UNIVERSIDADE PRESBITERIANA MACKENZIE

GABRIELLA LERIAM

DISCRIMINAÇÃO ALGORÍTMICA: PERSPECTIVAS ÉTICAS E GOVERNANÇA

São Paulo
2020

Classificação: Interna

GABRIELLA LERIAM

Trabalho de Graduação Interdisciplinar apresentado como requisito para obtenção do título de Bacharel no Curso de Direito da Universidade Presbiteriana Mackenzie.

ORIENTADOR: EDUARDO ARIENTE

Classificação: Interna

São Paulo
2020

GABRIELLA LERIAM

DISCRIMINAÇÃO ALGORÍTMICA: PERSPECTIVAS ÉTICAS E GOVERNANÇA

Trabalho de Graduação Interdisciplinar apresentado como requisito para obtenção do título de Bacharel no Curso de Direito da Universidade Presbiteriana Mackenzie.

Aprovada em:

BANCA EXAMINADORA

_____

Examinador(a):

_____

Examinador(a):

_____

Examinador(a):

Classificação: Interna

Discriminação algorítmica: perspectivas éticas e governança

Resumo: O artigo tem como objetivo analisar as diferentes formas de discriminação algorítmica, esclarecendo o potencial discriminatório das decisões automatizadas. Primeiro, discutiremos os conceitos de algoritmo, vieses e discriminação algorítmica, apontando as razões pelas quais esses conceitos são tão importantes na economia atual que é movida por dados. Em seguida, abordaremos como os algoritmos, em conjunto com o *Big Data*, alteram os processos de decisões e podem gerar desafios a respeito do risco de condutas discriminatórias. O último capítulo expõe possíveis soluções para o tema da discriminação algorítmica, apresentando diversos princípios debatidos entre os especialistas nas áreas, com enfoque na transparência como solução efetiva para resolver as questões colocadas durante o desenvolvimento do artigo.

Abstract: The aim of this article is to analyze the theories about algorithmic bias, explaining the discriminatory potential of automatic decisions. First, we will discuss the concepts of algorithm, biases and algorithmic bias, pointing out the reasons why these concepts are so important for today's data-driven economy. Then we will consider how algorithm, together with Big Data, change the decision-making processes and can create challenges regarding the risk of discriminatory conducts. The last chapter presents possible solutions for the algorithmic bias issue, showing various principles discussed by experts in this area, with a focus on transparency as an effective solution to the questions raised during the article development process.

Palavras-chave: Inteligência artificial; Vieses; Discriminação algorítmica; Governança algorítmica

Keywords: Artificial intelligence; Biases; Algorithmic bias; Control algorithm

Sumário: Introdução; 1 Algoritmo e discriminação: o problema da atribuição de função decisória às máquinas nos processos de decisão automatizadas; 1.1 Vieses adotados na automação dos processos decisórios; 1.2 Discriminação algorítmica; 2 Discriminação algorítmica e governança: uma discussão sobre as soluções expostas na literatura; Considerações finais; Bibliografia.

Introdução

Classificação: Interna

O uso da inteligência artificial vem crescendo nos mais diversos ramos, tendo em vista a eficácia, a simplificação e a precisão por ele proporcionado. Além disso, uma das mais importantes contribuições da inteligência artificial e do processamento de dados é o embasamento para tomada de decisões (DOSHI-VELEZ, 2017, p. 7).

Como busca explicar Doshi-Velez ao longo do seu artigo *Accountability of AI Under the Law: The Role of Explanation*, a inteligência artificial funciona por meio de sistemas de dados programados para fornecer respostas conforme as informações disponíveis. Esses sistemas são denominados de algoritmos que incorporam uma projeção do comportamento futuro de um indivíduo através de um procedimento automatizado.

Um algoritmo é comumente descrito como uma fórmula usada para alcançar um objetivo determinado através de uma ordem específica (ZIVIANI, NIVIO, 2011, p. 11). Por isso, qualquer pessoa pode criar um algoritmo para auxiliá-la a realizar diferentes tarefas, como escolher o melhor caminho para chegar a um destino, para fazer uma receita culinária, entre outras atividades. No entanto, para o objetivo deste artigo essa definição não nos fornece informações suficientes.

Thomas Cormen diferencia os algoritmos comuns dos usados em computadores. Computadores não processam palavras que impliquem em uma avaliação subjetiva, como "quase", "suficiente", "bom". Um computador é capaz de interpretar porcentagem, mas não de determinar o que o termo "quase acabando", por exemplo, significa. Segundo Cormen:

> Você pode ser capaz de tolerar quando um algoritmo é descrito de maneira imprecisa, mas um computador, não. [...] Assim, um algoritmo computacional consiste em uma série de etapas para completar uma tarefa que é descrita de maneira precisa o bastante para que um computador possa realizá-la. (CORMEN, 2013, p. 10)

Importante destacar que o objetivo dos algoritmos, como normalmente são utilizados hoje, é auxiliar a solucionar problemas e a tomar decisões. Aqui, encontramos uma outra característica dos algoritmos: o programa será mais útil e viável quando a informação obtida for precisa (BOYD, CRAWFORD, 2011, p. 15).

Assim, o crescente desenvolvimento no campo da análise de dados tem fornecido maiores incentivos para o uso dessas projeções nos setores público e privado, permitindo maior processamento de informações e novas conexões entre dados e comportamentos futuros que possam surgir.

Ainda, ao passo que os algoritmos se baseiam nas características de um conjunto de pessoas e na probabilidade delas agirem de uma forma determinada, é preciso entender se os

Classificação: Interna

critérios e processos utilizados para classificar indivíduos são igualitários, transparentes e corretos.

Inicialmente, a automação das tomadas de decisão visava superar as tendências de discriminação e de vieses. No entanto, com o passar do tempo, notou-se que na construção de programas de Inteligência Artificial fazem escolhas que são reflexos das opiniões dos programadores, as quais influenciam nas decisões dos sistemas (O'NEIL, 2016, p. 30).

Contudo, percebeu-se que os métodos utilizados para criar estatísticas e "outputs" objetivos são fatores que a experiência empresarial considerou como relevante para avaliação do risco de retorno do crédito concedido, como informações acerca do adimplemento das obrigações, idade, sexo, estado civil, renda, endereço, entre outros, e que poderia reproduzir vieses já existentes, uma vez que as correlações são definidas, na maioria das vezes, pelos programadores que acabam por transmitir as suas opiniões aos algoritmos. Corroborando tal fato, um relatório apresentado pela Federal Reserve ao Congresso dos Estados Unidos, apontou que negros e hispânicos têm um *credit score* significativamente inferior ao de brancos e asiáticos (Federal Reserve, 2020, p. 1).

Deste modo, ao se apresentarem como uma característica intrínseca do agir humano, pode-se concluir que um algoritmo criado por homens enviesados provavelmente também será enviesado, não de forma proposital, mas em decorrência dos dados fornecidos ao programa (BAROCAS, SELBST, 2016, p. 674). Assim, surgem os chamados vieses algorítmicos, que ocorrem quando as máquinas se comportam refletindo os valores humanos implícitos no processo da programação (O'NEIL, 2016, p. 35).

No mais, na medida que os programas de Inteligência Artificial se baseiam, na maioria das vezes, em discriminação estatística, ou seja, na diferenciação de indivíduos baseada na base de dados, nas opiniões dos programadores e na probabilidade de tal grupo agir da forma como eles pensam ser característica deste, torna-se indispensável entender em que medida o uso da Inteligência Artificial pode impactar nas desigualdades sociais. Para minimizar tal cenário é indispensável um controle interno das empresas, grandes usuárias dos programas e "outputs" extraídos desses processos.

Assim, as empresas devem buscar reduzir o impacto dos vieses dos programadores seja através de ações afirmativas, políticas internas mais claras e rigorosas quanto aos critérios utilizados como "inputs" e, até mesmo, através de uma governança racial de seus funcionários, buscando a diversidade de pensamentos e de contextos sociais para garantir que as classificações utilizadas sejam corretas, transparentes e, por fim, justas (PASQUALE, CIRON, 2014, p. 9).

Classificação: Interna

No entanto, o que se percebe é que a falta de legislação e fiscalização do uso de algoritmos nas tomadas de decisão automatizadas amplia as desigualdades já existentes e tão marcantes em nossa sociedade.

É preciso lembrar que as pessoas são afetadas pelo processo decisório e que, por isso, não deve haver isenção da responsabilidade civil pelo uso da Inteligência Artificial, sendo necessário oferecer alternativas de reparação de danos, tanto individuais quanto coletivos. Nesse sentido, por exemplo, a Comunidade Europeia criou um relatório de recomendações para responsabilidade civil por danos causados pela Inteligência Artificial (European Parlament, 2020, p. 1).

Este artigo, portanto, tem como objetivo analisar como a governança algorítmica pode ser aprimorada para que os vieses sejam minimizados nos inputs e nos resultados das decisões automatizadas, visando responder de que forma o direito e a centralidade humana podem auxiliar no combate à discriminação gerada através dessas decisões, a fim de confirmar a hipótese que de a transparência e o controle humano através de comitês, por exemplo, são elementos essenciais para o combate à discriminação algorítmica.

Para isso, no capítulo 1 iremos definir brevemente o conceito de vieses e de discriminação algorítmica, bem como explicar como tais conceitos impactam na geração de dados e na automação da tomada de decisões nos processos. O capítulo 2 apresenta propostas de soluções para lidar com vieses e a discriminação, examinando a bibliografia sobre o assunto abordado em questões de regulação das atividades dos desenvolvedores e programadores, além de uma governança algorítmica para revisão humana das decisões automatizadas.

1. Algoritmo e discriminação: o problema da atribuição de função decisória às máquinas nos processos de decisão automatizadas

1.1. Vieses adotados na automação dos processos

Inicialmente, devemos lembrar que um dos objetivos principais dos algoritmos é fazer previsões utilizando porcentagens. Os algoritmos não fornecem respostas precisas a todas as questões, mas eles podem analisar os dados fornecidos e oferecer suposições coerentes. Por isso, a quantidade e qualidade das informações disponibilizadas ao algoritmo devem ser grandes para aumentar a chance da sugestão ser próxima a realidade (MAYER-SCHONBERGER; CUKIER, 2014, p. 47),

Assim, quanto mais intenso for o uso de processamento de dados através dos algoritmos como base para tomadas de decisão, e quanto mais disponíveis e acessíveis as tecnologias para

Classificação: Interna

tornar esse cenário possível, mais urgente se torna a discussão sobre as consequências dos procedimentos para o indivíduo em si e os riscos a ele associados.

Os algoritmos necessitam de informações básicas para oferecer respostas relevantes. Essas informações são dadas através de dados. Por isso, a quantidade crescente de dados levou ao crescimento de sua utilização e de seu impacto no dia a dia. O termo "Big Data" foi criado para definir esse fenômeno. Como sugerem Mayer-Schönberger e Cukier, Big Data não é somente sobre tamanho, mas especialmente sobre "a habilidade de transformar em dados muitos aspectos do mundo que nunca foram quantificados antes" (MAYER-SCHÖNBERGER E CUKIER, 2014, p. 53).

O Big Data pode ser usado para elaborar prognósticos dos diversos setores da sociedade, como a política, a economia e, também, sobre o comportamento individual, sendo este de grande interesse, já que oferece base para tomada de decisões ao informar e conhecer sobre o comportamento de uma pessoa a partir de seus dados pessoais. Uma análise de Big Data pode, assim, afetar um indivíduo e produzir resultados discriminatórios que impactem sua vida (O'NEIL, 2016, p. 40).

Muitos desses dados são formados por representações abstratas de determinado processo, simplificando o mundo real. Ao criar um programa, os programadores devem selecionar os dados fornecidos que serão utilizados para prever sugestões e decisões. Desse modo, essa seleção faz com que haja pontos cegos nos algoritmos, refletindo os objetivos e crenças do criador, permeando os programas pela subjetividade dos desenvolvedores (O'NEIL, 2016, p. 47).

Além disso, como já mencionado anteriormente, a qualidade dos dados fornecidos aos programas impacta os resultados, pois os dados são coletados da sociedade que é imersa em discriminações e desigualdades. Conforme estudo realizado por pesquisadores da Universidade de Oxford:

> [...] o aprendizado de máquina pode confirmar padrões discriminatórios se eles forem encontrados no banco de dados, então, por conseguinte, um sistema de classificação exato irá reproduzi-lo. Desde modo, decisões enviesadas são apresentadas como resultado de um "algoritmo objetivo". (GOODMAN, B.; FLAXMAN, S. R, 2017, p. 7).

Nota-se, assim, que na construção de programas de Inteligência Artificial fazem escolhas que são reflexos das opiniões dos programadores, as quais influenciam nas decisões dos sistemas. Portanto, não há isenção completa nos programas tidos como universais e sem vieses, já que o ponto de partida deles é sempre uma atividade humana de escolha de dados,

Classificação: Interna

refletindo, também, o contexto social de seus criadores, ligando tal questão ao tema dos vieses cognitivos (NUNES, LUD, PEDRON, 2018, p. 11).

Os temas tratados neste trabalho adquirem, ainda, maior complexidade quando são considerados os novos avanços no campo da Inteligência Artificial. Estes voltam-se ao desenvolvimento de máquinas inteligentes e dentro dela há uma área, hoje de especial interesse para o ramo da ciência da computação, chamada *machine learning.* Pedro Domingos esclarece que:

> Todo algoritmo possui um input e um output: o dado ingressa no computador, o algoritmo faz o que seu código determina com esse dado, e, então, sai o resultado. O Machine Learning muda essa lógica: adentram na máquina tanto o dado como o resultado desejado, e o produto é algoritmo capaz de tornar a relação entre dado e resultado verdadeira. Algoritmos inteligentes – também conhecidos como learners – são algoritmos que criam outros algoritmos. Com machine learning, computadores escrevem seus próprios programas, para que nós não tenhamos que fazê-lo. (DOMINGOS, 2015, p. 6).

Ou seja, após a elaboração do modelo, são fornecidos dados para o sistema, possibilitando o *machine learning*. Através dele a máquina analisará as informações fornecidas, seguindo as instruções do algoritmo para encontrar padrões e conseguir prever resultados.

Com o aumento do uso de *machine learning* há outra questão decisiva com relação aos algoritmos que é a obscuridade destes em seus processos decisórios. O tema ganha maior relevância pois soluções algorítmicas vêm sendo amplamente adotadas tanto pelo setor privado quanto pelo setor público.

No caso analisado no REsp 1419697/RS, a parte autora teve a sua análise do risco de crédito. Nesta oportunidade esclareceu-se que o *SCPC Score Crédito* agrupa os consumidores em faixas de risco. A pontuação varia de 0 a 1000 e indica menor risco para a concessão de crédito à medida que se aproxima do 1000. Assim, a consumidora obteve a pontuação de 553, prestando-se, ainda, informações de que não constava nenhum protesto, cheque ou ação civil para o documento avaliado. A polêmica central do processo foi, exatamente, a verificação da licitude desse método de avaliação do risco de crédito (BRASIL. SUPERIOR TRIBUNAL DE JUSTIÇA. RECURSO ESPECIAL nº 1419697 – RS. Recorrente: Boa Vista Serviços S/A. Recorrido: Anderson Guilherme Prado Soares. Relator: Ministro Paulo de Tarso Sanseverino, 17/11/2014).

Outro exemplo é o chamado *CrimeRadar*, uma ferramenta brasileira que possui o objetivo de dar previsão de taxas e padrões de criminalidade na Cidade do Rio de Janeiro, como o chamado *PredPol*, que estão em uso pela polícia de diversos estados norte-americanos, mudando o modo de operar desse departamento. Ainda, não há informações sobre o uso do

Classificação: Interna

CrimeRadar pelas autoridades públicas no combate ao crime, mas, certamente, é uma tendência, conforme citado na reportagem feita pela *The Guardian* (MILHORANCE, 2018, p. 1).

O uso de ferramentas semelhantes no setor privado também é numeroso. Os algoritmos são usados no processo de recrutamento de muitas empresas, por exemplo. Como exposto pelo Wall Street Journal (GEE, 2017, p. 1), a Unilever está acabando com os processos seletivos tradicionais e confiando em *softwares* para selecionar seus candidatos, potencializando a obtenção de resultados consistentes e trazendo economia para a empresa, uma vez que os processos de contratação consomem muito tempo e dinheiro.

O objetivo, expressamente declarado pela Unilever, é diversificar o grupo de candidatos, aumentando a diversidade étnica e social dos seus colaboradores, minimizando, assim, as discriminações[1]. No entanto, quanto mais se recorre a essas ferramentas, mais difícil torna-se para os candidatos "fora dos padrões" a entrada no mercado de trabalho, dificultando o entendimento do porquê uma candidatura é menos atrativa que outras. Nesse sentido, Cathy O'Neil menciona em sua obra que pessoas "comuns" são selecionadas por máquinas e pessoas privilegiadas, por outras pessoas (O'NEIL, 2016, p. 96).

Como veremos no próximo capítulo, a falta de transparência é uma séria preocupação no que se refere às consequências legais da discriminação algorítmica. Isso porque, primeiramente, os algoritmos são obscuros, sendo difícil afirmar que tipo de discriminação ocorreu; em segundo lugar, é difícil prevenir que discriminações ocorram e, em terceiro, porque os algoritmos, se usados sem regulamentação, podem acabar reforçando resultados discriminatórios (O'NEIL, 2016, p. 54).

Sobre a questão da falta de transparência dos algoritmos, a Cathy O'Neil, matemática e autora do livro *Weapons of math destruction*, afirma que a ausência de transparência dos modelos algorítmicos torna seu funcionamento invisível para todos, salvo matemáticos e cientistas computacionais. Desse modo, mesmo quando equivocados, o veredito dos algoritmos se torna imune as discordâncias e reclamações, perpetuando por vezes as desigualdades e contribuindo, inclusive, para o seu crescimento, por meio do *feedback loop*[2]. Por isso, é essencial ter elevada transparência algorítmica, a fim de possibilitar que os afetados pelo modelo saibam o que determina o resultado alcançado pela Inteligência Artificial.

---

[1] Para mencionar somente um exemplo, um estudo da Northwestern University, conduzido por Lincoln Quillian, demonstrou que, entre 1990 e 2015, candidatos brancos receberam 36% mais telefonemas de retorno do que candidatos negros e 24% mais retorno do que latinos.

[2] O *feedback loop* perfila uma pessoa por suas circunstâncias pessoais e ajudam a criar o ambiente que justifica suas suposições. Este ciclo destrutivo gira e gira e o modelo se torna cada vez mais injusto.

Classificação: Interna

O Parlamento Europeu, em Resolução de 16 de fevereiro de 2017, atentou-se para essa questão e postulou um princípio ético para orientar a regulação da robótica, realçando o princípio da transparência.

Na Resolução, o Parlamento afirma que deve ser sempre possível fundamentar qualquer decisão com base na Inteligência Artificial, quando houver um impacto substancial sobre a vida de uma ou mais pessoas. Assim, toda a lógica que conduziu à formulação das decisões através de sistemas de IA deveria ser compreensível para os seres humanos, apresentando todos os dados sobre as operações realizadas pela máquina[3]. Nesse sentido, dispõe o art. 20 da Lei Geral de Proteção de Dados:

> 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 (BRASIL, 2018).

No mesmo sentido, o deputado Eduardo Bismarck em seu projeto de lei estabelece princípios, direito e deveres para o uso da Inteligência Artificial no Brasil, afirmando que o uso desta deve ter como fundamento o respeito aos direitos humanos e aos valores democráticos, visando à igualdade, à não discriminação, à pluralidade, entre outros, e ser norteada pelos princípios da transparência e explicabilidade:

> Garantia de transparência sobre o uso e funcionamento dos sistemas de inteligência artificial e de divulgação responsável do conhecimento de inteligência, observados os segredos comercial e industrial, e de conscientização das partes interessadas sobre suas interações com os sistemas, inclusive no local de trabalho. (BISMARCK, 2020, p. 1).

O deputado afirma, ainda, que os agentes de Inteligência Artificial devem fornecer informações claras e adequadas sobre os critérios e procedimentos utilizados pelos sistemas, observando sempre a Lei 13709 de 2018 (Lei Geral de Proteção de Dados).

Percebe-se, assim, que tornar os algoritmos acessíveis e transparentes para a população em geral, e não apenas para especialistas, é de extrema importância para o controle e fiscalização deles, tendo em vista a obscuridade e subjetividade que está presente no processo de tomada de decisão automatizadas e sua capacidade de afetar a igualdade entre os indivíduos.

1.2. Discriminação algorítmica

[3] Resolução do Parlamento Europeu, de 16 de fevereiro de 2017, que contém recomendações à Comissão sobre disposições de Direito Civil sobre Robótica (2015/2103(INL).

Classificação: Interna

Igualdade e justiça são noções que possuem relação íntima. Aristóteles, por exemplo, na obra Ética a nicômacos associou justiça e igualdade e sugeriu que os iguais devem ter tratamento igual e os diferentes, devem ser tratados de modo desigual.

No mais, a igualdade passou a ser valor central para o direito constitucional contemporâneo. De acordo com José Joaquim Gomes Canotilho e Vital Moreira, "o princípio da igualdade é um dos princípios estruturantes do sistema constitucional global, conjugando dialecticamente as dimensões liberais, democráticas e sociais inerentes ao conceito de Estado de direito democrático e social" (CANOTILHO, 2007, p. 336-7).

Já desde o início do constitucionalismo moderno, a igualdade perante a lei e a ideia de que "em princípio, direitos e vantagens devem beneficiar a todos; e os deveres e encargos devem impender sobre todos" (CANOTILHO, 2007, p. 338) passaram a estar presente nos textos constitucionais, presença que ganhou máxima expansão com a inserção do princípio da igualdade e dos direitos da igualdade no sistema internacional de proteção dos direitos humanos, a começar pela Declaração da ONU, de 1948, quando, no seu art. I afirmou que "todas as pessoas nascem livres e iguais em direitos", para, no art. VII, declarar, o ideal de igualdade material, expresso em "todos são iguais perante a lei e têm direito, sem qualquer distinção, a igual proteção da lei. Todos têm direito a igual proteção contra qualquer discriminação que viole a presente Declaração e contra qualquer incitamento a tal discriminação".

Ainda no plano internacional, verificou-se um amplo reconhecimento de direitos de igualdade mediante sua incorporação em vários tratados e convenções (Pacto Internacional de Direitos Civis e Políticos e a Convenção Americana de Direitos Humanos, por exemplo). Tais documentos, quando ratificados pelo Estado, assumem a condição de normas vinculantes que além de preverem cláusulas gerais, também preveem cláusulas especiais de igualdade a fim de combater as mais diversas modalidades de discriminação.

Entre elas está a discriminação algorítimica e para definí-la serão analisadas três principais fontes: o livro *Profiles, Probabilities and Stereotypes* de Frederick Schauer (2006); a teoria econômica de discriminação estatística; e, por fim, a obra de Gabriele Britz (2008). Após essa análise, quatro formas de discriminação algorítmica serão discutidas.

Ao refletir sobre o conceito de discriminação, normalmente imaginamos um cenário no qual a pessoa é excluída da sociedade por não possuir certas características. Alguém possui um *credit score* menor por causa de suas dívidas passadas. Outro, é considerado menos merecedor de um emprego por não ter em seu currículo uma universidade de primeira linha. No entanto, neste trabalho, o foco é que os resultados discriminatórios decorrem do fato de alguém pertencer

Classificação: Interna

a determinado grupo e ser julgado por isso, desconsiderando as características individuais da pessoa.

Schauer explica em sua obra que um dos problemas do conceito de preconceito - tomando o uso dessa palavra como sinônimo de discriminação neste trabalho - é a utilização generalista que fazemos dele. Ou seja, utilizamos o mesmo termo para descrever circunstâncias diferentes.

O autor afirma que há dois tipos de generalização: a consistente e a inconsistente. As generalizações consistentes podem ser universais ou não universais. O exemplo mais famoso de generalizações universais é a expressão "Todos os homens são mortais". Ou seja, a raça humana, com certeza. irá morrer um dia, tornando essa generalização verdadeira em 100% dos casos. Já as generalizações não universais descrevem uma característica compartilhada pela maioria dos indivíduos de um determinado grupo.

No entanto, em razão do modo como nos expressamos, Schauer define um terceiro tipo de generalização consistente: aquela que não é universal e nem descreve uma característica compartilhada pela maior parte dos membros de um grupo, mas que "retrata com precisão os membros de uma classe como possuindo uma maior prevalência de um traço do que a grande classe da qual o grupo é normalmente considerado parte, ainda que tal atributo não apareça na maioria dos membros de ambos os grupos" (SCHAUER, 2006, p. 11). Essa terceira categoria depende, portanto, de uma dimensão comparativa.

Já as generalizações inconsistentes falham em preencher os parâmetros definidos anteriormente. Uma conhecida generalização desse tipo no campo jurídico são os estudos de Cesare Lombroso acerca do "homem delinquente". Lombroso foi fundador da escola de Criminologia Positiva, cujo estudos foram dedicados em analisar como criminosos com certas características físicas que os identificariam como tanto. A sua pesquisa concluiu que o homem criminoso reuniria determinados traços, como braços excessivamente longos, crânio e rosto assimétricos etc. No entanto, até hoje, não existem evidências da veracidade de seus estudos.

Schauer ao abordar a questão do preconceito, registra que a generalização é usada em dois cenários diferentes: dizemos que algo é preconceituoso quando uma afirmação é baseada em generalizações estatísticas inconsistentes e, também, quando utilizamos generalizações consistentes, estatisticamente, mas não universais. Por exemplo: dizer que uma característica da pessoa é devido ao seu signo dos zodíacos é tão preconceituoso quanto afirmar que homossexuais possuem HIV. Por isso, Schauer afirma que "todos os seres humanos (...) merecem ser tratados como individuais, e não simplesmente como membros de um grupo, de

Classificação: Interna

modo que decisões atuariais sobre seres humanos são, na maioria das vezes, moralmente erradas" (SCHAUER, 2006, p. 19).

O grande problema é que decisões atuariais (baseadas em estatísticas) são extremamente comuns em qualquer sistema jurídico e, em grande parte, indispensáveis. A lei, por exemplo, define que pessoas acima de 18 anos podem votar ou limita a velocidade em rodovias. No entanto, certamente alguns indivíduos menores de 18 anos têm maturidade para votar e pessoas diferentes são capazes de dirigir de maneira segura em velocidades variadas.

Vale ressaltar que decisões atuariais não estão limitadas ao mundo jurídico. No dia a dia, em inúmeras situações, aplicamos a mesma lógica. Por exemplo, aplicamos provas nas faculdades para selecionar os alunos mesmo sabendo que elas deixam de fora muitos candidatos talentosos.

Assim, além de compreender que a discriminação é corriqueira na vida em geral, devemos entender o que é propriamente a discriminação estatística para a teoria econômica. Kenneth Arrow e Edmund Phelps elucidaram, através dessa teoria, como a desigualdade pode ser um problema mesmo quando não se está buscando qualquer resultado discriminatório. De acordo com esses autores, isso ocorre porque são usadas características de um grupo para validar a totalidade de indivíduos a ele pertencente (ARROW; PHELPS, 1973, p. 34).

Além disso, a teoria econômica supõe que tal forma de discriminação é racional e ocorre porque temos recursos e racionalidade limitados. Por isso, eles "tendem a utilizar características facilmente observáveis, como gênero, raça, educação etc., como proxies para características produtivas" (GOODMAN, 2016, p. 3). Para ilustrar essa situação, podemos tomar o mercado de trabalho como exemplo. Os entrevistadores talvez estejam mais propensos a selecionar homens ao invés de mulheres porque acreditam que a trajetória profissional da mulher é mais complicada. Elas precisam se dividir entre trabalho e família e, às vezes, não dão prioridade a profissão. O entrevistador não sabe nada em concreto da mulher que está entrevistando, mas pode adotar essa generalização quando for decidir quem irá contratar. Na teoria econômica, no entanto, não se discute se essa categorização é justa ou injusta, mas sim se seu uso é racional ou não. E, como já descrito, muitas vezes existe racionalidade na generalização.

Vale, ainda, observar outro importante aspecto da discriminação estatística. Ela pode ocorrer por diferenças exógenas ou endógenas entre grupos (MORO, 2009, p. 13). Na primeira situação, a variável de distinção é externa e, na segunda, é interna, podendo integrar a

Classificação: Interna

diferenciação[4]. Usando o exemplo do mercado de trabalho, historicamente as mulheres foram mais envolvidas nas tarefas domésticas do que os homens. Mas esse é um resultado do fato de que elas tiveram menos oportunidades profissionais e não uma característica do sexo feminino que torna as mulheres menos capacitadas para oportunidades profissionais. Portanto, a consequência da discriminação, nesse cenário, leva à confirmação da hipótese, já que estamos analisando uma variável endógena.

Por outro lado, observando o mercado de seguros de automóveis, a precificação guarda uma correlação entre gênero e taxas de acidentes. O fato de o seguro para jovens do sexo masculino ser mais caro do que para jovens do sexo feminino torna o gênero uma variável exógena, pois o fato dos homens pagarem mais por seguros não leva esse grupo, efetivamente, se envolver em mais acidentes.

Diante dessas observações, esclarece-se que o termo "discriminação algorítmica" engloba, neste artigo, tanto cenários que envolvam afirmações estatisticamente inconsistentes quanto cenários em que as afirmações, embora lógicas do ponto de vista estatístico, tomam o indivíduo apenas como parte de um grupo e não de forma efetivamente individualizada. Isso porque uma classificação, ainda que consistente do ponto de vista estatístico, pode se mostrar injusto, em algumas situações.

Tendo isso em mente, a seguir classificaremos os quatro tipos principais de discriminação algorítmica, segundo a Associação Brasileira de *Lawtechs* e *LegalTechs* (BABO, 2020, p. 1), que auxiliam na compreensão desse cenário. São eles: discriminação por erro estatístico, discriminação por generalização, discriminação pelo uso de informações sensíveis e, por fim, discriminação limitadora do exercício de direitos.

Discriminação por erro estatístico – é um tipo de discriminação que deriva de um erro cometido pelos programadores responsáveis pelo desenho do algoritmo. Ocorre quando o erro é genuinamente estatístico, abrangendo desde dados que não foram coletados corretamente até problemas no código do algoritmo.

Discriminação por generalização – deste tipo de discriminação, algumas pessoas são erroneamente classificadas em alguns grupos[5]. Se, por exemplo, uma pessoa mora em um local associado à pobreza e o modelo só utiliza seu endereço para decidir se ela é uma boa candidata

---

[4] Este resultado denomina-se *feedback loop*. Um feedback ocorre quando o *output* de um sistema é colocado de volta no sistema como um *input*. Portanto, o resultado de menos mulheres sendo contratadas retorna ao sistema decisório como um input para aquele que tomará a decisão, reforçando a conclusão que ele mesmo cria.

[5] O problema também pode advir do fato de que o sistema não possui informações suficientes do indivíduo, e, deste modo, classifica-o de acordo com as informações que possui, que acabam sendo insuficientes para refletir o que de fato é real.

Classificação: Interna

para obter crédito, ele a classificará como pertencente a um grupo que do qual talvez ela não faça parte, impedindo-a de receber o benefício. Assim, embora o algoritmo esteja correto e a informação também, o resultado é uma generalização incorreta, na medida em que um percentual de pessoas não se encaixam na média estabelecida mesmo que o resultado seja estatisticamente relevante.

Discriminação pelo uso de informações sensíveis – se baseia em dados legalmente protegidos. Por isso consideramos esta categoria como discriminatória. Um exemplo dessa situação é quando um algoritmo utiliza informações sobre identificação religiosa para designar o *credit score* do indivíduo. No Brasil, a Lei do Cadastro Positivo proíbe o uso dessa informação para a finalidade descrita. Para considerar um perfilhamento como discriminatório, nesse caso, deve-se levar em conta duas características: além do uso de dados sensíveis, a classificação baseia-se em características endógenas[6] ou destaca grupos historicamente discriminados.

Discriminação limitadora do exercício de direitos – ao contrário da discriminação anterior, o problema não é devido ao tipo de dado utilizado, mas da relação entre a informação e a realização de um direito. Se há uma relação estrita entre ambos e se o direito em questão é muito afetado, provável que o uso seja discriminatório[7].

É importante sabermos qual é a utilidade de cada tipologia, já que prover um parâmetro de análise para situações cotidianas é o maior objetivo de todas as categorizações. Em relação a primeira classificação, o resultado discriminatório é facilmente percebido, seja porque as informações usadas na decisão são incorretas, seja porque o procedimento, em si, é incorreto. Esse tipo de resultado equivocado é um problema muito comum em casos como o de *credit scoring*. A Comissão Federal do Comércio dos Estados Unidos (*Federal Trade Commission*), em um parecer ao Congresso, demonstrou que a margem de erro constatada em Relatórios de Crédito nos EUA está entre 10% e 21%, a depender da natureza do erro (BUCHNER, 2006, p. 3).

Já na segunda categoria, as decisões de gerenciamento de riscos, por exemplo, não levam em conta somente os dados pessoais, que indicam o risco de crédito ou de quebra de contrato pela pessoa. Normalmente, essas decisões se baseiam em outras características que, de

[6] Como dito anteriormente, os atributos utilizados para tomar decisões podem ser endógenos ou exógenos. Quando a característica possuir efeitos endógenos, sempre levará à discriminação. Se for exógena, ainda que sensível, o resultado pode não ser discriminatório.

[7] Na visão de Schauer, o problema não advém da discriminação, mas sim de um sentimento de exclusão do acesso a uma ferramenta que auxilia a realização de um direito essencial.

Classificação: Interna

acordo com a estatística, muitas vezes coincidem com o risco a ser analisado (BRITZ, 2008, p. 42).

A literatura econômica chama esse fato de discriminação estatística, como já abordado anteriormente, porque a correlação entre características que aparentam ser neutras e os dados-alvo é resultado de um método estatístico, conforme definido por Becker em *The Economics of Discrimination*. Do lado da perspectiva dogmático-jurídica, essa prática pode levantar questões sobre justiça individual e desigualdade, abrindo margem para que tratamentos diferenciados ocorram com base em características pessoais, já que essas são um aspecto importante para a tomada de decisões. Já que a característica que se está procurando é, frequentemente, difícil de se mensurar (credibilidade, solvência, produtividade laboral etc.), uma "característica proxy" é utilizada no lugar desta característica principal (MAYER-SCHONBERGER; CUKIER, 2014, p. 60).

No entanto, como abordado na obra de Mayer-Schonberger e Kenneth Cukier, apesar de possuir a característica *proxy*, uma pessoa pode não apresentar as demais qualidades esperadas do grupo e aí está o problema das injustiças causadas pela generalização. Um exemplo é o uso do endereço de um cliente como parte de uma análise de empréstimo, ao pressupor que os ativos do cliente podem ser derivados do local em que vive pode resultar em uma negativa no procedimento de crédito, sem qualquer avaliação posterior de solvência real e avaliações do solicitante.

Outro problema que gera reflexões em matéria de igualdade surge da forma como se trata diferentes pessoas baseando-se em características discriminatórias e estereotipadas clássicas, como nacionalidade, gênero, idade e até mesmo identidade sexual. Ao passo que essas características estão relacionadas a personalidade de cada um, mostrando-se invariáveis, esses atributos, por outro lado, representam diferenças históricas de tratamento de grupos e sua utilização como base para tomada de decisões pode trazer intensificar a discriminação de certos grupos na sociedade (O'NEIL, 2016, p. 57).

Aqui, todas as pessoas do grupo são afetadas porque não se trata simplesmente de uma violação decorrente de uma categorização incorreta de uma pessoa "atípica". Essa discriminação é resultando de informações sensíveis e, muitas vezes, pode ser estatisticamente consistente. No entanto, por serem grupos discriminados no decorrer da história, é um dos tipos mais bárbaro de discriminação ao reforçar o tratamento discriminatório e, ainda por cima, automatizá-lo, tornando mais complicado para os membros de tais grupos superarem uma situação prejudicial a que foram submetidos (O'NEIL, 2016, p. 61). Esse uso de características

Classificação: Interna

discriminatórias pode ser encontrado especialmente no setor de seguros, onde nacionalidade e gênero podem ser utilizados como critérios negativos em uma avaliação de risco de crédito.

Já o quarto tipo de discriminação, como relatado, é um tipo de cenário em que a circunstância estatística pode se verificar, mas o resultado pode ser discriminatório. Um importante exemplo desse tipo de discriminação ocorreu na Alemanha, onde se verificou que os bureaus de crédito consideravam o mero exercício do direito da pessoa de acesso ao seu *score* de crédito como um aspecto relevante e negativo para esse *score*: aquelas pessoas que efetivamente acessavam as suas informações tinham sua pontuação reduzida, já que teriam maior risco de inadimplemento (WYMAN, 2018, p. 26).

Quando essa prática foi descoberta, houve uma mudança legislativa: a antiga Lei de Proteção de Dados Alemã (hoje revogada pela entrada em vigor do Regulamento Europeu de Proteção de Dados) passou a prever, em seu parágrafo 6,3, que "exercício dos direitos dos titulares previstos na lei não poderia ser utilizado em seu detrimento".

Feitas essas considerações e essa tentativa de categorizar a discriminação algorítmica, debateremos formas de mitigar o problema, com foco na literatura de governança que se desenvolveu ao longo dos anos.

2. Discriminação algorítmica e governança: uma discussão sobre as soluções expostas na literatura

Para entender a questão da discriminação algorítmica e sua regulamentação, é importante demonstrar a literatura de governança algorítmica. Os autores que estão na discussão de como a discriminação é um fator determinante em processos decisórios algorítmicos perceberam a necessidade de haver soluções regulatórias e de política púbica para minimizar essa questão e, por isso, focam em oferecem uma visão geral de possíveis soluções.

Esse capítulo analisará as discussões mais importantes dessa área, enfatizando as soluções mais comumente discutidas, as discordâncias entre os autores a respeitos dessas propostas e a minha visão a respeito do assunto.

A literatura sobre governança algorítmica é vasta, embora seja recente. Muitos autores discutem se, quando ou como os algoritmos devem ser regulados. Uma das primeiras frentes que desejaram colaborar para o debate foi formada por acadêmicos que elaboraram princípios que poderiam regular os processos decisórios algorítmicos. A *Fairness, Accountability and Transparency in Machine Learning Organization* (FAT-ML) é uma dessas instituições. Ela organizou uma lista do que entendiam ser os princípios-chave que deveriam ser observados

Classificação: Interna

tanto pelo governo quanto pelo setor privado ao lidar com algoritmos. São eles: responsabilidade, explicabilidade, precisão, auditabilidade e justiça[8]. Nos Estados Unidos também surgiram associações que seguiram o mesmo caminho e desenvolveram seus próprios princípios, como a *Association for Computing Machinery* (ACM), acrescentando a lista os seguintes temas: conscientização, acesso e reparação, proveniência dos dados, validação e experimentação[9].

Responsabilidade, sob a ótica da FAT-ML, está ligada à ideia de que é preciso sempre ter em mente que as pessoas são afetadas pelos processos decisórios e que, por isso, é necessário oferecer alternativas caso haja necessidade de reparação de danos individuais ou coletivos. Esta ideia está relaciona ao princípio da conscientização da ACM, o qual torna os engenheiros e usuários dos algoritmos conscientes das consequências de seu uso, especialmente do enviesamento que pode vir a surgir e, consequentemente, os reguladores deveriam adotar medidas que permitissem que os indivíduos afetados pudessem questionar e serem reparados dos danos sofridos.

A ideia de explicabilidade, como abordado por Doshi-Velez, se refere às "razões ou justificativas para aquele resultado em participar, e não a uma descrição do processo decisório em geral" (DOSHI-VELZEZ, 2017, p. 2). Assim, os autores definiram como explicabilidade, e é esta definição que vamos adotar, é uma "descrição, compreensível por humanos, do processo por meio do qual aquele que toma a decisão, ao utilizar um certo grupo de inputs, atinge uma dada conclusão"[10] (DOSHI-VELEZ, 2017, p. 2-3). Ressalta-se que explicabilidade não é o mesmo que transparência, uma vez que entender o processo da decisão tomada não é a mesma coisa que conhecer os passos tomados para atingir a decisão.

Precisão, segundo Diakopoulos e Friedles, significa que "as fontes de erros e incertezas de um algoritmo e suas fontes de dados precisam ser identificadas, registradas e comparadas" (DIAKOPOULOS E FRIEDLES, 2016, p. 2). Eles afirmavam que somente entendendo de onde os erros surgem e possível minimizá-los. A ACM possui uma ideia parecida por meio do princípio da proveniência dos dados: "os designers dos algoritmos deveriam manter uma

[8] The Fairness, Accountability and Transparency in Machine Learning Organization. Os autores que subscrevem os princípios são: N. Diakopoulos, S. Friedler, M. Arenas, S. Barocas, M. Hay, B. Howe, H. V. Jagadish, K. Unsworth, A. Sahuguet, S. Venkatasubramanian, C. Wilson, Cong Yu e B. Zevenbergen.

[9] Association for Computing Machinery US Public Policy Council (USACM), "Statement on Algorithmic Transparency and Accountability".

[10] Eles vão além, para dizer que uma explicação deveria ser apta a responder ao menos uma das três perguntas seguintes: quais são os principais fatores em uma decisão; (a alteração de um fator teria mudado a decisão; e porque dois casos similares recebem decisões diferentes ou vice-versa.

Classificação: Interna

descrição do modo pelo qual os dados utilizados foram coletados, acompanhada de uma explanação das potenciais tendências induzidas pelo processo de coleta".

O princípio da auditabilidade também faz parte da discussão sobre governança algorítmica. Ele traduz a ideia de que um terceiro deve poder avaliar o método utilizado pelo algoritmo para chegar àquela conclusão. No entanto, há muita divergência a respeito de tal divulgação, se seria possível e se ela deve mesmo ocorrer em algumas situações.

Justiça, por mais comum que seja, é o princípio menos claro de todos os registrados. A ideia é que os algoritmos não podem gerar resultados discriminatórios. No entanto, analisar o que é um resultado discriminatório pode ser um desafio e há discordâncias sobre o assunto. A ACM, embora não expresse especificamente esse princípio, estabelece a validação e experimentação: "[instituições] deveriam realizar testes rotineiramente para avaliar e determinar se o modelo gera danos discriminatórios".

Um dos princípios amplamente discutidos pelos responsáveis pela elaboração de políticas públicas é o da transparência. Embora não esteja de maneira explícita em nenhum dos manifestos até aqui explorados, a transparência, normalmente, não está presente nos algoritmos que são chamados por Frank Pasquele, por exemplo, de "caixas pretas" por causa da falta de transparência dos processos decisórios, que acaba por transformá-los em uma fonte de desconfiança (PASQUELE, 2013, p. 21).

Não há uma combinação melhor para as propostas e os autores continuam discordando a respeito da sua utilidade e importância. Pasquale e Citron acreditam que a transparência é uma solução plausível para a discriminação algorítmica, especialmente quando se trata do *credit scoring*:

> Nós acreditamos que cada titular dos dados deveria ter acesso a todos os dados a ele pertencentes. Idealmente, a lógica dos sistemas de *scoring* preditivos também deveria ser aberta à inspeção pública. Há poucos indicativos de que a inabilidade de manter tais sistemas em segredo diminuiria a inovação. (PASQUELE, CITRON, 2014, p. 26).

Eles ainda afirmam que "ameaças à dignidade humana" são suficientes para ensejar a abertura do código e da modelação dos algoritmos ao público em geral (PASQUALE; CIRON, 2014, p. 30-31).

O autor Zersky também segue uma linha de raciocínio semelhante. Em resumo, ele afirma que "a justificativa mais básica e difundida para a transparência é que ela facilita a fiscalização das ações governamentais" (ZERSKY, 2013, p. 1533). Ele alega que, nesse cenário, transparência e *accountability* são praticamente sinônimos , mas argumenta que os termos em si são diferentes: *accountability* se refere à ideia de que "os indivíduos são

Classificação: Interna

eticamente responsáveis por seus atos, ao passo que a transparência é uma ferramenta voltada para facilitar a responsabilização" (ZERSKY, 2013, p. 1545).

No entanto, alguns especialistas afirmaram que a transparência é limitada. Lawrence Lessig no seu artigo *Against Transparency* publicado no *The New Republic* afirmou que direcionar o panóptico aos governantes, assim como construir uma onisciência cívica, traria problemas do ponto de vista da transparência governamental. Ele traduz seu pensamento explicando as ideias apresentadas por Brandeis na obra *Other People's Money*, quando o autor afirma que a total transparência das informações abriria espaço para o público julgar a qualidade, permitindo que as próprias pessoas regulassem o mercado. Contudo, como Lessig destaca "nem todos os dados são uma informação que os consumidores podem utilizar, apresentados de modo que eles possam utilizar" (LESSIG, 2009, p. 19). Embora o objetivo do artigo de Lessig não seja discutir a discriminação algorítmica, muitas das observações do autor mostram-se válidas para este contexto.

Nesse mesmo sentido, Krol aponta quatro argumentos centrais para justificar o porquê a transparência não é uma proposta de política suficiente (KROL, 2017, p. 628): a transparência é inatingível – sempre haverá razões públicas que afastam o direito à divulgação (segurança nacional, por exemplo) ou motivos privados voltados à prevenção de comportamentos estratégicos. O segundo argumento é que a transparência talvez seja insuficiente: "[os] métodos geralmente são insuficientes para verificar as propriedades dos sistemas de *softwares* se esses sistemas não tiverem sido projetados considerando avaliações e prestações de contas futuras" (KROL, 2017, p. 633). Em terceiro, uma vez que o algoritmo incorpora um tipo de aleatoriedade, que é uma função dos sistemas informatizados, a presença de transparência total não garante nada[11]. Por último, afirma que sistemas inteligentes estão em constante mudança e evolução, como é o caso do *machine learning*, e por isso, não podem ser compreendidos por meio da transparência. Por isso, afirmam que uma prestação de contas, uma regularização procedimental, seria uma proposta de política melhor e que deveria ter seu estudo aprofundado.

Veale e Edward, seguindo essa linha, destacam as limitações da transparência e da explicabilidade, focando no Regulamento Geral de Proteção de Dados (GDPR). Eles afirmam a inutilidade da transparência tanto pela sua inviabilidade quanto pela inadequação às necessidades dos usuários e sugerem uma estrutura criada para construir soluções algorítmicas

[11] Em suas palavras, "uma loteria oferece um excelente exemplo: um algoritmo perfeitamente transparente – utiliza um gerador de números aleatórios para atribuir um número para cada participantes e permite que os participantes com os números mais baixos vençam – produz resultados que não podem ser reproduzidos ou verificados, porque o gerador produzirá novos números aleatórios quando novamente solicitado" (J. Kroll et al., 2017, p.634).

Classificação: Interna

melhores desde o princípio (soluções efetivamente relacionadas aos objetivos da política pública) e em dar às instituições e agências o poder para supervisionar a integrada algorítmica (EDWARDS; VEALE, 2017, p. 22-23).

Sunstein colabora com a discussão ao identificar que os algoritmos possuem o potencial "de dar passos importantes no combate à discriminação e modificar a forma como o sistema jurídico a entender há muito tempo" (SUNSTEIN, 2019, p. 52). Embora os algoritmos forneçam novas ferramentas para as pessoas expressarem seus preconceitos e vieses, a implantação de um sistema regulatório adequado não é limitante apenas da possibilidade de discriminação algorítmica, mas tem também o potencial de transformar os próprios algoritmos um contrapeso à discriminação humana e uma força positiva para se alcançar o bem social.

Portanto, se a regulamentação garantir as salvaguardas apropriadas, as chances de detectar a discriminação em um mundo de design de algoritmos é muito maior quando comparada a um contexto de decisões tomadas por humanos. Para os autores aqui estudados, um processo regulamentado se destaca por sua transparência e especificidade. O acesso ao algoritmo nos permite fazer perguntas que não podemos fazer significativamente aos seres humanos (SUSTEIN ET AL., 2019, p. 4).

Considerações finais

A governança algorítmica é um assunto complexo e amplamente controvertido. Mesmo em jurisdições onde o debate sobre o assunto é mais presente, os problemas estão longe de serem solucionados. As soluções para a discriminação algorítmica apontadas pelos acadêmicos, como a transparência e a fiscalização, tratam de algumas preocupações, mas outras continuam sem solução, como o *learning machine* que trouxe diversas dúvidas sobre a efetividades dessas soluções.

Nesse sentido, o desafio da transparência e da explicabilidade é enorme e do ponto de vista científico tem que enfrentar alguns dilemas. Um deles é que a mesma complexidade que permite o desenvolvimento do *learning machine*, com previsão e antecipação, veda a transparência aos usuários e aos próprios programadores. A comparação com vieses e o erro humano pode até servir de consolo para os cientistas de dados. No entanto, ao final, quando se compara falha humana e dos algoritmos, a vantagem, certamente, ficará com a Inteligência Artificial. Assim, em áreas de sensibilidade elevada, por exemplo a saúde, a supervisão humana é mandatória (ARBIX, 2020, p.13).

Classificação: Interna

Diante disto, entendo que, independentemente da solução adotada (transparência, desenvolvimento de ferramentas de *accountability*, fiscalização, entre outras), o papel humano tem função fundamental no processo de automação. Isso não envolve apenas a revisão de decisões automatizadas por pessoas naturais, mas também de centralizar o elemento humano em todo o processo, desde o desenho até a implementação.

Esse papel deve iniciar desde a montagem de comitês diversos para desenvolver o design dos sistemas algorítmicos. Como aponta Cathy O"Neil, é muito importante que as pessoas responsáveis por desenvolver o programa sejam capacitadas não somente para entender os aspectos técnicos, mas também para prever os efeitos do uso do sistema no mundo real. Portanto, é fundamental que os programadores sejam treinados para compreender os aspectos éticos e morais do programa que estão desenvolvendo. Várias universidades dos Estados Unidos, por exemplo, estão seguindo esse caminho, oferecendo desde cursos sobre ética e governança da inteligência artificial até estruturando centros de pesquisa sobre a temática[12].

Aqui, no Brasil, ao analisar a grade curricular de cursos como Sistemas da Informação da Universidade de São Paulo (disponível no Jupiterweb), por exemplo, percebe-se que há poucas matérias de ciências humanas, o que dificulta a compreensão social dos efeitos dos sistemas criados. Uma reforma nesses cursos formadores dos cientistas de dados com a inclusão de matérias para desenvolver um senso crítico dos desenvolvedores, como o que se busca nas universidades americanas, seria um bom caminho para a governança.

Além disso, a ex-chefe de Inteligência Artificial do Google, Fei-Fei Lin, ressaltou, ainda, a necessidade de diversidade na constituição dos grupos dos cientistas de dados. Ela afirma que se os times não forem compostos por um grupo diversos de engenheiros e programadores a chance de os vieses presentes na sociedade serem passados aos algoritmos é muito grande (HEMPEL, 2018, p. 1).

A centralidade humana também precisa estar presente no processo de revisão das decisões automatizadas. Deve ser refletiva tanto no momento de teste, ou seja, antes de ser efetivamente utilizado para que os impactos reais sejam auferidos, quanto na análise de decisões tomadas e questionadas por pessoas que se sentiram, eventualmente, prejudicadas. Para que os resultados sejam revistos, é necessário que as pessoas responsáveis por esse processo sejam capazes de efetivas mudanças em uma situação concreta e estimularem uma eventual

[12] As universidades Harvard e MIT passaram a oferecer, conjuntamente, um curso sobre o tema. A Universidade de Stanford, por sua vez, inaugurou um centro de pesquisa.

Classificação: Interna

necessidade de adaptação do sistema, caso cheguem à conclusão que o resultado discriminatório ou erro do sistema possa ocorrer novamente.

Portanto, os valores que orientam a sociedade e o direito não podem ser afastados quando tratamos de inteligência artificial, discriminação algorítmica e automação das decisões. O desafio é, nesse sentido, traduzir para sistemas automatizados o que vem sido construído nas ciências humanas há muitos anos.

Referências

AFONSO DA SILVA, V. *A constitucionalização do direito*: direitos fundamentais e relações entre particulares. São Paulo: Malheiros, 2005.

ARROW, K. The theory of discrimination. In: ASHENFELTER, O.; REES, A. (Ed.). *Discrimination in Labor Markets*. Princeton: Princeton University Press, 1973.

ARBIX, G. *A transparência no centro da construção de uma IA ética*. São Paulo: CEBRAP. 2020.

BAROCAS, S.; SELBST, A. Big Data's Disparate Impact. *California Law Review*, v. 104, p. 671-732, 2016.

BENJAMIN, A. Herman et al. *Código brasileiro de Defesa do Consumidor comentado pelos autores do anteprojeto*. Rio de Janeiro: Forense Universitária, 2005.

BEWARE Spurious Correlations. *Harvard Business Review*, June 2015. Disponível em: <https://hbr.org/2015/06/beware-spurious-correlations>. Acesso em: 17 out. 2020.

BOYD, D.; CRAWFORD, K. Six Provocations for Big Data. In: *A decade in internet time*: symposium on the dynamics of the internet and society, 2011, 17 p. Disponível em: <https://ssrn.com/abstract=1926431>. Acesso em: 17 out. 2020.

BUCHNER, B. *Informationelle Selbstbestimmung im Privatrecht*. Tübingen: Mohr Siebeck, 2006.

BRITZ, G. *Freie Entfaltung durch Selbstdarstellung*. Tübingen: Mohr Siebeck, 2007.

BRASIL, Lei Geral de Proteção de Dados, Capítulo dos Direitos do Titular, Art. 20. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em 17 out. 2020

BURKELL et al. Facebook: public space, or private space? *Information, Communication & Society*, v. 17, p. 974-985, 2014.

BUTLER, D. When Google got flu wrong. *Nature*, 13 February 2013. Disponível em: <https://www.nature.com/news/when-google-got-flu-wrong-1.12413>. Acesso em: 17 out. 2020.

CANOTILHO, J. J. Gomes; MOREIRA, Vital. Constituição da República portuguesa anotada. 4. ed. Coimbra: Coimbra, v. I, 2007.

CORMEN, T. H. *Algorithms Unlocked*. MIT Press, 2013.

DANIELI, O.; HILLIS, A.; LUCA, M. How to Hire with Algorithms. *Harvard Business Review*, 17 October 2016. Disponível em: <https://hbr.org/2016/10/how-to-hire-with--algorithms>. Acesso em: 17 out. 2020.

DIAKOPOULOS, N.; FRIEDLER, S. How to Hold Algorithms Accountable, 17 november 2016. Disponível em: <https://www.technologyreview.com/s/602933/how-to-hold--algorithms-accountable>. Acesso em: 17 out.2020.

DOMINGOS, P. *The Master Algorithm*: How the Quest for the Ultimate Learning Machine Will Remake our World. New York: Basic Books, 2015.

DONEDA, D. *Da privacidade à proteção de dados pessoais*. Rio de Janeiro: Renovar, 2006.

_____; MENDES, L. Data Protection in Brazil: New Developments and Current Challenges. In: GURWIRTH, S.; LEENES, R.; DE HERT, P. (Ed.). *Reloading Data Protection*: Multidisciplinary Insights and Contemporary Challenges. Springer, 2014.

DOSHI-VELEZ, F. et al. Accountability of AI Under the Law: The Role of Explanation. *Harvard Public Law*, n. 18-07, 15 p., 2017.

EDWARDS, L.; VEALE, M. Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For. *Duke Law and Technology Review*, v. 16, n. 1, p. 18-84, 2017.

FEDERAL TRADE COMISSION. Report to Congress Under Section 319 of the Fair and Accurate Credit Transactions Act of 2003, December 2012. Disponível em: <https://www.ftc.gov/sites/default/files/documents/reports/section-319-fair-and-accurate-credit-transactions-act-2003-fifth-interim-federal-trade-commission/130211factareport.pdf>. Acesso em: 17 out. 2020.

FINLEY, S. I didn't even meet my potential employers. *BBC News*, 6 February 2018. Disponível em: <http://www.bbc.com/news/business-42905515>. Acesso em: 17 out. 2020.

GEE, K. In Unilever's Radical Hiring Experiment, Resumes are out, Algorithms are in'. *The Wall Street Journal*, 26 June 2017. Disponível em: <https://www.wsj.com/articles/in-unilevers-radical-hiring-experiment-resumes-are-out-algorithms-are--in-1498478400>. Acesso em: 17 out. 2020.

GINSBERG, J. et al. Detecting Influenza epidemics using search engine query data. *Nature*, v. 457, p. 1012-1014, 2009.

GOODMAN, Bryce W. Economic Models of (Algorithmic) Discrimination. *29th Conference on Neural Information Processing Systems*, Barcelona, Spain, p. 3, 2016.

HURLEY, M.; ADEBAYO, J., Credit Scoring in the Era Of Big Data. *Yale Journal of Law and Technology*, v. 18, n. 1, p. 148-216, 2016.

KROLL, J. et al. Accountable algorithms. *University of Pennsylvania Law Review*, v. 165, p. 633-705, 2017.

LAZER, D. et al. The Parable of Google Flu: Traps in Big Data Analysis. *Science*, v. 343, p. 1203-05, 2014.

LESSIG, L. Against Transparency. *The New Republic*, 9 October 2009. Disponível em:

MAYER-SCHÖNBERGER, V.; CUKIER, K. *Big Data*: A Revolution That Will Transform How We Live, Work, and Think. New York: First Mariner Books, 2014.

MENDES, L. *Privacidade, proteção de dados e defesa do consumidor*. Linhas gerais de um novo direito fundamental. São Paulo: Saraiva, 2014.

_____. *Schutz vor Informationsrisiken und Gewährleistung einer gehaltvollen Zustimmung*. Eine Analyse der Rechtmäßigkeit der Datenverarbeitung im Privatrecht. Berlin: De Gruyter, 2015.

MORO, A. Statistical Discrimination. In: DURLAUF, N. S.; LAWRENCE, E. (Ed.). *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, 2009.

O'NEIL, C. Personality Tests Are Failing American Workers. *Bloomberg View*, 18 January 2018. Disponível em: <https://www.bloomberg.com/view/articles/2018-01-18/personality-tests-are-failing-american-workers>. Acesso em: 17 out. 2020.

O'NEIL, C. Weapons of Math Destruction**:** how big data increases inequality and threatens democracy. New York: Crown Publishers, 2016.

PAPACHARISSI, Z. *A private sphere*: democracy in a digital age. Cambridge: Polity Press, 2010.

PASQUALE, F. The Emperor's New Codes: Reputation and Search Algorithms in the Finance Sector. *NYU "Governing Algorithms" Conference*, 2013.

_____; CITRON, D. The Scored Society: Due Process for Automated Predictions. *Washington Law Review*, v. 89, p. 1-33, 2014.

PHELPS, E. The Statistical theory of Racism and Sexism. *American Economic Review*, v. 62, p. 659-61, 1972.

Classificação: Interna

PINHO MELLO, J. M.; MENDES, M.; KANCZUK, F. Cadastro Positivo e democratização do crédito. *Folha de São Paulo*, mar. 2018. Disponível em: <https://www1.folha.uol.com.br/opiniao/2018/03/joao-manoel-pinho-de-mello-marcos-mendes-e-fabio-kanczuk-cadastro-positivo-e-democratizacao-do-credito.shtml>. Acesso em: 17 out. 2020.

SANDVIG, C. et al. An Algorithm Audit. In: GANGADHARAN, S. P. (Ed.). *Data and Discrimination*: Collected Essays. Open Technology Institute, 2014.

SCHAUER, F. *Profiles, Probabilities, and Stereotypes*. Cambridge: Harvard University Press, 2006.

SILVER, N. *The Signal and the Noise*. The Art and Science of Prediction. London, 2012.

STATE *v.* Loomis. 881 N.W.2d 749 (Wis. 2016). Disponível em: <https://harvardlawreview.org/2017/03/state-v-loomis/>. Acesso em: 17 out. 2020.

WYMAN, Oliver. Mercado de crédito. Série Panorama Brasil, 2018.

ZARSKY, T. Transparent Predictions. *University of Law Review*, v. 2013, n. 4, p. 1503-1570, 2013.

COORDENADORIA DE TRABALHO DE CONCLUSÃO DE CURSO (TCC)

TERMO DE AUTENTICIDADE E AUTORIZAÇÃO PARA PUBLICAÇÃO DO TRABALHO DE CONCLUSÃO DE CURSO

Eu, Gabriella Leriam

Aluno(a), regularmente matriculado(a), no Curso de Direito, na disciplina do TCC da 10ª etapa, matrícula nº 31623980, Período noturno, Turma S,

tendo realizado o TCC com o título: DISCRIMINAÇÃO ALGORÍTMICA: PERSPECTIVAS ÉTICAS E GOVERNANÇA

sob a orientação do(a) professor(a): Eduardo Altomare Ariente

declaro para os devidos fins que tenho pleno conhecimento das regras metodológicas para confecção do Trabalho de Conclusão de Curso (TCC), informando que o realizei sem plágio de obras literárias ou a utilização de qualquer meio irregular.

Declaro ainda que, estou ciente que caso sejam detectadas irregularidades referentes às citações das fontes e/ou desrespeito às normas técnicas próprias relativas aos direitos autorais de obras utilizadas na confecção do trabalho, serão aplicáveis as sanções legais de natureza civil, penal e administrativa, além da reprovação automática, impedindo a conclusão do curso.

São Paulo, 23 de outubro de 2020.

Gabriella Leriam

Assinatura do discente

Material Bibliográfico: (x) Artigo Científico ( ) Monografia

Classificação: Interna