Revista do Direito

https://online.unisc.br/seer/index.php/direito/index
ISSN on-line: 1982 - 9957
DOI: 10.17058/rdunisc.vi67.16337

DISCRIMINAÇÃO ALGORITMICA: INTELIGENCIA ARTIFICAL, VIESES HUMANOS E ALGORÍTIMICOS E A PROTEÇÃO CONSTITUCIONAL

ALGORITHMIC DISCRIMINATION: ARTIFICIAL INTELLIGENCE, HUMAN AND ALGORITHICAL BIASES AND CONSTITUTIONAL PROTECTION

Dirceu Pereira Siqueira [1]
Leonan Roberto de França Pinto [2]
Ernani José Pera Junior[3]

Recebido em: 23/12/2023
Aceito em: 11/12/2023

dpsiqueira@uol.com.br
leonan.roberto@gmail.com
ernanipera@hotmail.com

Resumo: O uso da Inteligência Artificial promete respostas simples para soluções complexas. Para além das relações de consumo, no Brasil o seu emprego avança rapidamente por meio de decisões que afetam os aspectos centrais de vida de milhões de brasileiros. Um dos efeitos indesejados, no entanto, é a discriminação algorítmica, que enseja a violação de direitos da personalidade. Quais os aspectos técnicos e jurídicos contribuem para uma decisão de Inteligência Artificial sem discriminação algorítmica? Para resposta a este problema, o presente artigo inicialmente demonstra que tanto as decisões humanas quanto da Inteligência Artificial possuem vieses que fitam a formação de generalizações, lugares comuns, estereótipos e estigmatização e expõe a taxonomia desses vieses. Em seguida, expor-se-á as fontes principais de preconceitos que podem resultar em discriminação por algoritmos, a definição do problema da "caixa preta" e o conceito de *profiling*. Ao final, apontar-se-á a necessidade de políticas de incentivo a empresas a desenvolverem modelos explicáveis de IA e a auditarem sistemas já em curso, bem como uma maior atenção aos decisores humanos ao chamado viés da automação quando da revisão humana das decisões automatizadas, tudo com foco na proteção constitucional da proteção de dados e direito à não discriminação. Quanto à metodologia, utilizou-se método de abordagem hipotético-dedutivo a partir de revisão bibliográfica.

Palavras-chave: Inteligência Artificial; direito fundamental; discriminação algorítmica; revisão humana e vieses.

Abstract: The use of Artificial Intelligence promises simple answers to complex solutions. In addition to consumer relations, in Brazil, employment is advancing rapidly through decisions that affect the central aspects of life for millions of Brazilians. One of the undesired effects, however, is algorithmic discrimination, which entails the violation of personality rights. What technical and legal aspects contribute to an Artificial Intelligence decision without algorithmic discrimination? To answer this problem, the present article initially demonstrates that both human and Artificial Intelligence decisions have biases that aim at the formation of generalizations, commonplaces, stereotypes and stigmatization and exposes the taxonomy of these biases. Then, the main sources of prejudice that can result in discrimination by algorithms, the definition of the "black box" problem and the concept of profiling will be exposed. In the end, it will be pointed out the need for policies to encourage companies to develop explainable models of AI and to audit systems already in place, as well as greater attention to human decision makers to the so-called automation bias when a human review of automated decisions , all focusing on the constitutional protection of data protection and the right to non-discrimination. As for the methodology, we used a hypothetical-deductive approach based on a literature review.

[1] Centro Universitário de Maringá - UNICESUMAR
[2] Universidade Cesumar - UNICESUMAR - Maringá/PR
[3] Universidade Cesumar - UNICESUMAR - Maringá/PR

Keywords: Artificial intelligence; fundamental law; algorithmic discrimination; human review and biases.

1. INTRODUÇÃO

Na atual era do Big Data, a Inteligência Artificial (IA) promete soluções simples para problemas sociais e econômicos complexos do mundo. Para além das relações de consumo, no Brasil ela tem avançado rapidamente em pautas acentuadamente em outras pautas. Durante o ápice da pandemia do COVID-19, por exemplo, a partir do cruzamento de diversos banco de dados, ela foi responsável por selecionar os brasileiros necessitados que atendiam os requisitos legais e deferir automaticamente o benefício do Auxílio Emergencial, garantindo rapidamente a segurança alimentar e dignidade a milhões de pessoas. Atualmente, a IA também atua via reconhecimento facial como validação da autodeclaração racial para candidatos que desejam concorrer nas cotas para pardos, negros e indígenas com vistas ao ingresso em universidade, contribuindo para a promoção das políticas afirmativas. Em aeroportos, a IA faz o reconhecimento biométrico e automaticamente libera o acesso de passageiros ao portão de embarque e aeronave, facilitando a liberdade ambulatorial.

Infelizmente, nem tudo funciona como o idealizado e um dos efeitos negativos é a discriminação algorítmica, altamente ofensiva aos direitos da personalidade. Assim como as decisões humanas são fortemente enviesadas por fatores psicológicos explicados pela econômica comportamental, as decisões da IA também experimentam vieses que a conduzem para escolhas disfuncionais e discriminatórias.

A par desse quadro, problematiza-se a seguinte questão: quais os aspectos técnicos e jurídicos contribuem para uma decisão pela Inteligência Artificial sem discriminação algorítmica?

Como hipótese para resolução desse problema, no primeiro tópico deste artigo mostrar-se-á a relevância do tema a partir de exemplos de como a Inteligência Artificial tem afetado a vida cotidiana dos brasileiros. Como se verá, para além das relações de consumo, como as já conhecidas análise de crédito e aquisição de produtos e serviços, a IA atua em aspectos centrais da vida humana, como a

escolha de brasileiros elegíveis para programas públicos de transferência de renda, cotas em universidade e ações na saúde pública e privada.

Na sequência, o segundo capítulo terá como objeto a taxonomia dos vieses humanos e algorítmicos que influenciam a formação de generalizações, lugares comuns, estereótipos, estigmatização e preconceitos. Em seguida, expor-se-á as fontes principais de preconceitos que podem resultar em discriminação por algoritmos, a definição do problema da "caixa preta" e o conceito de *profiling*.

Na conclusão, apontar-se-á a necessidade de políticas públicas de incentivo a empresas a desenvolverem modelos explicáveis de IA e auditarem sistemas já em curso, bem como a necessidade de protocolos para maior atenção aos gestores ao chamado viés da automação quando da revisão humana das decisões automatizadas, tudo com foco na proteção constitucional da proteção de dados e direito à não discriminação.

Quanto à metodologia, utilizou-se método de abordagem hipotético-dedutivo a partir de revisão bibliográfica da área jurídica, da ciência comportamental e da tecnologia da informação.

2. RELAÇÕES EXISTÊNCIAS NO BRASIL E A INTELIGÊNCIA ARTIFICIAL

Durante o pico da pandemia do COVID-19 no primeiro semestre do ano de 2020, com o fechamento do comércio decorrente do lockdown e a impossibilidade de as pessoas saírem de casa para buscar renda para a sua própria sobrevivência, o poder público federal editou a Lei 13982/2020 concedendo um auxílio emergencial no valor de R$ 600,00 (seiscentos reais) para brasileiros que atendessem alguns requisitos, entre eles, não possuir emprego formal ativo, não ser titular de benefício previdenciário ou assistencial, possuir renda per capita familiar de até 1/2 (meio) salário mínimo, não exercer atividade empresarial e não ter outro membro da família recebendo o auxílio.

Considerando o maior programa público de transferência de renda no Brasil e à vista da impossibilidade de uma análise humana do perfil socioeconômico de cada brasileiro requerente do benefício em um curto espaço de tempo, a solução encontrada foi implementar um processamento por Inteligência Artificial (IA) a partir de um complexo cruzamento de dados reunidos de diversos cadastros públicos,

entre eles, dados de CPF e CNPJ da Receita Federal do Brasil, da Relação Anual de Informações Sociais (RAIS) e Cadastro Geral de Empregados e Desempregados (CAGED) do Ministério da Economia, do Cadastro Único para Programas Sociais do Governo Federal (CadÚnico) do Ministério da Cidadania, do Cadastro Nacional de Informações Sociais (CNIB) do INSS e do Sistema de Controle de Óbitos (SISOB) do Dataprev.

Segundo o balanço divulgado pelo Ministério da Cidadania[4], em 2020, cerca de 68,3 milhões de pessoas foram elegíveis para o recebimento do referido benefício. O governo não revela quantos brasileiros tiveram seus pedidos indeferidos, mas desse total cerca de 60 mil tiveram seus benefícios implementados somente após decisão judicial dos Juizados Especiais Federais de todo o Brasil. Sem embargo, após muitas reclamações sobre as respostas eletrônicas do indeferimento por desatualização de dados ou por equívocos no processamento, visto que o algoritmo fazia muitas inferências erradas quanto à composição familiar ou negava o pedido do benefício se "os dados eram inconclusivos", o governo implementou uma possibilidade limitada de "contestar", via aplicativo, o indeferimento do pedido de um benefício assistencial.

De outro giro, a partir do segundo semestre de 2022, a Universidade Estadual de Campinas (UNICAMP) decidiu empregar no seu vestibular uma tecnologia de reconhecimento facial para validar a autodeclaração de candidatos que concorrerem às vagas reservadas para pretos, pardos e indígenas[5].

Segundo o noticiado, os desenvolvedores da tecnologia, batizada de *SmartFace*, esperam que ela evite fraude em cotas raciais. Pelas regras informadas, se o sistema reconhecer traços fenotípicos de pretos, partos ou indígenas, a banca não poderá vetá-lo[6], funcionando como uma espécie de "juiz eletrônico racial" definitivo. Por outro lado, se o reconhecimento facial não identificar as características fenotípicas do candidato, ele será submetido a uma videochamada com a comissão.

Outrossim, em agosto de 2022 começou a valer de forma definitiva e gradual o embarque de passageiros nos aeroportos de Congonhas (SP) e Santos Dumont (RJ)

[4]BRASIL. Ministério da Cidadania. Disponível em https://aplicacoes.mds.gov.br/sagi/vis/data3/?g=2 Acesso em 02/08/2022.
[5]Disponível em https://www1.folha.uol.com.br/educacao/2022/08/unicamp-vai-usar-reconhecimento-facial-para-evitar-fraude-em-cotas.shtml Acesso em 02/08/2022.
[6]Disponível em https://istoe.com.br/correcao-unicamp-tera-reconhecimento-facial-para-evitar-fraudes-em-cotas/ Acesso em 02/08/2022.

de forma 100% digital por reconhecimento facial biométrico. Segundo a Agência Nacional de Aviação Civil (ANAC), a análise dos dados e validação por biometria ocorre a partir da base de dados governamental, em especial o CNH digital e o Título de Eleitor Digital, e dispensa a apresentação de cartão de embarque e documento de identificação para acesso ao portão de embarque e à aeronave. Claro, em caso de falha, haverá um funcionário da companhia aérea para auxiliar, mas a decisão do algoritmo é cerca de 27% mais rápida se comparado à identificação pessoal documental[7].

Já na área da saúde, segundo dados da Pesquisa TIC Saúde 2021[8], do Comitê Gestor da Internet no Brasil (CGI.br), conduzida pelo Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br) do Núcleo de Informação e Coordenação do Ponto BR (NIC.br), em torno de 88% dos estabelecimentos de saúde no Brasil possuem informações dos pacientes em formato digital, parcela deles com funcionalidades mais avançadas como listas pacientes por tipo de diagnóstico, exames laboratoriais compilados e possibilidade de realização de prescrição médica. No entanto, apenas 4% dos estabelecimentos (cerca de4.268) realizam análise de Big Data, sendo 1% dos públicos (cerca de 635) e 6% dos privados (cerca de 3.633). A principal fonte foram os dados dos próprios estabelecimentos coletados das fichas cadastrais.

No campo securitário, o emprego da IA na saúde suplementar tem ganhado cada vez mais espaço no Brasil com a promessa de ser um diferencial competitivo para as operadoras de planos de saúde. A aplicação inclui, entre outros, rotina de trabalho dos colaboradores e otimização de autorizações dos procedimentos com regras de regulação automatizadas. Uma grande preocupação, à medida em que a ciência atuarial e estatística identifica a probabilidade de uma doença, é a falta de transparência sobre como o algoritmo decide se o risco é assegurável ou não e, em caso positivo, o respectivo prêmio, segregando-se, por exemplo, pessoas com características genéticas propensas a desenvolver certas doenças. Não por outra razão, o tratamento preditivo de dados genéticos para fins securitários teve lugar na Convenção de Strasbourg sobre Direitos Humanos e Biomedicina e Convenção de

[7] Disponível em https://www1.folha.uol.com.br/mercado/2022/08/como-funciona-a-ponte-aerea-com-embarque-biometrico-entre-rio-e-sao-paulo.shtml Acesso em 09/08/2022
[8]Disponível em https://cetic.br/pt/publicacao/pesquisa-sobre-o-uso-das-tecnologias-de-informacao-e-comunicacao-nos-estabelecimentos-de-saude-brasileiros-tic-saude-2021/ Acesso em 13/08/2022.

Oviedo, com recomendações específicas sobre dados genéticos (BIONI, 2021, p. 345-346).

Os exemplos acima citados revelam o rápido avanço da IA sobre os aspectos centrais da vida humana. Fortemente presente nas relações de consumo, como por exemplo na análise de crédito, aquisição de produtos e serviços por canais digitais e atendimento eletrônico ao consumidor por *chatbots*, hoje o tema facilmente atinge direitos da personalidade bastante sensíveis como direito ao patrimônio mínimo e segurança alimentar, dignidade, liberdade ambulatorial, saúde e acesso à meio de transporte. Segundo informações da Secretaria Especial de Desburocratização, Gestão e Governo Digital do Ministério da Economia, atualmente são 4015 (quatro mil e quinze) serviços digitalizados pelo Poder Público Federal que se utilizam de 261 (duzentos e sessenta e uma) base de dados diferentes, estando entre os mais comuns a obtenção da carteira de trabalho digital, a carteira digital de trânsito, a solicitação de aposentadoria ou de benefício assistencial e a inscrição em programa público de acesso às universidades federais[9].

Em boa medida, há uma crença de que algoritmos tomam decisões ou auxiliam os humanos a tomarem decisões mais isentas. No entanto, no próximo tópico falar-se-á como seres humanos e máquinas são influenciados por vieses e como isso pode resultar em estigmatização ou discriminação.

3. VIESES HUMANOS *VERSUS* VIESES ALGORÍTMICOS

As pesquisas desenvolvidas pelos psicólogos israelenses Amos Tversky e Daniel Kahneman desde a década de 70 já comprovaram a falácia da crença cega na racionalidade humana, assim compreendida como a tomada de decisão com coerência lógica e internamente consciente. Não se quer dizer que as decisões humanas são irracionais, o que conota impulsividade, emotividade e resistência ao argumento razoável, mas sim que são fortemente influenciadas por vieses cognitivos em grande parte do tempo em decisões relevantes da vida. Os autores citam a presença do viés da confirmação (busca de uma informação que ratifica uma pré-compreensão), viés da disponibilidade (tendência a interpretar a realidade com base em informações mais recentes ou abundantes), viés da ancoragem e ajuste (um

[9] Disponível em https://www.gov.br/governodigital/pt-br Acesso em 14/08/2022

dado inicial que serve como um “âncora” e comparação para os demais) e o viés da representatividade (influência por estereótipos ou semelhança) como fatores que pesam sobre a racionalidade (KAHNEMAN, 2012, p. 510-539).

Na obra *Nudge*, ao sugerir o implemento de uma forma suave de paternalismo nos políticas públicas, cunhado de “paternalismo libertário” (proteção dos indivíduos contra suas próprias decisões, mas preservando o núcleo de sua autonomia), o economista Richard Thaler e o jurista americano Cass Sunstein citam, como forte influência psicológica nas decisões humanas, o viés do otimismo irreal (superestimação que faz as pessoas deixarem de tomar decisões sensatas de prevenção), a teoria do prospecto (aversão à perda), o viés do *status quo* (tendência geral a permanecer na inércia), o efeito de enquadramento (as decisões dependem, em parte, pela forma com que os problemas são apresentados), o comportamento de manada (busca pela conformidade coletiva), o efeito holofote (tendência do indivíduo achar que está no centro da atenção) e o efeito *priming* (influências sutis ou indiretas que aumentam a facilidade com que determinada decisão se forma) (THALER e SUNSTEIN, 2019, p. 36-146).

Outrossim, ao explicar a razão pela qual as pessoas em geral fornecem suas informações na internet, inclusive dados pessoais sensíveis (origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico) sem grandes preocupações com a privacidade, Bruno Bioni (2021, p. 241/245) menciona a influência da teoria da utilidade subjetiva (tendência de focar nos benefícios imediatos) e o problema estrutural do *trade-off* da economia de dados (decisão que consiste na escolha de uma opção em detrimento de outra, ou seja, o câmbio do acesso ao serviço em troca da privacidade).

De um modo geral, todos esses vieses inconscientes direcionam os julgamentos humanos e fitam a formação de generalizações, lugares comuns, estereótipos pensamentos tendenciosos, intuições, pré-compreensão e preconceitos de cada pessoa. Até mesmo a racionalidade jurídica, mais refletida, complexa e guiada por princípios abstratos e regras da ciência do Direito, não goza de uma neutralidade asséptica de vieses. Em sua tese de doutorado na UERJ, o Procurador do Estado do Rio de Janeiro Francesco Conte concluiu que a sentença civil, em sua gênese, é ilógica, influenciada da mente inconsciente. Para ele, a decisão judicial

possui 2 (dois) momentos claramente distintos e traz pesquisas empíricas que evidenciam que em primeiro lugar vem o julgamento, chamado de “contexto de descoberta da decisão”, permeado por fatores extrajurídicos do psiquismo como a intuição, sentimento e emoção. Somente logo após vem o raciocínio discursivo, inferência lógica e dedutivismo, momento chamado de “contexto de justificação”. Para o referido autor, o Juiz não inicia, mas finda a racionalidade jurídica com silogismo, pois o juízo (julgamento) tem caráter inventivo, ao passo que o raciocínio silogístico tem natureza demonstrativa (CONTE, 2020, p. 509-601).

Expostos alguns dos vieses da inteligência “natural”, a Inteligência Artificial também possui fenômenos e fatores não intencionalmente programados que interferem no seu processamento, lógica e tomada de decisões ao ponto de produzir um resultado tendencioso, incorreto ou injusto. São os chamados vieses algorítmicos, também chamado de viés na IA.

Algoritmo pode ser descrito como um conjunto de instruções, organizadas de forma sequencial, que determina como algo deve ser feito. Seu objetivo é, sobretudo, solucionar problemas e auxiliar na tomada de decisões com previsões utilizando probabilidades e estatísticas. A partir do Big Data, termo que abrange, entre outros elementos, a capacidade computacional e a coleta, estoque, análise, tratamento e processamento de dados e informações, um algoritmo tem grande chance de apresentar um resultado próximo do real, desde previsão sobre a possível curva de uma pandemia até predizer o comportamento de um indivíduo (MENDES e MATTIUZZO, 2020, p. 430-432).

O professor da UNIJUI Mateus Fornasier, em obra dedicada a discutir questões ético-jurídicas sobre a inteligência artificial, reúne a taxonomia desses vieses a partir de pesquisas da tecnologia da informação. Entre eles, pode-se citar: (i) viés de dados de treinamento (dados de entrada tendenciosos que induz um modelo de aprendizado neutro a se desviar de estatísticas reais); (ii) viés do foco algorítmico (não inserção de certos dados de entrada por razões morais, legais, etc, mesmo que essas variáveis existem e estejam disponíveis); (iii) viés do processamento algorítmico (uso de um estimador estatisticamente tendencioso); (iv) viés do contexto de transferência (aplicação ou extensão injustificada e não declarada de um algoritmo fora do seu propósito específico); (v) viés de interpretação (interpretação incorreta das saídas pelo usuário ou pelo sistema autônomo em um contexto mais amplo no qual o algoritmo funciona); (vi) viés da co-ocorrência (classificação de um

conjunto de dados de forma desproporcional em uma categoria); (vii) viés epistemológico (geração de grau de descrédito na informação); (viii) viés de linguagem (desvio da imparcialidade, reputando opiniões como fatos); (ix) viés de cobertura (conclusão equivocada da representatividade do objeto); (x) viés da especificação (prejulgamento, geralmente do projetista do sistema); (xi) viés de amostragem/seleção (super-representação de observações de um segmento da população); (xii) viés herdado (vieses aprendidos do aprendizado de máquinas anteriores já enviesados) (FORNASIER, 2021, p. 76-84).

Como afirmado, verifica-se que alguns dos vieses algorítmicos acima citados são frutos dos vieses psicológicos de seus programadores e, portanto, reproduzem, em maior escala, suas impressões, preconceitos e subjetividades. Aliás, a cientista de dados americana Cathy O'Neil (2020, p. 10-229), na obra "algoritmos de destruição em massa", chega a dizer que uma grande questão da Inteligência Artificial é saber de fato há eliminação do viés humano ou simplesmente se ele foi camuflado com a tecnologia. Nessa obra, a aludida autora descreve com detalhes como aplicações matemáticas fomentando a economia de dados são baseadas em escolhas feitas por seres humanos falíveis. Mesmo desenvolvidos com as melhores das intenções, os modelos programam preconceitos, equívocos e vieses humanos nos sistemas de softwares para gerenciar a vida das pessoas como deuses, tudo de forma opaca e com mecanismos invisíveis. Segundo a aludida cientista de dados, os modelos punem os pobres e enriquecem mais os ricos, uma vez que seu sucesso é medido em termos de lucros, eficiência e taxa de adimplência. A obra explica como esses modelos agem na saúde (plano de saúde), consumo, educação, emprego (exclusão de candidatos), segurança pública (com mecanismos racistas de predição criminal) e vida cívica, tudo a provocar uma injustiça baseada na ganância e preconceito. A esses modelos ela apelidou de ADM – Armas de Destruição Matemática, em referência ao elevado poder bélico de destruição das conhecidas armas de destruição em massa.

Em suma, tal como as decisões humanas enviesadas, mesmo com boas intenções, os vieses algoritmos acima citados são potencialmente capazes de provocar falhas graves ofensivas aos direitos da personalidade em grande escala. Conhecer a taxonomia desses vieses e saber de forma atuam na formação de uma opinião ou decisão é imprescindível para a consciência correta das ações. No

próximo tópico explorar-se-á mais precisamente um dos resultados ofensivos, a discriminação algorítmica.

4. DISCRIMINAÇÃO ALGORÍTMICA: ESTIGMATIZAÇÃO, *PROFILING* E UM NOVO VIÉS NA REVISÃO HUMANA DA DECISÃO AUTOMATIZADA

Nos anos 2000, o jurista italiano Stefano Rodotà (2008, p. 15, 30, 83, 104 e 105) já alertava que a coleta de dados sensíveis em perfis sociais e individuais poderia levar à discriminação, sobretudo das diferentes minorias. Para ele, a defesa da privacidade no fornecimento desses dados supera o tradicional quadro individualista e dilata-se em uma dimensão coletiva, tendo em vista que não se leva em consideração o interesse do indivíduo enquanto tal, mas como integrante de um determinado grupo social. A categorização dos indivíduos e grupos ameaça anular a capacidade de perceber nuances sutis, o ocasionar a discriminação de pessoas que não correspondem ao modelo geral e a acentuação da estigmatização dos comportamentos desviantes. Identifica-se, ne quadro, um obstáculo ao pleno desenvolvimento da personalidade, cerceada por perfis determinados e comportamentos conformes aos perfis predominantes, a dificultar a criação de novas identidades coletivas, com potenciais riscos para a dinâmica social e organização democrática.

A categorização automática de indivíduos (não intencional) com cunho discriminatório noticiada por Rodotà teve um episódio altamente constrangedor em junho de 2015, quando ganhou as redes a notícia de que o *Google Photos*, famoso serviço de nuvem para armazenamento e organização de fotografias a partir de um algoritmo de reconhecimento de imagem, etiquetou com o termo “Gorilas” um álbum contendo fotos de um usuário negro e sua amiga[10]. Passados 7 (sete) anos desse constrangimento, até a presente data, a única saída encontrada pelo Google para o algoritmo não confundir seres humanos com macacos foi, simplesmente, a exclusão de qualquer resultado de buscas com as palavras “macaco”, “gorila” e “chimpanzé”[11]. Qualquer usuário pode verificar isso facilmente fazendo o upload de

[10]Disponível em https://www.tecmundo.com.br/google-fotos/82458-polemica-sistema-google-fotos-identifica-pessoas-negras-gorilas.htm Acesso em 10/08/2022

[11] Disponível em https://brasil.elpais.com/brasil/2018/01/14/tecnologia/1515955554_803955.html Acesso em 10/08/2022.

uma foto de um macaco e de um cachorro, por exemplo, via site[12] ou aplicativo. Depois basta promover a busca de fotos por palavras. A foto do cachorro será mostrada, mas relativamente ao macaco nada será encontrado, mesmo havendo esse animal no álbum de fotos.

Para entender os motivos que podem ter levado a essa discriminação involuntária pelo algoritmo de reconhecimento de imagem, é necessário compreender alguns aspectos da IA. Segundo explica Fornasier (2021, p. 70-72), 3 (três) são as fontes principais de preconceitos que podem resultar em discriminação por algoritmos: (i) a entrada (inputs) de dados no sistema sem representação ou já enviesados; (ii) o treinamento enviesado de algoritmos na categorização dos dados ou avaliação do resultado desejado; (iii) a programação enviesada do algoritmo, ocorrendo discriminação no designe ou no aprendizado, que se modifica com os contatos sucessivos com usuários humanos. Conquanto algoritmos inteligentes sejam conhecidos por sua precisão, eles possuem os chamados custos interpretativos e, para serem autônomos, os modos pelos quais escolhem, estudam e consideram as variáveis dentro de um conjunto massivo de dados é, por vezes, um mistério, nominado de “caixa preta”, assim definida como a incapacidade humana de entender o processo decisório de uma IA. Ainda, a mineração de dados e as provisões algorítmicas são bastante criticadas por sua opacidade/inexplicabilidade. A opacidade ou a falta de transparência em qualquer uma dessas etapas, em especial quanto às entradas de informações tendenciosas, disfarça resultados discriminatórios ou indesejáveis até os resultados negativos se tornarem visíveis, como no caso relatado.

Laura Mendes e Marcela Mattiuzzo (2020, p. 438-439), por seu turno, pontuam que, por muito tempo, a ciência e a descoberta científica funcionaram por busca de causalidade. O Big Data, contudo, fez a causalidade ceder espaço para a correlação, com forte base na estatística. A discriminação algorítmica, portanto, envolve tanto afirmações estatisticamente inconsistentes, quanto cenários em que as afirmações, conquanto estatisticamente lógicas, classificam as pessoas equivocadamente em certos grupos. Elas concluem que pode ocorrer discriminação na tomada de decisão por um algoritmo por um erro estatístico, por uso de dados sensíveis e por uma generalização injusta (correlação abusiva).

[12] Disponível em https://www.google.com/photos/about/ Acesso em 10/08/2022

Um passo a frente, Bruno Bioni (2021, p. 136) leciona sobre a prática discriminatória conhecida como *profiling,* consistente em se criar um perfil do consumidor e direcionar preços para ele de acordo com sua capacidade econômica (*price-discrimination*). Todas as decisões automatizadas são calibradas nesse estereótipo, desde a própria interação do usuário com outras pessoas em uma rede social até o acesso e a busca por informação na rede, criando-se uma bolha que impossibilita o contato com informações diferentes. As pessoas foram "datificadas", desde as suas relações de consumo até o acesso à informação, e suas informações valem ouro. Danilo Doneda (2021, p. 311) comenta sobre a existência da chamada *commodification* dos dados pessoais, isto é, transformar os dados pessoais em commodity e vende-los no mercado da informação.

De fato, incluir, excluir e classificar são um novo poder. Motores de busca fazem uma incursão profunda na cultura, economia, política, entre outros, para apresentar um resultado desejado. Seu mecanismo, no entanto, como acima citado, pode ser tão complexo que se torna uma "caixa preta" opaca e ininteligível por dois principais motivos: i) complexidade estrutural do algoritmo (redes neurais profundas baseada em modelos matemáticos que, a partir de um determinado ponto, são capazes de desenvolver uma "intuição de máquina" no acerto do resultado); ii) dimensionalidade (decisões baseadas em muitas variáveis de uma só vez e em padrões geométricos entre variáveis que os humanos não podem visualizar (FORNASIER, 2021, p. 43/46).

É certo que o art. 6º, IX da Lei Geral de Proteção de Dados Pessoais (Lei 13709/2018) veda a realização de tratamento dos dados pessoais para fins discriminatórios ilícitos, bem como o artigo 20 do mesmo diploma assegura que o titular dos dados tem direito a 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. Tal dispositivo, para boa parte da doutrina, assegura o chamado "direito à explicação", que merece um estudo à parte, sendo certa a posição majoritária de que a revisão da decisão deve ser humana. Para os fins do presente artigo, sem embargo, como contribuição para a solução da discriminação algorítmica, 2 (dois) pontos devem ser observados.

Primeiro, há algoritmos de aprendizagem tão profunda que não permitem uma compreensão por engenharia reversa ao ponto de ser tecnicamente complexo explicar como a IA chegou ao resultado, que informações foram determinantes e quais as variáveis processadas na ordem de classificação. No entanto, Mateus Fornasier (2021, p. 46/48) relata avanços significativos nas técnicas de interpretação algorítmica para tornar modelos de "caixa preta" em modelos interpretáveis, chamadas de IA explicável (XAI, no acrônimo em inglês). Nessa toada, Cathy O'Neil (2020, p. 232-236) sugere que é necessário mergulhar no código de programação do software, conduzindo uma auditoria no algoritmo, tal como uma iniciativa feita em Princeton, em Nova Jérsei/EUA, em que os pesquisadores lançaram o Web Transparency and Accountability Project. Por esse projeto, eles criaram softwares robôs que se mascaram *online* como pessoas de todos os tipos, ricos, pobres, homens, mulheres, etc. "Ao se estudar o tratamento que esses robôs recebem, os pesquisadores são capazes de detectar vieses ou parcialidades em sistemas automatizados, de mecanismos de busca a sites de colocação de empregos", de forma a aprimorar a transparência dos sistemas.

Assim, é impositiva a edição de políticas fiscais e econômicas que incentivem empresas a desenvolverem modelos explicáveis de IA e a auditarem sistemas já em curso, tudo para romper a lógica econômica que conduz o paradigma atual de aferição do sucesso da IA, medida hoje praticamente apenas em termos de taxa de adimplência e lucro.

Em segundo lugar, consoante exclama Ana Frazão (2021), a grande questão é saber em que medida a supervisão humana é realmente adequada e eficiente para resolver problemas de decisões algorítmicas enviesadas, discriminatórias ou disfuncionais, afinal, para que haja algum resultado prático da revisão humana, deve haver o efetivo controle e possibilidade de divergência diante dos resultados algorítmicos. Para tanto, a aludida autora cita uma pesquisa realizada pelo matemático Bem Green e pela cientista da computação Yiling Chen (2019) sobre sistemas algorítmicos utilizados para auxiliar juízes na dosimetria de penas por meio de cálculos sobre o potencial de reincidência dos réus sob julgamento. A pesquisa sugere que as ferramentas de automação influenciam psicologicamente a revisão humana pelo chamado viés da automação (*automation bias*), à medida que as pessoas não reconhecem quando um sistema automatizado erra, não conseguem

distinguir entre predições confiáveis e predições não confiáveis, bem como são incapazes de avaliar sua própria performance e a performance do algoritmo. Aliás, mesmo diante de sistemas mais acurados, as pessoas nem sempre incorporam suas decisões e preferem confiar no seu próprio julgamento, ou, pior, tendem a seguir um sistema automatizado sem considerar informações contraditórias. Para Ana Frazão (2019), é urgente se entenda toda a interação entre homem e algoritmo, uma vez que a forte confiança em sistemas automatizados pode ensejar a perda do senso de responsabilidade e da *accountability*, fazendo com o que a revisão humana prevista na lei não possa ser suficiente para resolver o problema da discriminação.

À vista de todos os vieses humanos e da IA citados neste artigo capazes de provocar discriminação, sugere-se estabelecer um parâmetro normativo no sentido de que a revisão humana da decisão automatizada, sempre que possível, seja realizada por pessoa que não teve contato com o resultado fornecido pelo sistema informatizado ou, ao menos, que se estabeleça protocolos cientificando o decisor humano que sua eventual intenção de concordar com a decisão automatizada pode estar sendo altamente influenciada pelo viés da automação.

Em arremate, é certo que o sistema jurídico brasileiro é bem robusto contra as mais diferentes formas de discriminação. A Constituição Federal prescreve expressamente, como uma dos objetivos fundamentais da República, promover o bem de todos, sem preconceitos de origem, raça, sexo, cor, idade e quaisquer outras formas de discriminação (art. 3º, IV); o repúdio, a inafiançabilidade e a imprescritibilidade do racismo (art. 4º, VIII e art. 5º, XLII); o dever de a lei punir qualquer discriminação atentatória dos direitos e liberdades fundamentais (art. 5º, XLI); a proibição de qualquer preconceito no tocante à salário e critérios de admissão, além da eliminação de todas as formas de discriminação contra pessoa com deficiência (art. 5º, XXXI e art. 227, §1º, II); a salvaguarda de crianças e adolescentes de todas as formas de discriminação (art. 227); e a proibição de discriminação quanto a filhos havidos dentro e fora do casamento (art. 227, §6º). Para além, em fevereiro de 2022, na condição de poder constituinte de reforma, o Congresso Nacional aprovou a Emenda Constitucional nº 115/2022, inserindo na Constituição Federal o direito fundamental à proteção dos dados pessoais, inclusive nos meios digitais (art. 5º, inciso LXXIX).

Lidos todos os dispositivos constitucionais conjunto, é possível concluir que a proibição da discriminação algorítmica, hoje já textualizado na legislação

infraconstitucional pelo artigo 20 da LGPD, possui, na verdade, status de matriz constitucional, como uma das vertentes do direito fundamental à proteção de dados. Por essa razão, ele goza de amplo espaço na organização política e espaço constitucional, maior juridicidade e força normativa vinculante aos 3 (três) poderes, que devem promover não somente medidas legislativas, mas políticas econômicas e sociais concretas para o seu implemento (FACHIN, 2022, p. 299-313), e atuar de forma vigilante contra a discriminação por meio da proteção de dados pessoais.

5. CONSIDERAÇÕES FINAIS

Para além das relações de consumo, a Inteligência Artificial tem avançado fortemente em aspectos centrais da vida humana, a proporcionar de forma mais célere soluções para os problemas econômicos e sociais, como a escolha de brasileiros elegíveis para programas públicos de transferência de renda, cotas em universidade e contrato securitário de saúde, mas, por outro lado, também tem potencialidade de provocar graves ofensas aos direitos da personalidade. Umas dessas ofensas é a discriminação algorítmica.

Tanto as decisões humanas quanto das máquinas possuem vieses que fitam a formação de generalizações, lugares comuns, estereótipos pensamentos tendenciosos, intuições, pré-compreensão e preconceitos. Conhecer a taxonomia desses vieses e saber a forma atuam na formação de uma opinião ou decisão é imprescindível para a consciência correta das ações.

Conquanto algoritmos inteligentes sejam conhecidos por sua precisão, os modos pelos quais escolhem, estudam e consideram as variáveis dentro de um conjunto massivo de dados é, por vezes, um mistério, nominado de “caixa preta”. A opacidade ou a falta de transparência na IA, em especial quanto às entradas de informações tendenciosas, disfarça resultados discriminatórios ou indesejáveis até os resultados negativos se tornarem visíveis.

É impositiva a edição de políticas fiscais e econômicas que incentivem empresas a desenvolverem modelos explicáveis de IA e a auditarem sistemas já em curso, tudo para romper a lógica econômica que conduz o paradigma atual de aferição do sucesso da IA, medida hoje praticamente apenas em termos de taxa de adimplência e lucro.

Além disso, à vista de todos os vieses humanos e da IA capazes de provocarem discriminação, deve-se estabelecer um parâmetro normativo no sentido de que a revisão humana da decisão automatizada, sempre que possível, seja realizada por pessoa que não teve contato com o resultado fornecido pelo sistema informatizado ou, ao menos, que se estabeleça protocolos cientificando o decisor humano que sua eventual intenção de concordar com a decisão automatizada pode estar sendo altamente influenciada pelo viés da automação.

Essas medidas asseguram, de forma mais concreta, o direito fundamental à proteção de dados, na vertente da proteção contra a discriminação algorítmica.

REFERÊNCIAS

BIONI, Bruno R. **Proteção de dados pessoais:** a função e os limites do consentimento. 3. ed. Rio de Janeiro: Forense, 2021. E-book (499 p.). ISBN 978-85-309-9409-9.

CONTE, Francesco. **A gênese ilógica da sentença civil:** intuição, sentimento e emoção na hora de julgar. Belo Horizonte: Fórum, 2020.

DONEDA, Danilo. **Da privacidade à proteção de dados pessoais**. 3. ed. São Paulo: Revista dos Tribunais, 2021.

FACHIN, Zulmar A. O direito fundamental à proteção de dados pessoais: análise da decisão paradigmática do STF na ADI 6.387-DF. **Revista Videre**, Dourados/MS, v. 14, n. 19, p. 298–313, Jul. 2022. Disponível em: https://ojs.ufgd.edu.br/index.php/videre/article/view/15629. Acesso em: 29 Jul. 2022.

FORNASIER, Mateus O. **Cinco questões ético-jurídicas fundamentais sobre a inteligência artificial**. Rio de Janeiro: Lumen Juris, 2021.

GREEN, Ben; CHEN, Yiling. Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments. **FAT* ’19: Conference on Fairness, Accountability, and Transparency (FAT* ’19)**, USA. ACM, New York, NY, USA, 29-31 Janeiro 2019. https://doi.org/10. 1145/3287560.3287563.

HOFFMANN-RIEM, Wolfgang. **Teoria Geral do Direito Digital:** transformação digital desafios para o Direito. Tradução de Ítalo Fuhrmann. 2. ed. Rio de Janeiro: Forense, 2022.

KAHNEMAN, Daniel. **Rápido e devagar:** duas formas de pensar. Tradução de Cássio de Arantes Leite. Rio de Janeiro: Objetiva, 2012.

LUIS, Fernando P. Algoritmos e decisões automatizadas: buscando a conformidade com a LGPD. In: ISZLAJU, Barbara, *et al.* **Estudos sobre privacidade e proteção**

de dados. São Paulo: Thompson Reuters Brasil, 2021. p. RB-5.1-RB-5.7. E-book. ISBN 978-65-5991-834-8.

MENDES, Laura S.; MATTIUZZO, Marcela. Discriminação algorítmica à luz geral de proteção de dados. In: MENDES, Laura, *et al.* **Tratado de proteção de dados pessoais**. Rio de Janeiro: Forense, 2020. Cap. 21, p. 429-454. E-book. ISBN 978-85-309-9219-4.

O'NEIL, Cathy. **Algorítimos de destruição em massa:** como o big data aumenta a desigualdade e ameaça a democracia. Tradução de Rafael Abraham. Santo André: Rua do Sabão, 2020. E-book (438 p.). ISBN 978-65-86460-02-5.

RODOTÀ, Stefano. **A vida na sociedade da vigilância:** a privacidade hoje. Rio de Janeiro: Renovar, 2008.

THALER, Richard H.; SUNSTEIN, Cass R. **Nudge**. Tradução de Ângelo Lessa. Rio de Janeiro: Objetiva, 2019. E-book (438 p.) ISBN 978-85-5451-342-9.

VERONESE, Alexandre. Os direitos de explicação e de oposição frente às decisões totalmente automatizadas: comparando o RGPD da União Europeia com a LGPD brasileira. In: TEPEDINO, Gustavo; FRAZÃO, Ana; OLIVA, Mlena **Lei geral de proteção de dados pessoais e suas repercussões no direito brasileiro**. São Paulo: Thomson Reuters Brasil, 2019. p. RB-14.1-RB-14.9. E-book. ISBN 978-65-5614-007-0.

.