

DISCRIMINAÇÃO E VIOLAÇÕES DE DIREITOS HUMANOS PELO ALGORITMO

DISCRIMINATION AND HUMAN RIGHTS VIOLATIONS BY THE ALGORITHM

Eduardo Tognetti¹

Resumo: o texto aborda como o uso crescente de algoritmos de inteligência artificial em diversas áreas tem gerado preocupações em relação à discriminação e às violações de direitos humanos, especialmente contra grupos vulneráveis. A metodologia utilizada será a da revisão bibliográfica. Discute-se a ideia de que, ao invés de serem neutros, os algoritmos podem reproduzir preconceitos históricos, como racismo e sexismo, refletidos nas próprias bases de dados utilizados para treiná-los. Além disso, exemplifica a perpetuação das desigualdades estruturais por meio de sistemas automatizados em processos como vigilância policial e concessão de benefícios sociais, que afetam desproporcionalmente minorias raciais e classes sociais mais pobres. A solução proposta passa por uma maior transparência e responsabilidade na criação desses sistemas, juntamente com a inclusão de princípios de justiça social desde o seu desenvolvimento.

Palavras-chave: Inteligência artificial; discriminação; violação de direitos humanos.

Abstract

The text discusses how the increasing use of artificial intelligence (AI) algorithms in various fields has raised concerns about discrimination and human rights violations, particularly against vulnerable groups. It highlights that, rather than being neutral, algorithms often reproduce historical biases such as racism and sexism, inherent in the data used to train them. It further illustrates how automated systems perpetuate structural inequalities in processes like police surveillance and social welfare distribution, disproportionately affecting racial minorities and lower social classes. The proposed solution involves greater transparency and accountability in developing these systems, along with the incorporation of social justice principles from their inception.

Keywords: artificial intelligence; discrimination; human rights violations.

INTRODUÇÃO

Nos últimos anos, o uso de algoritmos de inteligência artificial (IA) tornou-se uma prática comum em diversas áreas, incluindo segurança pública, saúde, educação e, sobretudo,

¹ Professor do Centro Universitário das Faculdades Metropolitanas Unidas (desde 2016). Doutor (2015) e Mestre (2012) pela Universidade de São Paulo (USP), sob orientação da Profa. Titular Maria Sylvia Zanella Di Pietro. Especialista em Direito do Estado pela PUC-SP (2009) e em Direito da Economia e da Empresa pela FGV-RJ (1998). Formado em direito pela PUC-RJ (1995). Advogado aposentado do Banco do Brasil (desde 2024). Parecerista do escritório Tognetti Advocacia - Sociedade de Advogados.

no setor privado. Embora as promessas tecnológicas ofereçam eficiência, precisão e soluções para problemas complexos, o uso de algoritmos também tem gerado preocupações profundas, especialmente no que tange à discriminação e violações de direitos humanos. Como consequência, a sociedade se vê diante de um dilema: os avanços tecnológicos podem agravar desigualdades já existentes, reforçando o preconceito e o racismo? Estaríamos repetindo os mesmos erros de Cesar Lombroso?

O presente trabalho aborda algumas das formas pelas quais algoritmos têm sido utilizados de forma discriminatória, contribuindo para a exclusão social de grupos vulneráveis e violando direitos humanos fundamentais, como o direito à igualdade, à privacidade e à não discriminação. A análise é fundamentada na obra de diversos autores e estudos recentes, que discutem como esses sistemas, em vez de serem neutros, podem reproduzir e amplificar preconceitos.

1 O HOMEM DELINQUENTE DE CESAR LOMBROSO

O Homem Delinquente (LOMBROSO, 2001) é uma obra originalmente publicada em 1876 pelo médico e criminologista italiano Cesare Lombroso, considerado um dos pioneiros da criminologia moderna. Seu trabalho é profundamente influenciado pelas teorias do positivismo, que buscam explicações científicas para o comportamento humano, especialmente no campo da criminalidade. O autor parte da premissa de que certos indivíduos possuem predisposições inatas para o crime, ideias que foram controversas à época e que continuam a ser criticadas até os dias atuais.

Na Itália do século XIX, quando Lombroso começou a desenvolver suas pesquisas, a Europa passava por uma transformação científica e social, com o avanço das ciências naturais e o crescimento do interesse por disciplinas como a psiquiatria e a antropologia. Nesse ambiente, Lombroso desenvolveu sua teoria da "atavismo" criminal (LOMBROSO, 2001, p. 266), que sugere que certos criminosos têm traços físicos, ou "estigmas", que os diferenciam do homem comum, remetendo a um estágio primitivo do desenvolvimento humano.

Lombroso argumenta que alguns indivíduos possuem características biológicas que os tornam mais propensos ao comportamento criminoso. Segundo ele, esses traços podem ser identificados através de uma análise minuciosa do corpo, particularmente da cabeça e do rosto. Seu conceito de "atavismo" refere-se a uma suposta regressão evolutiva, em que certas pessoas

apresentariam traços físicos que pertencem a estágios anteriores da evolução humana. Lombroso acreditava que esses indivíduos "atavísticos" eram uma espécie de "homem primitivo" preso em corpos contemporâneos, e que sua natureza violenta e impulsiva os levava a comportamentos delinquentes.

Entre os estigmas físicos descritos por Lombroso, encontram-se características como: mandíbulas proeminentes, arcos superciliares pronunciados, testa recuada, orelhas grandes ou malformadas, nariz achatado, além de braços longos e desproporcionais ao corpo. Essas características, segundo ele, seriam indicativas de uma predisposição ao crime.

Embora o trabalho de Lombroso tenha contribuído para a discussão de que o crime não deveria ser visto apenas como uma questão moral ou de livre-arbítrio, mas como um fenômeno multifacetado que envolvia elementos científicos, a ideia de que o comportamento criminoso poderia ser explicado exclusivamente por fatores biológicos levou a uma estigmatização dos indivíduos com certas características físicas, além de abrir precedentes para gravíssimas práticas discriminatórias.

O grave erro de Lombroso, de generalizar o perfil do criminoso a partir de análises individuais, pode estar sendo repetido pelos mecanismos de inteligência artificial, especialmente os que aprendem em pesquisas de sites na própria Internet, sem a interferência direta de um ser humano, que seja capaz de avaliar o juízo feito pela ótica moral e ética, como veremos.

2 DISCRIMINAÇÃO ALGORÍTMICA: O MITO DA NEUTRALIDADE TECNOLÓGICA

A ideia de que os algoritmos são "neutros" é amplamente contestada. Embora os algoritmos sejam criados para processar dados de maneira objetiva, os próprios dados nos quais eles se baseiam podem conter vieses históricos e sociais, como graves preconceitos de ordem racial. Tal desvio é constatado pela professora catedrática da Universidade da Califórnia em Los Angeles (UCLA), Safiya Umoja Noble (2018), segundo a qual os algoritmos de busca, como o do Google, refletem as desigualdades raciais e de gênero presentes na sociedade.

A autora explica que os algoritmos de busca são estruturados para maximizar o lucro das empresas por meio da publicidade segmentada, o que significa que as identidades e comportamentos dos usuários são transformados em produtos para serem vendidos a

anunciantes. Nesse contexto, grupos marginalizados são particularmente vulneráveis a serem explorados.

Sua pesquisa mostra que, quando os usuários pesquisavam termos como "garotas negras", os resultados frequentemente vinculavam essas mulheres a conteúdo sexualizado. Isso não é uma coincidência. Ao contrário, evidencia como os algoritmos, projetados por seres humanos e treinados em conjuntos de dados tendenciosos, podem reforçar estereótipos prejudiciais.

Outro exemplo discutido no livro envolve o uso de algoritmos em processos decisórios automatizados, como em admissões universitárias, processos seletivos de empregos e vigilância policial. Noble cita estudos que mostram que os algoritmos utilizados nesses contextos frequentemente prejudicam pessoas de cor, reforçando práticas discriminatórias existentes. O racismo algorítmico nesses sistemas não é apenas uma questão de preconceito consciente, mas de estruturas institucionais que codificam e perpetuam essas desigualdades, como afirma a autora:

[...]há um contexto social ausente nas plataformas de mídia digitais, e isso faz falta, particularmente para grupos marginalizados que são representado problemáticamente de forma estereotipa ou pornográfica, para aqueles que são vítimas de bullying, ou para aqueles que são consistentemente vítimas de abuso (...) Os resultados das buscas refletem os valores e normas dos parceiros e dos anunciantes das empresas de busca comercial e frequentemente refletem nossas crenças mais baixas e aviltantes, porque essas ideias circulam tão livremente e com tanta frequência que não normalizadas e extremamente lucrativas. (...) A noção dominante sobre mecanismos de pesquisa misóginos ou racistas são simplesmente um espelho da coletividade. Resultados de pesquisas problemáticos não apenas parecem 'normais' mas também inescapáveis, mesmo que essas ideias tenham sido desmentidas completamente por estudiosos" (NOBLE, 2018, p. 55-56).

A situação se agrava diante do mito da neutralidade da Inteligência Artificial, como afirma Noble:

O status de monopólio do Google, somado a suas práticas algorítmicas de enviar informação em direção aos interesses do capital neoliberal e das elites socioeconômicas dos Estados Unidos, resultou em um conjunto de informação que alega ter credibilidade, mas é na verdade um reflexo de interesses publicitários. (NOBLE, 2018, p. 57).

A discriminação algorítmica é particularmente visível em sistemas de reconhecimento facial. Estudo conduzido por Buolamwini e Gebru (2018) demonstrou que sistemas de reconhecimento facial de grandes empresas de tecnologia, como IBM e Microsoft, apresentam taxas de erro muito maiores ao identificar pessoas negras em comparação com pessoas brancas, especialmente mulheres negras, como se observa:

Using the dermatologist approved Fitzpatrick Skin Type classification system, we characterize the gender and skin type distribution of two facial analysis benchmarks, IJB-A and Adience. We find that these datasets are overwhelmingly composed of lighter-skinned subjects (79.6% for IJB-A and 86.2% for Adience) and introduce a new facial analysis dataset which is balanced by gender and skin type. We evaluate 3 commercial gender classification systems using our dataset and show that darker-skinned females are the most misclassified group (with error rates of up to 34.7%). The maximum error rate for lighter-skinned males is 0.8%. The substantial disparities in the accuracy of classifying darker females, lighter females, darker males, and lighter males in gender classification systems require urgent attention if commercial companies are to build genuinely fair, transparent and accountable facial analysis algorithms. (BUOLAMWINI; GEBRU, 2018, p. 1).

Esse viés coloca essas populações em maior risco de vigilância indevida e até prisões equivocadas, o que representa uma grave violação dos direitos humanos à privacidade e à não discriminação, reforçando o determinismo biológico de Lombroso.

3 O REFORÇO DAS DESIGUALDADES ESTRUTURAIS

Outra crítica comum aos algoritmos é a sua capacidade de reforçar desigualdades estruturais já existentes. Eubanks (2018) argumenta que as ferramentas de IA frequentemente são utilizadas para tomar decisões que afetam negativamente as populações mais vulneráveis, como a distribuição de benefícios sociais e a concessão de empréstimos bancários. A autora menciona que algoritmos usados em agências governamentais norte-americanas frequentemente rejeitam pedidos de assistência social ou determinam quais famílias são mais propensas a cometer fraudes, com base em padrões de dados enviesados.

Eubanks destaca três estudos de caso específicos para ilustrar como a automação impacta negativamente as pessoas marginalizadas. O primeiro em Indiana. A autora analisa a implementação de um sistema automatizado de elegibilidade para benefícios, que resultou na negação em massa de assistência a milhares de cidadãos. "The automation's impacts were devastating for poor and working-class Hoosiers. Between 2006 and 2008, the state of Indiana denied more than a million applications for food stamps, Medicaid, and cash benefits, a 54 percent increase compared to the three years prior to automation" (EUBANKS, 2018, p. 50). Esse sistema, conclui Eubanks, desconsidera as complexidades da vida dos mais pobres, tratando suas necessidades com frieza burocrática.

O segundo caso, ocorrido em Los Angeles, consiste no uso de ferramentas automatizadas de gestão de dados para lidar com a população sem-teto em Los Angeles. Em vez de abordar as causas estruturais da falta de moradia, o sistema de rastreamento é usado para

classificar as pessoas sem-teto com base em seu "merecimento", criando uma hierarquia de quem "merece" ajuda. Mais grave ainda é a sua utilização para criminalização dos sem-teto, como sugere a autora:

Some suspect that all data is being held for other purposes entirely: to surveil and criminalize the unhoused. (...) (EUBANKS, 2018, p. 114).

According to federal data standards, service providers may disclose protected personal information in HMIS to law enforcement 'in response to ... an oral request for the purpose of identifying or locating a suspect, fugitive, material witness or missing person'. The information that the LAPD can access is limited to name, address, date and place of birth, social security number, and distinguishing physical characteristics. But there is no mandatory review or approval process for oral requests. There is no requirement that the information released be limited in scope or specific to an ongoing case. There is no warrant process, no departmental oversight, no judge involved to make sure the request is constitutional (EUBANKS, 2018, p. 115).

No terceiro caso, da Pensilvânia, é feita uma análise do sistema de vigilância infantil, no qual algoritmos são utilizados para prever se uma criança corre o risco de abuso ou negligência. Ela critica esse sistema por penalizar desproporcionalmente as famílias pobres e negras, ao mesmo tempo em que falha em reconhecer os contextos complexos em que elas vivem. (EUBANKS, 2018, p. 127 e seg.).

Eubanks argumenta que os sistemas de automação fazem parte de uma tradição de "pobreza digital", onde as ferramentas tecnológicas são usadas para reforçar a desigualdade. O problema não é a tecnologia em si, mas sim a forma como ela é projetada e implementada, quase sempre visando os mais pobres com mais severidade do que os ricos:

As long as this invisible code is running in the background while we build new automated decision-making tools, the resulting technologies will profile, police, and punish the poor. This doesn't require bad intentions on the part of designers, data scientists, or program administrators. It only requires designing 'in neutral', ignoring the entirely predictable consequences of super-charging already unequal and discriminatory systems. The choice to design a system optimized for neutrality and objectivity is a choice to support the status quo. To build instead a world of justice and abundance, we will need to design for equity, on purpose, brick by brick and byte by byte (EUBANKS, 2018, p. 223).

A autora faz uma grave denúncia dessas ferramentas, que se baseiam em dados históricos que refletem preconceitos raciais e de classe, que inevitavelmente reproduzem esses padrões, tal como Lombroso o fez. Como resultado, comunidades já marginalizadas tornam-se ainda mais vulneráveis, sem acesso a recursos essenciais. Isso não só perpetua a exclusão, mas também exacerba as desigualdades sociais e econômicas, violando o direito fundamental à igualdade e à justiça social.

4 MONITORAMENTO E VIGILÂNCIA

Um aspecto crucial das discussões sobre a discriminação algorítmica é a questão da privacidade. O uso massivo de dados pessoais para alimentar algoritmos, especialmente em contextos como redes sociais e serviços financeiros, levanta preocupações significativas sobre como esses dados são coletados e utilizados. Shoshana Zuboff (2019), professora emérita da Harvard Business School, cunhou o termo "capitalismo de vigilância" para descrever a maneira como empresas de tecnologia monitoram, registram e exploram o comportamento dos usuários para fins comerciais. Segundo a autora, a exploração algorítmica dos dados coloca em risco o direito à privacidade, uma vez que as informações pessoais são coletadas e usadas sem o consentimento explícito dos indivíduos.

No estudo, o capitalismo de vigilância é definido como um sistema econômico que "reclama unilateralmente a experiência humana como matéria-prima gratuita para ser traduzida em dados comportamentais" (ZUBOFF, 2019, p. 9). Empresas como Google, Facebook e Amazon seriam pioneiras nesse modelo de negócio, que começou a tomar forma nos anos 2000, quando o Google descobriu que os dados de usuários podiam ser usados não apenas para melhorar seus serviços, mas também para prever e modificar o comportamento de consumidores. Esses "dados excedentes", segundo a autora, são a base sobre a qual o capitalismo de vigilância se sustenta, gerando um ciclo de coleta incessante de informações sobre todos os aspectos da vida humana.

Zuboff compara o capitalismo de vigilância com as formas anteriores de capitalismo, como o industrial, mas enfatiza que o primeiro é mais invasivo e perigoso. Enquanto o capitalismo industrial explorava o trabalho humano e recursos naturais, o capitalismo de vigilância explora o comportamento humano de maneiras sem precedentes, criando um novo tipo de mercado baseado no controle do futuro. Ela observa que "os capitalistas de vigilância reivindicam um domínio exclusivo sobre nossos dados comportamentais" (ZUBOFF, 2019, p. 12), destacando a unilateralidade desse processo, onde os usuários se tornam, muitas vezes, involuntariamente, parte de um ciclo de extração de dados.

Um dos conceitos centrais do livro é a "espiral de vigilância", na qual a coleta de dados se expande continuamente, ultrapassando fronteiras éticas e legais. A autora mostra como a exploração desses dados não é apenas passiva; ela envolve técnicas ativas para prever e, eventualmente, modificar o comportamento dos usuários. Ao discutir as práticas do Facebook,

ela menciona que a empresa busca "desenvolver mecanismos que podem não apenas prever, mas também influenciar o comportamento em direção a fins comerciais" (ZUBOFF, 2019, p. 134). Isto seria um passo crucial em direção ao que chama de "poder instrumental", onde empresas de tecnologia assumem um papel autoritário em nossas vidas, controlando não apenas o que fazemos, mas também o que pensamos e como agimos.

Zuboff explora como o capitalismo de vigilância impacta diversos aspectos da sociedade, desde a economia até a política. Ela argumenta que a estrutura desse novo sistema econômico cria uma assimetria de poder sem precedentes entre empresas de tecnologia e os cidadãos. Um exemplo claro é o papel que a coleta de dados desempenhou nas eleições políticas e na manipulação da opinião pública, como no caso do escândalo da Cambridge Analytica, que utilizou dados de milhões de usuários do Facebook para influenciar votações, como a eleição presidencial dos EUA em 2016 e o referendo do Brexit no Reino Unido. "Esses eventos mostraram como a tecnologia de vigilância pode ser instrumentalizada para distorcer o processo democrático" (ZUBOFF, 2019, p. 245).

Zuboff também examina a forma como o capitalismo de vigilância exacerba desigualdades. Ao coletar dados em massa, as grandes empresas de tecnologia detêm o poder de decidir quem pode acessar informações, quais anúncios são exibidos para determinados grupos e até mesmo quem tem acesso a serviços essenciais. Este tipo de discriminação algorítmica agrava as divisões sociais, ao passo que os mais vulneráveis são os mais explorados. Segundo Zuboff, "a segmentação de mercado baseada em dados pessoais perpetua preconceitos e desigualdades, à medida que grupos marginalizados são alvos mais frequentes de discriminação digital" (2019, p. 198).

A vigilância quando aplicada de maneira discriminatória, afeta desproporcionalmente certos grupos sociais. Em países como os Estados Unidos, por exemplo, dados sobre a movimentação e os comportamentos de comunidades negras e latinas são frequentemente utilizados para alimentar sistemas de monitoramento policial, exacerbando a violência policial e a criminalização dessas comunidades (O'NEIL, 2016).

O livro de O'Neil desafia a crença de que os algoritmos são imparciais. Ela afirma que, como esses modelos são criados por pessoas, eles refletem inevitavelmente os preconceitos e prioridades de seus criadores. Um exemplo claro é o uso de algoritmos de policiamento

preditivo, que muitas vezes direcionam a vigilância para bairros de baixa renda, perpetuando ciclos de criminalização, como segue:

Reading police chief William Heim had to figure out how to get the same or better policing out of a smaller force. So in 2013 he invested in crime prediction software made by PredPol, a Big Data start-up based in Santa Cruz, California. The program processed historical crime data and calculated, hour by hour, where crimes were most likely to occur. The Reading policemen could view the program's conclusions as a series of squares, each one just the size of two football fields. If they spent more time patrolling these squares, there was a good chance they would discourage crime. (O'NEIL, 2016, p. 76-77).

Além de serem tendenciosos, esses algoritmos frequentemente operam sem transparência ou supervisão, acabando por atingir majoritariamente negros e hispânicos, que vivem em bairros de mais baixa renda, área tidas como de maior ocorrência de crimes, como segue:

These nuisance crimes are endemic to many impoverished neighborhoods. In some places police call them antisocial behavior, or ASB. Unfortunately, including them in the model threatens to skew the analysis. Once the nuisance data flows into a predictive model, more police are drawn into those neighborhoods, where they're more likely to arrest more people. After all, even if their objective is to stop burglaries, murders, and rape, they're bound to have slow periods. It's the nature of patrolling. And if a patrolling cop sees a couple of kids who look no older than sixteen guzzling from a bottle in a brown bag, he stops them. These types of low-level crimes populate their models with more and more dots, and the models send the cops back to the same neighborhood.

This creates a pernicious feedback loop. The policing itself spawns new data, which justifies more policing. And our prisons fill up with hundreds of thousands of people found guilty of victimless crimes. Most of them come from impoverished neighborhoods, and most are black or Hispanic. So even if a model is color blind, the result of it is anything but. In our largely segregated cities, geography is a highly effective proxy for race. (O'NEIL, 2016, p. 78).

Aqui o “atavismo” de Lombroso atinge uma dimensão geográfica, mas igualmente busca identificar a ocorrência de futuros delitos a partir da generalização de fatos relacionados à violência, sob uma suposta neutralidade.

5 TRANSPARÊNCIA E RESPONSABILIDADE: DESAFIOS ÉTICOS E JURÍDICOS

Uma das maiores barreiras à justiça em casos de discriminação algorítmica é a falta de transparência. Muitos dos algoritmos utilizados por empresas e governos são considerados "caixas-pretas", já que operam com pouca ou nenhuma supervisão pública (CASTELVECCHI, 2016). Sem acesso aos critérios e decisões dos algoritmos, é quase impossível para as pessoas afetadas contestarem as decisões tomadas por essas tecnologias.

Diante desse cenário, torna-se necessária a criação de regulamentos específicos que exijam transparência e auditorias regulares dos sistemas de IA. Além disso, é fundamental que haja responsabilização legal para os desenvolvedores e utilizadores desses sistemas, quando comprovado que eles resultam em discriminação e violação de direitos. A proteção dos direitos humanos, neste contexto, requer não apenas políticas técnicas, mas também jurídicas e éticas, que garantam que os algoritmos estejam alinhados com os princípios fundamentais da dignidade humana.

6 INICIATIVAS E POSSÍVEIS SOLUÇÕES

Diante dos riscos apresentados pelos algoritmos, várias iniciativas têm surgido com o intuito de mitigar os impactos negativos dessas tecnologias. Uma das soluções mais promissoras é a incorporação de princípios de justiça social e equidade na fase de desenvolvimento dos algoritmos. Neste contexto, Reuben Binns (2018) aborda como as teorias de justiça da filosofia política podem ajudar a conceituar a equidade no aprendizado de máquina. O autor parte da ideia de que o problema da discriminação algorítmica, ou seja, quando os modelos de aprendizado de máquina resultam em decisões que afetam desigualmente grupos protegidos, como aqueles definidos por raça, gênero ou religião. Ele observa que, se houver padrões de discriminação no mundo real, os algoritmos podem acabar replicando essas desigualdades. Isso levanta a necessidade de definir a equidade nos sistemas de aprendizado de máquina de uma forma que possa ser operacionalizada.

O autor aborda várias definições de justiça em aprendizado de máquina que têm sido propostas na literatura recente. Essas abordagens tentam medir se os grupos protegidos são tratados de maneira diferente em relação aos não protegidos. Essas medidas podem ser limitadas, pois muitas vezes falham em capturar nuances importantes. Diferentes métricas de justiça podem ser mutuamente exclusivas, exigindo escolhas difíceis por parte dos programadores, como segue:

Ideally, a system might be expected to meet all of these intuitively plausible measures of fairness. But, somewhat problematically, certain measures turn out to be mathematically impossible to satisfy simultaneously except in rare and contrived circumstances Kleinberg et al. (2016), and therefore hard choices between fairness metrics must be made before the technical work of detecting and mitigating unfairness can proceed. (BINNS, 2018, p.2).

Binns faz uma conexão entre as discussões sobre justiça na filosofia política e as questões emergentes no AM. Ele argumenta que muitas definições de justiça usadas no AM

refletem debates filosóficos antigos, como as ideias de igualitarismo e meritocracia. Ele sugere que a filosofia política pode esclarecer os debates sobre a justiça nos sistemas automatizados, fornecendo uma base mais robusta para lidar com as diferentes concepções de equidade. Por exemplo, o igualitarismo, uma corrente que defende a igualdade de tratamento, tem implicações diretas no AM. Binns escreve: “Fairness as used in the fair machine learning community is best understood as a placeholder term for a variety of normative egalitarian considerations” (BINNS, 2018, p. 2). Isso significa que, na maioria das vezes, as discussões sobre justiça nos algoritmos são, na verdade, reflexos de debates sobre como reduzir desigualdades entre grupos.

CONCLUSÃO

A discriminação algorítmica representa um desafio urgente no campo dos direitos humanos, e sua resolução requer uma abordagem multidisciplinar que envolva tecnologia, direito, ética e políticas públicas. Enquanto algoritmos têm o potencial de contribuir para uma sociedade mais eficiente e inclusiva, sem a devida supervisão e regulamentação, eles correm o risco de perpetuar e ampliar as desigualdades sociais, incorrendo no mesmo erro de Lombroso. É imperativo que desenvolvedores, governos e a sociedade civil trabalhem juntos para criar sistemas que sejam justos e transparentes, garantindo que a tecnologia sirva aos interesses de todos os seres humanos, e não apenas de uma minoria privilegiada.

REFERÊNCIAS

- BINNS, Reuben. Fairness in machine learning: lessons from political philosophy. In: **Proceedings of Machine Learning Research**, v. 81, p. 1-15, 2018. Disponível em: <https://proceedings.mlr.press/v81/binns18a/binns18a.pdf>. Acesso em 18 Out. 2024.
- BUOLAMWINI, Joy; GEBRU, Timnit. Gender shades: intersectional accuracy disparities in commercial gender classification. **Proceedings of Machine Learning Research**, v. 81, p. 1-15, 2018. Disponível em <https://proceedings.mlr.press/v81/buolamwini18a.html#:~:text=Related%20Material-Download%20PDF,-Supplementary%20PDF>. Acesso em 10 Out. 2024.
- CASTELVECCHI, Davide. Can we open the black box of AI? **Nature**. v. 538, n. 7623, p. 20-23, 2016. Disponível em: <https://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731> . Acesso em 10 Out. 2024.
- CRAWFORD, Kate; CALO, Ryan. There is a blind spot in AI research. **Nature**, v. 538, n. 7625, p. 311-313, 2016. Disponível em: <https://www.nature.com/articles/538311a> . Acesso em 8 Out. 2024.

EUBANKS, Virginia. **Automating inequality**: how high-tech tools profile, police, and punish the poor. New York: St. Martin's Press, 2018. [Edição Kindle].

LOMBROSO, César. **O homem delinqüente**. Porto Alegre: Ricardo Lenz, 2001.

NOBLE, Safiya Umoja. **Algorithms of oppression**: how search engines reinforce racism. New York: NYU Press, 2018. [Kindle]

O'NEIL, Cathy. **Weapons of math destruction**: how big data increases inequality and threatens democracy. New York: Crown Publishing Group, 2016. Disponível em:
https://edisciplinas.usp.br/pluginfile.php/7574239/mod_resource/content/1/%28FFLCH%29%20LIVRO%20Weapons%20of%20Math%20Destruction%20-%20Cathy%20ONeal.pdf.
Acesso em 2 Mar. 2024.

ZUBOFF, Shoshana. **A era do capitalismo de vigilância**: a luta por um futuro humano na nova fronteira de poder. Rio de Janeiro: Intrínseca, 2020.