

## Segregação racial e vieses algorítmicos: máquinas racistas no âmbito do controle penal

## Racial segregation and algorithmic biases: racism machines in the context of criminal control

*Airto Chaves Junior(1); Bárbara Guasque(2); Thiago Santos Aguiar de Pádua(3)*

1 Universidade do Vale do Itajaí (Univali), Itajaí/SC, Brasil.

E-mail: oduno@hotmail.com | ORCID: <https://orcid.org/0000-0003-3016-5618>

2 Universidade do Vale do Itajaí (Univali), Itajaí/SC, Brasil.

E-mail: barbara@guasque.adv.br | ORCID: <https://orcid.org/0000-0003-0633-8363>

3 Universidade de Brasília (UNB), Brasília, DF, Brasil.

E-mail: thiago.padua@udf.edu.br | ORCID: <https://orcid.org/0000-0003-2229-2975>

**Revista Brasileira de Direito**, Passo Fundo, vol. 19, n. 2, e4768, maio-agosto, 2023 - ISSN 2238-0604

[Received/Recebido: agosto 8, 2022; Accepted/Aceito: setembro 20, 2023;

Publicado/Published: setembro 22, 2023]

DOI: <https://doi.org/10.18256/2238-0604.2023.v19i2.4768>

Como citar este artigo / How to cite item: [clique aqui/click here!](#)

## Resumo

A pesquisa tem por **objetivo** verificar se o emprego de Sistemas de Inteligência Artificial poderia abrandar a altíssima carga de seletividade do Controle Penal, notadamente aquelas fundadas no racismo e na discriminação. Para tanto, num primeiro momento, o texto trata da seletividade do poder punitivo como característica estrutural. Na sequência, dá atenção aos dados que são utilizados na construção de modelos de IA. Depois disso, elenca possíveis soluções para reduzir a incidência de algoritmos enviesados e mitigar a reprodução do racismo a partir do emprego da máquina. Por fim, propõe-se identificar algumas externalidades negativas oriundas da não observância de determinados padrões éticos em modelos de IA. **Justifica-se** a pesquisa em face da necessidade de se conhecer os perversos reflexos que algoritmos enviesados podem produzir nos segmentos de Controle Penal, catalisando preconceitos e segregando setores sociais eleitos pela cor da pele. Quanto ao **método de pesquisa**, revela-se o indutivo, subsidiado pela pesquisa bibliográfica e números oficiais extraídos de bases oficiais.

**Palavras-chave:** Racismo; Controle Penal; Inteligência Artificial; Vieses Algorítmicos; Justiça Criminal.

## Abstract

This paper aims to verify if the use of Artificial Intelligence Systems could reduce the very high load of Criminal Control's selectivity, especially those based on racism and discrimination. Initially, at first, for this purpose, the work deals with the selectivity of punitive power as a structural characteristic. Subsequently, we carefully observe the data that are used in the construction of AI models. After that, it lists possible solutions to reduce the incidence of biased algorithms and mitigate the reproduction of racism through the use of the machine. Finally, we propose to identify some negative externalities resulting from non-observance of certain ethical standards in AI models. The research is justified in view of the need to know the perverse reflexes that biased algorithms can produce in the Criminal Control segments, catalyzing prejudices and segregating social sectors chosen by skin color. The research method is the inductive one, supported by bibliographic research and official numbers extracted from official databases.

**Keywords:** Racism; Criminal Control; Artificial intelligence; Algorithmic Bias; Criminal Justice.

## 1 Introdução

Não é novidade alguma que o Controle Penal exercido pelo Estado comporta uma altíssima carga de seletividade. E, conforme lembra Zaffaroni, não se trata de um determinado mal-uso do sistema, mas de sua característica estrutural. As políticas penais fomentadas para o “combate da criminalidade” quase que sempre são direcionadas para grupos sociais bem delimitados identificados pela cor da pele ou mesmo, pelos espaços geográficos que ocupam nas regiões periféricas das cidades. A necropolítica, travestida de busca por segurança pública, é manifestada pelos segmentos de controle penal por práticas fundadas no racismo nas suas mais diversas acepções.

Por outro lado, permeando todas as esferas da vida em sociedade, a utilização de inteligência artificial está presente também no Sistema de Justiça. Até junho de 2020, já existiam 64 projetos de inteligência artificial em 47 tribunais brasileiros, em diferentes fases de implementação<sup>1</sup>. As novas tecnologias estão apontando no Judiciário brasileiro com o fito de automatizar atividades repetitivas, classificatórias e organizacionais, além de servir de apoio à tomada de decisão judicial, incrementando a produtividade, angariando eficiência e redução da morosidade processual e da insegurança jurídica – os grandes gargalos do Poder Judiciário nacional. As funcionalidades são muitas e o rol está se ampliando rapidamente. A inserção da disrupção nos tribunais brasileiros constitui uma grande aliada no aprimoramento do ambiente institucional judicial brasileiro, trazendo a esperança de uma justiça mais célere, efetiva e uniforme.

Um Poder Judiciário asfixiado por um acervo de 79,7 milhões de processos em tramitação, e com um custo aproximado de 91 bilhões de reais com servidores<sup>2</sup>, pode vislumbrar na tecnologia e na disrupção, uma esperança para tentar reverter o panorama de excessiva morosidade, alto custo e insegurança jurídica.

Em geral, todos os sistemas atuais desenvolvidos nos tribunais brasileiros efetuam tarefas simples e repetitivas, nas quais não se enquadra o uso problemático dos algoritmos e seus possíveis vieses. Todavia, como a relação entre cidadãos, Poder Judiciário e tecnologias disruptivas ainda é algo recente e crescente, existem algumas questões relevantes que necessitam ser abordadas. É o caso dos vieses algoritmos no âmbito do Controle Penal, que tem levantado preocupantes questionamentos de natureza ética.

Diante disso, o problema de pesquisa que se apresenta é o seguinte: a utilização de ferramentas de Inteligência Artificial no âmbito dos segmentos de Controle Penal (polícias, justiça, etc.) poderia abrandar a seletividade institucional/estrutural que permeia o Sistema Penal?

- 1 FGV – Centro de Inovação, Administração e Pesquisa do Judiciário. In: Inteligência Artificial: Tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário Brasileiro. Disponível em: [https://ciapj.fgv.br/sites/ciapj.fgv.br/files/estudos\\_e\\_pesquisas\\_ia\\_1afase.pdf](https://ciapj.fgv.br/sites/ciapj.fgv.br/files/estudos_e_pesquisas_ia_1afase.pdf). Acesso em: 15 out. 2020.
- 2 CONSELHO NACIONAL DE JUSTIÇA. Relatório Justiça em números 2020: Ano-base 2019. Brasília: CNJ, 2020. Disponível em: <https://www.cnj.jus.br>. Acesso em: 06 nov. 2021.

Para responder a esse questionamento, inicia-se a pesquisa com o estudo teórico-empírico do racismo no âmbito do Controle Penal. A partir de Foucault, verifica-se que as razões do racismo são separar e eliminar. Com base em números oficiais, mostrar-se-á que o Estado, pela via de seu braço penal, promove exatamente (e com extremo sucesso) essas duas características. Na sequência, dar-se-á atenção aos dados que são utilizados na construção de modelos de Inteligência Artificial, sobretudo, dos vieses algoritmos, os quais acontecem quando o algoritmo adquire e reflete os valores humanos implícitos nos dados que são utilizados para treinamento do modelo.

Depois disso, elencar-se-ão possíveis soluções para reduzir a incidência de algoritmos enviesados e mitigar a reprodução do racismo a partir do emprego da máquina. Por fim, propõe-se identificar algumas alternativas hábeis a atenuar a possibilidade de enviesamento dos modelos, mitigando as externalidades negativas oriundas de algoritmos enviesados para que não venham a contribuir com a manutenção e intensificação do racismo institucional/estrutural que permeia a sociedade e o Sistema de Justiça Criminal.

À vista disso, o presente estudo se justifica em razão da necessidade de se conhecer os perversos reflexos que algoritmos enviesados podem produzir nos segmentos de Controle Penal, catalisando preconceitos e segregando setores sociais eleitos pela cor da pele.

Quanto ao método de pesquisa, revela-se o indutivo, subsidiado pela pesquisa bibliográfica e números oficiais extraídos de bases oficiais.

## 2 Controle penal e racismo

Conforme já dizia Otto Kirchheimer<sup>3</sup>, “é bastante comum que todo regime político eleja seus inimigos preferenciais ou os crie ao longo do tempo”. E há preferências bem delimitadas, o que pode ser verificado pela contundente desigualdade racial programada no funcionamento das agências de repressão que compõe o Sistema Penal<sup>4</sup>, materialmente diagnosticada e percebida na maior severidade de tratamento nas sanções punitivas direcionadas a esse grupo de pessoas que, a grosso modo, pode

3 KIRCHHEIMER, Otto. *Political Justice: the use of legal procedure for political ends*. Princeton, New Jersey: Princeton University Press, 1961, p. 3.

4 Esse sistema engloba as atividades do legislador, do público, da polícia, do ministério público, do judiciário e dos funcionários que atuam na execução penal. Cfr. CHAVES JUNIOR, Airto; OLDONI, Fabiano. *Para que(m) serve o Direito Penal? Uma análise criminológica dos segmentos de controle social*. Rio de Janeiro: Lumen Juris, 2014, p. 22; Eugênio Raúl Zaffaroni e José Henrique Pierangeli destacam três grupos humanos que convergem na atividade institucionalizada do sistema penal: o policial, o judicial e o executivo penal. Cfr. ZAFFARONI, Eugenio Raúl; PIERANGELI, José Henrique. *Manual de direito penal brasileiro, vol. 1: parte geral*. 6 ed. São Paulo: Editora Revista dos Tribunais, 2006, p. 64-65.

ser identificado pela cor da pele. Ao que parece, “o chicote sobreviveu nos subterrâneos do Direito Penal”<sup>5</sup>.

Michel Foucault<sup>6</sup> sustentava que o racismo é a maneira pela qual, na esfera da vida, introduz-se uma separação que determina o que deve viver e o que deve morrer. Conforme o autor, duas são as suas funções. A primeira (a) é separar e desequilibrar. Desde que se passou a distinguir a espécie humana em raças (hierarquia, qualificação, etc.), fragmentou-se (desequilibrou-se) o campo do biológico e se passou a produzir um desequilíbrio entre os grupos que compõem a população.<sup>7</sup> Tem-se, portanto, uma forma de estabelecer censuras a partir do campo biológico. A segunda (b) função do racismo é permitir que se estabeleça uma relação bélica do tipo: “Quanto mais você mata ou mais faz morrer (ou deixa morrer), mais você vive”; ou, compatível com o exercício do *biopoder*, tal como “quanto mais indivíduos anormais são eliminados, menos degenerados haverá na espécie (...)”. Ou seja, raça e racismo são condições para aceitação do homicídio e, uma vez que o Estado funciona com base no *biopoder*, a função homicida do próprio Estado só pode ser assegurada para o racismo. Separar (a) e eliminar (b) são, portanto, as razões do racismo. A reflexão tem, pois, amparo empírico.

Com relação à esta função beligerante (b), estudos demonstram que o risco de ser morto por segmentos específicos que compõe o Sistema Penal é muitas vezes maior para negros do que para brancos. De acordo com informações divulgadas no 14º Anuário Brasileiro de Segurança Pública em outubro de 2020<sup>8</sup>, a comparação da taxa por 100 mil habitantes indica que a mortalidade entre pessoas negras em decorrência de intervenções policiais é 183,2% superior à taxa verificada entre brancos.<sup>9</sup> A polícia é o segmento que comporta grande poder seletivo, pois opera diretamente sobre o processo de “filtragem” do sistema. E a partir dessa “filtragem”, 79,1% das vítimas de intervenções policiais que resultaram em morte são pretas e pardas, indicando

5 BATISTA, Nilo. A violência do Estado e os aparelhos policiais. Discursos Sediciosos. Crime, direito e sociedade. Volume 2. 4. Ed. Rio de Janeiro, 1997, p. 146.

6 Aqui, a obra traduzida para o espanhol registra aulas proferidas pelo autor no Collège de France no início do ano de 1976 em que Foucault realiza uma análise acerca da fragmentação de suas pesquisas genealógicas para colocar em xeque os saberes instituídos como pretensos discursos elevados a nível de cientificidade. Cfr. FOUCAULT, Michel. Genealogia del Racismo. Traducción de Alfredo Tzveibel. Colección Caronte Ensayos. Buenos Aires: Editorial Altamira, 1993, p. 210-211.

7 A separação é característica fundante do racismo, o que também é partilhado como um dos seus efeitos no âmbito da psicanálise, aqui, sob a ótica da pessoa que sofre os efeitos do preconceito. Ver: Grada Kilomba, mais especificamente o Capítulo 10: “Segregação e Contágio Racial”. Cfr. KILOMBA, Grada. Memórias da Plantação: episódios de racismo cotidiano. Tradução de Jess Oliveira. Rio de Janeiro: Editora Cobogó, 2019, p. 167 e ss.

8 Anuário Brasileiro de Segurança Pública. Fórum Brasileiro de Segurança Pública, p. 90. Ano 14.2020. Disponível em: <https://forumseguranca.org.br/wp-content/uploads/2020/10/anuario-14-2020-v1-final.pdf>. Acesso em: 13 de março de 2022.

9 Enquanto entre brancos a taxa fica em 1,5 por 100 mil habitantes brancos, entre negros é de 4,2 por 100 mil negros.

a sobrerrepresentação de negros entre as vítimas da letalidade policial. Veja-se que este percentual é superior à média nacional verificada no total das mortes violentas intencionais, em que 74,4% de todas as vítimas são negras.

No que se refere à função de separação (a), as pesquisas (Anuário Brasileiro de Segurança Pública, 2020) indicam que os negros são também os alvos preferenciais das políticas de encarceramento do país: para cada três pessoas encarceradas no Brasil, duas são negras.<sup>10</sup> E, ao contrário da percepção que pode ser extraída do senso comum, a desproporcionalidade racial nas prisões brasileiras verificada a partir da predominância dos afrodescendentes povoando os presídios e penitenciárias não é um padrão que pode ser observado há muito tempo. Ocorre aqui, progressivamente, uma zona de inflexão em que a proporção de negros nas prisões cresce a cada ano. Se em 2005 os negros representavam 58,4% do total de presos, enquanto os brancos eram 39,8%; em 2019 essa proporção já chegava a 66,7% de negros e 32,3% de brancos.<sup>11-12</sup>

No âmbito das investigações criminais, a seletividade operada pelo sistema não se verifica diferente. No ano de 2011, Brandon L. Garrett publicou a obra *Convicting the Innocent: Where Criminal Prosecutions Go Wrong*<sup>13</sup>. O estudo do Prof. da Universidade da Virgínia examinou os primeiros 250 casos em que se restou demonstrado pelo exame de DNA que condenados eram, na verdade, inocentes. A partir da pesquisa, os exames de DNA deixaram nu a frequência com que esse sistema produz duas coisas: **condena inocentes e exonera culpados.**

Brandon estabeleceu um padrão desses inocentes condenados a partir da natureza das infrações penais que sustentaram as imputações e respectivas condenações. Verificou-se, neste ponto, que os primeiros 250 condenados inocentados a partir do exame de DNA haviam sido condenados principalmente por: estupro (68% dos casos); homicídio, (9% dos condenados); e crimes patrimoniais, a exemplo do roubo (21%). Desses, dezessete foram condenados à morte e oitenta foram condenados à prisão perpétua. Por fim, desses 250 inocentados, 155 eram negros, 20 latinos, 74 brancos e 1 asiático.

10 Além disso, conforme informações que podem ser extraídas do IBGE (Instituto Brasileiro de Geografia e Estatística), 56% da população é negra, percentual menor, portanto, do que aquele verificado nos presídios e penitenciárias.

11 A taxa de variação nesse período mostra o crescimento de 377,7% na população carcerária identificada pela raça/cor negra, valor bem superior à variação para os presos brancos, que foi de 239,5%. Conf.: *Anuário Brasileiro de Segurança Pública*, 2020. Além disso, tem-se que para cada não-negro preso no Brasil em 2019, 2 negros foram presos. Trata-se de pouco mais que o dobro, quando comparado aos não-negros, em especial os brancos.

12 Num recorte estadunidense, essa tendência já havia sido diagnosticada por Loïc Wacquant há cerca de 20 anos. Conforme o sociólogo, com base nas taxas de encarceramento do início dos anos 190, a probabilidade acumulada de, durante a vida, alguém ser “engaiolado” numa prisão estadual ou federal nos EUA era de 4% para brancos, 16% para latinos e 29% para os negros. Ver: WACQUANT, Loïc. *Punir os Pobres: a nova gestão da miséria nos Estados Unidos*. 3. Ed. Tradução de Sérgio Lamarão. Rio de Janeiro: Editora Revan, 2007, p. 334.

13 GARRETT, Brandon L. *Convicting the Innocent: where criminal prosecutions Go Wrong*. Cambridge, Massachusetts / London, England. Harvard University Press, 2011.



Conforme se verifica, apesar de o Direito Penal e o Processo Penal (especialmente quanto aos procedimentos de produção da prova) desenvolvidos nos últimos 150 anos estarem ancorados na ideia de cientificidade (veja-se, por exemplo, o positivismo como forma de explicar o mundo), a filtragem racial é flagrante. Abordagens policiais são, com muita frequência, realizadas por critérios raciais<sup>14</sup>, geográficos, e por estigmas fundados até mesmo na forma de vestimenta do “suspeito”. Ou seja, o gradiente de discricionariedade na ação de investigar é tão elástico que essa discricionariedade se transmuda em arbitrariedade seletiva.

O que ocorre aqui é o incremento para despir de humanidade determinados sujeitos que, previamente etiquetados pela cor ou mesmo, pelo espaço geográfico periférico que ocupam nas cidades, transformam-se em “vidas matáveis” (na forma da necropolítica diagnosticada pelo filósofo camaronês Achille Mbembe<sup>15</sup>) numa absoluta convergência ao sistema de “cidadania sacrificial” (percebido e criticado por Wendy Brown<sup>16</sup>), espaço em que alguns sujeitos são sacrificáveis pelo “bem da sociedade”.<sup>17</sup> E isso não parece ser um acidente de trajeto, mas parte da lógica estrutural das investigações policiais e, com muita frequência, atinge um grupo bem determinado da população e que já é, historicamente, marginalizado.

Desse modo, o Sistema Penal parece se dirigir “quase sempre contra certas pessoas, mais do que contra certas ações definidas como crime”<sup>18</sup>, pelo que, é de se concordar com Zaffaroni, que alerta que a seletividade do poder punitivo é a sua característica estrutural<sup>19</sup> e não um acidente de percurso ou produto do seu “mau funcionamento”. Por isso, a seletividade pode ser atenuada, jamais suprimida. A dúvida formulada a partir das premissas anotadas neste primeiro item é: o abrandamento desta altíssima carga de seletividade legitimada pelo Controle Penal poderia ser realizado com a utilização de sistemas de Inteligência Artificial? É o que se pretende investigar na sequência.

14 Em janeiro de 2022, chegou ao conhecimento da comunidade jurídica o fato de que uma foto do ator estadunidense Michael B. Jordan ter sido exibida num reconhecimento fotográfico realizado pela Polícia Civil do Estado do Ceará na investigação de uma chacina que teria ocorrido na cidade de Fortaleza no natal de 2021 e que teria deixado cinco mortos. O ator Michael Jordan era o “suspeito número dois” no termo de reconhecimento da Polícia Civil daquele Estado. Em miúdos, o racismo é tão manifesto que nem mesmo uma estrela hollywoodiana escapa. Ver: MATIDA, Janaína; CECCONELLO, Willian. O que há de errado no reconhecimento fotográfico de Michael B. Jordan? Conjur de 08 de janeiro de 2022.

15 MBEMBE, Achille. Necropolítica. São Paulo: N-1 Edições, 2018.

16 BROWN, Wendy. Undoing the Demos: Neoliberalism's Stealth Revolution. New York: Zone Books, 2015.

17 Do ponto de vista da psicanálise, o fenômeno da Violência de Estado pode ser verificado a partir da obra de Maria Ocariz. Cfr. OCARIZ, Maria Cristina (Org.). Psicanálise e Violência Social. São Paulo: Editora Escuta, 2018. Para um estudo da Violência Objetiva, ver: CHAVES JUNIOR, Airto. Além das Grades: a paralaxe da violência nas prisões brasileiras. Florianópolis: Tirant lo Blanch, 2018.

18 ANDREADE, Vera Regina Pereira de. A ilusão de segurança jurídica: do controle da violência a violência do controle penal. 2ª ed. Porto Alegre: Livraria do Advogado, 2003, p. 267.

19 ZAFFARONI, Eugênio Raul. O Inimigo no Direito Penal. Tradução de Sérgio Lamarão. Rio de Janeiro: Revan, 2007, p. 170.

### 3 Inteligência Artificial e vieses algorítmicos

De acordo com aquilo que anota Susana Navas Navarro, a inteligência artificial é um campo da ciência e engenharia, encarregado de compreender o comportamento inteligente do cérebro humano e, além disso, criar artefatos que simulem dito comportamento de maneira automatizada. Para a autora, a IA se ocupa de “emular as diversas capacidades do cérebro humano para apresentar comportamentos inteligentes, sintetizando e automatizando tarefas intelectuais”<sup>20</sup>.

Nesse sentido, pode-se delimitar a inteligência artificial como a capacidade das máquinas de mimetizar as habilidades humanas. Isto quer dizer que as máquinas, em alguma medida, imitam o processo cognitivo humano, após um processo de aprendizado baseado em dados que fornecem generalizações sobre dado assunto. Porém, a IA alcançará sempre resultados superiores aos que poderia conseguir qualquer ser humano. Isto porque o sistema não somente aceita, mas acessa uma quantidade de dados infinitamente maior que o cérebro humano. E, quanto mais dados, maior a possibilidade de relacioná-los, visualizar padrões ocultos e, portanto, obter melhores resultados<sup>21</sup>.

A palavra chave na IA é algoritmo<sup>22</sup>. É que um sistema de IA impende uma sequência de instruções que especifique as diferentes ações a serem executadas pelo computador a fim de resolver um determinado problema. Este esquema executivo contemplando as instruções, o caminho a ser percorrido, é desempenhado pela estrutura algorítmica<sup>23</sup>.

Um algoritmo é um esquema executivo, uma sequência de ações para resolver um problema ou responder uma questão. Assim, uma receita culinária é um algoritmo para humanos. Uma sequência de instruções para resolver o problema de como fazer determinada comida<sup>24</sup>. Todavia, computadores e máquinas precisam de algoritmos mais complexos, que os auxiliem a efetuar as tarefas definidas e que sempre produzam o mesmo resultado com base nos mesmos parâmetros. Neste sentido se encontra um subconjunto específico de algoritmos usados para aprendizado de máquina (*machine learning*)<sup>25</sup>.

20 NAVARRO, Susana Navas. Derecho e Inteligencia Artificial desde el Diseño. In: NAVARRO, Susana Navas, et al. Inteligencia Artificial: Tecnología, Derecho. Valencia: Tirant Lo Blanch, 2017, p. 24.

21 FENOLL, Jordi Nieva. Inteligencia Artificial y Proceso Judicial. Madri: Marcial Pons, 2018, p. 14.

22 FENOLL, Jordi Nieva. Inteligencia Artificial y Proceso Judicial. Madri: Marcial Pons, 2018, p. 15.

23 NAVARRO, Susana Navas. Derecho e Inteligencia Artificial desde el Diseño. In: NAVARRO, Susana Navas, et al. Inteligencia Artificial: Tecnología, Derecho. Valencia: Tirant Lo Blanch, 2017, p. 24.

24 RODRÍGUEZ, Pablo. Inteligencia Artificial. Como cambiará el mundo (y tu vida). Trad. Jorge Rizzo. Barcelona: Ediciones Deusto, 2018, p. 109.

25 RODRÍGUEZ, Pablo. Inteligencia Artificial. Como cambiará el mundo (y tu vida). Trad. Jorge Rizzo. Barcelona: Ediciones Deusto, 2018, p. 109-111.



A particularidade destes algoritmos é que eles aprendem por conta própria, fazendo inferências a partir dos dados. Portanto, o aprendizado de máquina é a capacidade destas de aprender com os dados, identificando tendências e padrões em eventos aparentemente aleatórios. Acaso estes padrões se baseiem em séries de dados longas o suficiente e dotados de suficiente qualidade, podem ser usados para prever o futuro “com uma base científica confiável”<sup>26</sup>. Desta maneira, “o aprendizado de máquina olha para o passado para encontrar padrões e tentar prever o futuro”. Consoante Rodríguez, “em essência, o aprendizado de máquina é pura predição”, tendo por base, os dados e as experiências do passado<sup>27</sup>.

Esses algoritmos vêm sendo utilizados em diversas tarefas dentro do Poder Judiciário. A título exemplificativo, pode-se citar os sistemas *Victoria* do TJRJ e *Elis* do TJPE, que automatizaram o rito das execuções fiscais; o sistema *Mandamus*, que utiliza de técnicas de inteligência artificial para auxiliar na automação do processo de elaboração, distribuição e gerenciamento do cumprimento de mandados judiciais; e, ainda, o mais conhecido deles, *Victor*, utilizado no Supremo Tribunal Federal para otimizar a análise da Repercussão Geral<sup>28</sup>.

Em que pese os modelos de IA desenvolvidos nos Tribunais brasileiros sirvam para apoiar a automação em atividades repetitivas, classificatórias e como ajuda à tomada de decisão judicial, importa destacar os efeitos nocivos que o aprendizado de máquina alimentado com *big datas* pode gerar na sociedade e no sistema de justiça, sobretudo, quando portadores de preconceitos e discriminações. Isso porque esses sistemas catalisam injustiças sociais com a falsa aparência de neutralidade matemática e geram, consoante preleciona O’Neil, “ciclos destrutivos de *feedback*”<sup>29</sup>. Desse modo,

26 RODRÍGUEZ, Pablo. *Inteligencia Artificial. Como cambiará el mundo (y tu vida)*. Trad. Jorge Rizzo. Barcelona: Ediciones Deusto, 2018, p. 133.

27 RODRÍGUEZ, Pablo. *Inteligencia Artificial. Como cambiará el mundo (y tu vida)*. Trad. Jorge Rizzo. Barcelona: Ediciones Deusto, 2018, p. 105.

28 ROSA, Alexandre; GUASQUE, Bárbara. O avanço da Disrupção nos Tribunais Brasileiros. In: NUNES, Dierle; LUCON, Paulo Henrique; WOLKART, Erik Navarro (org.). *Inteligência Artificial e Direito Processual: os impactos da virada tecnológica no Direito Procesual*. 2ªed. rev. atual. e ampl. Salvador: JusPODVIM, 2021. p. 93-121.

29 Conforme Cathy O’Neil, um modelo nada mais é do que a representação abstrata de algum processo, seja um jogo de *beisebol*, a cadeia logística de uma petroleira, as ações de um governo estrangeiro, ou o público de um cinema. Esteja ele rodando dentro de um computador ou na nossa cabeça, o modelo pega o que sabemos e usa isso para prever respostas em situações variadas. Todos nós carregamos milhares de modelos em nossas cabeças. Eles nos dizem o que esperar e guiam nossas decisões. [...] Nenhum modelo consegue incluir toda a complexidade do mundo real ou as nuances da comunicação humana. Inevitavelmente alguma informação importante fica de fora. [...] Para criar um modelo, então, fazemos escolhas sobre o que é importante o bastante para ser incluído, simplificando o mundo numa versão de brinquedo que possa ser facilmente entendida, e a partir da qual possamos inferir fatos e ações importantes. Esperamos que o modelo lide com apenas um trabalho e aceitamos que irá ocasionalmente agir como uma máquina ignorante com enormes pontos cegos. [...] Os pontos cegos de um modelo refletem o julgamento e prioridades de seus criadores. Cfr. O’NEIL. Cathy. *Algoritmos de*

os vieses algoritmos ocorrem quando o algoritmo adquire e reflete os valores humanos, os mesmos desvios culturais que estão implícitos nos dados que são utilizados para treinamento do modelo. E isso acaba por enviesar o resultado final que é obtido.

Isto quer dizer que os modelos algorítmicos, apesar de desfrutarem de uma reputação de imparcialidade, são “opiniões embutidas em matemática”, de modo que expressam os objetivos e ideologias de seus criadores e “seus pontos cegos refletem as prioridades e o julgamento de quem os alimentou”. Assim, a depender de quem construa esses modelos, quais variáveis leve em conta, e com que dados os alimente, o resultado será um ou outro<sup>30</sup>. Em miúdos, algoritmos não são neutros, não são puramente matemáticos.

É claro que não se está a dizer que o algoritmo cria preconceitos. No entanto, é certo que ele repete um padrão contemplado nos dados utilizados para treinamento do modelo. Desse modo, partindo da premissa de que um algoritmo comporta uma sequência de regras para se chegar a um resultado, se, embutido nesta sequência há um desvio, ter-se-á um problema no resultado final. Conforme se verifica, o preconceito não brota da sequência de regras (que, aliás, costuma ser objetiva), mas dos dados que reproduzem padrões e preconceitos sociais.

Mas, por que isso ocorre? Isso acontece porque os modelos algoritmos são, em grande medida, estatísticos. Então, a princípio, eles não fazem nenhum pré-julgamento, não estabelecem regras, tampouco efetuam valorações. Eles somente aprendem com os dados mediante análises estatísticas, ou seja, eles apontam o que acontece habitualmente. Em outras palavras, os modelos algoritmos refletem uma dinâmica social predominante.

#### **4 Algoritmos e racismo: perpetuação e massificação de práticas sociais discriminatórios**

Em 2018, a Amazon decidiu descartar um modelo que a empresa utilizava para fazer a seleção dos currículos para novas contratações. Ele apresentava um viés sexista que desprezava os currículos femininos, atribuindo *score* alto apenas para currículos masculinos. A discriminação da ferramenta, contra candidatas do sexo feminino, acontecia porque os dados utilizados para treinamento do modelo foram currículos enviados para a empresa nos últimos 10 anos, os quais são, em sua imensa maioria, de homens, como acontece na maior parte da indústria de tecnologia<sup>31</sup>. Consoante o *AI*

---

*Destruição em Massa: como o Big Data aumenta a desigualdade e ameaça à democracia*. Tradução de Rafael Abraham. 1. ed. Santo André/SP: Editora Rua do Sabão, 2020, p. 22-31.

30 O'NEIL. Cathy. *Algoritmos de Destruição em Massa: como o Big Data aumenta a desigualdade e ameaça à democracia*. Tradução de Rafael Abraham. 1. ed. Santo André/SP: Editora Rua do Sabão, 2020, p. 30-35.

31 REUTERS. Amazon scraps secret AI recruiting tool that showed bias. Disponível em: <https://www.>

*Index*, relatório independente sobre o setor da tecnologia, os homens representam 71% dos candidatos a vagas de empregos na área da IA nos Estados Unidos<sup>32</sup>. Diante deste contexto, o algoritmo passou a entender que os candidatos homens eram mais aptos para as vagas.

O'Neil explica que, como o aprendizado de máquina se baseia nos dados do passado, o algoritmo constatou, estatisticamente, que não somente os homens eram maioria, como eram promovidos com mais frequência e recebiam mais aumento salarial. Em contrapartida, as mulheres eram minoria, tendiam a sair rapidamente e recebiam menos aumento. Logo, o algoritmo entendeu que os homens eram as melhores contratações, perpetuando e ampliando, por meio do modelo algoritmo, um viés histórico, sexista e misógino<sup>33</sup>.

Outro estudo, agora efetuado pelo NIST – *National Institute of Standards and Technology*, agência governamental norte-americana, testou 189 algoritmos de 99 desenvolvedores. A constatação foi que a tecnologia de reconhecimento facial é, para a correspondência um-para-um, entre 10 a 100 vezes menos eficaz para rostos asiáticos e afro-americanos em relação às imagens de caucasianos. Para a correspondência um-para-muitos a equipe observou taxas mais altas de falsos positivos para mulheres afro-americanas<sup>34</sup>. Diferenciais em falsos positivos, na correspondência um-para-muitos, são particularmente importantes, porque as consequências podem incluir acusações falsas<sup>35</sup> – o que vem acontecendo com indesejável frequência<sup>36</sup>.

---

reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G. Acesso em: 10.dez. 2021.

32 THE GUARDIAN. Disastrous' lack of diversity in AI industry perpetuates bias, study finds.

Disponível em: <https://www.theguardian.com/technology/2019/apr/16/artificial-intelligence-lack-diversity-new-york-university-study>. Acesso em: 15 abr. 2022.

33 O'NEIL. Cathy. Amazon's Gender-Biased Algorithm Is Not Alone. Bloomberg. Disponível em:

<https://www.bloomberg.com/opinion/articles/2018-10-16/amazon-s-gender-biased-algorithm-is-not-alone>. Acesso em: 10 set. 2021.

34 NIST. NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software. Disponível

em: <https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software>. Acesso em: 15 dez. 2021.

35 O estudo do NIST se concentrou em quão bem cada algoritmo individual executa uma das duas tarefas diferentes que estão entre os aplicativos mais comuns de reconhecimento de rosto. A primeira tarefa, confirmar que uma foto corresponde a uma foto diferente da mesma pessoa em um banco de dados, é conhecida como correspondência “um para um” e é comumente usada para trabalhos de verificação, como desbloquear um smartphone ou verificar um passaporte. A segunda, que determina se a pessoa na foto tem alguma correspondência em um banco de dados, é conhecida como correspondência “um para muitos” e pode ser usada para identificar uma pessoa de interesse. NIST. *NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software*. Disponível em: <https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software>. Acesso em: 15 dez. 2020.

36 THE NEW YORK TIMES. *Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match*.

Disponível em: <https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail>.

Num recorte de gênero, estes casos possuem, como raízes e denominador comum, a predominância maciça de homens brancos na indústria da tecnologia e a consequente ausência de diversidade nas equipes desenvolvedoras. Uma pesquisa publicada pelo *AI Now Institute*, instituto de pesquisa que estuda as implicações sociais da inteligência artificial, constatou que a ausência de diversidade dentro das equipes vem contribuindo para a criação de sistemas falhos e que perpetuam preconceitos de gênero e raça. Apenas 15% dos pesquisadores de IA do Facebook e, 10% do Google, são mulheres. As mulheres representam apenas 18% dos graduados em ciência da computação nos Estados Unidos<sup>37</sup>.

Com relação a população negra, os números são ainda menores, pois correspondem a apenas 2,5% dos funcionários do Google, 4% no Facebook e na Microsoft e 6% no Twitter. A força de trabalho da Apple contempla 9% de negros, todavia, esta estatística inclui funcionários de varejo. Assim como a Amazon, que possui 26,5% de funcionários negros, porém a maioria ocupando empregos de baixa remuneração, sendo que apenas 8,3% deles estão em cargos de gerência<sup>38</sup>.

A participação de negros na indústria da tecnologia ainda é inexpressiva. De acordo com a *U.S. Equal Employment Opportunity Commission*, no ano de 2014, apenas 8,6% dos graduados com bacharelado em ciência da computação e informática eram negros<sup>39</sup>. Evidências sugerem que a IA, como um campo, é ainda menos diversificada do que a ciência da computação como um todo<sup>40</sup>.

As estatísticas revelam um setor dominado por homens brancos e asiáticos. Sendo que os demais grupos permanecem sub-representados na seara da tecnologia e da IA. A limitação de raça, gênero, de áreas e contextos, no desenvolvimento do modelo, conduz a uma cegueira do algoritmo quanto aos elementos que não estão presentes naquele enquadramento. Então, por exemplo, nos algoritmos de reconhecimento facial, como as equipes de desenvolvimento são majoritariamente compostas por homens brancos, empreende-se uma incapacidade do algoritmo em identificar pessoas negras, mormente mulheres negras. E isso ocorre porque não há componentes negros dentro

---

html; UOL. *Sistema de reconhecimento facial erra, e homem negro é preso por engano*. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2020/06/25/homem-e-presos-apos-erro-de-tecnologia-de-reconhecimento-facial-nos-eua.htm>; AGENCIA BRASIL. *Defensoria tenta libertar jovem preso no RJ por reconhecimento facial*. Disponível em: <https://agenciabrasil.ebc.com.br/radioagencia-nacional/direitos-humanos/audio/2020-10/defensoria-tenta-libertar-jovem-presos-no-rj-por-reconhecimento-facial>. Acesso em: 20 dez. 2020.

37 AI NOW INSTITUTE. *AI Now Report 2018*. Disponível em: <https://ainowinstitute.org/>. Acesso em: 16 dez. 2021.

38 AI NOW INSTITUTE. *AI Now Report 2018*. Disponível em: <https://ainowinstitute.org/>. Acesso em: 16 dez. 2021.

39 U.S. Equal Employment Opportunity Commission. *Diversity in High Tech*. Disponível em: <https://www.eeoc.gov/special-report/diversity-high-tech>. Acesso em: 15.dez.2021.

40 AI NOW INSTITUTE. *AI Now Report 2018*. Disponível em: <https://ainowinstitute.org/>. Acesso em: 16 dez. 2021.

das equipes, tampouco nos bancos de dados, os quais poderiam auxiliar a treinar os respectivos modelos. Não há, pois, representatividade estatística.<sup>41</sup>

Conforme os cientistas da computação constroem e treinam os algoritmos, eles se concentram nas características faciais que são mais visíveis em determinada raça, mas não nas demais. Ademais, normalmente se baseiam em bibliotecas de códigos pré-existentes, normalmente escritas por outros pesquisadores brancos. O resultado final é voltado para focar em rostos caucasianos. Isto restou evidente no estudo do NIST<sup>42</sup>, que demonstrou que algoritmos de reconhecimento facial que são desenvolvidos por empresas asiáticas são muito mais precisos em rostos asiáticos, comprovando que, as características de gênero e raça dos criadores dos modelos, detêm forte influência sobre o seu resultado final.

Mesmo que o software seja desenvolvido para aprender com a experiência e tornar-se mais preciso com técnicas de aprendizado de máquina, os conjuntos de dados de treinamento são geralmente compostos por pessoas brancas. Portanto, o algoritmo especializa-se e aprimora-se somente na identificação de indivíduos brancos.

Isto pode acontecer com relação à raças e gêneros, mas também pode acontecer com determinados contextos e classes sociais que, se não forem bem representados no treinamento do modelo, não terão uma representatividade estatística. Desse modo, formar-se-á, provavelmente, um ponto cego no modelo, capaz de carrear danosas consequências sociais e, conseqüentemente, lesar Direitos Fundamentais constitucionalmente assegurados.

O *software Rekognition*, desenvolvido pela Amazon para identificar pessoas em vídeos e imagens, é utilizado por inúmeros órgãos policiais dos Estados Unidos. Contudo, este *software* possui falhas graves no reconhecimento de pessoas como falso positivo. Um estudo efetuado pela União Americana pelas Liberdades Civas, em 2018, testou o sistema submetendo fotos dos congressistas norte-americanos. O sistema identificou 28 dos deputados e senadores como criminosos procurados pela polícia. Ainda, “as falsas correspondências eram desproporcionalmente de pessoas de cor”. Cerca de 40% dos congressistas, erroneamente identificados como criminosos, eram negros, embora eles representassem apenas 20% dos membros do Congresso<sup>43</sup> (ACLU,

41 Representatividade estatística refere-se à “determinação da qualidade de uma amostra constituída de modo a corresponder à população no seio da qual ela é escolhida” Cfr. FREITAS, Cinthia Obladen de A. A obscuridade dos algoritmos e a LGPD. O Estado de São Paulo, 4.ago.2020. Disponível em: <https://politica.estadao.com.br/blogs/fausto-macedo/a-obscuridade-dos-algoritmos-e-a-lgpd/>. Acesso em: 10 dez. 2020.

42 NIST. *NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software*. Disponível em: <https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software>. Acesso em: 15 dez. 2020.

43 ACLU ORG. *Amazon’s Face Recognition Falsely Matched 28 Members of Congress*. Disponível em: <https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28>. Acesso em: 15 dez. 2021.



2018). Ocorre que, por ser um sistema muito acessível, o *Rekognition* vêm sendo facilmente adquirido e comercializado para utilização na segurança pública, o que deve exacerbar o falso reconhecimento positivo de pessoas negras, intensificando o racismo que permeia no Sistema Penal.

Também é importante destacar os efeitos nocivos que o aprendizado de máquina alimentado com *big datas* que retratam preconceitos e discriminações podem gerar na sociedade e no sistema de justiça, catalisando injustiças sociais com a falsa aparência de neutralidade matemática<sup>44</sup>.

Um dos exemplos mais conhecido é o sistema *COMPAS*, utilizado nos EUA, sistema que escancarou a prática de racismo refletido no emprego dos Sistemas de Inteligência Artificial. O *COMPAS* (*Correctional Offender Management Profiling for Alternative Sanctions*, ou Perfil de Gerenciamento Corretivo de Infratores para Sanções Alternativas) é uma ferramenta desenvolvida e comercializada pela *Northpointe* e foi ela elaborada com o intuito de prever a reincidência criminal de condenados (quase que um trabalho de futurologia) e, por consequência, auxiliar os juízes na definição de quais programas de tratamento ou de condicional o réu seria submetido. Em casos de prisões em flagrante, o sistema é utilizado para decidir se a pessoa deve ser mantida na prisão ou colocada em liberdade sob o recolhimento de fiança; uma vez condenada, o sistema auxilia em decisões tais como se poderia o réu apelar em liberdade ou se teria que se recolher a prisão para recorrer. Curiosa é a maneira de como isso se dá. Os réus são submetidos a um questionário. Suas respostas são inseridas no *COMPAS* e geram um *score*, incluindo previsões de “risco de reincidência” e “risco de reincidência violenta”. São várias as perguntas que avaliavam (num gradiente de 0 a 10) o quanto o sujeito poderia ser capaz de voltar a cometer um crime futuramente (um tipo de exercício de futurologia).<sup>45</sup>

Contudo, o algoritmo passou a ser utilizado não somente para negar liberdade condicional, com base na alta possibilidade de reincidência demonstrada no *score* do algoritmo, como passou a servir de base para o aumento da pena, com base neste mesmo critério.

O primeiro grande problema desta circunstância, é que a utilização deste algoritmo para justificar um aumento de pena solapa importantes Direitos

44 O'NEIL. Cathy. Algoritmos de Destruição em Massa: como o Big Data aumenta a desigualdade e ameaça a democracia; tradução Rafael Abraham. 1. ed. Santo André, São Paulo: Editora Rua do Sabão, 2020.

45 Os problemas, é claro, não estão nas respostas, mas nas perguntas. O Sistema perguntava “se alguém na família já havia sido preso; se a pessoa vivia numa área com alto índice de criminalidade; se tinha amigos que fazem parte de gangues; tratava do histórico escolar; se exercia atividade remunerada lícita; se usava drogas ou se convivia com alguém que faz uso, etc.” E também tinham perguntas sobre aquilo que o sistema chamava de “pensamentos criminosos”, como por exemplo, se a pessoa concorda ou não com a afirmação: “é aceitável que alguém que passe fome pratique furtos”. Cfr. PROPUBLICA. How we analyzed the compas recidivism algorithm. 23. Mai. 2016. Disponível em: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Acesso em: 20 out. 2020.



Fundamentais, tais como a legalidade e o Devido Processo Legal. Como bem delineou O'Neil, o questionário utilizado para avaliar o réu inclui informações sobre o local de moradia do réu, seus familiares e amigos. São detalhes que passam ao largo do fato típico e das circunstâncias previstas legalmente para justificar um aumento de pena. Nesta conjuntura, o réu tem sua liberdade condicional negada, assim como a sua pena base aumentada, não por circunstâncias legais ou devido ao crime por ele praticado, mas devido ao seu local de moradia, suas relações de amizade, seus familiares, etc. Por meio do *score* determinado pelo algoritmo, os réus são julgados por serem quem são e no meio em que vivem, e não pela conduta praticada<sup>46</sup>. Ou seja, a partir dessas perguntas, o sistema realiza aquilo que a gente conhece por Direito Penal de Autor, na medida em que o sujeito não é criminalizado por aquilo que ele supostamente fez, mas pelo que ele é, ou mesmo, pelo registro histórico das coisas que a ele são atribuídas.

Mas, a teratologia vai além. Um estudo feito pela *ProPublica*, analisou mais de dez mil réus criminais no Condado de *Broward*, Flórida. A pesquisa comparou as taxas de reincidência previstas pelo algoritmo com as taxas reais de reincidência dos réus nos dois anos após terem sido pontuados. Os resultados da análise demonstraram que os réus negros tinham 45% mais chances de receber pontuações de risco mais altas do que os réus brancos. Os réus negros também tinham duas vezes mais probabilidade de serem classificados erroneamente com um risco maior de reincidência violenta<sup>47</sup>.

Ou seja, o algoritmo detinha um viés que considerava réus negros mais perigosos que os demais. Isto porque ele extraiu esse padrão dos dados que foram utilizados para treinar o sistema. No entanto, não existiam erros nos dados fornecidos. Eles estavam corretos. Em que pese constituam somente 13% da população, os negros ocupam 40% das vagas em presídios nos Estados Unidos<sup>48</sup>. Diante disso, o modelo efetuou uma análise estatística, levando-o a adotar o padrão de que os negros eram mais perigosos. Por consequência, ao receberem *scores* mais altos do que o de réus brancos, os réus negros não somente têm negado a liberdade condicional, mas também lhes são atribuídas condenações com penas mais altas. De acordo com a Associação Americana para Liberdades Civis (ACLU), as sentenças impostas a homens negros são cerca de 20% maiores do que os condenados brancos por crimes similares<sup>49</sup>.

46 O'NEIL. Cathy. Algoritmos de Destruição em Massa: como o Big Data aumenta a desigualdade e ameaça a democracia; tradução Rafael Abraham. 1. ed. Santo André, São Paulo: Editora Rua do Sabão, 2020, p. 43.

47 PROPUBLICA. How we analyzed the compas recidivism algorithm. 23. Mai. 2016. Disponível em: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Acesso em: 20 out. 2021.

48 O'NEIL. Cathy. Algoritmos de Destruição em Massa: como o Big Data aumenta a desigualdade e ameaça a democracia; tradução Rafael Abraham. 1. ed. Santo André, São Paulo: Editora Rua do Sabão, 2020, p. 39.

49 O'NEIL. Cathy. Algoritmos de Destruição em Massa: como o Big Data aumenta a desigualdade e ameaça a democracia; tradução Rafael Abraham. 1. ed. Santo André, São Paulo: Editora Rua do Sabão, 2020, p. 40.

Em que pese o modelo não contemplasse o aspecto raça no questionário, constavam entre as perguntas o local de moradia do réu. E este fator, diante do contexto social segregador dos Estados Unidos, era um forte indicativo da raça do acusado. Então, ainda que a raça não seja um parâmetro utilizado no julgamento, as perguntas efetuadas acabaram por extrair uma informação racial, e os dados estatísticos refletem o racismo estrutural, prejudicando o resultado final do modelo, tornando-o racista. Perpetua-se, desta maneira, o racismo que contamina o Sistema de Justiça Criminal estadunidense, e com ares de cientificidade.

Outro retrato pernicioso que estampa esta conjuntura se encontra na ferramenta *Ring*, da *Amazon*. O *Ring* consiste em campainhas inteligentes que são conectadas a câmeras de vigilância e que, tem por fito, o combate ao crime. Referido sistema é disponibilizado para ser utilizado em bairros residenciais, valendo-se da rede cada vez maior de câmeras de segurança para alertar quando um indivíduo considerado suspeito for enquadrado por alguma câmera do bairro. No entanto, a *Amazon* estabeleceu “parcerias secretas” com departamentos de polícia dos Estados Unidos, munindo as polícias com “câmeras onipresentes” e um sistema de vigilância muito mais amplo e poderoso<sup>50</sup>. Algumas cidades criaram incentivos pecuniários para colocar as câmeras do *Ring* em bairros periféricos e naqueles em que os índices de notificações criminais são mais altos, permitindo aos departamentos de polícia amplo acesso às filmagens. Resultado: a utilização desta ferramenta pelas forças policiais permite uma vigilância excessiva e invasiva em comunidades de baixa renda e de maioria negra, gerando um policiamento desproporcional que aumentou, por consequência, a violência policial nestas comunidades<sup>51</sup>.

Ainda, a ferramenta permite e incentiva que policiais, em conjunto com proprietários dos bairros e condomínios, criem listas de pessoas indesejáveis circulando pelas áreas residenciais nobres. Permite que os moradores de bairros de alta renda façam julgamentos sobre pessoas que consideram desagradáveis, convocando a polícia para confrontá-los<sup>52</sup>.

O aplicativo *Neighbours*, do *Ring* já firmou parceria com mais de 1.300 forças policiais nos Estados Unidos - um aumento de 300% em relação a agosto de 2019<sup>53</sup>. Novamente, uma pesquisa efetuada pelo *Motherboard* revelou que pessoas negras são

50 SADOWSKI, Jathan. *Too Smart: How Digital Capitalism is Extracting Data, Controlling Our Lives*. Massachusetts: The MIT Press, 2020, p. 153.

51 SADOWSKI, Jathan. *Too Smart: How Digital Capitalism is Extracting Data, Controlling Our Lives*. Massachusetts: The MIT Press, 2020, p. 154.

52 SADOWSKI, Jathan. *Too Smart: How Digital Capitalism is Extracting Data, Controlling Our Lives*. Massachusetts: The MIT Press, 2020, p. 154.

53 THE INTERCEPT. Amazon's Ring Planned Neighborhood Watch lists built on facial recognition. Disponível em: <https://theintercept.com/2019/11/26/amazon-ring-home-security-facial-recognition/>. Acesso em: 15 dez. 2020.

mais propensas a serem vigiadas pelo aplicativo *Neighbours*<sup>54</sup>.

Segundo Kelley e Guariglia, da *Electronic Frontier Foundation*,<sup>55</sup> as campanhas de vigilância de *Ring* e o aplicativo *Neighbours*, possuem papel protagonista em “permitir e perpetuar o assédio policial aos negros americanos”. Estas ferramentas incentivaram “os piores instintos de muitos residentes”, incitando-os a espionar pedestres, vizinhos e trabalhadores<sup>56</sup>.

Nesta mesma linha está o *software PredPol*, utilizado por departamentos de polícia para prever lugares onde crimes futuros podem acontecer. Programas como este vêm sendo amplamente utilizados em departamentos de polícia que dispõem de reduzido orçamento nos Estados Unidos. O *PredPol* é baseado em software sísmico, ou seja, à medida que ele vê um crime acontecer em dado local, ele incorpora em padrões de histórico a fim de prever os locais aonde novos crimes poderão acontecer futuramente. Diferentemente do COMPAS, este tipo de software não se concentra no indivíduo, levando em conta apenas a localização geográfica. No entanto, ao configurar o sistema *PredPol*, os departamentos de polícia inserem, como dados de treinamento, relatórios policiais de crimes anteriores. E estes relatórios contemplam majoritariamente crimes de perturbação (já que os policiais foram todos treinados na ortodoxia da política da “tolerância zero” e na ideologia da Defesa Social<sup>57</sup>).

É fato que esses tipos de notificações de práticas delitivas são endêmicos em bairros de baixa renda, o que alimenta o sistema com mais casos, gerando alta

---

54 MOTHERBOARD. Amazon's Home Security Company Is Turning Everyone Into Cops. Disponível em: <https://www.vice.com/en/article/qvyvzd/amazons-home-security-company-is-turning-everyone-into-cops>. Acesso em: 15 dez. 2020.

55 Fundada em 1990, a Electronic Frontier Foundation é a principal organização sem fins lucrativos que defende as liberdades civis no mundo digital (EFF, 2020).

56 Cfr. Amazon Ring Must End Its Dangerous Partnerships With Police, disponível em: <https://www.eff.org/deeplinks/2020/06/amazon-ring-must-end-its-dangerous-partnerships-police>. Acesso em: 21 dez. 2021.

57 A ideologia da Defesa Social foi construída pelo “saber oficial” (desde a Escola Clássica, passando pela Escola Positiva e chegando a Técnico-Jurídica) e filtrada através do seu debate escolar, veio a se constituir não apenas ideologia dominante na Ciência Penal, na Criminologia e nos representantes do sistema penal, mas no “saber comum do homem de rua” sobre a criminalidade e a pena (Cfr. ANDREADE, Vera Regina Pereira de. A ilusão de segurança jurídica: do controle da violência a violência do controle penal. 2ª ed. Porto Alegre: Livraria do Advogado, 2003, p. 135-136). De acordo com Alessandro Baratta (Cfr. BARATTA, Alessandro. Criminologia crítica e crítica do direito penal: introdução à sociologia do direito penal. Tradução de Juarez Cirino dos Santos. 3. ed. Rio de Janeiro: Revan, 2002, p. 42-43), os princípios que dão sustentação à ideologia da defesa social são: o princípio do “bem e do mal”, o princípio da culpabilidade, o princípio da legitimidade (do contrato social), o princípio da prevenção, o princípio da igualdade e o princípio do interesse social. É bom esclarecer que os teóricos da Criminologia Crítica e da Sociologia do Direito Penal têm desconstruído cada uma dessas pilastras, mostrando que o que há, na realidade, é uma ilusão de combate à criminalidade, notadamente, através da violência seletiva das instituições oficiais.

pontuação no *PredPol*<sup>58</sup>. A consequência é que o sistema acaba prevendo como pontos críticos para cometimento de crimes sempre os mesmos lugares, mormente bairros de baixa renda e maioria negra, angariando um ciclo contínuo de policiamento excessivo e abordagens violentas nestes bairros. Logo, o *PredPol* acaba por automatizar os preconceitos e o racismo com ares de evidência matemática e a falsa percepção de legitimidade, pois opera mediante algoritmos “imparciais”.

Diante destas circunstâncias, os presídios estão repletos de condenados por crimes sem vítimas determinadas (crimes de perturbação), a grande maioria provenientes de bairros de baixa renda, negros e imigrantes hispânicos. De outra parte, os crimes de fraude contra o mercado imobiliário e o mercado de ações (que notadamente trouxeram profundas repercussões nefastas para milhares de pessoas que perderam suas casas, empregos e planos de saúde em decorrência da crise financeira de 2008) não preenchem os relatórios policiais nem as estatísticas do *PredPol*. Conforme já se verificou nas linhas anteriores, as raías do aparato coercitivo do Estado sempre foram voltadas e ampliadas para a população negra e pobre<sup>59</sup>. Mas, agora, conforme anota O’Neil<sup>60</sup>, criminaliza-se esse grupo preferencial se apoiando em ferramentas que se acredita não apenas serem científicas, mas justas.

Recentemente, impulsionados pela onda de protestos que tomou os Estados Unidos da América em decorrência da morte de George Floyd, a *Amazon* decidiu suspender, por um ano, o seu sistema de reconhecimento facial *Rekognition*. Porém, manteve intacta as parcerias para a utilização do aplicativo *Neighbours*. Sem prestar explicações detalhadas sobre a real motivação da iniciativa, a companhia afirmou apenas que defende que os governos implementem regulamentações mais rígidas para o uso ético da tecnologia de reconhecimento facial, principalmente nos Estados Unidos. A IBM aproveitou a ocasião para declarar que cancelou, de maneira definitiva, o desenvolvimento de *software* de reconhecimento facial<sup>61</sup>. Traduz-se em forte indicativo de que as ferramentas de reconhecimento facial são ineficientes, além de produzirem danos sociais seletivos e exclusão.

Além da problemática da ausência de diversidade, todos estes exemplos de modelos enviesados também se devem ao fato de que os *softwares* não são auditados,

58 O’NEIL. Cathy. Algoritmos de Destruição em Massa: como o Big Data aumenta a desigualdade e ameaça a democracia; tradução Rafael Abraham. 1. ed. Santo André, São Paulo: Editora Rua do Sabão, 2020, p. 135-137.

59 WACQUANT, Loïc. Punir os Pobres: a nova gestão da miséria nos Estados Unidos. 3. Ed. Tradução de Sérgio Lamarão. Rio de Janeiro: Editora Revan, 2007.

60 O’NEIL. Cathy. Algoritmos de Destruição em Massa: como o Big Data aumenta a desigualdade e ameaça a democracia; tradução Rafael Abraham. 1. ed. Santo André, São Paulo: Editora Rua do Sabão, 2020, p. 140-144.

61 THE GUARDIAN. Amazon to ban police use of facial recognition software for a year. Disponível em: <https://www.theguardian.com/technology/2020/jun/10/amazon-rekognition-software-police-black-lives-matter>. Acesso em: 14 dez. 2021.

ou seja, não estão expostos a legislação que os regulamente e fiscalize. Também não são obrigados a submeterem-se a revisão para verificar se há enviesamento embutido. Tampouco são abertos para serem revisados e auditados por pessoas de fora. Na maioria dos casos, inclusive, sequer é dado ciência à população de que está sendo submetida a um algoritmo.

A empresa *CyberExtruder*, por exemplo, criadora de um *software* de reconhecimento facial para ser utilizado na segurança pública, admite não haver realizado testes para aferir vieses em sua ferramenta. Apenas confirma que certas cores de pele são simplesmente mais difíceis para o *software* lidar, dadas as limitações atuais da tecnologia<sup>62</sup>.

A *Northpointe*, que fornece o software COMPAS, por sua vez, contestou as conclusões do relatório efetuado pela ProPublica, mas se recusou a revelar o funcionamento interno do programa, o qual considera comercialmente sensível<sup>63</sup>.

## 5 Inteligência artificial e racismo no Brasil

A criticável experiência do uso de sistemas de Inteligência Artificial no âmbito dos segmentos de Controle Penal estadunidenses nos últimos anos se apresenta como referência quase que obrigatória ao Estado Brasileiro, notadamente, pela via da utilização de instrumentos que promovem o reforço ao etiquetamento de uma criminalidade proveniente de setores sociais bem determinados.

É que, mesmo diante do enviesamento dos modelos e da baixa efetividade das opções disponíveis, as Secretarias de Segurança Pública de vários estados brasileiros adquiriram e fazem uso de tecnologia de reconhecimento facial para auxiliar na prisão de procurados pela polícia. Hoje, os softwares operam em estados como Bahia, Rio de Janeiro, Ceará, São Paulo, Santa Catarina e Paraíba. Tecnologias que, sabidamente, possuem as maiores falhas operacionais relativas ao falso reconhecimento positivo de pessoas negras.

Veja-se que, após cerca de um ano de utilização, as estatísticas relativas ao enviesamento do modelo já começaram a aparecer. Levantamento realizado pela Rede de Observatórios da Segurança constatou que, das 151 prisões pela via do sistema de reconhecimento facial que aconteceram no país, 90% eram de pessoas negras<sup>64</sup>.

62 THE GUARDIAN. How white engineers built racist code – and why it’s dangerous for black people. Disponível em: <https://www.theguardian.com/technology/2017/dec/04/racist-facial-recognition-white-coders-black-people-police>. Acesso em: 17 abr. 2022.

63 THE GUARDIAN. Rise of the racist robots – how AI is learning all our worst impulses. Disponível em: <https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses>. Acesso em: 15 abr. 2022.

64 REDE DE OBSERVATÓRIOS DA SEGURANÇA. Retratos da Violência. Cinco meses de monitoramento, análises e descobertas. Disponível em: <http://observatorioseguranca.com.br/wp->



Minas Gerais, Espírito Santo, Pará e Distrito Federal são Estados que também já contrataram esta tecnologia para utilização na Segurança Pública. Mas, brevemente, essas tecnologias devem ser utilizadas em todos os estados brasileiros. Isto se deve ao fato de que o Governo Federal aprovou o uso do dinheiro do Fundo Nacional de Segurança Pública para a implantação de reconhecimento facial, além de outras práticas de Segurança Pública envolvendo tecnologias.

Assim, em que pese o intenso movimento mundial que vem escancarando a discriminação algorítmica e requerendo o banimento da utilização de *softwares* de reconhecimento facial,<sup>65</sup> a adoção dessas ferramentas está se alastrando pelos Estados e municípios brasileiros, incentivada e financiada pelo Governo Federal.<sup>66</sup>

E muitos problemas decorrem do emprego dessas tecnologias. Não há, por exemplo, regulamentação que exija e fiscalize os protocolos dos *softwares* utilizados. Não há regulamentação que estabeleça os cuidados com relação aos dados para treinamento dos modelos, nível de acurácia, possibilidade de auditoria externa, transparência com relação ao uso e, nem mesmo, como se dá proteção dos dados coletados. Apesar de o art. 4º da LGPD (Lei Geral de Proteção de Dados) estabelecer expressamente a necessidade de aprovação de lei específica para Segurança Pública e investigação criminal, até o presente momento isso não se materializou.

Essa ausência de regulamentação, responsabilidade, auditoria e transparência na utilização da IA, vem angariando consequências nefastas do ponto de vista jurídico-social, por perpetua e amplia as discriminações e a seletividade que permeia o controle penal e as agências que o integram. E tudo sob a ilusória impressão de que se baseia em modelos matemáticos imparciais, precisos e objetivos. Em verdade, os resultados estão carregados de subjetividade e vieses racistas tendentes a catalisar setores sociais bem delimitados para a marginalização e a discriminação. Conforme bem apontado na exposição de motivos do Anteprojeto de Lei de Proteção de Dados para Segurança Pública e Persecução Penal:

[...] há um enorme déficit de proteção dos cidadãos, visto que não há regulação geral sobre a licitude, a transparência ou a segurança do tratamento de dados em matéria penal, tampouco direitos estabelecidos ou requisitos para utilização de novas tecnologias que possibilitam um grau de vigilância e monitoramento impensável há alguns anos. Apesar do crescimento vertiginoso de novas técnicas de vigilância e de investigação, a ausência de regulamentação sobre o tema gera uma assimetria de poder muito grande entre os atores envolvidos (Estado e cidadão). Nesse contexto, o titular

content/uploads/2019/11/1relatoriorede.pdf. Acesso em: 15 abr. 2022.

65 Cfr. A título exemplificativo, a organização *Algorithmic Justice League*, o *Big Brother Watch*, a *Liberty Human Rights*, a campanha *Ban Facial Recognition* e a *Internet Freedom Foundation*.

66 Cfr. Portaria nº 793 de 24/10/2019.



dos dados é deixado sem garantias normativas mínimas e mecanismos institucionais aplicáveis para resguardar seus direitos de personalidade, suas liberdades individuais e até a observância do devido processo legal.

Desse modo, há que se voltar uma atenção rigorosa sobre essas tecnologias, em especial, para que não contribuam ou intensifiquem aquilo que o Sistema Penal já tem de pior. Partindo-se da premissa (já bastante segura) de que os sistemas não são comprovadamente efetivos e confiáveis, parece bastante óbvio que o emprego desses instrumentos livres de regulamentação e transparência irá agravar ainda mais o controle penal e a violência que é ínsita à sua instrumentalização direcionado à população negra e pobre (a exemplo de prisões ilegais e arbitrárias de pessoas negras, mediante o falso reconhecimento positivo). Modelos como estes, que possuem pontos cegos que os tornam discriminatórios, servem apenas para reforçar padrões racistas socialmente difundidos, potencializados e perpetuados em virtude do viés de automação.<sup>67</sup>

Por isso, no Brasil, a proibição da utilização de reconhecimento facial para fins de Segurança Pública é de importância primeira. Os falhos *softwares* de reconhecimento facial constituem tecnologias sobrepostas em alicerces discriminatórios, capazes de dinamizar a seletividade do Sistema Penal.

Por aqui, da forma que se apresentam, o seu emprego significa muitos retrocessos, notadamente diante da violação de Direitos Fundamentais marcada pela dificuldade de se contraditar autoria de fato diante de um falso reconhecimento positivo. De igual modo, os modelos preditivos, utilizados para auxiliar o policiamento ou sentenças penais, a exemplo do *Compas* e do *PredPol*, não podem ser permitidos, ainda que dotados de supervisão humana.<sup>68</sup> Modelos que forem treinados e alimentados a partir de estatísticas originalmente racistas servirão para retroalimentar essas discriminações e, conforme exposto, sob um manto de “cientificidade matemática”.

Com relação aos demais modelos algorítmicos, é fundamental uma preocupação especial com os dados que são utilizados para treinamento dos modelos, assim como as

67 Cfr. “O viés de automação se apresenta como uma das espécies dos vieses cognitivos humanos que ocorre pela propensão de favorecer sugestões de sistemas automatizados de tomada de decisão. Isso ocorre quando o humano sobrevaloriza a resposta da máquina e passa a não refletir acerca da correção de seus resultados. Tal viés conduz as pessoas a não reconhecerem quando os sistemas automatizados erram e a seguirem seus resultados quando apresentadas informações contraditórias”. Cfr. NUNES, Dierle. A supervisão humana das decisões de inteligência artificial reduz os riscos? Consultor Jurídico. Disponível em: <https://www.conjur.com.br/2021-jun-25/nunes-supervisao-humana-decisoes-ia-reduz-riscos>. Acesso em: 26 jun. 2022.

68 Vide: NUNES, Dierle. A supervisão humana das decisões de inteligência artificial reduz os riscos? Consultor Jurídico. Disponível em: <https://www.conjur.com.br/2021-jun-25/nunes-supervisao-humana-decisoes-ia-reduz-riscos>. Acesso em: 26 jun. 2022.

variáveis que são levadas em conta. Afinal, o algoritmo não cria o preconceito, apenas repete os padrões contemplados nos dados de treinamento e validação dos modelos (matemáticos ou probabilísticos). Portanto, o desafio está em controlar a base de dados e fiscalizar as correlações realizadas pelo aprendizado de máquina. É imprescindível, assim, uma conscientização e engajamento para o desenvolvimento de modelos que não perpetuem os preconceitos inseridos nos dados sociais, contaminando, assim, desde os segmentos de polícia até o sistema de justiça.

## 6 Alternativas hábeis a atenuar a possibilidade de enviesamento dos modelos

Apresenta-se, então, um grande desafio: impedir que as ferramentas tecnológicas reproduzam as limitações, falhas e preconceitos já enraizados e normalizados no âmbito da sociedade e nos segmentos de controle penal.

Para tanto, é imperioso lançar uma atenção rigorosa sobre a base de dados utilizada quando do treinamento e validação dos modelos para que os impactos fundados em diferenças biológicas não se perpetuem com as práticas de separação e eliminação diagnosticados por Foucault, mas, agora, instrumentalizados em novas tecnologias com ares de neutralidade matemática.

Os dados constituem a matéria prima para o desenvolvimento dos algoritmos de aprendizado de máquina. Se o objetivo é melhorar o resultado final, faz-se necessário livrar esses dados, o quanto mais seja possível, de enviesamentos humanos. Necessita-se fiscalizar e controlar os dados utilizados para treinamento e validação dos modelos.

Talvez um importante critério a ser implementado na busca de uma maior neutralidade do algoritmo seja a remoção de todos os atributos sensíveis presentes nos dados. Atributos como raça, gênero, posição social, religião, bem como indicativos que forneçam esses atributos, devem ser extirpados do treinamento do modelo. Neste passo, as perguntas efetuadas devem ser cuidadosamente pensadas para que não aborem, tampouco indiquem sinais, de atributos sensíveis. A eliminação de informações sensíveis dos dados de treinamento conduziria, a princípio, a uma maior neutralidade algorítmica e resultaria em um modelo livre de discriminação.

Além disso, é preciso atenção com os dados e variáveis que são escolhidos para treinar o algoritmo, pois a eleição equivocada pode resultar em um padrão enviesado. Tome-se como exemplo o que aconteceu com o *PredPol*. Lum e Isaac, do *Human Rights*, demonstram que, para sugerir prováveis locais de crimes, o algoritmo se valia de dados sobre prisões anteriores, impulsionando o policiamento excessivo em bairros de maioria negra. No entanto, quando os pesquisadores alteraram os dados de treinamento inserindo dados relativos ao uso geral de drogas da cidade, com base

nas estatísticas nacionais, o policiamento foi distribuído de maneira muito mais uniforme<sup>69</sup>.

Outro critério que deve ser observado é a promoção da máxima transparência do modelo. Esta transparência deve abranger os dados que foram utilizados, os resultados obtidos, nível de acurácia, além de informações sobre a equipe desenvolvedora. A transparência destas informações permite a auditoria do algoritmo e a adoção de medidas corretivas.

É certo que muitos algoritmos são opacos, ou seja, ainda que exista uma regra de aprendizagem, a maneira como se dá esse aprendizado não é objeto de controle, nem mesmo pelos desenvolvedores – o denominado *black box*. Então, não há, aqui, como se obter 100% de transparência. Por outro lado, a transparência relativa aos dados utilizados no treinamento do modelo, aos resultados obtidos, nível de acurácia e equipe desenvolvedora, já permitiria a constatação de dados sensíveis, de vieses, de falhas relativas à composição da equipe, permitindo correções.

Ainda, é imperioso que os sistemas sejam projetados, possuam *designs*, que permitam auditoria e a prestação de contas. A opacidade de alguns modelos não afasta a possibilidade de auditoria e a prestação de contas relativas aos dados, ao resultado final, à acurácia, etc.

Questão que também merece especial atenção se refere à diversidade nas equipes desenvolvedoras. É preciso uma equipe multidisciplinar e heterogênea. Diversidade envolvendo tanto profissionais de distintas áreas, mas também diversidade racial e de gênero. Isso amplia as perspectivas e contextos representados e possibilita a construção de modelos mais abertos e democráticos.

Ainda, é preciso ampla publicidade para que haja ciência inequívoca, pelos usuários, de que estão diante de um sistema de IA. A publicidade também permite maior controle por parte de agentes externos. Esta é, inclusive, uma exigência contida no artigo 20 da LGPD, quando trata de algoritmos de tomada de decisão e que interfiram na esfera jurídica dos cidadãos.

Sem embargo, em que pese a extrema relevância desses desafios éticos na utilização da inteligência artificial, certo é que se trata de um caminho inexorável a ser percorrido por todas as ciências e sociedades. Cogitar que a inteligência artificial não pode ser utilizada em quaisquer das esferas do Poder Público é querer, de maneira obtusa e estéril, frear a realidade e aprisionar-se às amarras da tradição. Os desafios de natureza ética impendem vigilância, fiscalização, regulamentação, mas não afastam a circunstância de que a automação e a inteligência artificial podem ser utilizadas em inúmeras funções de auxílio, mormente no âmbito judicial, proporcionando uma

69 LUM, Kristian; ISAAC, William. To predict and serve? Significance Magazine, oct, 2016. Disponível em: <https://rss.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1740-9713.2016.00960.x>. Acesso em: 15 jun. 2022.

grande evolução no “ecossistema da Justiça”<sup>70</sup>.

A absorção de inovações tecnológicas nos Tribunais brasileiros perfaz um poderoso aliado a fim de melhorar o ambiente institucional judicial brasileiro, trazendo celeridade, efetividade e segurança jurídica. A utilização de modelos algoritmos pode, inclusive, eliminar alguns dos vieses cognitivos humanos, por não estarem sujeitos, por exemplo, à fadiga, humor, nem à fome – elementos externos a atividade de julgar, mas que sabidamente interferem de maneira determinante na tomada de decisão judicial.

Neste sentido cabe aventar a pesquisa efetuada por Danziger et al. Os pesquisadores analisaram 1.112 decisões, de 8 juízes israelenses, incumbidos de julgar pedidos de concessão de liberdade condicional. A conclusão foi de que quando os juízes fazem decisões sequenciais repetidas, eles tendem a decidir mais a favor do *status quo* ao longo do tempo, mas eles podem superar essa tendência fazendo uma pausa para refeição. O estudo mostrou que, cerca de 65% das decisões foram a favor do pedido de liberdade condicional nos julgamentos que ocorreram no início de cada sessão (pela manhã, após o café da manhã, e após a pausa para o almoço). Contudo, ao final da sessão de julgamento, menos de 10% dos pedidos foram concedidos. Os autores concluíram que, portanto, fatores alheios à tomada de decisão judicial, como o cansaço, esgotamento mental e a fome dos julgadores, influenciam diretamente na tomada de decisão judicial<sup>71</sup>.

Convém, por fim, enaltecer, que as plataformas de IA desenvolvidas para o Judiciário brasileiro restringem-se, em sua maioria, ao apoio a atividades de automação. Com relação ao auxílio na tomada de decisão judicial, os algoritmos, hoje, operam no modo sugestivo e não decisório. A supervisão humana é imprescindível, não só pelo nível de IA que está operando no Judiciário, mas também por ser um dos requisitos exigidos pelo Conselho Nacional de Justiça para o funcionamento dos modelos de IA (Resolução nº 332/2020 que dispõe sobre a ética no uso de IA no Poder Judiciário).

Apesar de tudo isso, atribuir à supervisão humana a solução dos problemas de enviesamento algorítmico é indesejado e pueril. O viés algoritmo acontece, exatamente, porque os algoritmos reproduzem os vieses cognitivos humanos, ou seja, os valores e preconceitos inerentes aos indivíduos. Logo, não é suficiente uma rigorosa supervisão humana em decisões sensíveis que interferem de maneira contundente em Direitos Fundamentais dos cidadãos. Necessita-se de prudência para evitar o desenvolvimento de modelos que possam aprofundar as já abissais mazelas sociais do preconceito e da segregação<sup>72</sup>, notadamente aquelas marcadas pelas desigualdades étnicas. O racismo

70 PICCOLI, Ademir Milton. Judiciário Exponencial: sete premissas para acelerar a inovação e o processo de transformação do ecossistema da justiça. São Paulo: Vidaria Livros, 2018.

71 DANZIGER, Shai; LEVAV, Jonathan; AVNAIM-PESSO, Liora. Extraneous factors in judicial decisions. PNAS, vol. 108, n. 17, 2011.

72 NUNES, Pablo. O Algoritmo e o Racismo nosso de cada dia. Revista Piauí. Disponível em: <https://piaui.folha.uol.com.br/o-algoritmo-e-racismo-nosso-de-cada-dia/>. Acesso em: 02 jan. 2021.

estrutural presente nas sociedades contemporâneas autorizou e autoriza as mais diversas práticas de violência voltadas à população negra. Se ao invés de mitigar o avanço seletivo do braço penal do Estado a essas classes historicamente perseguidas, os sistemas de Inteligência Artificial fomentarem o preconceito e o racismo (e sob o verniz de uma pseudo “neutralidade matemática”), manter-se-á a exclusão social legitimada, dirigindo-se o controle a determinados grupos de pessoas e não às ações por eles praticadas.

## 7 Considerações Finais

O presente estudo se dedicou a analisar a característica estrutural mais latente do Controle Penal, a sua seletividade. O recorte da pesquisa foi realizado a partir da segregação racial da população afrodescendente fundada no racismo institucional dos segmentos de controle, mas, agora, sob o verniz de uma pseudo “neutralidade matemática” na utilização de sistemas de Inteligência Artificial.

Cientes de que não há Sistema Penal que não seja seletivo, o problema de pesquisa formulado na parte introdutória do trabalho e que impulsionou o estudo foi o seguinte: a seletividade dos segmentos de Controle Penal poderia ser abrandada com a utilização de sistemas de Inteligência Artificial?

Para tanto, forneceu-se um breve apanhado acerca dos conceitos de inteligência artificial, algoritmos para explicar o uso problemático dos algoritmos e seus possíveis vieses. Essa abordagem foi necessária para se avançar à parte seguinte do estudo, a qual se ocupou de demonstrar alguns dos perniciosos impactos que modelos algoritmos enviesados vêm produzindo, lesando Direitos Fundamentais e intensificando o racismo e a segregação que já são bastante presentes em nossa sociedade e no sistema de Justiça Criminal.

A partir disso, fora possível apresentar alguns modelos que vêm sendo utilizados e demonstrar de que maneira esses sistemas contribuem para a manutenção do racismo, da misoginia e da segregação. Restou claro, neste ponto, que é fundamental voltar a atenção para a utilização de inteligência artificial em situações que podem interferir na esfera das liberdades individuais das pessoas, em modelos algoritmos que possam ferir garantias fundamentais, sobretudo, porque trazem consigo enviesamentos potencialmente lesivos por detrás de uma falsa aparência de neutralidade.

No que diz respeito a utilização de reconhecimento facial na segurança pública, restou evidente que se trata de uma tecnologia que deve ser banida. São inúmeras as limitações atuais da tecnologia e a manifesta possibilidade de afronta a Direitos Fundamentais marcada pela identificação equivocada de pessoas, notadamente aquelas pertencentes aos grupos mais vulneráveis da população. Não há maturidade tecnológica, efetividade, tampouco transparência e ciência à população, alvo desses sistemas.

A partir da análise dessa problemática, parte seguinte do estudo procurou elencar algumas medidas necessárias e hábeis a atenuar a possibilidade de enviesamento dos modelos, tais como a necessidade de regulamentação e fiscalização. É essencial que as leis acompanhem a tecnologia e regulamentem a utilização de modelos de IA, mormente com relação a utilização dos segmentos de Controle Penal. Para se reduzir danos, deve-se garantir transparência, prestação de contas, diversidade dentro das equipes, além de necessária auditoria externa. Essas são premissas imprescindíveis para a construção de modelos mais amplos e democráticos.

Na seara penal, que constitui o ponto nevrálgico da temática, principalmente em relação à tutela de Direitos Fundamentais, o art. 23º da resolução 332 do CNJ, que regulamenta a utilização de inteligência artificial no âmbito do Poder Judiciário, prevê expressamente esta preocupação. Referido dispositivo estipula que a utilização de modelos de IA em matéria penal não deve ser estimulada, sobretudo com relação à sugestão de modelos de decisões preditivas. Isso é muito importante, pois restou demonstrado que a utilização desses sistemas vem ampliando e perpetuando as discriminações e a seletividade que permeia o controle penal e as agências que o integram. E, pior. A máquina opera sob a ilusória impressão de que se baseia em modelos matemáticos imparciais, precisos e objetivos. Na verdade, os resultados estão carregados de subjetividade e vieses racistas tendentes a catalisar setores sociais bem delimitados para a marginalização e a discriminação.

Precisa-se, então, de regulamentação e fiscalização. É fundamental que haja transparência, auditoria e prestação de contas de cada modelo que interfere na esfera jurídica dos cidadãos. Isso poderia permitir ciclos positivos de *feedbacks* e o aprimoramento dos modelos com vistas a contornar possíveis falhas de funcionamento e melhorar o seu resultado geral, reduzindo a incidência de danosos enviesamentos algorítmicos.

A inserção e o aumento da utilização de IA, em todos os aspectos da vida social é uma tendência inexorável. Ao invés de resistir à mudança inevitável, deve-se descobrir a melhor maneira de trilhar essa senda, mas sempre cientes do grande desafio que está colocado: impedir que as ferramentas tecnológicas reproduzam as limitações, falhas e preconceitos já enraizados e normalizados no âmbito da sociedade e nos segmentos de controle penal. Desse modo, além de voltar os olhos para o futuro, faz-se necessário empreender atenção a um problema mais imediato e urgente, qual seja, os meios de se evitar que os algoritmos reproduzam e ampliem nossas mazelas sociais do presente.



## Referências

- ACLU ORG. *Amazon's Face Recognition Falsely Matched 28 Members of Congress*. Disponível em: <https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28>. Acesso em: 15 dez. 2020.
- AI NOW INSTITUTE. *AI Now Report 2018*. Disponível em: <https://ainowinstitute.org/>. Acesso em: 16 dez. 2020.
- ANDREADE, Vera Regina Pereira de. *A ilusão de segurança jurídica: do controle da violência a violência do controle penal*. 2ª ed. Porto Alegre: Livraria do Advogado, 2003.
- Anuário Brasileiro de Segurança Pública*. Fórum Brasileiro de Segurança Pública. Ano 14.2020. Disponível em: <https://forumseguranca.org.br/wp-content/uploads/2020/10/anuario-14-2020-v1-final.pdf>. Acesso em: 13 março 2022.
- BARATTA, Alessandro. *Criminologia crítica e crítica do direito penal: introdução à sociologia do direito penal*. Tradução de Juarez Cirino dos Santos. 3. ed. Rio de Janeiro: Revan, 2002.
- BATISTA, Nilo. *A violência do Estado e os aparelhos policiais. Discursos Sediciosos*. Crime, direito e sociedade. Volume 2. 4. Ed. Rio de Janeiro, 1997.
- BRASIL. *Anteprojeto de Lei de Proteção de Dados*. Disponível em: <https://www2.camara.leg.br>. Acesso em: 20 dez. 2020.
- BROWN, Wendy. *Undoing the Demos: Neoliberalism's Stealth Revolution*. New York: Zone Books, 2015.
- CHAVES JUNIOR, Airto. *Além das Grades: a paralaxe da violência nas prisões brasileiras*. Florianópolis: Tirant lo Blanch, 2018.
- CHAVES JUNIOR, Airto; OLDONI, Fabiano. *Para que(m) serve o Direito Penal? Uma análise criminológica dos segmentos de controle social*. Rio de Janeiro: Lumen Juris, 2014.
- CONSELHO NACIONAL DE JUSTIÇA. *Relatório Justiça em números 2020: Ano-base 2019*. Brasília: CNJ, 2020. Disponível em: <https://www.cnj.jus.br>. Acesso em: 06 nov. 2020.
- DANZIGER, Shai; LEVAV, Jonathan; AVNAIM-PESSO, Liora. Extraneous factors in judicial decisions. *PNAS*, vol. 108, n. 17, 2011.
- FENOLL, Jordi Nieva. *Inteligencia Artificial y Proceso Judicial*. Madri: Marcial Pons, 2018.
- FGV – Centro de Inovação, Administração e Pesquisa do Judiciário. In: *Inteligência Artificial: Tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário Brasileiro*. Disponível em: [https://ciapj.fgv.br/sites/ciapj.fgv.br/files/estudos\\_e\\_pesquisas\\_ia\\_1afase.pdf](https://ciapj.fgv.br/sites/ciapj.fgv.br/files/estudos_e_pesquisas_ia_1afase.pdf). Acesso em: 15.out.2020.
- FOUCAULT, Michel. *Genealogia del Racismo*. Traducción de Alfredo Tzveibel. Colección Caronte Ensayos. Buenos Aires: Editorial Altamira, 1993.
- FREITAS, Cinthia Obladen de A. *A obscuridade dos algoritmos e a LGPD. O Estado de São Paulo*, 4.ago.2020. Disponível em: <https://politica.estadao.com.br/blogs/fausto-macedo/a-obscuridade-dos-algoritmos-e-a-lgpd/>. Acesso em: 10 dez. 2020.

- GARRETT, Brandon L. *Convicting the Innocent: where criminal prosecutions Go Wrong*. Cambridge, Massachusetts / London, England. Harvard University Press, 2011.
- KILOMBA, Grada. *Memórias da Plantação: episódios de racismo cotidiano*. Tradução de Jess Oliveira. Rio de Janeiro: Editora Cobogó, 2019.
- KIRCHHEIMER, Otto. *Political Justice: the use of legal procedure for political ends*. Princeton, New Jersey: Princeton University Press, 1961.
- LUM, Kristian; ISAAC, William. To predict and serve? *Significance Magazine*, oct, 2016. Disponível em: <https://rss.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1740-9713.2016.00960.x>. Acesso em: 15 jun. 2021.
- MATIDA, Janaína; CECCONELLO, Willian. O que há de errado no reconhecimento fotográfico de Michael B. Jordan? *Consultor Jurídico (CONJUR)*. 08 de janeiro de 2022.
- MBEMBE, Achille. *Necropolítica*. São Paulo: N-1 Edições, 2018.
- METROPOLES. *Sistema de reconhecimento facial nas ruas é regulamentado no DF*. Disponível em: <https://www.metropoles.com/distrito-federal/sistema-de-reconhecimento-facial-nas-ruas-e-regulamentado-no-df>. Acesso em: 17 dez. 2020.
- MOTHERBOARD. *Amazon's Home Security Company Is Turning Everyone Into Cops*. Disponível em: <https://www.vice.com/en/article/qvyvzd/amazons-home-security-company-is-turning-everyone-into-cops>. Acesso em: 21 set. 2023.
- NAVARRO, Susana Navas. Derecho e Inteligencia Artificial desde el Diseño. In: NAVARRO, Susana Navas, et al. *Inteligencia Artificial: Tecnología, Derecho*. Valencia: Tirant Lo Blanch, 2017.
- NIST. *NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software*. Disponível em: <https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software>. Acesso em: 15 dez. 2020.
- NUNES, Dierle. A supervisão humana das decisões de inteligência artificial reduz os riscos? *Consultor Jurídico*. Disponível em: <https://www.conjur.com.br/2021-jun-25/nunes-supervisao-humana-decisoes-ia-reduz-riscos>. Acesso em: 26 jun. 2021.
- NUNES, Pablo. O Algoritmo e o Racismo nosso de cada dia. *Revista Piauí*. Disponível em: <https://piaui.folha.uol.com.br/o-algoritmo-e-racismo-nosso-de-cada-dia/>. Acesso em: 02 jan. 2021.
- O'NEIL, Cathy. *Algoritmos de Destruição em Massa: como o Big Data aumenta a desigualdade e ameaça a democracia; tradução Rafael Abraham*. 1. ed. Santo André, São Paulo: Editora Rua do Sabão, 2020.
- O'NEIL, Cathy. Amazon's Gender-Biased Algorithm Is Not Alone. *Bloomberg*. Disponível em: <https://www.bloomberg.com/opinion/articles/2018-10-16/amazon-s-gender-biased-algorithm-is-not-alone>. Acesso em: 10 set. 2021.
- OCARIZ, Maria Cristina (Org.). *Psicanálise e Violência Social*. São Paulo: Editora Escuta, 2018.
- PICCOLI, Ademir Milton. *Judiciário Exponencial: sete premissas para acelerar a inovação e o*

processo de transformação do ecossistema da justiça. São Paulo: Vidaria Livros, 2018.

PROPÚBLICA. *How we analyzed the compas recidivism algorithm*. 23. Mai. 2016. Disponível em: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Acesso em: 20 out. 2020.

REDE DE OBSERVATÓRIOS DA SEGURANÇA. *Retratos da Violência*. Cinco meses de monitoramento, análises e descobertas. Disponível em: <http://observatorioseguranca.com.br/wp-content/uploads/2019/11/1relatoriorede.pdf>. Acesso em: 15 dez. 2020.

REUTERS. *Amazon scraps secret AI recruiting tool that showed bias*. Disponível em: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>. Acesso em: 10 dez. 2020.

RODRÍGUEZ, Pablo. *Inteligencia Artificial*. Como cambiará el mundo (y tu vida). Trad. Jorge Rizzo. Barcelona: Ediciones Deusto, 2018.

ROSA, Alexandre; GUASQUE, Bárbara. O avanço da Disrupção nos Tribunais Brasileiros. In: NUNES, Dierle; LUCON, Paulo Henrique; WOLKART, Erik Navarro (org.). *Inteligência Artificial e Direito Processual: os impactos da virada tecnológica no Direito Processual*. 2ªed. rev. atual. e ampl. Salvador: JusPODVIM, 2021.

SADOWSKI, Jathan. *Too Smart: How Digital Capitalism is Extracting Data, Controlling Our Lives*. Massachusetts: The MIT Press, 2020.

THE GUARDIAN. *Amazon to ban police use of facial recognition software for a year*. Disponível em: <https://www.theguardian.com/technology/2020/jun/10/amazon-rekognition-software-police-black-lives-matter>. Acesso em: 14 dez. 2020.

THE GUARDIAN. *Disastrous' lack of diversity in AI industry perpetuates bias, study finds*. Disponível em: <https://www.theguardian.com/technology/2019/apr/16/artificial-intelligence-lack-diversity-new-york-university-study>. Acesso em: 15 dez. 2020.

THE GUARDIAN. *How white engineers built racist code – and why it's dangerous for black people*. Disponível em: <https://www.theguardian.com/technology/2017/dec/04/racist-facial-recognition-white-coders-black-people-police>. Acesso em: 17 dez. 2020.

THE GUARDIAN. *Rise of the racist robots – how AI is learning all our worst impulses*. Disponível em: <https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses>. Acesso em: 15 dez. 2020.

THE INTERCEPT. *Amazon's Ring Planned Neighborhood Watch lists built on facial recognition*. Disponível em: <https://theintercept.com/2019/11/26/amazon-ring-home-security-facial-recognition/>. Acesso em: 15 dez. 2020.

THE NEW YORK TIMES. *Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match*. Disponível em: <https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html>. Acesso em: 10 dez. 2020.

U.S. Equal Employment Opportunity Commission. *Diversity in High Tech*. Disponível em: <https://www.eeoc.gov/special-report/diversity-high-tech>. Acesso em: 15 dez. 2020.

UOL. *Sistema de reconhecimento facial erra, e homem negro é preso por engano*. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2020/06/25/homem-e-preso-apos-erro-de-tecnologia-de-reconhecimento-facial-nos-eua.htm>. Acesso em: 10 dez. 2020.

WACQUANT, Loïc. *Punir os Pobres: a nova gestão da miséria nos Estados Unidos*. 3. Ed. Tradução de Sérgio Lamarão. Rio de Janeiro: Editora Revan, 2007.

ZAFFARONI, Eugênio Raul. *O Inimigo no Direito Penal*. Tradução de Sérgio Lamarão. Rio de Janeiro: Revan, 2007.

ZAFFARONI, Eugenio Raúl; PIERANGELI, José Henrique. *Manual de direito penal brasileiro*, vol. 1: parte geral. 6 ed. São Paulo: Editora Revista dos Tribunais, 2006.