esses materiais foram avaliados sistematicamente, considerando critérios como relevância, qualidade metodológica e aplicabilidade para o estudo. Eles forneceram o embasamento necessário para compreender a natureza dos vieses e seus efeitos em diversos contextos.

Embora a pesquisa se concentre em uma revisão da literatura, os dados coletados a partir de fontes bibliográficas foram tratados de maneira sistemática, combinando insights qualitativos. Esses procedimentos garantiram uma análise abrangente e fundamentada, permitindo uma compreensão clara das causas e consequências dos vieses na IA, bem como das possíveis soluções para mitigá-los.

## **8 ANÁLISE DE RESULTADOS**

Com a crescente presença da inteligência artificial em nossa vida cotidiana, esta pesquisa foi motivada pela necessidade de identificar e expor os impactos causados pelos vieses algorítmicos. Para isso, foi realizada uma busca sistemática em diversas plataformas acadêmicas, como Google Scholar e ACM Digital Library. Além disso, utilizou-se a técnica de pesquisa em cascata, em que as referências de artigos previamente encontrados foram exploradas para identificar novas fontes de informação relevantes.

Os estudos analisados abordam os desafios sociais, culturais e técnicos associados ao uso da IA. Hagerty e Rubinov, por exemplo, destacam que o impacto social da IA varia entre diferentes regiões do mundo, sendo influenciado por particularidades culturais e sociais locais.

Como explorado anteriormente, a IA se baseia em algoritmos capazes de aprender e tomar decisões com base em padrões extraídos dos dados (TOTVS, 2024). No entanto, o caso COMPAS exemplifica como dados tendenciosos podem levar a decisões enviesadas, refletindo padrões históricos de discriminação, conforme alertado por Dieterich et al. e ProPublica.

O caso COMPAS, explorado por ProPublica, Vieira, Dieterich, Mendoza e Brennan, ilustra de forma concreta como algoritmos podem amplificar desigualdades existentes, demonstrando o impacto real de sistemas enviesados em processos de justiça criminal. Embora, a ProPublica, diga que o COMPAS preveja corretamente 61% dos casos, no gráfico abaixo podemos ver uma comparação, que demonstra que a previsão do software falha de forma diferente para os réus negros. Ver figura 6.

Fig.6. A previsão falha de forma diferente para os réus negros

Prediction Fails Differently for Black Defendants		
	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Fonte: Machine Bias; There's software used across the country to predict future criminals. And it's biased against blacks.; Angwin, et al., (2016).

A Figura 6 apresenta uma tabela que destaca o percentual de previsões errôneas feitas pelo algoritmo (COMPAS) para réus brancos e negros. Entre os indivíduos classificados como alto risco, mas que não voltaram a cometer infrações, os percentuais foram de 23,5% para réus brancos e 44,9% para réus negros. Isso indica que, apesar de serem classificados como de alto risco, uma proporção significativamente maior de réus negros (44,9%) não reincidiu em comparação aos réus brancos (23,5%).

Por outro lado, ao analisar os réus classificados como baixo risco que voltaram a cometer crimes, os percentuais foram de 47,7% para réus brancos e 28% para réus negros. Esses dados mostram que, embora considerados de baixo risco, quase metade dos réus brancos reincidiu, enquanto apenas 28% dos réus negros voltaram a cometer infrações.

Esses resultados evidenciam como o algoritmo apresenta falhas de desempenho que variam entre os grupos analisados, sugerindo a presença de vieses algorítmicos. O modelo demonstra uma precisão desigual, tratando réus brancos e negros de forma inconsistente. Essa disparidade reforça a necessidade de investigações mais aprofundadas sobre o desempenho dos modelos em diferentes grupos demográficos, tema que será abordado nos próximos trechos.

Além de casos reais, a pesquisa também analisou um exemplo hipotético apresentado por Rajkomar et al., que discute o viés em sistemas de monitoramento na área da saúde, sugerindo que a falta de representatividade nos dados de treinamento compromete a precisão de algoritmos, particularmente para minorias, reforçando a necessidade de abordar a 'igualdade de desempenho', conforme destacado nos eixos de justiça distributiva. Esse exemplo foi utilizado como uma ferramenta reflexiva, incentivando o leitor a considerar os impactos éticos e sociais da IA em contextos sensíveis.

A análise desses exemplos permite responder ao questionamento inicial da pesquisa, confirmando que os vieses algorítmicos representam um desafio significativo para o uso responsável da IA. Os resultados obtidos confirmam a hipótese inicial de que os vieses algorítmicos impactam majoritariamente grupos vulneráveis. Os exemplos analisados evidenciam que, sem intervenções proativas, a IA pode perpetuar desigualdades sociais, econômicas e culturais.

Atualmente o desenvolvimento de algoritmos tende a priorizar a performance técnica, muitas vezes à custa de introduzir vieses ou comprometer a qualidade da implementação devido à pressão para o lançamento. Para mitigar esses problemas, sistemas com humanos no processo decisório final (*human-in-the-loop*) podem ser usados como mecanismo de controle.

Em vosso artigo Rajkomar et al., inspirados pelos princípios de justiça distributiva sugerem três eixos centrais que podem auxiliar na redução dos vieses. Eles destacam que esses eixos podem ser conflitantes e que não existe uma solução

única para todos os casos. A escolha da abordagem mais adequada depende do contexto específico e requer um raciocínio ético cuidadoso, envolvendo clínicos, especialistas em aprendizado de máquina e representantes dos grupos afetados. Ver Quadro 1.

Quadro 1: Síntese da “justiça distributiva em aprendizado de máquina” para Rajkomar et al.

<b>Igualdade de Resultados</b>	<b>Igualdade de Desempenho</b>	<b>Igualdade de Alocação</b>
Assegura que grupos protegidos (ex.: minorias étnicas, pacientes de baixa renda) se beneficiem igualmente da implantação de modelos de aprendizado de máquina em termos de resultados de saúde. Isso pode envolver não apenas benefícios semelhantes para todos os grupos, mas também a redução ativa de disparidades existentes.	O modelo deve ter a mesma acurácia e performance para todos os grupos, independente de suas características demográficas. Isso significa que métricas como sensibilidade, especificidade e valor preditivo positivo devem ser semelhantes entre os grupos. Existem técnicas para forçar os modelos a atingirem esse tipo de equidade.	Quando recursos limitados são alocados com base em previsões de um modelo (ex.: direcionamento de intervenções), a alocação deve ser proporcional entre os grupos, levando em consideração possíveis diferenças em fatores relevantes.

Fonte: Autoria própria, baseado no artigo “Ensuring Fairness in Machine Learning to Advance Health Equity” de Rajkomar et al.

O Quadro 1 apresenta os eixos centrais que orientam a justiça distributiva em aprendizado de máquina. No caso COMPAS, a falta de “igualdade de desempenho” é evidente, com previsões menos precisas para réus negros. Já no exemplo de saúde, a “igualdade de alocação” poderia garantir uma distribuição mais justa de recursos, como alertas médicos, reduzindo disparidades nos cuidados.

As propostas de Munro, como o uso de sistemas “*human-in-the-loop*”, e os eixos de justiça distributiva de Rajkomar et al., são ferramentas viáveis para mitigação dos vieses em setores críticos como saúde e justiça. A adoção dessas técnicas, aliada a uma maior representatividade nos dados de treinamento, pode reduzir disparidades e promover um uso mais ético da IA.

A análise apresentada reforça que os vieses algorítmicos não são meras limitações técnicas, mas desafios éticos e sociais que afetam diretamente grupos vulneráveis. Exemplos como o caso COMPAS e os sistemas de saúde ilustram as consequências dessas falhas, enquanto abordagens como os eixos de justiça distributiva e o uso de sistemas ‘*human-in-the-loop*’ destacam caminhos promissores para mitigá-los. Essa discussão evidencia a necessidade de soluções éticas e proativas para garantir que a IA seja utilizada de forma justa e inclusiva.

## 9 CONSIDERAÇÕES FINAIS

Este artigo estabelece uma base sobre a inteligência artificial, abordando os princípios de seu funcionamento, as diferentes formas de aprendizado e principalmente os impactos causados pelos vieses algorítmicos. A análise de casos reais e simulações, permitiu não apenas responder ao questionamento inicial, mas também confirmar a hipótese de que os vieses representam um desafio significativo para o uso responsável da IA. Por fim, a pesquisa destaca potenciais métodos para mitigar esses vieses, evidenciando que estratégias como a justiça distributiva e sistemas "*human-in-the-loop*" podem promover uma IA mais ética e inclusiva.

As considerações finais desta pesquisa reforçam a relevância do tema "Vieses Algorítmicos na Inteligência Artificial" no contexto de um mundo cada vez mais automatizado. A análise realizada confirmou que os vieses nos algoritmos representam um desafio crítico, afetando majoritariamente grupos minoritários e vulneráveis. Exemplos como o caso COMPAS demonstram como algoritmos podem reproduzir e amplificar desigualdades sociais, enquanto análises teóricas, como as de Rajkomar et al., incentivam uma reflexão sobre as implicações éticas da IA em áreas sensíveis como a saúde.

Este trabalho contribui para a literatura, ao destacar a importância de métodos como o "*human-in-the-loop*" e a representatividade dos dados no treinamento de algoritmos. No entanto, a ausência de dados empíricos próprios, limita o escopo da pesquisa.

Para pesquisas futuras, sugere-se a aplicação prática das soluções propostas em setores como a educação e as finanças, ampliando a discussão para contextos ainda pouco explorados. Assim, espera-se que esta pesquisa inspire novas iniciativas voltadas para o desenvolvimento de tecnologias de IA mais justas e inclusivas.

Em última análise, garantir que a IA seja uma força para o progresso humano requer não apenas soluções técnicas, mas também um compromisso ético contínuo de pesquisadores, desenvolvedores e tomadores de decisão.

## REFERÊNCIAS

ANGWIN, J.; LARSON, J.; MATTU, S.; KIRCHNER, L. Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks [online]. ProPublica, 23 mai. 2016. Disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Acesso em: 13 mai. 2024.

Anhanguera. Exemplos de Inteligência Artificial e suas mil e uma aplicações. Artigo online publicado no site da Anhanguera. Disponível em: [https://blog.anhanguera.com/exemplos-de-inteligencia-artificial/#Exemplos de inteligencia artificial na saude](https://blog.anhanguera.com/exemplos-de-inteligencia-artificial/#Exemplos_de_inteligencia_artificial_na_saude). Acesso em: 07 de maio de 2024.

AWAN, Abid Ali. What is Algorithmic Bias? Artigo online publicado no site datacamp. Disponível em: <https://www.datacamp.com/blog/what-is-algorithmic-bias>. Acesso em: 07 maio. 2024.

BERRY, David M. The Limits of Computation: Joseph Weizenbaum and the ELIZA Chatbot. Journal of the Digital Society, v. 3, n. 3, art. w3.3.2, p. 11-06, 11 jun. 2023. Disponível em: <https://ojs.weizenbaum-institut.de/index.php/wjds/article/view/106/96> Acesso em 22 abr. 2024.

Bíblia. (1994, 1995, 2007, 2011). Mateus 25:29. *Almeida Corrigida Fiel*. Sociedade Bíblica Trinitariana do Brasil (SBTB). Disponível em: <https://www.bibliaonline.com.br/acf/mt/25/29>. Acesso em: 10/12/2024.

BISHOP, Christopher M. Pattern recognition and machine learning. Springer google scholar, v. 2, p. 645-678, 2006.

BRANDÃO, Pedro. Alan Turing: da necessidade do cálculo, a máquina de Turing até à computação. Revista de Ciências da Computação, p. 73-88, 2017. DOI: <https://doi.org/10.34627/rcc.v12i0.8>. Disponível em: <http://hdl.handle.net/10400.2/7683>" Acesso em: 22 abr. 2024.

CHARNIAK, Eugene; MCDERMOTT, Drew. A Bayesian Model of Plan Recognition. Massachusetts: Addison-Wesley, 1985.

CHRISLEY, Ronald; BEGEER, Sander (Org.). Artificial Intelligence: Critical Concepts, Volume 2. Edição ilustrada. Taylor & Francis, 2000. 586 p. ISBN: 978-0-415-19331-3.

DIETERICH, William; MENDOZA, Christina; BRENNAN, Tim. COMPAS risk scales: Demonstrating accuracy equity and predictive parity. Northpointe Inc, v. 7, n. 4, p. 1-36, 2016. Disponível em: [http://go.volarisgroup.com/rs/430-MBX-989/images/%20ProPublica\\_Commentary\\_Final\\_070616.pdf](http://go.volarisgroup.com/rs/430-MBX-989/images/%20ProPublica_Commentary_Final_070616.pdf). Acesso em: 13 maio. 2024.

GALLANT, S. (1990). Perceptron-based learning algorithms. IEEE transactions on

neural networks / a publication of the IEEE Neural Networks Council, 1(2), 179-191. DOI: 10.1109/72.80230. Disponível em: <https://www.ling.upenn.edu/courses/cogs501/Gallant1990.pdf> Acesso em 22 abr. 2024

HAGERTY, Alexa; RUBINOV, Igor. Global AI ethics: a review of the social impacts and ethical implications of artificial intelligence. **arXiv preprint arXiv:1907.07892**, 2019. HAUGELAND, John. Artificial Intelligence: The Very Idea. Massachusetts: The MIT Press, 1985.

HERZOG, Lisa. Algorithmic bias and access to opportunities. In: **The Oxford Handbook of Digital Ethics**. Oxford: Oxford Academic, 2021.

IBM. O que é inteligência artificial?. Artigo online publicado no site da IBM Brasil. Disponível em: <https://www.ibm.com/br-pt/topics/artificial-intelligence>. Acesso em: 27 abr. 2024.

IBM. O que é machine learning? Artigo online publicado no site da IBM Brasil. Disponível em: <https://www.ibm.com/br-pt/topics/machine-learning>. Acesso em: 06 de maio de 2024.

KURZWEIL, Ray. The Age of Spiritual Machines. Massachusetts: The MIT Press, 1990.

LU, Jie et al. Transfer learning using computational intelligence: A survey. Knowledge-Based Systems, v. 80, p. 14-23, 2015. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0950705115000179>

MONARD, Maria Carolina; BARANAUKAS, José Augusto. Aplicações de Inteligência Artificial: Uma Visão Geral. São Carlos: Instituto de Ciências Matemáticas e de Computação de São Carlos, 2000. Disponível em: <https://www.academia.edu/download/51841234/49-148-1-PB.pdf> Acesso em 07 de maio de 2024

MUNRO, Rob. Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI. Manning, 2021.

NASTESKI, Vladimir. An overview of the supervised machine learning methods. Horizons. b, v. 4, p. 51-62, 2017.

ONGSULEE, Pariwat. Artificial intelligence, machine learning and deep learning. In: 2017 15th international conference on ICT and knowledge engineering (ICT&KE). IEEE, 2017. p. 1-6.

PANCH, Trishan; MATTIE, Heather; ATUN, Rifat. Artificial intelligence and algorithmic bias: implications for health systems. Journal of global health, v. 9, n. 2, 2019.

POOLE, D.; MACKWORTH, A. K.; GOEBEL, R. Computational Intelligence: A Logical Approach. Oxford: Oxford University, 1998.

RAJKOMAR, Alvin et al. Ensuring fairness in machine learning to advance health equity. *Annals of internal medicine*, v. 169, n. 12, p. 866-872, 2018.

RUSK, Nicole. Deep learning. *Nature Methods*, v. 13, n. 1, p. 35-35, 2016.

SEGOVIA SPADINI, Allan. O que é Inteligência Artificial? Como funciona uma IA, quais os tipos e exemplos. Artigo online publicado no site Alura. Disponível em: <https://www.alura.com.br/artigos/inteligencia-artificial-ia>. Acesso em: 27 abr. 2024.

SUTTON, Richard S.; BARTO, Andrew G. Reinforcement learning: An introduction. MIT press, 2018.

SZCZEPANSKI, M. Economic impacts of artificial intelligence (AI). EPRS: European Parliamentary Research Service, Belgium, 2019. Disponível em: <https://policycommons.net/artifacts/1334867/economic-impacts-of-artificial-intelligence-ai/1940719/>. Acesso em: 07 maio. 2024.

TOTVS. Conheça 13 exemplos de aplicações da inteligência artificial nas empresas. Artigo online publicado no site da TOTVS. Disponível em: <https://www.totvs.com/blog/inovacoes/aplicacoes-da-inteligencia-artificial/>. Acesso em: 07 de maio de 2024.

VIEIRA, Leonardo Marques. A problemática da inteligência artificial e dos vieses algorítmicos: caso COMPAS. In: Brazilian Technology Symposium. 2019. Disponível em: <https://lcv.fee.unicamp.br/images/BTSym-19/Papers/090.pdf>. Acesso em: 13 maio. 2024.

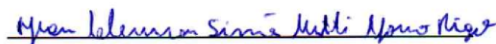


**UNIVERSIDADE DO ESTADO DA BAHIA - UNEB**  
**DEPARTAMENTO DE CIÊNCIAS EXATAS E DA TERRA – CAMPUS II**  
**CURSO: BACHARELADO DE SISTEMAS DE INFORMAÇÃO**  
**COMPONENTE CURRICULAR: TRABALHO DE CONCLUSÃO DE CURSO**

ATA DA SESSÃO DE DEFESA PÚBLICA DE TRABALHO DE CONCLUSÃO DE CURSO, DO CURSO DE BACHARELADO DE SISTEMAS DE INFORMAÇÃO DO SEGUNDO SEMESTRE 2024.

No dia **vinte de dezembro de dois mil e vinte quatro**, às **doze horas**, no auditório do Departamento de Ciências Exatas e da Terra – Campus II, Universidade Estado da Bahia - UNEB, reuniu-se a Banca Examinadora composta pelo(a) professor(a) **Jean Cleverson Simões Mutti Afonso Rego** (presidente da mesa e orientador(a)), professor(a) **Leandro dos Santos Gonzalez** (professor(a) convidado(a)) e professor(a) **Larissa Rocha Soares Bastos** (professor(a) convidado(a)), para avaliar o Trabalho de Conclusão de Curso (artigo acadêmico), do(a) discente **Ruan Tiago Santana de Oliveira** intitulado “**Enviesamento na inteligência artificial: um estudo dos seus impactos**”. O presidente da Banca Examinadora abriu a sessão com os cumprimentos ao(a) candidato(a), aos demais membros da banca, esclarecendo, também, o caráter do evento e respectivas normas. A seguir, foi concedida a palavra ao autor do trabalho para apresentação por vinte minutos. Após esta exposição, os membros da Banca Examinadora realizaram suas considerações emitindo sugestões ao trabalho apresentado e em seguida à palavra foi devolvida ao(a) candidato(a). Após as necessárias considerações ao trabalho, a banca examinadora reuniu-se e o (a) professor(a) **Jean Cleverson Simões Mutti Afonso Rego** atribuiu nota 9,4 (NOVE vírgula quatro -), o(a) professor(a) **Leandro dos Santos Gonzalez** atribuiu nota 9,0 (NOVE -) e o(a) professor(a) **Larissa Rocha Soares Bastos** atribuiu nota 8,0 (oito -). Para registro e finalidades legais, eu **Prof. Jean Cleverson Simões Mutti Afonso Rego**, presidente da banca, lavrei a presente Ata que será assinada por mim e demais membros da Banca Examinadora.

Alagoinhas, 20 de dezembro de dois mil e vinte quatro.



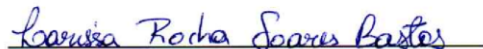
**Prof. Jean Cleverson Simões Mutti Afonso Rego**

Presidente da mesa e orientador(a)



**Prof. Leandro dos Santos Gonzalez**

Professor(a) Convidado(a)



**Prof. Larissa Rocha Soares Bastos**

Professor(a) Convidado(a)