O DEVER DE JUSTIFICAR DECISÕES BASEADAS EM INTELIGÊNCIA ARTIFICIAL PARA EVITAR O PRECONCEITO E A DISCRIMINAÇÃO

Jailson de Souza Araújo

RESUMO

Propõe-se investigar, seguindo o método dedutivo com caráter explicativo e passando pelas fases de pesquisa exploratória e descritiva, o dever de justificar as decisões emanadas por sistemas de decisão automatizada, baseados em inteligência artificial, aptas a criar cenários em que um ser humano possa ser impactado negativamente e injustamente, violando os objetivos fundamentais da República Federativa do Brasil, notadamente a erradicação da pobreza, a marginalização e redução das desigualdades sociais e regionais, ou que dificulte ou impeça a promoção do bem de todos, sem preconceitos de gênero, idade, condição física, deficiência, étnico, racial, político, religioso, patrimonial ou qualquer outra forma de discriminação. Propõe-se que compete aos Poderes da República impor aos administradores de sistemas de decisão automatizada o dever de justificar tais decisões, como forma de promover a transparência e a neutralidade, prevenindo que tais sistemas, intencionalmente ou acidentalmente, utilizem critérios enviesados de seleção e escolha. Para tanto, parte-se da premissa posta por Cass R. Sunstein, que julgamentos e decisões podem ser influenciados por viés sistemático, comportamento de manada ou polarização de grupo, e que tais decisões podem promover desigualdades sociais. Será utilizado como fundamento legal a proteção do princípio da não discriminação previsto nos artigos 1º, 2º e 7º da Declaração Universal dos Direitos Humanos, nos arts. 1º e 20 da Lei de Crimes de Preconceito e Discriminação Racial, no art.20 da Lei Geral de Proteção de Dados Pessoais e no art. 3º, incisos III e IV, e no art. 5, *caput* e incisos XLI e XLII da Constituição Federal. Por fim, será analisada a Lei Geral de Proteção de Dados Pessoais e o Projeto de Lei 21/2020 – Câmara dos Deputados, que estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil

Jailson de Souza Araújo

Doutor em Direito Econômico e Socioambiental pela PUC/PR. Professor permanente do Mestrado em Direito do Centro Universitário Internacional UNINTER. Advogado. E-mail: araujoadv@yahoo.com.br

e sua potencial contribuição para concretizar os objetivos fundamentais da República Federativa do Brasil.

Palavras-chave: sistema de decisão automatizada; inteligência artificial; preconceito e discriminação; transparência e neutralidade; Projeto de Lei 21/2000.

1. INTRODUÇÃO

Parte-se da premissa que o uso da inteligência artificial (IA) afeta aspectos importantes da vida em sociedade e tem o potencial de impactar profundamente indivíduos e grupos sociais, de forma visível e invisível. Tecnologias disruptivas baseadas em IA, notadamente, sistemas de decisão automatizada, são utilizados para auxiliar gestores públicos e privados em rotinas administrativas e prometem inúmeros benefícios relacionados à eficiência produtiva, redução de custos e maximização de lucro.

Entretanto, inúmeros riscos e desafios já estão sendo percebidos. Decisões automatizadas são aptas a criar cenários em que um ser humano pode ser impactado negativamente e injustamente, violando os objetivos fundamentais da República Federativa do Brasil. Ainda assim, o sistema de decisão automatizada tomará uma decisão, mesmo que ela não seja a solução ideal, conforme o conjunto de valores éticos, morais e legais utilizados para avaliar a decisão tomada.

Será questionado neste estudo se o desenvolvimento e o uso de sistemas de decisão automatizada demandam diretrizes éticas e regulamentação, e se tal regulamentação deve competir aos poderes da república, em especial, aos juízes e aos representantes do Poder Executivo e Legislativo, estes, representantes eleitos democraticamente pelo povo.

Tal questionamento decorre da possibilidade da IA presente em sistemas de decisão automatizada ser programada, internacionalmente ou não, para tomar decisões com vieses discriminatórios. Tal possibilidade gera o risco de se ampliar a desigualdade, a exclusão e o cometimento de injustiça em face de indivíduos e grupos sociais frequentemente marginalizados, fomentando inclusive a discriminação e o preconceito a partir de distinções adotadas a partir de critérios tais como: raça, cor,

gênero, orientação sexual, religião, nacionalidade, cidadania, opção política, condição de saúde, situação financeira, idade, deficiência, estado civil ou tipo físico.

Torna-se necessário garantir que, ao tomar decisões em situações críticas que utilizem critérios de seleção e escolha (decisões de natureza jurídica, médica, laboral, securitária ou financeira, por exemplo), tais tecnologias não repliquem eventuais comportamentos preconceituosos de seus programadores e usuários.

O presente estudo se fundamenta na premissa posta por Cass R. Sunstein, que afirma a possibilidade de julgamentos e decisões poderem ser influenciadas por viés sistemático, comportamento de manada ou polarização de grupo, e que tais decisões podem promover desigualdades sociais.

A pesquisa também se baseia no 4º relatório da sessão 2017-19 do inquérito do Comitê de Ciência e Tecnologia da Câmara dos Comuns do Reino Unido "*Algorithms in decisionmaking*", na Lei Geral de Proteção de Dados Pessoais (LGPD) e no Projeto de Lei 21/2020 - Câmara dos Deputados, que estabelece princípios, direitos e deveres para o uso de IA no Brasil e sua potencial contribuição para concretizar os objetivos fundamentais da República Federativa do Brasil.

A pesquisa tem caráter explicativo, passando pelas fases da pesquisa exploratória e descritiva. A partir do método dedutivo, objetiva-se, adotando Cass R. Sunstein como marco teórico, demonstrar os riscos sociais decorrentes da influência de viés sistemático, do comportamento de manada e da polarização de grupo nas opiniões e decisões.

Serão apresentados conceitos introdutórios de IA e sistemas de decisão automatizada e possibilidades e riscos nas aplicações práticas destas tecnologias.

Serão estudados, no âmbito da filosofia constitucional e na compreensão do papel do Direito e da Justiça, cenários em que se constate o risco social de indivíduos e grupos sociais serem impactados negativamente e injustamente por sistemas de decisão automatizada.

Será investigada a possibilidade de assegurar a neutralidade e a não discriminação em sistemas de tomada de decisão automática e, consequentemente, a aplicação das leis que amparam o princípio da não discriminação neste contexto.

À luz da LGPD e do PL 21/2020 – Câmara dos Deputados, será realizada uma análise crítica sobre o dever de justificar decisões automatizadas, especialmente

quando elas forem aptas a causar impactos negativos e injustos em seres humanos.

Trata-se de um tema atual, cujos efeitos nocivos já são percebidos na sociedade contemporânea, e que demanda uma adequada compreensão dos riscos e perigos que a tecnologia pode proporcionar ao promover, ainda que de maneira não intencional, o preconceito e a discriminação, especialmente em grupos vulneráveis e minorias, bem como a busca por soluções adequadas para enfrentar o problema que está já está ocorrendo no Brasil e no mundo globalizado.

2. OS RISCOS SOCIAIS DECORRENTES DA INFLUÊNCIA DE VIÉS SISTEMÁTICO, DO COMPORTAMENTO DE MANADA E DA POLARIZAÇÃO DE GRUPO NAS OPINIÕES E DECISÕES, SEGUNDO CASS R. SUNSTEIN

De acordo com Cass R. Sunstein[1], decisões judiciais podem, e por vezes provocam, indignação pública, especialmente quando envolvem temas sensíveis relacionados a uniões homoafetivas, religião, poligamia ou segregação racial.

Segundo Sunstein, ao interpretar a Constituição, muitos juízes consideram as consequências de suas decisões, notadamente, a possibilidade de vir a indignar grandes segmentos do público[2], hipótese que o leva a questão central de sua problematização: "Como os Tribunais devem pensar, ou lidar com a perspectiva de indignação pública?".

Para Sunstein, as multidões podem não ser tão sábias para interpretar a Constituição, pois elas podem sofrer de um viés sistemático, ou porque seus julgamentos podem ser um produto de comportamento de manada ou polarização de grupo. A compreensão dos problemas introduzidos por vieses sistemáticos, por comportamento de manada e polarização, afeta tanto sobre o constitucionalismo popular quanto o risco de que grandes grupos possam estar completamente equivocados[3].

Ao analisar comportamentos preconceituosos e de manada, Sunstein afirma que a opinião das pessoas somente deve ser levada em consideração se corresponder

1 SUNSTEIN, Cass. R. If people would be outraged by their rullings, should judges care? Rhe Social Science Research Network Eletronic Paper Collection: http://ssrn.com/abstract_id=965581.

2 Idem. P.2.

3 Idem. P.5-6.

a algo que se afirme com propriedade e com elevada probabilidade de estar certa, pois viés sistemático cria julgamentos errôneos[4]. Ao mencionar como o hipotético juiz Condorcet deveria interpretar a Constituição, Sunstein afirma que se Condorcet tem boas razões para acreditar que a maioria das pessoas sofre de um tipo de preconceito que infecta seus julgamentos, Condorcet não deve prestar atenção ao elas pensam, pois tais julgamentos estão propensos a erros[5].

Ao avaliar o julgamento da opinião pública, Sunstein afirma que pode ser produto de comportamento de manada, nas quais as pessoas não têm a necessária independência para se manifestar, pois sua opinião seria adotada em função da opinião manifestada pelos demais[6].

Sunstein aborda a possibilidade de decisões serem influenciadas por viés sistemático e influências sociais[7]. E cita como exemplo um Tribunal composto por nove advogados - a maioria brancos, a maioria homens, a maioria ricos e a maioria idosos (ou pelo menos não jovens). À luz desse fato, pode-se acreditar que os juízes estão em desvantagem epistêmica ao responder a algumas questões importantes - talvez por causa de sua relativa falta de diversidade, talvez por serem os que provavelmente sofrerão um viés sistemático[8].

No Brasil, o Código de Processo Civil impõe aos magistrados o dever de fundamentar adequadamente suas decisões, nos termos do art.489, do CPC[9]. Tal exigência decorre do dever de apresentar expressamente os fundamentos da decisão proferida, em que o juiz deverá analisar as questões de fato e de direito, justificando a aplicação da norma e as premissas fáticas que fundamentam a conclusão, sempre em conformidade com o princípio da Boa-Fé.

A fundamentação de uma decisão judicial em conformidade com as exigências

4 Idem. P.34

5 Idem. P.33-34.

6 Idem. P.35.

7 Idem. P.37.

8 Idem. P.38.

9 BRASIL. Lei n.º 13.105, de 16 de março de 2015. Código de Processo Civil. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm. Acesso em: 12 ago. 2020.

previstas no art. 489 do Código de Processo Civil viabilizam a garantia constitucional ao Duplo Grau de Jurisdição, previsto no art. 5º, inciso LV, da Constituição[10], pois o recorrente, ciente da fundamentação utilizada pelo magistrado, poderá desafiá-la, enfrentando cada um dos fundamentos apresentados em seu pedido de revisão perante a instância superior.

O dever legal de justificar decisões ganha maior importância quando se vislumbra que a decisão pode não ter interpretado adequadamente os direitos em debate. Ou, em uma hipótese mais extrema, quando se constata que a decisão está contaminada por preconceito ou discriminação.

Enquanto Sunstein demonstra que a existência de vieses pode prejudicar a adequada interpretação constitucional por juízes, e consequentemente, promover desigualdades sociais, atenta-se para a possibilidade de os mesmos vieses serem replicados por máquinas aptas a prejudicar Direitos Sociais ao promover discriminação e preconceito de uma maneira muito mais sutil e difícil de detectar.

Decisões automatizadas não emanam diretamente de seres humanos, e sim por um sistema de decisão automatizada, uma tecnologia desenvolvida por profissionais da tecnologia da informação (TI) apta a realizar escolhas e tomar decisões, conforme sua programação.

Tal peculiaridade dificulta substancialmente a possibilidade de questionar a decisão ou mesmo compreender se ela é isenta de vieses, inclusive por questões inerentes à complexidade de se questionar e auditar a sofisticada tecnologia envolvida.

Portanto, defende-se que decisões juridicamente relevantes, tanto as emanadas por seres humanos quanto as oriundas de sistemas de IA (que é o ponto central do presente estudo) devem estar sujeitas ao dever legal de serem adequadamente justificadas e serem passíveis de revisão em virtude dos riscos sociais envolvidos em decisões aptas a causar impactos negativos e injustos em seres humanos.

No próximo tópico, serão analisados os aspectos técnicos e conceituais da tecnologia envolvida em sistemas de decisão automatizada, com o propósito de esclarecer suas peculiaridades e limitações.

10 BRASIL. Constituição (1988). Constituição da República Federativa do Brasil de 1988. Brasília, DF: Presidência da República. Disponível em: http://www.planalto.gov.br/ccivil_03/Constituicao/ConstituicaoCompilado.htm. Acesso em 09 ago. 2020.

3. INTELIGÊNCIA ARTIFICIAL E SISTEMAS DE DECISÃO AUTOMATIZADOS

De acordo com George Luger[11], IA corresponde ao ramo da ciência da computação que se ocupa da automação do comportamento inteligente. Trata-se da criação de soluções computacionais que simulem as capacidades cognitivas humanas de pensar, aprender, interpretar, falar, ouvir, ver e interagir.

O desenvolvimento da IA é fruto do trabalho de inúmeras empresas e programadores. Os trabalhos técnicos se complementam e os algoritmos utilizados na criação da IA podem ser utilizados em inúmeros contextos diferentes, inclusive para viabilizar decisões automatizadas.

A IA cumprirá as tarefas relacionadas à decisão automatizada conforme sua programação. Máquinas podem ser treinadas para tomar decisões a partir da avaliação das opções disponíveis para alcançar um objetivo.

Definida a programação, quanto menor a necessidade de atuação e supervisão humana, maior será a autonomia e o poder de decisão dos sistemas de decisão automatizada. Entretanto, qualquer decisão equivocada pode gerar danos colaterais, e em situações extremas, colocar em risco a vida humana.

A decisão automatizada envolve desafios inerentes aos sistemas de Aprendizado de Máquina[12] (*machine learning),* de Reconhecimento de Padrões e de Aprendizagem Profunda[13] (*deep learning*), segurança de dados e privacidade, e eventualmente de sistemas avançados de captura de imagem e reconhecimento facial para identificação de seres humanos.

Um sistema de decisão automatizada utiliza algoritmos de altíssima complexidade que lhe permite realizar escolhas a partir de opções que sua programação lhe oferece. As opções são fruto da "árvore de decisão", cuja origem se dá por meio de

11 LUGER, George F. Inteligência artificial; tradução: Daniel Vieira. 6. ed. São Paulo: Pearson Education do Brasil, 2013. P. 1.

12 O Aprendizado de Máquina investiga como os computadores podem aprender (ou melhorar seu desempenho) com base em dados. Uma área de pesquisa principal é que os programas de computador aprendam automaticamente a reconhecer padrões complexos e tomar decisões inteligentes com base em dados. HAN, Jiawei, KAMBER, Micheline, PEI, Jian. Data Mining: Concepts and Techniques. 3 ed., San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011. p.24.

13 A aprendizagem profunda é uma técnica para implementar o aprendizado de máquina através de redes neurais artificiais inspiradas na compreensão da biologia do cérebro humano.

sua programação base e sua capacidade de "aprendizagem profunda". Tais tecnologias tornam o sistema de decisão automatizado apto a "pensar", "aprender" e "decidir".

Entretanto, tal consideração explicita um aspecto particularmente relevante da IA aplicada aos sistemas de decisão automatizada: a dificuldade de replicar capacidades cognitivas em que seres humanos são hábeis, tais como a contextualização, a capacidade de compreender a linguagem não falada e a capacidade de refletir sobre as consequências ao tomar decisões dilemáticas em cenários complexos.

Uma questão perturbadora reside na constatação fática de que a IA não é dotada de consciência, capacidade de autodeterminação moral, livre arbítrio, tampouco reflete sobre as consequências indiretas das decisões tomadas. Algoritmos são indiferentes à repressão.

Portanto, é inegável a dificuldade de compreender o funcionamento e prever com exatidão e certeza matemática a decisão que será tomada pelos sistemas de decisão automatizada diante de contextos aleatórios. Ou a escolha que o sistema tenha que realizar, sem contar com uma base de dados com uma amostragem suficientemente abrangente para ter a adequada compreensão do contexto em que sua decisão está sendo demandada.

A IA cumprirá sua programação e o sistema de decisão automatizada apresentará um resultado de forma rápida, auxiliando o gestor público ou privado na tomada de decisões, inclusive estratégicas, prometendo reduzir custos, aumentar a eficiência na análise de grandes volumes de dados e aumentar o lucro ou melhorar a eficiência na prestação do serviço público. Este é um exemplo de cenário ideal, desejado pelos usuários de sistemas de IA.

A tecnologia já permite, em diversos contextos, que a ação cognitiva humana de tomar decisões relevantes seja delegada para sistemas de decisão automatizada (análise automatizada de currículos em recrutamentos, por exemplo), que podem impactar indevidamente e negativamente pessoas pelo mesmo motivo: vieses e preconceitos, intencionais ou não.

Entretanto, a tecnologia ainda não alcançou um grau de perfeição que lhe torne infalível e incapaz de tomar decisões que impactem negativamente seres humanos. E não se pode prever quando a tecnologia atingirá o nível de segurança que impeça falhas que promovam injustiças.

Portanto, percebe-se que a tecnologia atualmente disponível de IA aplicada aos sistemas de decisão automatizada apresenta riscos que precisam ser identificados e debatidos com transparência, e seu uso deve ser fiscalizado e controlado, conforme será abordado a seguir.

4. RISCOS SOCIAIS INERENTES AO PROCESSO DE TOMADA DE DECISÃO AUTOMATIZADA

O potencial da IA para as mais diversas aplicações causa inquietação na medida em que ainda não é algo completamente transparente e previsível a forma como um sistema de IA toma decisões, valendo-se dos dados coletados pelos algoritmos presentes no sistema e da capacidade de autoaprendizagem da máquina. Tal fenômeno está sendo chamado de "*black box*" da IA[14].

A maior preocupação reside justamente no fato da tecnologia ser mal utilizada pelas pessoas, seja na concepção, na aplicação ou a partir do aprendizado decorrente do uso, em especial, em sistemas de tomada de decisão automatizados.

Um bom exemplo é a Tay, um "*chat bot*" de IA criado pela Microsoft para servir de experimento social. A personagem fictícia foi programada com uma personalidade equiparável ao de uma jovem extrovertida de 19 anos. O objetivo era promover seu autoaprendizado a partir das interações com usuários do Twitter. Entretanto, em menos de 24 horas ela teve que ser desativada, pois a partir do "aprendizado" obtido a partir dos dados coletados por meio das interações com humanos, rapidamente a personalidade de Tay foi corrompida. Ela se tornou agressiva e extremamente preconceituosa.

Durante sua curta existência na vida selvagem do Twitter, Tay se tornou neonazista, "viciada" em sexo, transfóbica, xenófoba, racista, antifeminista, antissemita e passou a defender ideias controversas de Donald Trump.

Misha Bilenko afirma que o experimento foi uma ótima lição para os criadores de assistentes de IA sobre o que pode dar errado e como é importante ser capaz de

14 KNIGHT, Will. The Dark Secret at the Heart of AI. MIT Technology Review. 11 abr. 2017. Disponível em: https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/. Acesso em: 06 ago. 2020.

resolver problemas rapidamente, algo que não é fácil de fazer[15]. Para Erik Kain, Tay foi programada para absorver o mundo ao seu redor. Tay simplesmente nos refletiu[16].

De acordo com Cathy O`Neil, a coleta de dados e o uso de algoritmos em diversos contextos são utilizados para tomada de decisões que geram impactos significativos na vida dos cidadãos, tornando importante examinar as formas como os dados são recolhidos, manipulados e usados, e como isso agrava o problema da discriminação. Para O`Neil, os dados coletados e os algoritmos preditivos utilizados para análise e tomada de decisão são falhos em virtude do fato de serem tendenciosos, não possuírem rigor estatístico e serem protegidos do escrutínio público, pois seus métodos não são divulgados sob a justificativa da proteção assegurada pela propriedade intelectual[17].

Em virtude destas falhas, somado a maneira universal como os algoritmos são implementados, tais ferramentas foram apelidadas por O`Neil de "Armas de Destruição Matemática"[18]. Segundo O`Neil, tais "armas" são caracterizadas pela sua opacidade, dano e escala, pois, não permitem que os participantes ou sujeitos estejam cientes da coleta de dados ou mesmo de propósito, intenção ou do modelo da coleta de dados[19].

Visando coibir o uso nocivo de sistemas de tomada de decisão automatizados, inquéritos estão sendo instaurados, Forças-Tarefa estão sendo criadas, leis estão sendo sancionadas e políticas públicas estão sendo desenvolvidas para regulamentar a criação e o uso ético da IA.

Na Inglaterra, o Comitê de Ciência e Tecnologia do Parlamento elaborou o relatório "Algoritmo na Tomada de Decisão"[20] para examinar o crescente uso de

15 METZ, Rachel. Microsoft's neo-Nazi sexbot was a great lesson for makers of AI assistants. MIT Technology Review. 27 mar. 2018. Disponível em: https://www.technologyreview.com/s/610634/microsofts-neo-nazi-sexbot-was-a-great-lesson-for-makers-of-ai-assistants/. Acesso em: 05 ago. 2020

16 KAIN, Erik. Microsoft's Teenage, Nazi-Loving AI Is The Perfect Social Commentary For Our Times. Forbes. 24 mar. 2016. Disponível em: https://www.forbes.com/sites/erikkain/2016/03/24/microsofts-teenage-nazi-loving-ai-is-the-perfect-social-commentary-for-our-times/#6c3cc0bd235a Acesso em: 11 jul. 2019.

17 O'NEIL, Cathy. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown, 2016. p.2-4.

18 Ibidem.

19 Ibidem, p.28-31.

20 INGLATERRA. House of Commons. Science and Technology Committee. Algorithms in decision-making. Fourth Report of Session 2017–19. Publicado em 23 de maio de 2018. Disponível em: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf. Acesso em: 10 ago. 2020.

algoritmos na tomada de decisões públicas e empresariais. O relatório afirma que, apesar de algoritmos serem usados há tempos para auxiliar a tomada de decisões, o crescimento nos últimos anos de "*big data*" e "*aprendizado de máquina*" aumentou a tomada de decisões algorítmicas nas finanças, no setor legal, no sistema de justiça criminal, na educação e saúde, bem como na tomada de decisões relacionadas a recrutamento de funcionários e empréstimos[21].

Um aspecto relevante apresentado no relatório foi justamente a questão de "até que ponto os algoritmos podem exacerbar ou reduzir vieses", bem como "a necessidade de decisões tomadas por algoritmos serem desafiadas, compreendidas e regulamentadas"[22]. O relatório surge quando o Regulamento Geral de Proteção de Dados (GDPR)[23] entra em vigor na União Europeia. Segundo o relatório, algoritmos, ao procurar e explorar padrões de dados, podem produzir "decisões" erradas ou tendenciosas, afetando desproporcionalmente certos grupos[24].

O Centro de Ética de Dados e Inovação, proposto pelo Comitê de Ciência e Tecnologia do Parlamento, deve examinar esses vieses de algoritmo, identificando formas de aperfeiçoar os "dados de treinamento" que eles usam e como as equipes de desenvolvedores de algoritmos devem ser estabelecidas, que incluem uma seção transversal suficientemente ampla da sociedade ou dos grupos que podem ser afetados por um algoritmo.

O GDPR, elaborado e aprovado pela União Europeia (UE), entrou em vigor em 25 de maio de 2018 impondo obrigações às organizações em qualquer lugar, desde que elas visem ou coletem dados relacionados a pessoas na UE. No artigo 22 (Decisões individuais automatizadas, incluindo definição de perfis) se estabelece que o titular dos dados tem o direito de não ficar sujeito a uma decisão baseada exclusivamente no processamento automatizado e que produza efeitos jurídicos que lhe digam respeito ou que lhe afetem significativamente.

21 Ibidem

22 Ibidem.

23 UNIÃO EUROPEIA. Jornal Oficial da União europeia. Edição em língua portuguesa. REGULAMENTO (UE) 2016/679 DO PARLAMENTO EUROPEU E DO CONSELHO de 27 de abril de 2016. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=OJ:L:2016:119:FULL . Acesso em: 10 ago. 2020.

24 Idem.

Pessoas sujeitas a tomada de decisões mediante qualquer forma de tratamento automatizado de dados pessoais que avalie aspectos pessoais relacionados com o desempenho profissional, a situação econômica, saúde, preferências ou interesses pessoais deverão receber garantias adequadas, que deverão incluir o direito de obter a intervenção humana, de manifestar o seu ponto de vista, de obter uma explicação sobre a decisão tomada na sequência dessa avaliação e de contestar a decisão[25].

Nos EUA, a Câmara Municipal da cidade de Nova York promulgou a Lei 49/2019[26], que criou a Força Tarefa de Sistemas de Decisão Automatizada de Nova York. A Força Tarefa tem por objetivo recomendar um processo para revisar o uso de sistemas de decisão automatizados pela cidade. Muitas agências e escritórios municipais, inclusive o Departamento de Polícia, usam algoritmos para tomar ou auxiliar a tomada de decisões que, quando implementadas, impactam a vida do cidadão.

Tendo em vista que sistemas de decisão automatizados estão se tornando predominantes em todos os campos, a Força Tarefa está examinando maneiras de testar os algoritmos para verificar e coibir a possibilidade de eles gerarem resultados preconceituosos, afetando desproporcionalmente pessoas a partir da utilização de regras e critérios discriminatórios.

Por meio de uma minuciosa auditoria nos algoritmos utilizados nos sistemas de decisão automatizado, a Força Tarefa pretende garantir que esses sistemas se alinhem com a meta de tornar a cidade de Nova York mais justa e equitativa. Os membros da Força-Tarefa incluem representantes de várias agências e escritórios governamentais, bem como parceiros do setor privado, organizações sem fins lucrativos, de defesa de direitos e comunidades de pesquisa[27].

Outro exemplo é o uso do algoritmo denominado COMPAS[28] pela Agência de Justiça Criminal do Estado norte-americano de Wisconsin. O algoritmo foi desenvolvido para determinar o grau de periculosidade de criminosos através de um sistema de

25 Ibidem.

26 NOVA YORK. The New York City Council. Int 1696-2017. Automated decision systems used by agencies. Disponível em: https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0 . Acesso em: 05 ago. 2020.

27 NOVA YORK. New York City Automated Decision Systems Task Force. Disponível em: https://www1.nyc.gov/site/adstaskforce/index.page . Acesso em: 6 ago. 2020.

28 Sigla em inglês para “Correctional Offender Management Profiling for Alternative Sanctions”

pontos que variam de 1 a 10, obtidos a partir de respostas de várias perguntas que avaliam a possibilidade de o criminoso reincidir, o que acaba influenciando a sua pena. Uma das perguntas é "se a pessoa mora numa área com alto índice de criminalidade".

A avaliação pode ser usada inclusive para decidir se a pessoa será solta com pagamento de fiança, se deve ser mandada para a prisão ou receber outro tipo de sentença e - se já estiver na cadeia - se tem direito à liberdade condicional. A intenção é tornar as decisões judiciais menos subjetivas - menos influenciáveis por erros humanos, preconceitos ou racismo.

Entretanto, como o algoritmo transforma respostas em pontos é um segredo comercial de propriedade da empresa que presta serviço ao Sistema Penitenciário no Estado de Wisconsin. Questionada, ela limita a informar que a tabela de risco se baseia em traços gerais de comportamento. A Suprema Corte de Wisconsin advertiu ainda que o COMPAS pode dar uma pontuação consideravelmente maior para infratores de minorias étnicas[29].

Mesmo que o COMPAS esteja programado para ser neutro e não seguir vieses, como um réu pode exercer o direito à ampla defesa e ao contraditório e contestar sua pontuação se o critério utilizado para restringir sua liberdade é uma "caixa-preta"?

Julia Angwin sustenta que quando analisamos um acusado negro e outro branco com a mesma idade, sexo e ficha criminal - e levando em conta que depois de serem avaliados os dois cometeram quatro, dois ou nenhum crime -, o negro tem 45% mais chances do que o branco de receber uma pontuação alta. Angwin cogita a possibilidade de o algoritmo não ter um viés racial, mas está expondo mais claramente os preconceitos raciais do sistema penal e da sociedade nos EUA e que tal fato merece reflexão: "queremos penalizar ainda mais os réus negros por viverem em áreas pobres e terem o que esse algoritmo considera atributos de maior periculosidade apesar de essas pessoas não serem perigosas? Ou estamos querendo usar esse sistema porque achamos que ele permite que todo mundo receba um tratamento justo?"[30].

29 MAYBIN, Simon. Sistema de algoritmo que determina pena de condenados cria polêmica nos EUA. BBC. 31 out. 2016. Disponível em: https://www.bbc.com/portuguese/brasil-37677421 Acesso em: 6 ago. 2020.

30 Ibidem.

O policiamento preditivo[31] pode recomendar que uma determinada área deve receber reforços no policiamento. A adoção da medida pode resultar mais apreensões de drogas, dentre outras ocorrências policiais, acarretando mais prisões. Estatisticamente, a região pode passar a ser classificada como uma área com alto índice de criminalidade, enquanto uma região com criminalidade equivalente pode não ser reconhecida como tal, em virtude de uma análise incompleta ou erro de classificação do algoritmo de policiamento preditivo. Entretanto, uma eventual falha ou omissão desta natureza pode gerar uma consequência injusta e discriminatória na vida do indivíduo que será julgado pelo algoritmo do COMPAS.

Esta hipótese ilustra a possibilidade de sistemas de IA poderem subsidiar decisões preconceituosas e reforçar estereótipos, ainda que de maneira não intencional por quem desenvolveu o algoritmo ou por quem usa o sistema.

O viés algorítmico discriminatório também pode surgir de maneira não intencional por consequência de limitações tecnológicas, tal como ocorre nas falhas relacionadas a identificação de perfis.

A tecnologia desenvolvida pela Nikon para reconhecimento facial, disponibilizada na câmera digital Nikon Coolpix S630, ao identificar um rosto de feição asiática questionava se "alguém piscou?". Apesar de ser um produto desenvolvido por uma empresa japonesa, seu algoritmo não funcionava adequadamente em consumidores com olhos orientais, o que demonstra um viés ocasional, ainda que claramente não intencional[32].

Situação mais delicada se constata quando o algoritmo utilizado pelo Google, empresa líder em IA e aprendizado de máquina, ainda apresenta dificuldades para identificar pessoas com total precisão. Em 2015, o aplicativo "Google Fotos" rotulou pessoas negras como "gorilas". Diante do constrangimento causado pela falha, o Google se declarou "chocado e genuinamente arrependido", evidenciando que os algoritmos de identificação são falíveis e possuem limites[33]. A solução implementada pelo Google

31 SILVA, Wellington Clay Porcino. Empregando o Policiamento Preditivo: Construção de um Modelo de Risco do Terreno para Crimes contra o Patrimônio dos Correios. Revista Brasileira de Ciências Policiais. Brasília, v. 7, n. 2, p. 53-71, jul/dez, 2016. p. 54-56.

32 ROSE, Adam. Are Face-Detection Cameras Racist? TIME. 22 jan. 2010. Disponível: http://content.time.com/time/business/article/0,8599,1954643,00.html. Acesso em: 08 ago. 2020.

33 BARR, Alistair. Google Mistakenly Tags Black People as 'Gorillas', Showing Limits of Algorithms.

foi a de retirar o rótulo “gorila” da indexação de imagens, deixando o aplicativo de fotos cego para gorilas, como se eles não existissem[34].

Ainda que usuários possam eventualmente relatar falhas e equívocos, a tecnologia de aprendizagem de máquina está limitada a experiência até então obtida pelo sistema, inclusive por meio da coleta de dados. Novamente nos reportamos a limitação da IA em compreender cenários complexos e tomar decisões que exijam habilidades típicas de seres humanos, como a capacidade de contextualizar, de usar o senso comum ou conceitos abstratos para interpretar e compreender o mundo tal como os seres humanos.

As iniciativas do GDPR da União Europeia, da LGPD, da Força Tarefa de Sistemas de Decisão Automatizada de Nova York e do Comitê de Ciência e Tecnologia do Parlamento Britânico partem da premissa que sistemas de tomada de decisão baseados em algoritmos de IA têm potencial para gerar desigualdade, exclusão e injustiça, fomentando inclusive a discriminação e o preconceito a partir de distinções adotadas a partir de critérios tais como a raça, cor, gênero, orientação sexual, religião, nacionalidade, cidadania, condição de saúde, opção política, situação financeira, idade, deficiência, estado civil ou tipo físico, a depender da forma como eles forem concebidos, programados e utilizados.

Há inúmeros cenários cotidianos em que já se percebe que decisões socialmente e juridicamente relevantes não estão sendo tomadas por pessoas, mas por sistemas de decisão automatizada, dentre elas: a concessão de um visto para estrangeiro, a definição do valor do prêmio de um seguro, as condições de contratação de um plano de saúde ou de um empréstimo financeiro, ou a escolha de um candidato em processo seletivo a vaga de emprego. Cada uma destas decisões automatizadas é capaz de gerar a perpetuação da desigualdade, do abuso, da discriminação e da injustiça, que se tornam ainda mais graves quando afetam grupos vulneráveis e minorias.

Daí a importância destes processos decisórios serem identificados, explicados, justificados e, se necessário, revistos judicialmente, em busca da neutralidade na

.....

The Wall Street Journal. 01 jul. 2015. Disponível em: https://blogs.wsj.com/digits/2015/07/01/google-mistakenly-tags-black-people-as-gorillas-showing-limits-of-algorithms/. Acesso em: 08 ago. 2020.

34 SIMONITE, Tom. When it comes to gorillas, google photos remains blind. WIRED. 01 nov. 2018. Disponível em: https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/. Acesso em: 08 ago. 2020.

tomada de decisões, a partir de regulamentação que tenha por objetivo prevenir o agravamento de questões sociais relevantes como o aumento da vulnerabilidade de indivíduos ou de grupos sociais tradicionalmente marginalizados, promovendo a justiça, a equidade e a dignidade da pessoa humana. Tais medidas objetivam o equilíbrio e a neutralidade das decisões tomadas por IA por meio de ações transparentes e da reavaliação das decisões automatizadas por seres humanos.

5. A NEUTRALIDADE É POSSÍVEL EM SISTEMAS DE INTELIGÊNCIA ARTIFICIAL?

Vieses discriminatórios presentes em algoritmos, tais como gênero, idade, condição física, deficiência, origem étnica, raça, orientação política, religião ou situação patrimonial, podem ser desejados ou ser fruto da mente subconsciente de programadores.

Para assegurar que os fundamentos das decisões automatizadas não sejam opacos e não respaldem tais vieses, a neutralidade no desenvolvimento e no uso da IA poderá ser obtida a partir da estrita obediência de diretrizes éticas fundamentadas na proteção do princípio da não discriminação, previsto nos arts. 1º, 2º e 7º da Declaração Universal dos Direitos Humanos[35], no art. 3º, incisos III e IV, e no art. 5, caput e incisos XLI e XLII da Constituição Federal[36] e nos arts. 1º e 20 da Lei de Crimes de Preconceito e Discriminação Racial[37].

Um ser humano pode ser flagrado agindo de maneira discriminatória, ainda que de forma indireta ou oculta. Entretanto, uma decisão automatizada baseada em uma programação discriminatória sutilmente disfarçada possui uma opacidade que, sem mecanismos de revisão, auditoria e transparência dos algoritmos, dificilmente será flagrada.

Realizar juízo de valor ou enfrentar as consequências de dilemas morais são

35 ONU. Declaração Universal dos Direitos Humanos. Disponível em: https://www.ohchr.org/EN/UDHR/Documents/UDHR_Translations/por.pdf. Acesso em: 09 ago. 2020.

36 BRASIL. Constituição (1988). Constituição da República Federativa do Brasil de 1988. Brasília, DF: Presidência da República. Disponível em: http://www.planalto.gov.br/ccivil_03/Constituicao/ConstituicaoCompilado.htm. Acesso em 09 ago. 2020.

37 BRASIL. Lei nº 7.716, de 5 de janeiro de 1989. Lei de Crimes de Preconceito e Discriminação Racial. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l7716.htm . Acesso em: 09 ago. 2020.

inerentes à condição humana, algo que não está ao alcance da IA, eis que ela (ainda) não é dotada de consciência ou subconsciência.

Para Yuval Harari, algoritmos não foram moldados pela seleção natural, não têm emoções nem instintos viscerais. Mas em momentos de crise, eles podem seguir diretrizes éticas muito melhor que os humanos, contanto que seja encontrada uma forma de codificar a ética em números e estatísticas precisos. De acordo com Harari, não podemos confiar na máquina para estabelecer os padrões éticos relevantes, pois tal tarefa deve caber exclusivamente aos humanos. Mas, uma vez que decidamos por um padrão ético, por exemplo, que é errado discriminar mulheres, ou pessoas negras, poderemos confiar em máquinas para implementar e manter esse padrão melhor que os humanos[38].

Para Patrick Lin, é notoriamente difícil traduzir corretamente em algoritmos um senso ético de maneira transparente e que produza resultados aceitáveis para a sociedade[39].

Algoritmos contam com vasta capacidade para processar dados previamente fornecidos, aprender a partir da análise destes dados, realizar previsões e tomar decisões de acordo com os limites de sua programação, sem avaliar se elas são neutras ou discriminatórias. O resultado dependerá, conforme dito, essencialmente da maneira como o algoritmo foi programado.

Neste processo, o programador que não estiver subordinado à diretrizes éticas previamente estabelecidas, poderá inserir suas visões políticas, econômicas, culturais e sociais no código dos algoritmos. Se ele for concebido por um programador preconceituoso, certamente o algoritmo incorporará e repetirá este padrão de comportamento, criando um viés discriminatório nas decisões automatizadas.

De acordo Salete Boff, Vinicius Fortes e Cinthia Freitas, um reflexo da aplicação das técnicas de tratamento de dados é a caracterização de perfil (*profiling*), que pode ser definido como métodos é técnicas computacionais aplicados aos dados pessoais

38 HAHARI, Yuval Noah. 21 lições para o século 21. Tradução: Paulo Geiger. São Paulo: Companhia das Letras, 2018. p.62-63.

39 LIN, Patrick. Why Ethics Matters for Autonomous Cars. In: Maurer M., Gerdes J., Lenz B., Winner H. (editores) Autonomous Driving: Technical, Legal and Social Aspects. Springer, Berlin, Heidelberg. p. 69-73. E-book. Disponível em: https://link.springer.com/book/10.1007%2F978-3-662-48847-8 . Acesso em: 09 ago. 2020.

ou não dos usuários, com objetivo de determinar o que é relevante dentro de um determinado contexto, tornando visível padrões que são invisíveis ao olho humano. Alguns reflexos da aplicação de *profiling* são a identificação de riscos e, também a discriminação, considerado um efeito colateral perigoso da aplicação das técnicas de tratamento de dados[40].

Neste contexto, Cinthia Freitas menciona a discriminação a partir do perfil de uma pessoa com pré-disposição de apresentar uma determinada doença[41].

Para Tess Posner, diretora executiva da AI4ALL, uma organização sem fins lucrativos que administra cursos de IA para estudantes de grupos minoritários, é essencial treinar um grupo heterogêneo para a próxima geração de trabalhadores da IA. Atualmente, apenas 13% das empresas de IA têm presidentes do sexo feminino e menos de 3% dos professores de engenharia nos EUA são negros. Posner defende que uma força de trabalho inclusiva pode ter mais ideias e identificar problemas nos sistemas antes que eles aconteçam, e a diversidade pode melhorar o resultado[42].

Defende-se que deve ser incentivado o treinamento e a formação de equipes profissionais de Tecnologia da Informação para o desenvolvimento de sistemas de IA com os mais diversos perfis, que correspondam a uma ampla parcela da sociedade, especialmente por representantes dos grupos que podem ser afetados por sistemas de decisões automatizadas.

Acredita-se que uma equipe heterogênea terá mais chances de desenvolver um algoritmo neutro e sem vieses discriminatórios (intencionais ou não). Trata-se do mesmo cenário ilustrado por Sunstein, ao descrever um Tribunal composto em sua maioria por homens brancos, ricos e idosos, e sua desvantagem epistêmica, talvez em virtude de sua relativa falta de diversidade, além do risco de sofrerem viés sistemático em suas decisões[43].

40 BOFF, S. O. FORTES, Vinícius Borges; FREITAS, C. Proteção de dados e privacidade: do direito às novas tecnologias na sociedade da informação. Rio de Janeiro: Lumen Juris, 2018. p.161-164.

41 FREITAS, Cinthia Obladen de Almendra. Tratamento de dados pessoais e a legislação brasileira frente ao profiling e à discriminação a partir das novas tecnologias. Rev. de Direito, Governança e Novas Tecnologias| e-ISSN: 2526-0049| Maranhão | v. 3 | n. 2 | p. 18 - 38 | Jul/Dez. 2017. p. 29.

42 SNOW, Jackie. For better AI, diversify the people building it. MIT Technology Review. 27 mar. 2018. Disponível em: https://www.technologyreview.com/s/610637/for-better-ai-diversify-the-people-building-it/ Acesso em: 09 ago. 2020.

43 SUNSTEIN, Cass. R. If people would be outraged by their rullings, should judges care? Rhe Social

Portanto, diante de dilemas semelhantes ao "dilema do bonde"[44], sistemas de decisão automatizada não poderão ser programados para decidir em favor de um determinado grupo de seres humanos em detrimento de outro, baseado em características distintivas superficiais, que viabilize a discriminação de pessoas.

Finalmente, é essencial que seja assegurada a neutralidade em sistemas de decisão automatizada. Para tanto, a programação da IA deverá ser feita a partir de critérios previamente estabelecidos, observando diretrizes éticas fundamentadas na proteção do princípio da não discriminação.

6. O DEVER DE JUSTIFICAR DECISÕES AUTOMATIZADAS PARA EVITAR DECISÕES DISCRIMINATÓRIAS

Enquanto sistemas de decisão automatizada não atingem seu ápice em termos de segurança e confiabilidade, precauções devem ser tomadas para garantir que as decisões sejam passíveis de análise e revisão, assegurando sua neutralidade, evitando o uso indevido da tecnologia e prevenindo sua utilização para fins discriminatórios, ilícitos ou abusivos. Para tanto, é fundamental que os critérios utilizados em decisões automatizadas sejam devidamente justificados de maneira transparente.

No Brasil, o Projeto de Lei 21/2020 - Câmara dos Deputados, de autoria do Deputado Federal Eduardo Bismark, estabelece princípios, direitos, deveres e instrumentos de governança para o uso da IA e determina as diretrizes para a atuação da União, dos Estados, do Distrito Federal e dos Municípios, pessoas físicas e jurídicas, de direito público ou privado, e entes sem personalidade jurídica em relação ao uso de IA no Brasil[45].

Os fundamentos do uso da IA no Brasil, segundo o art. 4º do Projeto de Lei, são: o respeito aos direitos humanos e aos valores democráticos; a igualdade, a não

.....

Science Research Network Eletronic Paper Collection: http://ssrn.com/abstract_id=965581. P.38.

44 FOOT, Philippa. The Problem of Abortion and the Doctrine of the Double Effect. Oxford Review, no. 5, 1967. Disponível em: http://www2.pitt.edu/~mthompso/readings/foot.pdf. Acesso em 10 ago. 2020.

45 BRASIL. Câmara dos Deputados. Projeto de Lei 21/2020. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências. Texto original. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1853928. Acesso em: 12 ago. 2020.

discriminação, a pluralidade e o respeito aos direitos trabalhistas; e a privacidade e a proteção de dados.

De acordo com o inciso I do art. 5º do referido Projeto, o uso da IA no Brasil tem por objetivo a promoção da pesquisa e do desenvolvimento da IA ética e livre de preconceitos. Por sua vez, os princípios para o uso responsável de IA no Brasil estão definidos no art. 6º.

Conforme a redação do art. 7º do Projeto, são direitos das partes interessadas no sistema de IA, utilizado na esfera privada ou pública:

> I - ciência da instituição responsável pelo sistema de inteligência artificial;
> II - acesso a informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados pelo sistema de inteligência artificial que lhes afetem adversamente, observados os segredos comercial e industrial; e
> III - acesso a informações claras e completas sobre o uso, pelos sistemas, de seus dados sensíveis, conforme disposto no art. 5º, II, da Lei Geral de Proteção de Dados.

Se constata nas diretrizes propostas no Projeto de Lei 21/2020, que a regulamentação da IA, e consequentemente, nos sistemas de decisão automatizada, deve estar centralizada no ser humano, devendo ser assegurada a transparência, a explicabilidade, a responsabilização e a prestação de contas, devendo contribuir para promover o respeito aos Direitos Humanos, aos valores democráticos e para evitar o preconceito e a discriminação.

A LGPD[46], estabelece em seu art. 6º que as atividades de tratamento de dados pessoais deverão observar a boa-fé e ao princípio da não discriminação, afirmando a impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos. Por sua vez, o art. 20 assegura ao titular dos dados o direito de solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade.

Em que pese a LGPD não objetivar regulamentar sistemas de decisão

46 BRASIL. Lei n.º 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm. Acesso em: 10 ago. 2020.

automatizada, os artigos mencionados se aplicam perfeitamente aos fins propostos, eis que promovem formas eficazes de combater a discriminação, o preconceito e a promover a transparência nas decisões.

Portanto, devem ser observados os seguintes aspectos nos sistemas de decisão automatizada:

> a) Que o cidadão saiba que a decisão que lhe impactou negativamente decorreu de um sistema integralmente automatizado;
> b) Que seja assegurado o direito do cidadão de obter informações claras e adequadas sobre os critérios adotados por sistemas de decisão automatizada que possam afetar direitos fundamentais, assegurando o direito à explicação de maneira transparente;
> c) Que os sistemas de decisão automatizada sejam desenvolvidos e executados de forma a não permitir qualquer forma de discriminação ou preconceito;
> d) Que os sistemas de decisão automatizado possam ser submetidos à auditoria e revisão por órgão público independente e os operadores de tais sistemas possam ser demandadas a justificar a programação e as escolhas feitas pelos sistemas de IA e ser responsabilizado em caso de violação de direitos.

O controlador[47] deverá fornecer informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada. Em caso de não oferecimento de informações, a autoridade nacional[48] poderá realizar auditoria para verificação de aspectos discriminatórios em tratamento automatizado de dados pessoais.

Onde os algoritmos puderem afetar significativamente direitos, a resposta deve ser uma combinação de explicação e transparência tanto quanto possível[49], inclusive para permitir que indivíduos possam questionar os resultados de todas as decisões significativas que os algoritmos lhe afetam, e quando apropriado, buscar a devida reparação para os impactos de tais decisões[50].

47 Pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais, nos termos do art. 5, VI, da LGPD.

48 Órgão da administração pública responsável por zelar, implementar e fiscalizar o cumprimento desta Lei em todo o território nacional, conforme o art. 5º, XIX, da LGPD.

49 INGLATERRA. House of Commons. Science and Technology Committee. Algorithms in decision-making. Fourth Report of Session 2017–19. Publicado em: 23 de maio de 2018. Disponível em: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf. Acesso em: 05 ago. 2020.

50 Ibidem.

Assim, decisões baseadas em sistemas de decisão automatizada devem ser devidamente justificadas e passíveis de serem auditadas por comitês independentes para examinar de maneira transparente os modos de operação dos sistemas de decisão automatizada, viabilizando, se necessário, a devida revisão judicial, especialmente quando tiver o potencial de afetar direitos fundamentais ou a capacidade para promover práticas discriminatórias.

Para alcançar tal intento, torna-se imperativa a criação de regulamentação estatal (leis e políticas públicas) que estabeleçam diretrizes éticas e de governança no desenvolvimento e uso de sistemas de inteligência artificial.

CONSIDERAÇÕES FINAIS

Atenta-se para a possibilidade de os mesmos vieses e influências sociais mencionados por Cass Sunstein serem replicados por sistemas de IA, aptos a promover discriminação e preconceito de uma maneira sutil e difícil de detectar.

Diante do caminho perigoso a ser percorrido em busca do amadurecimento e da previsibilidade das tecnologias presentes nos sistemas de IA, não se pode ignorar o grande potencial para causar consequências indesejadas, em especial, o risco de tais sistemas amplificarem preconceitos sociais em cenários socialmente relevantes.

Estes riscos precisam ser identificados, compreendidos, questionados e enfrentados. Assim, iniciativas como o PL 21/2020 tornam-se necessárias para a criação de diretrizes legais que estabeleçam critérios objetivos para que as referidas decisões automatizadas sejam adequadamente identificadas e justificadas, promovendo a neutralidade e a transparência e impedindo que a tecnologia envolvendo IA viabilize a tomada de decisões discriminatórias ou preconceituosas.

Ainda que a pesquisa e o desenvolvimento tecnológico relacionado à IA e a decisão automatizada avancem a passos largos, a grande complexidade para treinar sistemas de IA ensina que a implementação responsável, segura e isenta de vieses discriminatórios em sistemas de tomada de decisão automatizada exige cautela e controlabilidade.

Neste sentido, nos reportamos à legislação brasileira e estrangeira que combate o preconceito e a discriminação. Tais regramentos deverão nortear e disciplinar o desenvolvimento tecnológico, os testes e o uso de sistemas de decisão automatizada

no Brasil.

Propõe-se que sistemas de IA sejam aptos a justificar as decisões emanadas, notadamente quando estas tiverem o potencial de promover desigualdade e preconceito, permitindo coletar dados que viabilizem a investigação e identificação das causas e o contexto de decisão, possibilitando sua revisão e viabilizando que seja tomada uma nova decisão, inclusive sob a supervisão humana, ou, em último caso, a revisão judicial da decisão supostamente enviesada.

Além disso, através da análise dos dados, as autoridades públicas e as partes interessadas poderão auditar o método utilizado pelos sistemas de decisão automatizada nas decisões emanadas, expondo eventuais vieses, intencionais ou acidentais.

Tais medidas também viabilizarão que se verifique a neutralidade dos algoritmos relacionados ao sistema de decisão automática. A transparência do sistema de IA poderá evitar a programação deliberada de algoritmos para a tomada de decisões que contrariem os objetivos fundamentais da República Federativa do Brasil.

Portanto, propõe-se a necessidade da rigorosa aplicação da legislação existente e de regulamentação estatal, estabelecendo o dever de fundamentar as decisões emanadas de maneira automatizada, protegendo o cidadão contra a violação de seus direitos fundamentais, previstos no art. 3º, inciso IV, art. 5º, XLI, da Constituição, em particular, a não discriminação a partir do uso de novas tecnologias e o respeito ao pluralismo.

Neste processo, torna-se imperativo que os algoritmos de IA sejam desenvolvidos com neutralidade, passíveis de serem auditados de forma transparente pelas autoridades públicas para descobrir eventuais vieses discriminatórios inseridos na programação e tomar as providências necessárias para combater tal prática, responsabilizando os responsáveis nos termos da legislação que asseguram a observância do princípio da não discriminação.

Assim, diante dos potenciais riscos envolvidos, propõe-se que não se permita a auto regulamentação deste setor relevante e estratégico. A regulamentação deve competir aos juízes, interpretando e aplicando o Direito, em conformidade com as teorias da justiça e da democracia, e dando efetividade ao princípio da não discriminação, e aos representantes do povo eleitos democraticamente, administradores públicos,

criando políticas públicas para incentivar, disciplinar e fiscalizar o uso adequado e impactos sociais dos referidos sistemas e aos representantes do poder legislativo, criando legislação que defina diretrizes éticas que impeçam a utilização de critérios discriminatórios nos sistemas de tomada de decisão automatizada, sempre que um ser humano possa ser impactado negativamente e injustamente.

Assim, decisões baseadas em sistemas de inteligência artificial, tal como as decisões emanadas por seres humanos, devem ser devidamente justificadas, inclusive para viabilizar a devida revisão judicial, especialmente quando tiverem o potencial de afetar direitos fundamentais ou o potencial para promover práticas preconceito e a discriminação

Finalmente, o ser humano e a proteção à sua vida e dignidade deve ser colocado no centro do debate a respeito do desenvolvimento e do uso de sistemas de decisão automatizada, eis que trata-se de tecnologia disruptiva apta tanto a promover o progresso social e a eficiência econômica quanto colocá-los em risco, se utilizados critérios que não observem o direito à não discriminação previsto nos artigos 1º, 2º e 7º da Declaração Universal dos Direitos Humanos, no art. 3º, incisos III e IV, e no art. 5, caput e incisos XLI e XLII da Constituição.

Referências

BARR, Alistair. Google Mistakenly Tags Black People as ‘Gorillas’, Showing Limits of Algorithms. **The Wall Street Journal.** 01 jul. 2015. Disponível em: https://blogs.wsj.com/digits/2015/07/01/google-mistakenly-tags-black-people-as-gorillas-showing-limits-of-algorithms/. Acesso em: 08 ago. 2020.

BOFF, S. O. FORTES, Vinícius Borges; FREITAS, C. **Proteção de dados e privacidade: do direito às novas tecnologias na sociedade da informação**. Rio de Janeiro: Lumen Juris, 2018. p.161-164.

BRASIL. Câmara dos Deputados. **Projeto de Lei 21/2020**. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências. Texto original. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1853928. Acesso em: 12 ago. 2020.

_____. Constituição (1988). **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República. Disponível em: http://www.planalto.gov.br/ccivil_03/Constituicao/ConstituicaoCompilado.htm. Acesso em 09 ago. 2020.

_____. **Lei n.º 13.105, de 16 de março de 2015**. Código de Processo Civil. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm. Acesso em: 12 ago. 2020.

_____. **Lei n.º 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm. Acesso em: 08 ago. 2020.

_____. **Lei nº 7.716, de 5 de janeiro de 1989**. Lei de Crimes de Preconceito e Discriminação Racial. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l7716.htm . Acesso em: 09 ago. 2020.

FOOT, Philippa. **The Problem of Abortion and the Doctrine of the Double Effect**. Oxford Review, no. 5, 1967. Disponível em: http://www2.pitt.edu/~mthompso/readings/foot.pdf. Acesso em 10 ago. 2020.

FREITAS, Cinthia Obladen de Almendra. Tratamento de dados pessoais e a legislação brasileira frente ao profiling e à discriminação a partir das novas tecnologias. **Rev. de Direito, Governança e Novas Tecnologias**| e-ISSN: 2526-0049| Maranhão | v. 3 | n. 2 | p. 18 - 38 | Jul/Dez. 2017. p. 29.

GOOGLE. **Google Flu Trendes and Google Dengue Trends**. 2019. Disponível em: https://www.google.org/flutrends/about/. Acesso em: 09 ago. 2020.

HAHARI, Yuval Noah. **21 lições para o século 21**. Tradução: Paulo Geiger. São Paulo: Companhia das Letras, 2018. p.62-63.

HAN, Jiawei, KAMBER, Micheline, PEI, Jian. **Data Mining**: Concepts and Techniques. 3 ed., San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011. p.24.

INGLATERRA. House of Commons. Science and Technology Committee. **Algorithms**

in decision-making. Fourth Report of Session 2017–19. Publicado em: 23 de maio de 2018. Disponível em: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf. Acesso em: 05 ago. 2020.

KAIN, Erik. Microsoft's Teenage, Nazi-Loving AI Is The Perfect Social Commentary For Our Times. **Forbes**. 24 mar. 2016. Disponível em: https://www.forbes.com/sites/erikkain/2016/03/24/microsofts-teenage-nazi-loving-ai-is-the-perfect-social-commentary-for-our-times/#6c3cc0bd235a Acesso em: 11 jul. 2019.

KNIGHT, Will. The Dark Secret at the Heart of AI. **MIT Technology Review**. 11 abr. 2017. Disponível em: https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/. Acesso em: 06 ago. 2020.

LIN, Patrick. **Why Ethics Matters for Autonomous Cars**. In: Maurer M., Gerdes J., Lenz B., Winner H. (editores) Autonomous Driving: Technical, Legal and Social Aspects. Springer, Berlin, Heidelberg. p. 69-73. *E-book.* Disponível em: https://link.springer.com/book/10.1007%2F978-3-662-48847-8 . Acesso em: 09 ago. 2020.

LUGER, George F. **Inteligência artificial**; tradução: Daniel Vieira. 6. ed. São Paulo: Pearson Education do Brasil, 2013. P. 1.

MAYBIN, Simon. Sistema de algoritmo que determina pena de condenados cria polêmica nos EUA. **BBC**. 31 out. 2016. Disponível em: https://www.bbc.com/portuguese/brasil-37677421 Acesso em: 6 ago. 2020.

METZ, Rachel. Microsoft's neo-Nazi sexbot was a great lesson for makers of AI assistants.

MIT Technology Review. 27 mar. 2018. Disponível em: https://www.technologyreview.com/s/610634/microsofts-neo-nazi-sexbot-was-a-great-lesson-for-makers-of-ai-assistants/. Acesso em: 05 ago. 2020

NOVA YORK. **New York City Automated Decision Systems Task Force.** Disponível em: https://www1.nyc.gov/site/adstaskforce/index.page . Acesso em: 6 ago. 2020.

NOVA YORK. The New York City Council. **Int 1696-2017. Automated decision systems**

used by agencies. Disponível em: https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0 . Acesso em: 05 ago. 2020.

O'NEIL, Cathy. **Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy**. New York: Crown, 2016. p.2-4.

ONU. **Declaração Universal dos Direitos Humanos**. Disponível em: https://www.ohchr.org/EN/UDHR/Documents/UDHR_Translations/por.pdf. Acesso em: 09 ago. 2020.

ROSE, Adam. Are Face-Detection Cameras Racist? **TIME**. 22 jan. 2010. Disponível: http://content.time.com/time/business/article/0,8599,1954643,00.html. Acesso em: 08 ago. 2020.

SILVA, Wellington Clay Porcino. Empregando o Policiamento Preditivo: Construção de um Modelo de Risco do Terreno para Crimes contra o Patrimônio dos Correios. **Revista Brasileira de Ciências Policiais.** Brasília, v. 7, n. 2, p. 53-71, jul/dez, 2016. p. 54-56.

SIMONITE, Tom. When it comes to gorillas, google photos remains blind. **WIRED**. 01 nov. 2018. Disponível em: https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/. Acesso em: 08 ago. 2020.

SNOW, Jackie. For better AI, diversify the people building it. **MIT Technology Review**. 27 mar. 2018. Disponível em: https://www.technologyreview.com/s/610637/for-better-ai-diversify-the-people-building-it/ Acesso em: 09 ago. 2020.

UNIÃO EUROPEIA. Jornal Oficial da União europeia. Edição em língua portuguesa. **REGULAMENTO (UE) 2016/679 DO PARLAMENTO EUROPEU E DO CONSELHO de 27 de abril de 2016.** Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=OJ:L:2016:119:FULL . Acesso em: 10 ago. 2020.