

ALGORITMIZAÇÃO DA JUSTIÇA CRIMINAL: UMA ANÁLISE DO APLICATIVO COMPAS E SEUS VIESES

ALGORITHMIZATION OF CRIMINAL JUSTICE: AN ANALYSIS OF THE SOFTWARE COMPAS AND ITS BIASES

Acesse o QR Code para ler
este artigo em Visual Law



VICTOR SAMPAIO GONDIM

Mestre em Direito Constitucional (2021) pela Universidade de Fortaleza. Especialista (LLM) em Direito Empresarial (2014) pela Fundação Getúlio Vargas-RJ. Graduado em Direito (2011) pela Universidade Federal do Ceará. Advogado e membro da Comissão de Direito da Tecnologia da Informação da OAB-CE.
Lattes: [lattes.cnpq.br/5137636318659623].
ORCID: [orcid.org/0000-0002-8393-4962].
vsgondim@gmail.com

GUSTAVO RAPOSO PEREIRA FEITOSA

Doutor em Ciências Sociais (2005) pela Universidade Estadual de Campinas. Mestre em Sociologia (2000) e graduado em Direito (1997) pela Universidade Federal do Ceará. Professor Titular do Programa de Pós-Graduação em Direito Constitucional e do Centro de Ciências Jurídicas da Universidade de Fortaleza. Professor adjunto de Direito Processual Civil na UFC. Coordenador do Mestrado Profissional em Direito e Gestão de Conflitos da Universidade de Fortaleza.
Lattes: [lattes.cnpq.br/4608410457474387].
ORCID: [orcid.org/0000-0002-3766-0112].
DOI: [doi.org/10.54415/rbccrim.v188i188.93].
gfeitosa@terra.com.br

Recebido em: 19.07.2021

Aprovado em: 23.11.2021

Última versão dos autores: 06.12.2021

ÁREAS DO DIREITO: Processual; Penal

RESUMO: O presente estudo analisa o funcionamento do *software* COMPAS, aplicativo que concede notas a infratores penais para subsidiar decisão judicial sobre a prisão ou liberdade provisória, referência a partir da qual se busca descrever como se dá atualmente a relação entre a

ABSTRACT: The present study analyzes the functioning of COMPAS, an application that awards notes to criminal offenders to support the judicial decision on arrest or provisional release, a reference from which it seeks to describe the relationship between artificial intelligence and law,

inteligência artificial e o Direito, explicar o que é a justiça algorítmica e avaliar a existência de vieses, mesmo em sistemas supostamente objetivos e neutros. Utiliza-se metodologia de investigação do tipo bibliográfica-documental, com pesquisa pura de abordagem qualitativa e quantitativa, descritiva e exploratória quanto aos objetivos. Conclui-se que o uso de algoritmos exigirá familiaridade com conceitos da lógica, da matemática e da estatística para fundamentar ou se contrapor às decisões, que é possível observar vieses nos softwares ainda que respeitem os parâmetros da justiça algorítmica, e que a reprodução desses vieses em diferentes cenários indica que essas distorções estão fora dos algoritmos, podendo representar uma manifestação de racismo estrutural.

PALAVRAS-CHAVE: Decisão judicial – Inteligência artificial – Justiça algorítmica – Viés racial.

explain what is algorithmic justice and evaluate the existence of biases, even in supposedly objective and neutral systems. The methodology used was bibliographic-documentary investigation, with qualitative, quantitative, descriptive, and exploratory approach regarding the objectives. The conclusion is that the use of algorithms will require familiarity with concepts of logic, mathematics and statistics to substantiate or oppose decisions, that can be observed bias in software even if they respect the parameters of algorithmic justice, and that the reproduction of these biases in different scenarios indicates that these distortions are outside the algorithms and may represent a manifestation of structural racism.

KEYWORDS: Judicial decision – Artificial intelligence – Algorithmic justice – Algorithmic fairness – Racial bias.

SUMÁRIO: 1. Introdução. 2. Inteligência artificial e aprendizado de máquina. 3. O aplicativo COMPAS. 4. Os vieses do algoritmo. 5. A justiça algorítmica. 6. Precisão e vieses para além dos algoritmos. 7. Considerações finais. 8. Referências.

1. INTRODUÇÃO

A rotina de juízes, promotores e policiais nos Estados Unidos envolve avaliação diária sobre relaxamento ou não de prisões em flagrante, fixação de condições para a soltura de réus que aguardam julgamento e concessão de fiança. Comumente, os parâmetros utilizados nessas análises sofrem críticas por serem subjetivos ou anti-isonômicos, como no caso da libertação mediante fiança em dinheiro, o chamado *cash bail*, cujo valor sobrecarrega especialmente acusados pobres e negros.

De maneira inovadora e polêmica, um condado norte-americano buscou superar esses aspectos e passou a classificar a periculosidade de um indivíduo e a sua “propensão” à delinquência e à reincidência por meio de algoritmos de análise de risco. Em tese, o sistema permitiria a aplicação de fatores mais claros e isentos nas decisões que tratassem de liberdade de locomoção e estabelecimento de fiança. Este cenário, que poderia ornamentar um filme de ficção científica,

há quase duas décadas é realidade no estado de Nova Iorque, por intermédio de um projeto piloto que foi estendido para todo o seu território (exceto a cidade de mesmo nome) em 2010 (ANGWIN et al., 2016; LANSING, 2012).

Trata-se de apenas um dos exemplos que demonstram as rápidas transformações pelas quais o Poder Judiciário tem passado decorrentes da adoção de inovações tecnológicas. Consolidaram-se ao longo da última década, sob a promessa de maior eficiência e comodidade, a implantação maciça do processo eletrônico, a utilização de *softwares* de inteligência artificial e a tomada de decisões estratégicas a partir da análise de grandes massas de dados (o *big data*). Nasce aí um novo paradigma digital que se apresenta para os profissionais do Direito e para os usuários da Justiça.

Tais mudanças implicam melhorias, mas também perdas. Questões sobre privacidade, transparência, acessibilidade, justiça, custos, equidade, discriminação, entre outras, surgem e suscitam dúvidas, sem que se tenham soluções ou respostas apropriadas para os desafios que se apresentam. A cultura do *fail fast*, oriunda do Vale do Silício, que sugere colocar em prática as inovações de modo rápido e corrigir suas falhas durante o uso, parece influenciar até a tradicional administração judiciária, que não deixa de compartilhar com toda a sociedade a avidez por se inserir no atual movimento de modernização tecnológica.

Os novos impasses resultantes da tentativa de solução de antigos problemas exigem, muitas vezes, conhecimentos e competências que extrapolam o Direito. É o caso da inteligência artificial, cujo funcionamento envolve lógica, matemática e estatística, disciplinas sobre as quais o profissional do Direito não costuma sequer ter noções básicas. Na América Latina, tais *softwares* vêm sendo utilizados essencialmente em atividades burocráticas e administrativas, como triagem de processos e indicação de possíveis precedentes aplicáveis, aplicações ainda introdutórias se considerado o potencial desse tipo de ferramenta.

No entanto, propagam-se pelos Estados Unidos experiências com aplicativos cujo emprego se dá no momento da prestação jurisdicional, como o COMPAS. O programa, desenvolvido por uma empresa privada, concede notas para o risco de recidiva dos infratores detidos e seus resultados podem ser utilizados para fundamentar prisões preventivas. Recentemente, o algoritmo foi acusado de ter um viés racista, sendo mais rigoroso com negros e leniente com brancos (ANGWIN et al., 2016).

A iminência da disseminação de programas do tipo nos sistemas judiciários e as muitas incertezas que eles provocam tornam relevante e premente o debate acerca da inteligência artificial no âmbito do Judiciário. O presente artigo utiliza o modelo de funcionamento do *software* COMPAS como pano de fundo para

a descrição da relação entre a inteligência artificial e o Direito. Pretende-se assim explicar o que é a “justiça algorítmica” e avaliar a existência de vieses na aplicação da lei penal, mesmo em sistemas supostamente objetivos e neutros. Questiona-se como se dá a abordagem técnica da justiça em aplicativos como o COMPAS, sob a hipótese de que os procedimentos adotados não são suficientes para evitar a introdução de preconceitos e vieses nos resultados.

A pesquisa engloba a revisão da literatura sobre o tema e adota fontes bibliográficas e documentais, a exemplo de livros, artigos em periódicos, teses e dissertações disponíveis nas bases BDTD, SciELO, EBSCO, SSRN, JSTOR, ResearchGate, Google Scholar e Portal Periódicos CAPES, entre outras, somadas a dados e informações disponibilizados em sites jornalísticos e de instituições públicas e privadas da área jurídica. O estudo tem caráter descritivo e exploratório, por meio de abordagem qualitativa e quantitativa e de natureza teórica e empírica. Foram utilizadas como palavras-chave para buscas as expressões “COMPAS”, “algorithmic justice”, “algorithmic fairness”, “artificial intelligence” e afins, agregadas aos termos “law” e “justice”, além de seus correspondentes em português, para, em seguida, selecionar os resultados conforme a adequação ao escopo da pesquisa.

2. INTELIGÊNCIA ARTIFICIAL E APRENDIZADO DE MÁQUINA

A inteligência artificial e o aprendizado de máquina, assim como uma internet onipresente e sensores cada vez menores, mais potentes e mais baratos, caracterizam a denominada Quarta Revolução Industrial, iniciada na virada para o século XXI, segundo Klaus Schwab (2016). O autor compreende que o mundo passará por uma transformação sistêmica e profunda, a partir de uma revolução integral promovida pela tecnologia e pela digitalização.

Richard Susskind (2013) assenta que é ao mesmo tempo excitante e desconcertante vislumbrar que não há limites para a tecnologia digital e notar que ela não apenas automatiza, mas também inova, permitindo a realização de tarefas que antes eram impossíveis ou mesmo inimagináveis. Em paralelo, aponta que juristas precisam ter a mente aberta para vivenciar esta época de mudança tecnológica sem precedentes e inovar na prática jurídica de modo irreproduzível no passado, como por meio do uso da inteligência artificial. Um relevante ponto colocado para o profissional do Direito seria o de adotar cedo novos sistemas, além de identificar e aproveitar as oportunidades surgidas a partir das tecnologias emergentes. Cabe, portanto, compreender melhor no que consistem essas tecnologias.

As definições de inteligência artificial (*artificial intelligence* ou AI), para Stuart Russell e Peter Norvig (2010), historicamente variam entre duas dimensões, raciocínio (emulação de pensamento) e comportamento (emulação de conduta), e duas medidas de resultado, o desempenho humano (averiguado mediante comparação empírica) e o desempenho ideal (obtido conforme a lógica racional). A combinação desses fatores resultaria em quatro abordagens diversas: máquinas que pensam como humanos, agem como humanos, pensam racionalmente e/ou agem racionalmente.

Osonde Osoba e William Welser IV (2017) descrevem que a noção de inteligência artificial começa a partir dos algoritmos – uma sequência finita de instruções precisas implementável em sistemas computacionais. Os pioneiros da inteligência artificial teriam buscado empoderar os sistemas computacionais com o dom do intelecto, formulando algoritmos que pudessem aprender ou adaptar-se conforme a experiência (isto é, a sua exposição a dados). Os esforços seguiram no sentido de produzir algoritmos que atualizassem seu comportamento iterativamente, com base nos dados analisados pelos próprios sistemas e nas suas métricas de performance, permanecendo como grande desafio a capacidade de rivalizar com ou exceder a inteligência humana (OSOBA; WELSER IV, 2017).

Denota-se que o conceito da inteligência artificial está ligado à aprendizagem de máquina ou *machine learning*, definida por Peter Flach (2012) como o estudo sistemático de algoritmos e sistemas que melhoram seu conhecimento ou performance por meio da experiência. De acordo com Dan McQuillan (2018), esse “aprendizado” significa que, ao ser alimentado com uma grande massa de dados (*input*) e seus respectivos resultados esperados (*output*), o sistema pode encontrar padrões reproduzíveis e, com isso, passar a prever resultados para um dado novo e desconhecido. Ressalta-se que o conhecimento gerado pela máquina estaria intrinsecamente ligado aos mecanismos computacionais específicos e aos dados utilizados no treinamento.

Com essas técnicas, David Vladeck (2014) acredita que, de certa forma, começa a existir uma parceria entre humanos e máquinas extremamente inteligentes em prol da realização de tarefas extraordinárias. No campo do Direito, destacam-se o ROSS, sistema de inteligência artificial voltado para advogados baseado no Watson da IBM (ELEMENTARY, 2015)¹; o Projeto Victor, derivado de

1. A plataforma ROSS teve suas atividades encerradas em janeiro de 2021 em virtude de um litígio judicial (ARRUDA; OVBIAGELE; DAL'OGGIO, 2020).

uma parceria entre o Supremo Tribunal Federal (2018) e a Universidade de Brasília; o Prometea, na Argentina (CORVALÁN, 2018); e o COMPAS, nos Estados Unidos. Este último será analisado mais detidamente na seção seguinte.

3. O APLICATIVO COMPAS

O aplicativo COMPAS é um instrumento de avaliação de riscos, desenvolvido pela empresa Northpointe Institute for Public Management (hoje denominada Equivant) e baseado em sistemas utilizados no mercado de seguros. Seu nome é um acrônimo formado a partir do termo *Correctional Offender Management Profiling for Alternative Sanctions*². O programa funciona atribuindo pontuação aos infratores numa escala de possibilidade de reincidência e serve para auxiliar decisões de autoridades judiciárias, tornando-se conhecido a partir de seu emprego em algumas localidades nos Estados Unidos (NORTHPOINTE, 2015).

A adoção desse *software* vincula-se ao contexto da aplicação do modelo *Risk-Need-Responsivity* ou Risco-Necessidade-Responsividade (RNR) em países anglófonos como Reino Unido, Estados Unidos, Canadá e Nova Zelândia (ANDREWS; BONTA; WORMITH, 2006; NORTHPOINTE, 2015; POLASCHEK, 2012). Trata-se de um modelo de reabilitação baseado num arcabouço teórico que descreve tanto as causas principais de comportamento delituoso persistente como princípios amplos de combate ao envolvimento na criminalidade (POLASCHEK, 2012). Um destes princípios é o do risco, segundo o qual o comportamento criminoso poderia ser previsto e o nível de tratamento dado ao infrator deveria ser compatível com o seu grau de risco, ou seja, pessoas mais “perigosas” demandariam um tratamento mais intenso para que não reincidissem, enquanto pessoas que oferecem “risco mais baixo” requereriam intervenção mínima ou mesmo nenhuma (ANDREWS; BONTA, 2010).

O sistema já avaliou mais de um milhão de infratores desde sua criação em 1998 até 2018 (DRESSSEL; FARID, 2018) e considera fatores mediadores, de risco e de necessidade estabelecidos por pesquisas científicas como causas de reiteração penal. Ele calcula quatro diferentes modelos de equação de risco que sopesam quatro dimensões de risco: a violência, a recidiva, o não comparecimento (perante as cortes judiciais) e as falhas sociais (BRENNAN; DIETERICH; EHRET, 2007).

2. Perfil de gestão penitenciária de infratores para sanções alternativas, em tradução livre.

GONDIM, Victor Sampaio; FEITOSA, Gustavo Raposo Pereira. Algoritmização da justiça criminal: uma análise do aplicativo Compas e seus vieses. *Revista Brasileira de Ciências Criminais*. vol. 188. ano 30. p. 447-470. São Paulo: Ed. RT, fevereiro 2022. DOI: [doi.org/10.54415/rbccrim.v188i188.93].

O aplicativo faz uso de escalas³ divididas em duas categorias: as de necessidade, que buscam descrever o infrator, e as de risco, que tentam prever os infratores que reincidirão no crime. Tais escalas são desmembradas em 10 etapas, cada uma contemplando cerca de 10% dos resultados, do menor para o maior, e posteriormente convertidas em uma nota de um (mais baixo risco de reincidência) a 10 (mais alto risco). Uma nota 1 indica que a pontuação está na primeira etapa (10% mais baixos) de todas as pontuações do grupo normativo; uma nota 2 indica que a pontuação está na segunda etapa (acima dos 10% e abaixo dos 20%) do grupo normativo; e assim por diante.

Em geral, as notas são divididas em baixo risco (1-4), médio risco (5-7) e alto risco (8-10) em relação a outros infratores do grupo normativo. O grupo normativo resulta de aproximadamente 30.000 avaliações realizadas entre janeiro de 2004 e novembro de 2005 acerca de casos de detenção, reclusão, liberdade provisória e livramento condicional nos Estados Unidos, contemplando 7.381 infratores (NORTHPOINTE, 2015). Por questões comerciais, a empresa não revela os detalhes do seu algoritmo de avaliação (CORBETT-DAVIES et al., 2016; DRESSEL; FARID, 2018).

4. OS VIESES DO ALGORITMO

A emissão da nota por um *software*, sem intervenção humana aparente, pode trazer a sensação de isenção e impessoalidade num primeiro momento. Percepções desse tipo são estimuladas pelas grandes plataformas de tecnologia, que costumam propagandear uma suposta neutralidade, segundo Anupam Chander e Vivek Krishnamurthy (2018), colaborando para formar um senso comum equivocado sobre o tema. Essa visão superficial tem sido confrontada sob aspectos diversos, entre os quais a atuação carregada de subjetivismo na elaboração dos *softwares* e na produção e exame dos dados que alimentam as máquinas, de acordo com Lyria Moses e Janet Chan (2014).

Para o primeiro caso, Frank Pasquale (2015) registra que os parâmetros de mineração de dados, os métodos de tomada de decisão e os modelos de predição

3. O sistema utiliza as 21 escalas a seguir: Risco de reincidência violenta, risco de reincidência em geral, risco de soltura pré-julgamento, envolvimento criminal, histórico de inadequação, histórico de violência, violência atual, parceiros no crime, abuso de substâncias, pobreza/problemas financeiros, problemas educacionais/vocacionais, pensamento criminal, criminalidade na família, problemas sociais, lazer e recreação, instabilidade residencial, problemas de ajuste social, falhas de socialização, oportunidade criminal, personalidade criminal e isolamento social.

são definidos por engenheiros de *software*. Logo, cada passo do desenvolvimento de programas estaria impregnado de vieses e valores humanos – o que poderia servir para impulsionar discriminação preexistente. No segundo, Osonde Osoba e William Welser IV (2017) assentam que a qualidade de um algoritmo depende essencialmente dos dados que lhe serviram de apoio para o aprendizado. A geração de tais dados, por sua vez, decorreria de manifestações sociais imbuídas de tendências e preconceitos. Logo, uma análise algorítmica procedimentalmente correta de informações com essas características levaria a inteligência artificial a reproduzir os mesmos preconceitos encontrados nos dados processados (OSOBA; WELSER IV, 2017).

O fenômeno é denominado pelos autores como o paradoxo da atuação artificial (*the paradox of artificial agency*). O conceito resulta da observação empírica de que os sistemas artificiais, ao aprender autonomamente por dados derivados da atuação humana, tendem a replicar os vieses humanos, sejam eles bons ou ruins (OSOBA; WELSER IV, 2017). Sob tal paradigma, informações obtidas de uma comunidade machista podem originar uma inteligência artificial que marginalize mulheres, enquanto dados oriundos da atividade de uma força policial racista podem levar a um aprendizado de máquina que valoriza excessivamente os riscos de criminalidade e reincidência quanto a pessoas negras. Como atenta Frank Pasquale (2015), a análise da ação policial não é capaz de fornecer dados objetivos de comportamento apartados do contexto social ou dos vieses dos observadores e ignorar esse fato conduziria à perseguição de minorias.

Questões do gênero foram levadas ao debate público de modo mais amplo a partir de matéria da agência de jornalismo investigativo ProPublica, que analisou o funcionamento do *software* COMPAS observando a pontuação conferida a cerca de 7.000 pessoas no condado de Broward, Flórida (EUA). Julia Angwin et al. (2016), ao confrontar as previsões algorítmicas e a pesquisa empírica a respeito da conduta posterior dos infratores, com uso dos mesmos parâmetros adotados pela desenvolvedora Equivant, identificaram uma taxa de acerto de apenas 20% em relação à previsão de cometimento de crimes violentos. A imprecisão constatada também possuía uma grave tendência racial. A classificação equivocada de alto risco de reincidência ocorria em uma taxa duas vezes maior para indivíduos negros em comparação aos brancos; ao mesmo tempo, brancos recebiam indevidamente notas baixas (significando menor propensão à reincidência) de modo mais frequente do que negros (ANGWIN et al., 2016). Por conta da série de reportagens envolvendo jornalismo de dados e algoritmos, os autores foram finalistas do Prêmio Pulitzer na categoria Reportagem Explicativa no ano de 2017 (THE PULITZER PRIZES, 2017).

Um dos casos citados refere-se a Brisha Borden, uma jovem de 18 anos que foi presa com uma amiga por furtar uma bicicleta e um patinete infantis quando corria para buscar a filha de seu padrinho na escola, os quais foram abandonados tão logo a mãe da criança reagiu correndo atrás delas. Os itens custavam 80 dólares. Em paralelo, Vernon Prater, de 41 anos, foi detido por furtar 86 dólares em ferramentas de uma loja. Prater já havia sido condenado por roubo e tentativa de roubo, pelo que passou cinco anos preso, além de ter contra si uma outra acusação de roubo. Borden possuía registro de quatro delitos que cometera na adolescência (ANGWIN et al., 2016).

Ao avaliar o risco de reincidência de ambos, o *software* atribuiu nota oito a Borden, o que corresponde a um alto risco, e nota três a Prater, um risco baixo. Dois anos depois, o algoritmo seria desmentido: Borden não havia sido acusada de nenhum outro crime, enquanto Prater estava cumprindo oito anos de prisão por invadir um depósito e roubar milhares de dólares em eletrônicos. Os pesquisadores atribuem o erro ao fato de que Brisha Borden é negra, enquanto Vernon Prater é branco, de forma que o algoritmo teria um viés racista. Foi percebido que, entre os que não reincidiram, 23,5% dos brancos foram classificados como de alto risco, enquanto a mesma coisa ocorreu com 44,9% dos negros, o que representa os falsos positivos. Entre os que foram classificados como de baixo risco, mas vieram a reincidir no crime (falsos negativos), 47,7% eram brancos e 28% eram negros (ANGWIN et al., 2016).

A matéria foi tecnicamente refutada pela Northpointe, em artigo científico publicado por William Dieterich, Christina Mendoza e Tim Brennan (2016), em nome do Departamento de Pesquisa da empresa. Os pesquisadores sustentaram que a agência ProPublica concentrou-se em classificações estatísticas que não levavam em conta as diferentes taxas básicas de reincidência para negros e brancos; quando utilizadas as classificações corretas, os dados não corroborariam as acusações de viés racista. A interpretação adequada dos resultados das amostras utilizadas pela agência demonstraria que as escalas utilizadas pelo *software* são igualmente precisas para negros e brancos, ou seja, que uma determinada nota implica chances idênticas de recidiva, independentemente da raça do infrator.

Avaliando os resultados de ambos os estudos, Sam Corbett-Davies et al. (2016) explicam que as distorções observadas pela agência ProPublica são decorrências matemáticas do modelo adotado pela Northpointe e que qualquer modelo adotado implicaria algum tipo de distorção. Considerando que, em cada categoria de risco (baixo, médio ou alto), por definição, a proporção de infratores reincidentes é aproximadamente a mesma independentemente de raça e que a taxa de reincidência para infratores negros é maior do que a de brancos (52%

contra 39%), então haverá mais propensão para classificar negros como de médio ou alto risco e, conseqüentemente, negros que não reincidirem estarão sujeitos a receber notas mais altas do que brancos na mesma situação. Diga-se que o sistema não utiliza classificação por raça; contudo, na prática, muitos dos fatores utilizados pelo algoritmo têm correlação racial, a exemplo da antecedência criminal, que é maior entre os negros (CORBETT-DAVIES et al., 2016).

5. A JUSTIÇA ALGORÍTMICA

O debate travado entre ProPublica e Northpointe suscitou questões sobre como definir justiça no tocante a aprendizado de máquina, o que tem recebido a denominação de *algorithmic justice*, *algorithmic fairness* ou “justiça algorítmica” (DOLMAN; SEMENOVICH, 2018; WOLFE, 1990; ZAVRSNIK, 2019). Ben Green e Lily Hu (2018) indicam que as duas principais abordagens técnicas em relação à justiça quando se trata de aprendizado de máquina contemplam a categoria procedimental e a categoria estatística. Na primeira, considera-se justa a tarefa de aprendizado de máquina quando ela é aderente a uma máxima procedimental, como a de que “semelhantes devem ser tratados de maneira semelhante”. Nesse caso, o procedimento busca resolver o conflito entre os princípios gerais de justiça no campo da ética e o modo particular de operacionalização da justiça usado na engenharia para substituir a “máxima da semelhança” por uma “métrica de semelhança” (GREEN; HU, 2018).

Por sua vez, na categoria estatística, admite-se justa a tarefa que satisfaça vários tipos de “equilíbrio”, assim entendido como uma igualdade aproximada nas estatísticas baseadas em resultado. Cada um desses tipos de equilíbrio recebe um rótulo: a “igualdade de oportunidades” (*equal opportunity*), por exemplo, descreve uma ferramenta de aprendizado de máquina que equaliza taxas de verdadeiro positivo⁴ entre os grupos, enquanto “discriminação desproporcional” (*disparate mistreatment*) refere-se à igualdade nas probabilidades de erro (GREEN; HU, 2018; ZAFAR et al., 2017).

4. O conceito traz um paralelo entre a previsão/classificação prévia (correta ou errada, verdadeira ou falsa) e a ocorrência ou não (positiva ou negativa) do respectivo evento/condição. “Verdadeiro positivo” remete a uma previsão correta e à efetiva ocorrência do evento. Em paralelo, “falso positivo” retrata prognóstico equivocado de ocorrência do evento; “verdadeiro negativo” reflete a precisa identificação da não ocorrência do evento; e “falso negativo” implica previsão errada de não ocorrência do evento (DEAN, 2014). Para o caso do COMPAS, “verdadeiro positivo” significa um criminoso corretamente classificado como futuro reincidente; “falso positivo” indica um criminoso

Em ambas as categorias, repara-se que são considerados importantes fatores, como a imparcialidade do procedimento no primeiro caso e a proteção contra efeitos adversos no segundo. Contudo, nenhuma delas mostra-se suficiente, pois, ainda que consideradas em conjunto, limitam-se a satisfazer requisitos técnicos, mas não contemplam uma análise mais ampla acerca do contexto social e moral que envolve a justiça. São desprezados aspectos como o propósito da ferramenta, a sorte ou o mérito do indivíduo, o respeito ao devido processo legal, entre outros (GREEN; HU, 2018).

Richard Berk et al. (2018) descrevem seis tipos de justiça algorítmica: equalização de precisão geral (*overall accuracy equality*), paridade estatística (*statistical parity*), equalização de precisão de procedimento condicional (*conditional procedure accuracy equality*), equalização de precisão de uso condicional (*conditional use accuracy equality*), equalização de tratamento (*treatment equality*) e justiça total (*total fairness*). Para melhor compreendê-los, transcreve-se a denominada “matriz de confusão” (*confusion matrix*), utilizada para a avaliação da precisão de algoritmos:

Tabela 1 – Matriz de confusão para avaliação dos resultados empíricos em comparação com as previsões de algoritmos

Verdade	Previsão de falha	Previsão de sucesso	Erro de procedimento condicional
	<i>a</i>	<i>b</i>	$b / (a + b)$
Falha - positivo	Verdadeiros positivos	Falsos negativos	Taxa de falso negativo
	<i>c</i>	<i>d</i>	$c / (c + d)$
Sucesso - negativo	Falsos positivos	Verdadeiros negativos	Taxa de falso positivo
	$c / (a + c)$	$b / (b + d)$	$(c + b) / (a + b + c + d)$
Erro de uso condicional	Erro de previsão de falha	Erro de previsão de sucesso	Erro de procedimento geral

Fonte: Richard Berk et al. (2018).

A matriz compara as previsões algorítmicas (resultados previstos, *predicted outcome*) com os dados de fato observados (resultados reais, *actual outcome*), sendo cada variável representativa da contagem da célula respectiva. A letra *a*, por ilustração, indica a quantidade de verdadeiros positivos, isto é, o número de casos em que o algoritmo previu um resultado “falha” e obteve um resultado real “falha” (BERK et al., 2018). Os autores explicitam que, em sua notação, uma falha é denominada “positivo” porque ela incrementa o risco, a exemplo de uma prisão por crime violento. Por sua vez, o sucesso é descrito como “negativo”, pois

erroneamente classificado como futuro reincidente; “verdadeiro negativo” representa um criminoso justamente identificado como não reincidente; e “falso negativo” descreve um criminoso falsamente identificado como não reincidente.

ele implica menor risco, como na ausência de novas infrações durante o livramento condicional.

A partir das variáveis estabelecidas, Richard Berk et al. (2018) expõem as características, resumidas na tabela a seguir, de cada tipo de justiça algorítmica, que é obtido por meio da equalização de certas variáveis das previsões e dos resultados para cada classe (como negros e brancos) de um determinado grupo (infratores, no caso do COMPAS).

Tabela 2 – Tipos de justiça algorítmica e representação das respectivas condições conforme a matriz de confusão

Tipo de justiça algorítmica	Variáveis idênticas para cada uma das classes
equalização de precisão geral (<i>overall accuracy equality</i>)	$(a + d) / (a + b + c + d)$
paridade estatística (<i>statistical parity</i>)	$(a + c) / (a + b + c + d)$ e $(b + d) / (a + b + c + d)$
equalização de precisão de procedimento condicional (<i>conditional procedure accuracy equality</i>)	$a / (a + b)$ e $d / (c + d)$
equalização de precisão de uso condicional (<i>conditional use accuracy equality</i>)	$a / (a + c)$ e $d / (b + d)$
equalização de tratamento (<i>treatment equality</i>)	c / b ou b / c
justiça total (<i>total fairness</i>)	todas as anteriores

Fonte: adaptado de Richard Berk et al. (2018).

Na primeira hipótese, o resultado de $(a + d) / (a + b + c + d)$, definido como precisão geral (*overall accuracy*) por representar a taxa de resultados previstos corretamente, deveria ser o mesmo para o grupo de negros e para o grupo de brancos, e assim por diante. No entanto, Richard Berk et al. (2018) esclarecem que a pertinência da aplicação de cada tipo de justiça algorítmica deve ser avaliada no caso concreto, sendo impróprio acreditar que a justiça total é possível na prática ou que ela significa um algoritmo perfeito. Adotar a equalização de precisão geral, por exemplo, implicaria a assunção de que a ocorrência de verdadeiros positivos (a) é tão desejável quanto a de verdadeiros negativos (d), eis que essas duas variáveis são tratadas com o mesmo peso. Isso, porém, não seria correto em análises como a do COMPAS, uma vez que se ambiciona uma taxa maior de verdadeiros negativos, os não reincidentes, em relação aos verdadeiros positivos, reincidentes (BERK et al., 2018). Reflexões semelhantes, dizem os autores, poderiam ser feitas acerca de cada tipo de justiça, resultando na conclusão de que, exceto em casos triviais, seria impossível maximizar ao mesmo tempo a precisão e a justiça dos algoritmos, como também atender simultaneamente todos os tipos de justiça, pois alguns deles são incompatíveis uns com os outros e com a precisão. Na

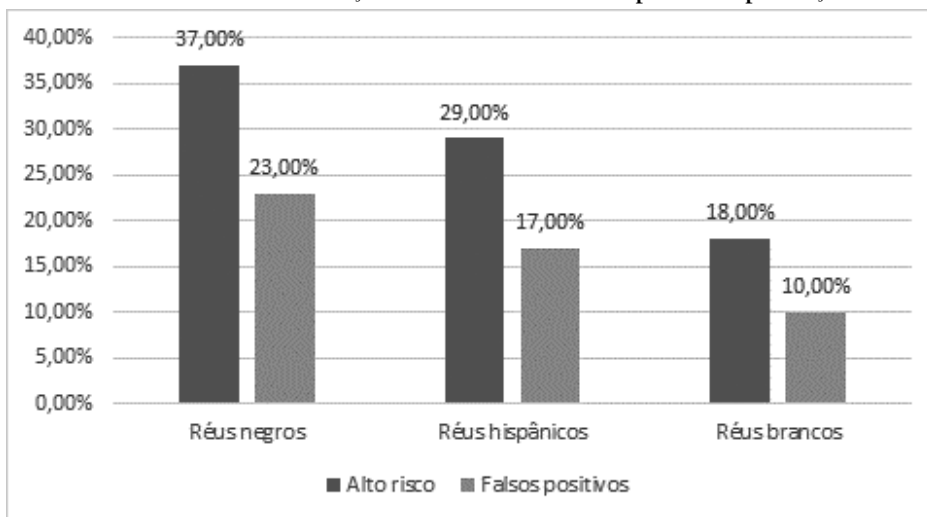
prática, tornar-se-ia necessário privilegiar alguns fatores em detrimento de outros, numa escolha desafiadora (BERK et al., 2018).

A respeito do COMPAS, Chris Dolman e Dimitri Semenovitch (2018) observam a argumentação da empresa Northpointe no sentido de que o resultado fornecido (a nota na escala de risco) pelo algoritmo deve significar o mesmo em valores absolutos, independentemente da raça do infrator. O algoritmo, portanto, não faz calibração referente a grupos raciais e, se o fizesse, poderia ser acusado de possuir um viés racista do mesmo modo. Já a matéria de ProPublica defende que os falsos positivos (notas altas para não reincidentes) e falsos negativos (notas baixas para reincidentes) são desigualmente distribuídos entre os grupos raciais. Ambos os argumentos são considerados observáveis e de difícil contestação pelos autores, que sustentam a impossibilidade de que um algoritmo como esse contemple simultaneamente todas as definições de justiça.

Para testar a hipótese de que as distorções constatadas não se restringiriam a um determinado programa, Sarah Picard et al. (2019), do *Center for Court Innovation*⁵ (CCI), desenvolveram um algoritmo preditivo apenas para fins de pesquisa, utilizando uma amostra anônima de 175 mil réus originária da cidade de Nova Iorque. Neste caso, obtiveram-se vieses similares aos observados em relação ao COMPAS, pois o novo *software* manteve um nível de precisão equivalente entre grupos étnicos e raciais, mas, em relação aos tipos de erros cometidos, houve uma tendência maior de classificar incorretamente infratores negros como de alto risco⁶ quando comparados com réus hispânicos ou brancos. Os dados foram consolidados no gráfico a seguir:

5. O *Center for Court Innovation* (2019) é um *think tank* novaiorquino, oriundo de uma parceria público-privada entre o Sistema Unificado de Tribunais do Estado de Nova Iorque e o Fundo para a Cidade de Nova Iorque, com o objetivo de realizar pesquisas, implantar programas e prestar assessoria no âmbito de reformas judiciais.
6. A pesquisa inicialmente dividiu as categorias de risco em cinco (mínimo, baixo, moderado, moderado-alto e alto). Porém, para fins de análise dos resultados, as categorias moderado-alto e alto foram condensadas em uma única categoria de alto risco, pois presumiu-se que, na prática, as duas categorias mais altas implicariam um dado contra o réu no momento do julgamento.

Gráfico 1 – Classificação de alto risco e falsos positivos por raça



Fonte: Sarah Picard et al. (2019).

Registra-se ainda a crítica de Ales Zavrsnik (2019), para quem a automatização de decisões judiciais pelo uso de inteligência artificial com a finalidade de incrementar a eficiência do Poder Judiciário tem lugar em um contexto neoliberal de “fazer mais com menos”, o que tornaria *softwares* como o COMPAS um avatar algorítmico do neoliberalismo. O autor ressalta que a concepção de um método científico puro para a justiça criminal remonta a Cesare Beccaria, cujas ideias permitiram uma humanização do procedimento criminal e ajudaram a legitimar as classes então emergentes e o estado moderno. O mesmo estaria ocorrendo agora, segundo ele, com a introdução da inteligência artificial na justiça criminal, colaborando para reforçar o poder social de uma elite digital, que capitaliza a ideologia de imparcialidade dos algoritmos e do *big data*.

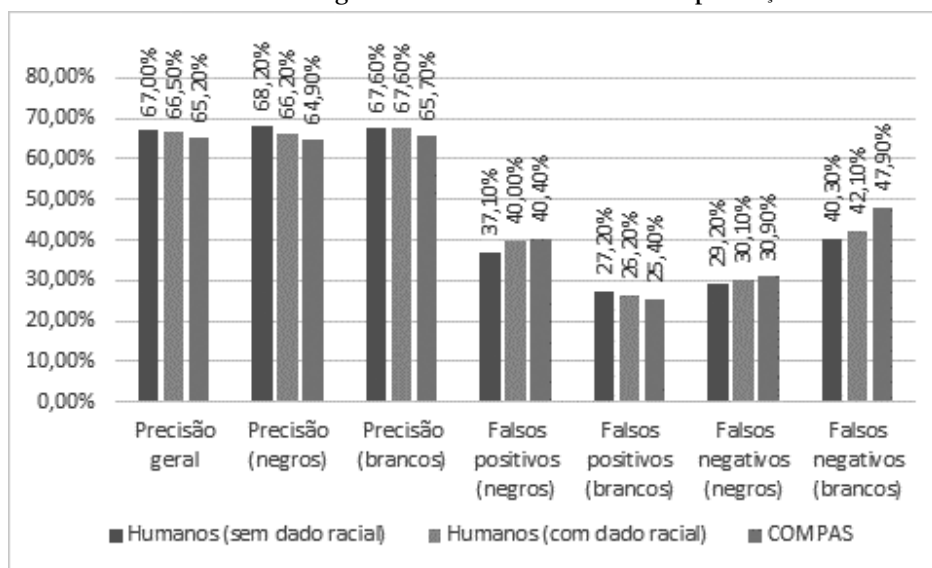
6. PRECISÃO E VIESES PARA ALÉM DOS ALGORITMOS

Dos estudos citados, percebe-se que é possível demonstrar tecnicamente tanto a neutralidade do programa COMPAS como o seu viés racial. Isso porque a sua implementação envolve escolhas que não podem ser classificadas como certas ou erradas, e tais escolhas implicam melhorias em algum aspecto dos resultados, em detrimento de outros elementos, mantendo-se, em qualquer caso, a integridade técnica do algoritmo.

Pesquisa empírica realizada por Julia Dressel e Hany Farid (2018), por outro lado, demonstrou que um grupo de indivíduos com acesso a apenas sete dados

sobre os infratores (contra 137 fatores utilizados pelo COMPAS) chegou a resultados com precisão análoga e até ligeiramente melhor do que a oferecida pelo *software* estadunidense. A pesquisa foi repetida informando aos avaliadores a raça dos infratores, que havia sido omitida na experiência anterior, e não houve impacto relevante nos resultados obtidos. Os pesquisadores concluem que a complexa avaliação realizada pelo COMPAS não obtém resultados melhores do que um grupo de pessoas sem orientação prévia nem experiência com a justiça criminal respondendo um questionário *online*. Além disso, viu-se que um classificador linear simples⁷ com apenas dois fatores (idade e número de condenações prévias) é suficiente para obter a mesma precisão do COMPAS, sugerindo que regras mais básicas, transparentes e inteligíveis de previsão seriam tão eficientes quanto o algoritmo privado. Alguns dos dados do estudo foram resumidos no gráfico seguinte e atestam a manutenção do viés racial em quaisquer dos casos:

Gráfico 2 – Prognósticos humanos x COMPAS por raça



Fonte: Julia Dressel e Hany Farid (2018).

7. Um classificador linear simples é uma ferramenta cujo objetivo é encontrar uma função linear – função de primeiro grau $f(x) = ax + b$, com $a \in \mathbb{R}$, $a \neq 0$ e $b = 0$ – capaz de separar as amostras em suas respectivas classes. No caso, o intuito é separar de maneira matematicamente trivial os infratores em grupos de reincidentes e não reincidentes, de acordo com suas características (DRESSEL; FARID, 2018; SOUZA, 2018).

Deste modo, a análise de risco oferecida por programas como o COMPAS mostra-se válida, mas não pode ser considerada isoladamente, muito menos quando se trata de valores como a liberdade e a dignidade humana. Faz-se necessária uma avaliação crítica permanente dos resultados do algoritmo, bem como a sua ponderação com outros aspectos objetivos e subjetivos capazes de subsidiar decisões judiciais. Nesse mesmo sentido, a Suprema Corte do Estado de Wisconsin, EUA, em recente decisão, sustentou que, se utilizada com as devidas cautelas, a avaliação de risco do programa COMPAS não fere o devido processo legal (WISCONSIN, 2016).

Também vale destacar que o algoritmo e a tecnologia não são necessariamente neutros nem imparciais (STRATE, 2012; ZAVRSNIK, 2019), e, por isso, introduzem vieses na análise jurídica. Nesse sentido, para um uso apropriado do *software*, exige-se que o julgador tenha em mente as possíveis distorções presentes nos dados obtidos, sob pena de tratar os resultados como verdades absolutas e reproduzir inadvertidamente os vieses ali contidos. Matheus Fornasier e Norberto Knebel (2020, p. 213) sugerem até mesmo que sejam incorporados vieses e valores de forma proposital em sistemas de inteligência artificial, investindo-se em debate público que possa direcionar e, assim, democratizar a tomada de decisões por meio de algoritmos.

Eduardo Magrani (2019, p. 162) ressalta a necessidade de melhor compreensão do tema para fins de regulação. O autor considera que os algoritmos dotados de algum grau de inteligência artificial, entre outros elementos não humanos, estão influenciando profundamente a sociedade, notadamente na tomada de decisões e na interação comunicativa, e possuem capacidade para produzir efeitos materiais de cunho político-democrático significativos, algo corroborado pela análise do *software* COMPAS.

Por fim, um aspecto que não pode ser negligenciado está além da qualidade técnica dos algoritmos e das previsões. A validade estatística e a semelhança dos resultados oriundos de algoritmos comerciais, algoritmos acadêmicos, questionários *online* respondidos por leigos e classificadores lineares simples, com uso de base de dados diversas, no sentido de tratar negros com maior rigor, sugere que os eventuais vieses, distorções ou imprecisões verificados não decorreriam exatamente das máquinas, dos desenvolvedores, dos estatísticos ou das informações que alimentam as análises, mas sim do próprio contexto social que os envolve.

Sobrevém, portanto, a hipótese de que os vieses percebidos são manifestações de um racismo estrutural, assim entendido por Silvio Almeida (2019, p. 50) como “[...] uma decorrência da própria estrutura social, ou seja, do modo ‘normal’

com que se constituem as relações políticas, econômicas, jurídicas e até familiares, não sendo uma patologia social e nem um desarranjo institucional.”

O autor explica que, numa sociedade cujo racismo é regra e não exceção, surgem comportamentos individuais e processos institucionais que reproduzem o racismo. Nessa conjuntura, os vieses dos algoritmos poderiam configurar mera consequência da maior presença de policiais em bairros de predominância negra, ou da maior probabilidade de detenção de um negro quando comparada com a de um branco na mesma situação (CORBETT-DAVIES et al., 2016).

Ilustra essa possibilidade a reportagem de Ben Poston e Cindy Chang (2019), que identificou, na cidade de Los Angeles, um volume muito maior de abordagens da polícia contra negros ou latinos no trânsito em relação a brancos, ainda que estes últimos sejam mais comumente flagrados portando itens ilegais. Por intermédio de uma nova lei californiana sobre questões raciais, os jornalistas receberam dados indicando que um negro em um veículo tem quatro vezes mais chance de ser revistado pela polícia do que um branco. Para os latinos, essa chance é três vezes maior. Apesar disso, foram encontradas drogas, armas ou contrabando em 20% das buscas contra brancos, enquanto esse índice foi de 17% quanto aos negros e 16% para latinos.

A matéria afirma que os números não necessariamente permitem a conclusão de que há um viés racial nas taxas de abordagem. Contudo, o maior número de interpelações contra negros em termos absolutos pode se tornar um gatilho para o surgimento de dados desfavoráveis a esse grupo no sistema criminal, o que introduziria o viés racial nos algoritmos, conforme observado no presente artigo.

Ainda tratando dos Estados Unidos, Michelle Alexander (2010) cita a Guerra às Drogas como exemplo de adoção institucional de métodos e práticas contra inimigos racialmente definidos, empregados quase exclusivamente em comunidades pobres, resultando no encarceramento em massa de afro-americanos e latinos. A autora registra o argumento oficial de que se combate algo, as drogas, e não um grupo de pessoas, embora isso seja negado pelos fatos, na sua visão. Entre muitos outros casos concretos, ela menciona relatório da Human Rights Watch (2000), com dados do ano de 1996, atestando a alta proporção de negros entre presos por delitos relacionados às drogas (na média nacional, 63%, superando os 80% em sete estados) em contraste com a sua diminuta participação na população (12% da média nacional, com um máximo de 36% no estado do Mississippi).

No mesmo tema, David W. Koch, Jaewon Lee e Kyunghye Lee (2016), usando controle estatístico de dados de pesquisa nacional realizada em 1997, verificaram que os afro-americanos são mais comumente presos por uso ou tráfico de drogas do que os brancos, embora as evidências indicassem que estes últimos são

mais propensos ao comportamento delituoso referente ao consumo. O estudo apresenta prova empírica de que o combate às drogas é discriminatório, produzindo resultados contaminados por um viés racial sistêmico e reforçando a teoria do combate às drogas enviesado (*biased drug enforcement theory*), no sentido de que a discriminação racial atua nos níveis individual e estrutural, causando maior encarceramento de infratores negros (KOCH; LEE; LEE, 2016).

A tendência de policiais, promotores e juízes nos EUA por uma ação mais intensa sobre a população negra cria um mecanismo que fortalece o encarceramento e a reincidência. Logo, ao extrair seu aprendizado desses dados, os algoritmos acabam por produzir indicadores condicionados potencialmente por um racismo estrutural presente na dinâmica do sistema de justiça criminal norte-americano. A aparência de neutralidade e imparcialidade estatística esconde o risco do uso de dados resultantes de ações humanas sujeitas a uma ampla gama de condicionantes sociais, culturais, econômicas etc. As mesmas variáveis interferem não apenas nos algoritmos, mas na sociedade em geral, o que parece explicar a observação contínua de vieses discriminatórios, independentemente do sistema escolhido para a avaliação de risco.

De modo concomitante, a forma de uso dos dados também pode condicionar os resultados subsequentes e influenciar de maneira contínua um viés nesses algoritmos, retroalimentando a discriminação. Uma avaliação de riscos mais rigorosa contra acusados negros importa em maior encarceramento desta população e isso tende a produzir efeitos negativos severos sobre renda, empregabilidade, relações familiares, entre outros fatores aptos a auxiliar na saída do ciclo de crime e detenção. A aplicação do cálculo de risco para reforçar o aprisionamento pouco contribui para desenvolver soluções mais efetivas para o problema real da reincidência, em particular para os acusados negros.

A necessária cautela com a algoritmização da justiça encontra, assim, paralelo com a exigida reflexão diante da atuação policial e judicial cotidiana. Seja no campo analógico ou digital, percebem-se fatores que corroboram com a discriminação e mantêm a população negra marginalizada. Nesse sentido, o uso acrítico dos *softwares* poderia implicar a naturalização dos preconceitos ali contidos, diante de seu ocultamento nos códigos e dados que permitem o funcionamento dos programas.

A análise refere-se à realidade norte-americana, mas certamente encontra aplicabilidade em outros países. Inobstante não se conheçam experiências de justiça algorítmica no Brasil, as semelhanças entre as duas nações, notadamente quanto ao regime escravocrata, a replicação da política criminal antidrogas, relatada por Luís Carlos Valois (2017), e a maior proporção de negros na população

carcerária (FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA, 2016) permitem a compreensão de que, em tese, a adoção local de *softwares* como o COMPAS ensejaria ponderações análogas.

7. CONSIDERAÇÕES FINAIS

As discussões apresentadas atestam que a revolução digital tem participação relevante nas mudanças e nos desafios que se apresentam para a justiça criminal e para os profissionais do Direito. O uso de novas tecnologias, como a inteligência artificial, para subsidiar os julgamentos exige familiaridade com conceitos da lógica, da matemática e da estatística para entender como se dá a avaliação promovida pelas máquinas e fundamentar as decisões e eventuais contraposições aos resultados.

A polêmica em torno do *software* COMPAS e seu suposto viés racial retrata essa nova realidade, irresponsável apenas com as ferramentas do Direito. Ainda que se admita a existência do viés, as notas dadas pelo algoritmo, em tese, permanecem tecnicamente fiáveis e aptas a embasar decisões judiciais, desde que se compreendam as nuances do aprendizado de máquina e os efeitos concretos do uso das informações produzidas. Isso porque os aplicativos desta natureza não conseguem implementar ao mesmo tempo todos os parâmetros de justiça algorítmica, sendo necessário privilegiar alguns em detrimento de outros, o que causa inevitáveis distorções nos resultados obtidos, sem necessariamente invalidá-los.

Por outro lado, é notável a reprodução dos mesmos vieses raciais percebidos no programa COMPAS por outros algoritmos, por classificações lineares simples, por questionários *online*, por compilações de dados de diferentes cidades, sugerindo que a origem da distorção se encontra não apenas nos *softwares* e nos dados, mas também fora deles. Abre-se a hipótese de que se trata de uma manifestação de racismo estrutural, que, por vários mecanismos naturalizados dentro do contexto social, produz dados indicativos da discriminação racial existente, posteriormente reeditada no momento do processamento e uso desses dados.

A atuação policial e judicial diante da população negra pode, de modo inadvertido e artificial, gerar dados que sobrevalorizam a participação de indivíduos negros na criminalidade. Tais informações, quando utilizadas no *machine learning*, incutem nos programas vieses preconceituosos, ocultados pela aparente neutralidade e imparcialidade das máquinas. Neste sentido, o uso incauto dos *softwares* poderia perpetuar a discriminação racial, retroalimentando-a de maneira subliminar e menos sujeita ao escrutínio público.

Embora a informação utilizada retrate regiões dos Estados Unidos, o estudo pode trazer importantes alertas para o Brasil, diante da similitude dos problemas enfrentados por ambos os países na questão racial. A pesquisa pode ter continuidade mediante o levantamento de dados no plano brasileiro e de uma avaliação acerca da efetiva caracterização ou não de racismo estrutural em virtude dos vieses raciais percebidos.

8. REFERÊNCIAS

- ALEXANDER, Michelle. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*. Nova Iorque: The New Press, 2010.
- ALMEIDA, Silvio Luiz de. *Racismo estrutural*. São Paulo: Pólen, 2019.
- ANDREWS, Donald Arthur; BONTA, James; WORMITH, J. Stephen. The Recent Past and Near Future of Risk and/or Need Assessment. *Crime & Delinquency*, v. 52, n. 1, p. 7-27, jan. 2006. DOI: [10.1177/0011128705281756]. Disponível em: [www.researchgate.net/profile/J_Wormith/publication/249718755_The_Recent_Past_and_Near_Future_of_Risk_andor_Need_Assessment/links/0c960530f834036e07000000/The-Recent-Past-and-Near-Future-of-Risk-and-or-Need-Assessment.pdf]. Acesso em: 04.12.2021.
- ANDREWS, Donald Arthur; BONTA, James. *The psychology of criminal conduct*. 5. ed. New Providence: Anderson Publishing, 2010.
- ANGWIN, Julia et al. Machine bias. *ProPublica*, 23 maio 2016. Disponível em: [www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing]. Acesso em: 04.12.2021.
- ARRUDA, Andrew; OVBIAGELE, Jimoh; DAL'OGGIO, Pargles. Announcement. *ROSS Intelligence*, 11 dez. 2020. Disponível em: [blog.rossintelligence.com/post/announcement]. Acesso em: 04.12.2021.
- BERK, Richard; HEIDARI, Hoda; JABBARI, Shahin; KEARNS, Michael; ROTH, Aaron. Fairness in Criminal Justice Risk Assessments: The State of the Art. *Sociological Methods & Research*, p. 1-42, jul. 2018. DOI: [10.1177/0049124118782533]. Disponível em: [arxiv.org/pdf/1703.09207.pdf]. Acesso em: 04.12.2021.
- BRENNAN, Tim; DIETERICH, Bill; EHRET, Beate. *Research synthesis: reliability and validity of COMPAS*, set. 2007. Disponível em: [www.northpointeinc.com/files/research_documents/reliability_validity.pdf]. Acesso em: 04.12.2021.
- CENTER FOR COURT INNOVATION. *About*. Nova Iorque: Center for Court Innovation, 2019. Disponível em: [www.courtinnovation.org/about]. Acesso em: 04.12.2021.
- CHANDER, Anupam; KRISHNAMURTHY, Vivek. The myth of platform neutrality. *Georgetown Law Technology Review*, v. 2, n. 2, p. 400-416, 2018. Disponível

- em: [georgetownlawtechreview.org/wp-content/uploads/2018/07/2.2-Chander-Krishnamurthy-pp-400-16.pdf]. Acesso em: 04.12.2021.
- CORBETT-DAVIES, Sam; PIERSON, Emma; FELLER, Avi; GOEL, Sharad. A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear. *The Washington Post*, 17 out. 2016. Disponível em: [www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/]. Acesso em: 04.12.2021.
- CORVALÁN, Juan Gustavo. Inteligencia artificial: retos, desafíos y oportunidades – Prometea: la primera inteligencia artificial de Latinoamérica al servicio de la Justicia. *Revista de Investigações Constitucionais*, Curitiba, v. 5, n. 1, p. 295-316, abr. 2018. Disponível em: [ref.scielo.org/g6pxrc]. Acesso em: 04.12.2021.
- DEAN, James. *Big data, data mining, and machine learning: value creation for business leaders and practitioners*. Hoboken: Wiley, 2014.
- DIETERICH, William; MENDOZA, Christina; BRENNAN, Tim. COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity, 8 jul. 2016. Disponível em: [go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf]. Acesso em: 04.12.2021.
- DOLMAN, Chris; SEMENOVICH, Dimitri. Algorithmic Fairness: Contemporary Ideas in the Insurance Context. In: GIRO CONFERENCE, 23-26 out. 2018, Birmingham. *Proceedings* [...]. Birmingham: Institute and Faculty of Actuaries, 2018. p. 1-13. Disponível em: [www.actuaries.org.uk/documents/b9-algorithmic-fairness-contemporary-ideas-insurance-context-paper]. Acesso em: 04.12.2021.
- DRESSEL, Julia; FARID, Hany. The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, v. 4, n. 1, jan. 2018. DOI: [10.1126/sciadv.aao5580]. Disponível em: [advances.sciencemag.org/content/advances/4/1/eaao5580.full.pdf]. Acesso em: 04.12.2021.
- ELEMENTARY. *The Economist*, São Francisco, 1 out. 2015. Science & technology. Disponível em: [www.economist.com/science-and-technology/2015/10/01/elementary]. Acesso em: 04.12.2021.
- FLACH, Peter. *Machine learning: the art and science of algorithms that make sense of data*. Cambridge: Cambridge University Press, 2012.
- FORNASIER, Matheus de Oliveira; KNEBEL, Norberto. Inteligência artificial: desafios e riscos ético-jurídicos. *Direito & Paz*, São Paulo, ano XIV, n. 43, p. 207-228, jul./dez. 2020. Disponível em: [www.revista.unisal.br/lo/index.php/direitoepaz/article/view/1279/546]. Acesso em: 04.12.2021.
- FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. *Pesquisa e análise de dados vinculados ao Campo da segurança pública e sistema penitenciário*. Termo de Parceria nº 817052/2015. Meta 04 – Elaboração de relatórios semestrais

- descritivos e analíticos, com base nos dados e informações do DEPEN. Etapa 01 – Relatório descritivo e analítico com base nos dados e informações do DEPEN, considerando a coleta referente à situação em dezembro de 2014. [São Paulo], 15 de abril de 2016. Disponível em: [dspace.mj.gov.br/handle/1/3045]. Acesso em: 04.12.2021.
- GREEN, Ben; HU, Lily. The Myth in the Methodology: Towards a Recontextualization of Fairness in Machine Learning. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 35., 2018, Estocolmo. *Proceedings* [...]. Estocolmo: International Conference on Machine Learning, 2018. Disponível em: [econcs.seas.harvard.edu/files/econcs/files/green_icml18.pdf]. Acesso em: 04.12.2021.
- HUMAN RIGHTS WATCH. Punishment and Prejudice: Racial Disparities in the War on Drugs. *HRW Reports*, v. 12, n. 2, maio 2000. Disponível em: [www.hrw.org/reports/2000/usa/]. Acesso em: 04.12.2021.
- KOCH, David W.; LEE, Jaewon; LEE, Kyunghye. Coloring the War on Drugs: Arrest Disparities in Black, Brown, and White. *Race and Social Problems*, v. 8, n. 4, dez. 2016. Disponível em: [link.springer.com/article/10.1007/s12552-016-9185-6]. Acesso em: 04.12.2021.
- LANSING, Sharon. *New York State COMPAS-Probation Risk and Need Assessment Study: Examining the Recidivism Scale's Effectiveness and Predictive Accuracy*. Set. 2012. Disponível em: [www.northpointeinc.com/downloads/research/DCJS_OPCA_COMPAS_Probation_Validity.pdf]. Acesso em: 04.12.2021.
- MAGRANI, Eduardo. *Entre dados e robôs: ética e privacidade na era da hiperconectividade*. 2. ed. Porto Alegre: Arquipélago, 2019.
- MCQUILLAN, Dan. People's Councils for Ethical Machine Learning. *Social Media + Society*, v. 4, n. 2, abr. 2018. DOI: [10.1177/2056305118768303]. Disponível em: [journals.sagepub.com/doi/10.1177/2056305118768303]. Acesso em: 04.12.2021.
- MOSES, Lyria Bennett; CHAN, Janet. Using big data for legal and law enforcement decisions: Testing the new tools. *University of New South Wales Law Journal*, v. 37, n. 2, p. 643-678, 2014. Disponível em: [ssrn.com/abstract=2513564]. Acesso em: 04.12.2021.
- NORTHPOINTE. *Practitioner's guide to COMPAS Core*, 2015. Disponível em: [www.northpointeinc.com/downloads/compas/Practitioners-Guide-COMPAS-Core-_031915.pdf]. Acesso em: 04.12.2021.
- OSOBA, Osonde; WELSER IV, William. *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. Santa Monica: RAND, 2017.
- PASQUALE, Frank. *The black box society: the secret algorithms that control money and information*. Cambridge: Harvard University Press, 2015.
- PICARD, Sarah; WATKINS, Matt; REMPEL, Michael; KERODAL, Ashmini. *Beyond the Algorithm: Pretrial Reform, Risk Assessment, and Racial Fairness*. Nova Iorque: Center for Court Innovation, 2019.

- POLASCHEK, Devon L. L. An appraisal of the risk–need–responsivity (RNR) model of offender rehabilitation and its application in correctional treatment. *Legal and Criminological Psychology*, v. 17, n. 1, p. 1-17, fev. 2012. DOI: [10.1111/j.2044-8333.2011.02038.x]. Disponível em: [onlinelibrary.wiley.com/doi/abs/10.1111/j.2044-8333.2011.02038.x]. Acesso em: 04.12.2021.
- POSTON, Ben; CHANG, Cindy. LAPD searches blacks and Latinos more. But they're less likely to have contraband than whites. *Los Angeles Times*, 8 out. 2019. Disponível em: [www.latimes.com/local/lanow/la-me-lapd-searches-20190605-story.html]. Acesso em: 04.12.2021.
- RUSSELL, Stuart J.; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Upper Saddle River: Prentice-Hall, 2010.
- SCHWAB, Klaus. *The fourth industrial revolution*. Genebra: World Economic Forum, 2016.
- SOUZA, Nahim Alves de. *Aumentando o poder preditivo de classificadores lineares através de particionamento por classe*. 2018. Dissertação (Mestrado em Ciência da Computação) – Centro de Ciências em Gestão e Tecnologia, Universidade Federal de São Carlos, Sorocaba, 2018. Disponível em: [repositorio.ufscar.br/bitstream/handle/ufscar/9530/SOUZA_NahimAlvesDe_dissertacao.pdf]. Acesso em: 04.12.2021.
- STRATE, Lance. If It's Neutral, It's Not Technology. *Educational Technology*, v. 52, n. 1, p. 6-9, jan./fev. 2012. Disponível em: [www.jstor.org/stable/44429982]. Acesso em: 04.12.2021.
- SUPREMO TRIBUNAL FEDERAL. *Inteligência artificial vai agilizar a tramitação de processos no STF*. 30 maio 2018. Disponível em: [portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=380038]. Acesso em: 04.12.2021.
- SUSSKIND, Richard. *Tomorrow's lawyers: An Introduction to Your Future*. Oxford: Oxford University Press, 2013.
- THE PULITZER PRIZES. *Finalist*: Julia Angwin, Jeff Larson, Surya Mattu, Lauren Kirchner and Terry Parris Jr. of ProPublica. Nova Iorque, [2017]. Disponível em: [www.pulitzer.org/finalists/julia-angwin-jeff-larson-surya-mattu-lauren-kirchner-and-terry-parris-jr-propublica]. Acesso em: 04.12.2021.
- VALOIS, Luís Carlos. *O direito penal da guerra às drogas*. 2. ed. Belo Horizonte: D'Plácido, 2017.
- VLADECK, David C. Machines without principals: liability rules and Artificial Intelligence. *Washington Law Review*, n. 89, p. 117-150, 2014. Disponível em: [digitalcommons.law.uw.edu/cgi/viewcontent.cgi?article=4800&context=wlr]. Acesso em: 04.12.2021.
- WISCONSIN. Supreme Court of Wisconsin. *Case no. 2015AP157-CR*. Recorrente: Eric L. Loomis. Recorrido: Estado do Wisconsin. 13 de julho de 2016. Disponível em: [www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690]. Acesso em: 04.12.2021.

- WOLFE, Alan. Algorithmic Justice. *Cardozo Law Review*, Nova Iorque, v. 11, p. 1409-1434, 1990. Disponível em: [is.muni.cz/el/law/podzim2017/MV735K/um/virtualizace_autoritativni_aplikace_prava_-_6/AlanWolfeAlgorithmicJusti.pdf]. Acesso em: 04.12.2021.
- ZAFAR, Muhammad Bilal; VALERA, Isabel; RODRIGUEZ, Manuel Gomez; GUMMADI, Krishna P. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 26., 2017, Genebra. *Proceedings* [...]. Genebra: International World Wide Web Conferences Steering Committee, 2017. p. 1171-1180. DOI: [doi.org/10.1145/3038912.3052660]. Disponível em: [arxiv.org/pdf/1610.08452.pdf]. Acesso em: 04.12.2021.
- ZAVRSNIK, Ales. Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology*, [S. l.], p. 1-20, set. 2019. DOI: [10.1177/1477370819876762]. Disponível em: [journals.sagepub.com/doi/pdf/10.1177/1477370819876762]. Acesso em: 04.12.2021.



PESQUISAS DO EDITORIAL



ÁREAS DO DIREITO: Processual; Penal

Veja também Doutrinas relacionadas ao tema

- Algoritmos de predição no sistema penal: as profecias tecnopolíticas que se autorrealizam no Século XXI, de Felipe da Veiga Dias – *RBCCrim* 183/99-124;
- Algoritmos e direitos fundamentais: riscos, transparência e *accountability* no uso de técnicas de automação decisória, de João Paulo Lordelo – *RBCCrim* 186/205-236;
- Compreendendo algoritmos aplicados ao sistema de justiça criminal – Ilegibilidade, acesso, compreensão, verdade e computabilidade no 'eu' identificado por algoritmos, de Rafael de Deus Garcia e Evandro Piza Duarte – *RBCCrim* 183/199-226; e
- O devido processo na era algorítmica digital: premissas iniciais necessárias para uma leitura constitucional adequada, de Georges Abboud e João Sérgio dos Santos Soares Pereira – *RT* 1026/125-145.