

Todos São Iguais Perante o Algoritmo? Uma Resposta Cultural do Direito à Discriminação Algorítmica

Is Everyone Equal Before the Algorithm? A Cultural Response of Law to Algorithmic Discrimination

ALAN DUARTE¹

Universidade Federal do Ceará (UFC).

RAMON DE VASCONCELOS NEGÓCIO²

Centro Universitário 7 de Setembro (UNI7).

RESUMO: São cada vez mais notórios os casos em que algoritmos de aprendizado de máquina que operam em rede apresentam comportamentos discriminatórios, de modo a evidenciar uma semântica da discriminação que perpassa instituições e estruturas sociais, incorpora-se aos aspectos técnicos da programação e ameaça direitos fundamentais. Evidencia-se, nesse sentido, a importância de se discutir, no âmbito das ciências sociais, sobretudo do Direito, como se construir um modelo de observação, identificação e busca de soluções capaz de lidar com o problema da discriminação perpetrada por algoritmos de aprendizado de máquina, os quais operam majoritariamente por meio da internet. Para solucionar esse problema, o trabalho apresenta, em um primeiro momento, a igualdade compreendida como não discriminação. Questiona-se, em seguida, por meio da análise do funcionamento dos algoritmos, como esses sistemas com capacidade de aprendizado tratam – e como isso ocorre – grupos de maneira desfavorável. Por fim, partindo da compreensão do Direito como uma semântica social, como cultura imbricada nas mais variadas relações sociais, e não apenas nas instituições formais, apresentam-se mecanismos de construção de um modelo de observação, identificação e busca de soluções voltada ao combate da discriminação por algoritmos. A partir de uma pesquisa essencialmente bibliográfica e do método hipotético-dedutivo, conclui-se que a normatividade do Direito está na comunidade de usuários e nas práticas dos programadores, o que revela a importância de impregnar a semântica jurídica da igualdade na seleção e formação de programadores, como possibilidades de solução do problema posto.

1 Orcid: <http://orcid.org/0000-0002-0762-1710>.

2 Orcid: <http://orcid.org/0000-0003-4724-9658>.

PALAVRAS-CHAVE: Discriminação por algoritmos; algoritmos de aprendizado de máquina; Direito como cultura.

ABSTRACT: There are increasingly notorious cases in which machine learning algorithms that operate in a network reveal discriminatory behaviors. They demonstrate a semantics of discrimination that permeates social institutions and structures, and end up being incorporated into technical aspects of programming, thus threatening fundamental rights. In this sense, it is fundamental to discuss, in the realm of social sciences, and especially in the realm of law, how to build a model of observation, identification, and search for solutions capable of dealing with the problem of discrimination perpetrated by machine learning algorithms, which operate mostly on the internet. To solve this problem, this paper deals, at first, with the concept of equality in the sense of non-discrimination. By analyzing the functioning of algorithms, I then turn to the question of how these systems with learning capacity treat groups in an unfavorable way, and how this occurs. Finally, taking law as a social semantics, as a culture intertwined with all sorts of social relations, and not only with formal institutions, we present mechanisms to build a model of observation, identification, and search for solutions aimed at combating discrimination by algorithms. From an essentially bibliographic methodology and the hypothetical-deductive method, we conclude that the normativity of law is to be found in the community of users and in the practices of programmers, which reveals the importance of embedding the legal semantics of equality in the selection and training of programmers as one of the possible solutions for the problem here identified.

KEYWORDS: Discrimination by algorithm; machine learning algorithms; Law as culture.

SUMÁRIO: Introdução; Igualdade como não discriminação; A (des)igualdade por algoritmos e pela inteligência artificial; Direito como cultura; A semântica jurídica na comunidade de usuários e no programador: uma resposta para o problema; Conclusão; Referências.

INTRODUÇÃO

Joy Buolamwini, quando era estudante de Ciência da Computação em Georgia Tech, trabalhava com robôs sociais, tendo como um de seus projetos construir um desses robôs para jogar *peek-a-boo*³; para a execução dessa tarefa, Joy utilizou um *software* genérico de reconhecimento facial. Ocorre que ela encontrou um problema nessa tarefa: o sistema que ela desenvolveu não conseguia identificar seu rosto, muito embora conseguisse identificar o rosto de sua colega de quarto. A diferença entre ambas e o fator determinante para a falha no reconhecimento é que Joy é negra (Buolamwini, 2016). Esse problema, todavia, não ocorreu apenas de forma isolada com Joy. Repetiu-se em outros contextos e com repercussões mais dramáticas,

3 *Peek-a-boo* é um jogo, comumente jogado com crianças, em que uma pessoa esconde o rosto, principalmente com as mãos, e de repente descobre o rosto e diz “*peek-a-boo*”, o que, em tradução livre, seria algo como “te achei”.

pois um sistema de reconhecimento facial levou ao etiquetamento de Jacky Alcine e sua amiga, ambos negros, como “gorilas” (Google, 2015), e o uso de outro pelo sistema policial resultou na prisão de Robert William, rapaz também negro, por um delito que ele não cometeu (Hill, 2020).

Situações como essas demonstram que a exclusão de determinados grupos pode e é incorporada por algoritmos de tomada de decisão, de modo que tais sistemas não são neutros ou mais precisos quanto pretendem ou como advogam aqueles que defendem o seu uso irrestrito. Isso porque os algoritmos carregam em sua programação, além de dados brutos e cálculos estatísticos, crenças, vieses e equívocos humanos, pois as escolhas que levam à construção desses sistemas são feitas por seres humanos falíveis⁴. Além disso, os próprios dados que são carregados nesses sistemas são, em certa medida, enviesados, tendenciosos e/ou não representam adequadamente todos os setores da sociedade que deveriam representar.

Isso revela que as formas de discriminação se apresentam de maneiras mais sofisticadas, sendo incorporadas em processos técnicos e perpetuadas por algoritmos de computador, considerados ainda, por muitos, como mecanismos exatos e isentos de subjetividades. Diante desse cenário, os mecanismos jurídicos clássicos, sobretudo as normas de cunho antidiscriminatório, apresentam-se como insuficientes, principalmente por carregarem, em seu bojo, a ideia de intencionalidade na conduta discriminatória e por não compreenderem adequadamente as estruturas socioculturais e a semântica da discriminação que permeia os constructos sociais e técnicos.

A partir disso, o presente trabalho tem como pergunta principal a seguinte: como é possível construir um modelo de observação, identificação e busca de soluções capaz de lidar com o problema da discriminação perpetrada por algoritmos de aprendizado de máquina, os quais operam majoritariamente por meio da internet? Para solucionar esse problema, o trabalho parte de uma pesquisa essencialmente bibliográfica e do método hipotético-dedutivo e divide-se em três partes, além de introdução e conclusão. Em um primeiro momento, apresenta-se a igualdade compreendida como não discriminação, concretamente aduzida e não apenas abstratamente consi-

4 Sobre esse ponto há uma vasta literatura, dentre a qual se destaca: BAROCAS, Solon; SELBST, Andrew D. Big data's disparate impact. *Calif. L. Rev.*, v. 104, p. 671, 2016; O'NEIL, Cathy. *Weapons of math destruction: how big data increases inequality and threatens democracy*. New York: Crown Publishers, 2016; EUBANKS, Virginia. *Automating inequality: how high-tech tools profile, police, and punish the poor*. St. Martin's Press, 2018.

derada. Em seguida, questiona-se como os sistemas algorítmicos, sobretudo os que operam de forma autônoma, isto é, com capacidade de aprendizado, tratam – e como isso ocorre – grupos de maneira desfavorável. Tal questionamento é feito por meio da análise do funcionamento desses sistemas, bem como pela exploração das vias capazes de fazer o algoritmo redundar em comportamentos discriminatórios. Por fim, a partir da compreensão do Direito como uma semântica social, como cultura imbricada nas mais variadas relações sociais, e não apenas nas instituições formais, apresentam-se mecanismos de construção de um modelo de observação, identificação e busca de soluções voltado ao combate da discriminação perpetrada por sistemas algorítmicos.

IGUALDADE COMO NÃO DISCRIMINAÇÃO

O debate acerca da discriminação algorítmica tem como plano de fundo, do ponto de vista jusfilosófico, a antiga, relevante e profunda discussão sobre a igualdade. Em um primeiro momento, tem-se a igualdade formal, a qual é reduzida à fórmula “todos são iguais perante a lei”. Essa ideia de igualdade é adotada e disseminada dentro da primeira geração de direitos, a qual valoriza o homem-singular, de liberdades abstratas (Bonavides, 2017). A compreensão baseada nas ideias do liberalismo clássico (de que todos são iguais perante a lei), embora insuficiente nos dias de hoje, possibilitou a eliminação de privilégios de certos grupos a partir da premissa de que todos possuem uma identidade comum como cidadãos (Moreira, 2017, p. 50). O indivíduo, concebido como uma abstração singular e não corporificada, era a categoria universal do ser humano.

Todavia, essa igualdade formal passa a ser um problema exatamente naquilo que ela apresentava como mais positivo à época de sua instauração, na medida em que se pressupõe que se escolhe um padrão específico de indivíduo ao qual os demais, para serem considerados sujeitos de direitos, deveriam se igualar. Nesse sentido, “essa igualdade formal se corrompe ao eleger como parâmetro pressuposto um sujeito social nada abstrato: masculino, branco, europeu, cristão, heterossexual, burguês e proprietário” (Rios, 2012, p. 172-173). A ideia de igualdade jurídica, portanto, passa a estar mais associada à neutralização das desigualdades no exercício dos direitos do que à igualdade de fato (Neves, 2008, p. 166).

A partir dessa especificação do paradigma de indivíduo detentor de direitos fundamentais, exclui-se desse âmbito de proteção o outro, o diferen-

te. A diferença, portanto, era visibilizada apenas para conceber esse outro como um ser esvaziado de qualquer dignidade, descartável, uma verdadeira *res*, objeto de compra e venda (como na escravidão) ou de extermínio sob o pretexto de purificação do eu (que se contrapõe ao outro), como foi no nazismo (Piovesan, 2008). Nessa perspectiva, destaca Joan Scott (2005, p. 18):

[...] nesses momentos – quando exclusões são legitimadas por diferenças de grupo, quando hierarquias econômicas e sociais favorecem certos grupos em detrimento de outros, quando um conjunto de características biológicas ou religiosas ou étnicas ou culturais é valorizado em relação a outros – que a tensão entre indivíduos e grupos emerge. Indivíduos para os quais as identidades de grupo eram simplesmente dimensões de uma individualidade multifacetada descobrem-se totalmente determinados por um único elemento: a identidade religiosa, étnica, racial ou de gênero.

Emerge, portanto, em contraposição e complementação a esse cenário, a compreensão de um direito, igualmente fundamental, à diferença, à diversidade, ao lado do direito à igualdade. Isso não se trata de um esforço constitucional por homogeneização social. Muito pelo contrário: o princípio da igualdade busca o respeito entre sistemas sociais de modo recíproco e com simetria às diferenças (Neves, 2008, p. 167; Luhmann, 2016, p. 93). Assim, o princípio da igualdade se revela como uma moeda, cujos lados apresentam a neutralização da desigualdade no acesso ao direito e a manutenção de uma esfera pública plural (Neves, 2008, p. 170).

Assim, enquanto o princípio da igualdade se preocupa em formular uma compreensão do conteúdo e a extensão dessa cláusula constitucional, o conceito de discriminação “aponta para a reprovação jurídica das violações do princípio isonômico, atentando para os prejuízos experimentados pelos destinatários de tratamentos desiguais” (Rios, 2008, p. 19). Ressalta-se, portanto, a compreensão da igualdade material – não como substitutivo, mas como complementação à igualdade formal –, correspondente tanto ao ideal de justiça social e distributiva, quanto ao reconhecimento de identidades, como necessária para que a própria noção de igualdade não se corrompa e justifique ofensas graves à dignidade humana.

Destaque-se, desde já, que as discriminações não são, *per se*, ruins ou proibidas, uma vez que elas, em diversos casos, fazem-se necessárias para a adequada realização do princípio da igualdade. Existe uma assimetria entre a norma de tratamento igual (tratar os iguais de forma igual) e a norma de

tratamento desigual (tratar os desiguais de forma desigual), cuja consequência é a possibilidade de compreender o princípio da igualdade como norma que, *prima facie*, exige um tratamento igual e que permite um tratamento desigual apenas se isso for justificado por princípios contrapostos (Alexy, 2017, p. 411). Busca-se, com as normas antidiscriminatórias, a eliminação de discriminações que mitiguem, eliminem ou prejudiquem, em alguma medida, o exercício de direitos fundamentais de modo injusto.

Sendo assim, o direito fundamental à não discriminação decorre da compreensão do princípio da igualdade, visto sob a perspectiva substancial, e não apenas pela compreensão de uma igualdade formal perante a lei. Essa compreensão substancial destaca uma dimensão concreta do princípio da igualdade – pois concebe os indivíduos dentro de uma realidade social contingente, e não apenas como sujeitos abstratamente considerados aos quais e para os quais deve ser reconhecido o mesmo valor e direcionado um mesmo tratamento –, e se traduz na busca da inclusão social daqueles sujeitos excluídos e subordinados em razão de certas características, como cor, gênero, orientação sexual, dentre outras. Isto é, verifica-se com o direito à não discriminação uma dimensão emancipatória do princípio da igualdade, na medida em que se pretende a eliminação de padrões de exclusão social (Moreira, 2017).

Nesse sentido, o Direito, cuja função vai para além de uma mera estabilização das expectativas normativas, vincula-se a uma pretensão de correção (Alexy, 2007) e, portanto, deve proteger tais características, já que é em razão delas que indivíduos são impedidos de se desenvolverem na vida social⁵ (Moreira, 2017, p. 49). Nesse sentido, visualiza-se a igualdade substancial como uma proibição de discriminação. O princípio da igualdade, por conseguinte, só pode ser um princípio jurídico eficaz na medida em que as instituições sociais identificam os processos responsáveis pelas diferentes formas de discriminação⁶ (Moreira, 2017).

5 Para maior compreensão do argumento, ver os seguintes casos: *Brown v. Board of Education* (EUA), no qual escolas impediam o acesso de crianças em razão da cor de sua pele, dividindo escolas de pessoas brancas e escolas de pessoas negras. No âmbito brasileiro, destaque-se o caso de Simone André Diniz (posteriormente levado à Comissão Interamericana de Direitos Humanos – OEA), a qual foi impedida de participar de seleção para uma vaga de empregada doméstica em razão da sua cor de pele, pois a divulgação da vaga explicitava que a pessoa a ser empregada deveria, preferencialmente, ser branca.

6 A capacidade que alguém tem para se desenvolver, como argumentam Grant-Thomas e Powell (2006), depende diretamente do acesso à oportunidade na medida em que as oportunidades são produzidas e reguladas como fruto da interação entre instituições e indivíduos – isto é, o modo como as instituições lidam com as contingências naturais e sociais –, de modo a fornecer e negar conjuntamente o acesso com base em cor, gênero, classe social e outros marcadores de diferença social (Rawls, 2008; Grant-Thomas; Powell,

Embora, como afirma Moreira (2017), haja diversas normas jurídicas que pretendem prevenir e/ou punir práticas discriminatórias, ainda persiste uma semântica discriminatória, decorrente, em grande parte, de uma discriminação institucional (Pager; Shepherd, 2008) e, mais profundamente, de uma discriminação estrutural (Almeida, 2020). Nas palavras de Barocas e Selbst (2016, p. 673-674 – tradução livre), “[i]nconscientes, preconceitos implícitos e inércia nas instituições da sociedade, em vez de escolhas intencionais, são responsáveis por grande parte dos efeitos discriminatórios observados”.

Além disso, muitas dessas normas partem de uma compreensão da discriminação apenas como discriminação direta, isto é, implicam os conceitos de intencionalidade e arbitrariedade (Moreira, 2017, p. 17). Todavia, é necessária também uma perspectiva institucional e estrutural da discriminação, uma vez que se preocupa com a gênese da discriminação não intencional levadas a cabo por indivíduos, grupos e organizações – fenômeno inadequadamente respondido por abordagens que enfatizam a intencionalidade como fator determinante para a constatação de práticas discriminatórias, denominadas como discriminação indireta (Rios, 2008). A ideia de discriminação engloba como prática qualquer distinção, exclusão, restrição ou preferência, cuja finalidade seja impedir que os indivíduos desfrutem, em pé de igualdade, de direitos fundamentais, ainda que referidas ocorram despidas de intencionalidade. Ou seja, a discriminação está não só na causalidade, mas também nas correlações (Tischbirek, 2019).

Em uma perspectiva do princípio da igualdade, pode-se entender que, quanto mais se sedimenta a discriminação social negativa – historicamente obstáculo ao reconhecimento de determinados direitos a certos grupos sociais –, mais se valida a discriminação positiva de direitos, orientando o princípio da igualdade à realidade social (Neves, 2008, p. 174). Não é só, todavia, um compromisso de o Estado Democrático de Direito incluir pelo princípio da igualdade. A esfera pública – com a mudança dos paradigmas dos meios de comunicação, isto é, tornando a internet o principal meio de comunicação da sociedade – passa a ser mediada não somente pela imprensa tradicional como também por atores privados que administram intermediários e aplicativos da internet, o que os vincula a um compromisso de

2006). Em razão disso, é preciso “pensar a igualdade jurídica como a igualação jurídica [...] a igualdade posta em movimento, em processo de realização permanente; a igualdade provocada pelo Direito segundo um sentido próprio a ela atribuído pela sociedade” (Rocha, 1996, p. 287).

inclusão pela administração de sua normatividade técnica⁷. Assim, a garantia da igualdade pela heterogeneidade é um compromisso mais amplo que inclui e vincula esses agentes privados e sua programação de algoritmos.

A (DES)IGUALDADE POR ALGORITMOS E PELA INTELIGÊNCIA ARTIFICIAL

A busca por um termo por meio de um motor de busca, a indicação de amigos em redes sociais e o aparecimento de notícias direcionadas ao interesse do usuário são típicos resultados de um processo algorítmico. O termo algoritmo tem entrado crescentemente em análises sociais, especialmente políticas e jurídicas, seja pelos crescentes usos de serviços de busca na internet, seja pelos riscos da perda de privacidade e pelo crescente poder de informação de atores privados. Sob o risco de se tornar um termo performativo para os juristas, os algoritmos precisam de uma definição, quando não, ao menos o entendimento de sua função na internet. Quando se trata de algoritmos neste artigo, refere-se a “um procedimento de tomada de decisão e tecnologia da informação digital”, de modo que se diferenciam algoritmos de “avaliação dos dados, geração de conhecimento e, portanto, determinação dos critérios de decisão e os critérios para a decisão em si” (Ernst, 2017, p. 1026 – tradução livre). A partir dessa definição, é possível entender que os algoritmos são um sistema de execução de uma programação de código digital, que pode ser determinada por homens e mulheres, assim como por inteligência artificial, isto é, independente de uma programação humana (Hoffmann-Riem, 2017, p. 3-4). Por sua forma adaptativa, os algoritmos contribuem na individualização e precisão em sistemas de busca e sugestão de decisões a partir dos dados indicados e pelo histórico existente do usuário⁸.

Algoritmos “são mecanismos de solução de problemas”, sendo “essencialmente definida pela atribuição automatizada de relevância a certas informações selecionadas” (Just; Latzer, 2016, p. 2). Há vários tipos de algoritmos, tais que vão desde os mais conhecidos mecanismos de busca, como o Google Search, passando pelos de produção de conteúdo e de produção

7 Mais à frente será tratada a normatividade como marcação positiva de uma possibilidade. Nesse caso, a normatividade técnica está demarcada positivamente por uma tipificação de padrões técnicos, normas-artefatos, indicadores de comportamento etc. Para mais detalhes, confira: ARTOSI, Alberto. Technical normativity. In: *Italian Philosophy of Technology*, Springer, Cham, p. 149-160, 2021.

8 Tamanha é a influência dos algoritmos, que, segundo Luciano Floridi (2015), não vivemos em um mundo *online* ou *offline*, mas sim *Onlife*. Isso significa que não estão claros os limites do que está ou não no mundo digital e fora dele.

no âmbito jornalístico (ex.: Quill e Quakebot), até os de recomendações (ex.: Amazon e Netflix). Independentemente da finalidade oferecida pelos programas, eles “geram, coletam, processam e agregam numerosos dados por meio de métodos algorítmicos de modo a extrair valores econômicos disso” (Just, 2018, p. 3). Assim, as informações, os comportamentos e as comunicações se individualizam conforme o uso do usuário.

A solicitação do usuário mais as características disponíveis dele na plataforma fazem os algoritmos selecionar elementos de um conjunto de dados básico (*Big Data Set 1*), classificando e gerando relevância contextualizada sob uma avaliação estatística automatizada: o modelo *input-output* de uso dos algoritmos possui três fases: (1) a inserção de dados (*input*), (2) o tratamento e (3) a emissão de dados (*output*). Ao lado da inserção de uma requisição por parte do usuário, o programa já possui uma base de dados que servirá para auxiliar na precisão máxima do requisitado com base em outros dados do usuário. Todo resultado retroalimenta o sistema para uma nova busca: a base de dados do usuário e a requisição (1) servirá de fonte para a seleção do algoritmo de escolha que encaminhará a uma base de dados reduzida, que seguirá para o algoritmo de classificação (2). Este encaminhará a base de dados de saída, que, por fim, servirá para a utilização dos dados de emissão (3)⁹.

O que foi explicado acima é uma forma geral da lógica de tratamento de dados por meio de algoritmos. Nem todo tratamento de dados, porém, se refere à inteligência artificial (IA). A despeito de o conceito de inteligência artificial ter se popularizado nos últimos anos, principalmente em razão do avanço e dos usos no dia a dia, seu conceito, ou pelo menos a ideia por trás, é mais antiga e discutida na literatura por várias décadas (Russel; Norvig, 2016), alternando momentos de grande progresso e euforia com momentos de frustrações e esquecimentos (Maini; Sabri, 2017). As pesquisas sobre esse assunto, conforme a maioria dos estudiosos, são inauguradas pelo artigo *Computing machinery and intelligence*, publicado na revista *Mind* em outubro de 1950 pelo matemático britânico Alan Turing, considerado o pai da computação moderna.

9 É possível encontrar um quadro explicativo do funcionamento desse sistema em (Schulz; Dankert, 2016, p. 65); para exemplos, cf. GRIMMELMANN, James. The structure of search engine law. *93 Iowa L. Rev.*, 1 2007-2008, s. 7-11; SAURWEIN, Florian; JUST, Natascha; LATZER, Michael (2017). Algorithmische Selektion im Internet: Risiken und Governance automatisierter Auswahlprozesse. *Kommunikation@gesellschaft*, 18:online., s. 1-2.

O termo inteligência artificial é cunhado pelo *Dartmouth Summer Research Project on Artificial Intelligence*, organizado por, dentre outros, J. McCarthy, M. L. Minsky e C. Shannon (1955). Com isso, a IA passa a ser um campo de estudo da Ciência da Computação que parte da ideia de que os aspectos da inteligência humana, sobretudo o aprendizado e a resolução de problemas, podem ser precisamente descritos ao ponto de uma máquina ser capaz de simular tais características, tendo como problema central “fazer uma máquina se comportar de maneiras que poderiam ser chamadas de inteligentes se um ser humano assim se comportasse” (McCarthy *et al.*, 1955, p. 11 – tradução livre)¹⁰. Enquanto aplicação prática, IA pode ser definida como o conjunto de técnicas, métodos e sistemas que objetivam fazer um computador exibir comportamentos inteligentes, sobretudo o aprendizado, a partir da coleta e interpretação de dados externos, a fim de atingir objetivos e tarefas específicas por meio de uma adaptação flexível (Kaplan; Haenlein, 2018).

Dentre as diversas técnicas abarcadas pelo conceito de IA, aquela que é responsável por uma grande parcela dos avanços desse campo nos últimos e também aquela em que são concentrados os maiores esforços (Cortiz, 2020, *online*) é a área de *machine learning*. O desenvolvimento e aperfeiçoamento dos processadores, a crescente disponibilidade de dados¹¹ e a possibilidade de coletá-los e armazená-los, associada ao desenvolvimento tecnológico e maior uso da internet (Goldschmidt; Passos, 2005), permitiram que muitas pesquisas de cunho teórico desenvolvidas desde a década de 1950 (Peixoto; Silva, 2019, p. 87) fossem postas em prática, mediante o desenvolvimento de algoritmos de aprendizado de máquina.

Kevin Murphy define *machine learning* (2012, p. 1) como conjunto de métodos que podem automaticamente detectar padrões em dados, e então usá-los para prever dados futuros ou desempenhar outras formas de tomada de decisão. Em razão disso, os mecanismos decisórios deixaram de ser exclusividade de seres humanos e passaram a ser delegados aos algoritmos aprendizes, uma vez que eles são mais eficientes e eficazes que os humanos em determinadas atividades, como análise de dados e tarefas preditivas. Os algoritmos inteligentes, desenvolvidos pela área de *machine*

10 O termo, contudo, sofre críticas, tais como as de Elena Espósito (2017), que, sob uma perspectiva da teoria dos sistemas, trata o termo como “comunicação artificial”.

11 Estima-se que são gerados cerca de 2,5 quintilhões de bytes por dia, e essa quantidade tende a aumentar ainda mais (Marr, 2018).

learning, adquirem certo grau de autonomia, na medida em que aperfeiçoam sua *performance* em determinada tarefa, ou seja, aprendem (Mitchell, 1997, p. 3)¹².

Entretanto, a despeito desses benefícios, durante a construção, a programação e o treinamento de algoritmos, é possível que o processo apresente resultados desproporcionalmente adversos, concentrados em grupos historicamente desfavorecidos¹³ (Barocas; Selbst, 2016, p. 673). Esses algoritmos apresentam essas tendências, ainda que não sejam programados para isso, tendo em vista a capacidade de aprendizado autônomo desses sistemas. Como ressaltam Barocas e Selbst (2016), a discriminação pode ser um artefato próprio do processo de aprendizagem dos algoritmos, em vez de um resultado decorrente do fato de os desenvolvedores atribuírem a determinados fatores pesos inadequados. Isso revela que a exclusão de determinados grupos pode e é incorporada pelo processo técnico que constitui a programação algorítmica, de modo que tais sistemas não são neutros ou mais precisos o quanto pretendem ou como advogam aqueles que defendem o seu uso irrestrito.

A programação de algoritmos não inteligentes pode ser direcionada propositalmente para a discriminação de grupos ou possuir algum processo negligente em não adequar seus dados para evitar efeitos discriminatórios. É na IA, todavia, que reside maior complexidade sobre o tema. Solon Barocas e Andrew D. Selbst (2016) elencam alguns mecanismos da construção técnica dos algoritmos de IA que podem redundar em efeitos discriminatórios. O primeiro diz respeito ao desenho do modelo (ou *design*). Com o advento e a modernização das técnicas de *machine learning*, tornou-se possível a criação de algoritmos aprendizes, algoritmos que não precisam de instruções completas e exaustivas para que ele possa executar uma tarefa, pois eles apresentam uma capacidade de aprendizado autônomo a partir dos dados externos. Esses algoritmos, conforme explica Pedro Domingos (2017),

12 Em razão disso, o setor privado utiliza massivamente esses mecanismos visando, pelo menos, a quatro finalidades mercadológicas principais: i) previsibilidade e diminuição de riscos inerentes à atividade empresarial; ii) interação direta com o consumidor, de modo a prover as necessidades individuais de cada um e assim conseguir captar e fidelizar a maior quantidade de clientes possível; iii) diferenciação de produtos e iv) de serviços, diferente do modelo anterior, que direcionava o mesmo produto/serviço a várias pessoas (Mendes, 2014).

13 Exemplos claros disso são os casos destacados por Ryan Calo (2017, p. 9): um sistema de tradução instantânea que associa a profissão de engenheiro como masculino e a de enfermeira como feminino (mesmo que o idioma traduzido não faça distinção de gênero) e uma câmera que se recusou a fotografar uma blogueira de Taiwan por “acreditar” que ela estava de olhos fechados.

criam novos algoritmos, os modelos. Mas, para criar esse modelo, é preciso que seja especificado para o computador, na forma de um algoritmo de treinamento, o que ele precisa aprender e a partir do que ele irá aprender. Os desenvolvedores, portanto, devem traduzir um certo problema amorfo em uma pergunta que possa ser expressa em termos mais formais de modo que os computadores possam analisar (Barocas; Selbst, 2016, p. 678). Essa tarefa de tradução é essencialmente subjetiva, na medida em que exige a representação, em linguagem que o computador possa compreender, de um determinado problema.

Se o desenvolvedor rotula um conjunto de dados, os quais são usados para treinar o modelo, como “bom pagador” e outro como “mau pagador” e faz isso com base em critérios próprios, ou seja, naquilo que ele acredita ser a definição mais adequada de “bom/mau pagador”, ele estará expondo ao modelo um conjunto de dados cujos rótulos refletem uma perspectiva específica e subjetiva sobre um fenômeno maior e o sistema irá inferir desse conjunto de dados um certo padrão, o qual será aplicado para outros dados, distintos daqueles do *dataset training*. Essas novas classes criadas pelo desenvolvedor apresentam alguns pontos cegos (O’Neil, 2016), em razão da limitação epistêmica – ou seja, a definição do que é bom adotada refletirá apenas um aspecto dentro de um amplo espectro de possibilidades –, os quais poderão ser irrelevantes ou, ao contrário, apresentar efeitos nefastos.

O segundo problema diz respeito à seleção/tratamento dos dados utilizados pelos algoritmos de treinamento para treinarem os modelos que serão exportados. A maioria dos problemas em que são utilizados algoritmos aprendizes depende, para a adequada solução do problema, de uma grande base de dados da qual o modelo irá extrair os padrões acerca daquele fenômeno objeto da previsão, isto é, os exemplos por meio dos quais o modelo será treinado para agir de determinada maneira. Como destacam os cientistas da computação: dados tendenciosos (ou enviesados) levam a modelos discriminatórios (Custers, 2013).

Se a aprendizagem de máquina trata determinados casos nos quais o preconceito desempenhou algum papel como exemplos válidos para aprender – por exemplo, se o banco de dados utilizado foi construído com base em decisões humanas diretamente influenciadas por preconceitos, como a contratação mais frequente de homens brancos e alta rotatividade de pessoas pertencentes a grupos minoritários –, essa regra pode simplesmente reproduzir o preconceito envolvido nesses casos anteriores (Barocas; Selbst, 2016, p. 680-681). Em poucas palavras, os dados refletem situações passa-

das e, ao serem incorporados nos algoritmos de tomada de decisão, definirão comportamentos futuros.

Há ainda, conforme os autores, um outro problema decorrente do uso de dados enviesados: se o processo de aprendizagem de máquina extrai certas inferências de uma amostra tendenciosa da população, qualquer decisão que se apoie nessas inferências pode sistematicamente prejudicar aqueles que estão sub ou super-representados (ou sub- ou sobreintegrado¹⁴) no conjunto de dados. O problema aqui reside no processo de coleta de dados, ou seja, não decorre das intenções que motivaram a tomada de decisão e que gerou aquele *dataset*, mas a capacidade que aquela base de dados tem de representar adequadamente determinados grupos sociais. No primeiro caso, tem-se um banco de dados alimentado por dados tendenciosos, ou seja, que foram gerados com base em preconceitos; ao passo que, no segundo, há um *dataset* com sub ou super-representação de certos grupos, o qual, de igual modo, pode conduzir a tomada de decisões discriminatórias.

Jonas Lerman (2013), nesse sentido, revela uma preocupação particular acerca do fenômeno chamado por ele de *big data's exclusion*, consistente na (sub)representação de certos grupos nos *datasets* com base nos quais grandes empresas e setores governamentais tomam decisões estratégicas por meio de algoritmos de aprendizagem de máquinas. A preocupação refere-se “a omissão sistêmica e não aleatória de pessoas que vivem às margens do *big data*, seja devido à pobreza, geografia ou estilo de vida, e cujas vidas são menos ‘datadas’ do que as da população em geral” (Lerman, 2013, p. 57 – tradução livre). A marginalização em setores-chave corre o risco de distorcer conjuntos de dados e, por conseguinte, distorcer as decisões tomadas com base naqueles dados.

Por fim, esses algoritmos se utilizam e, em certa medida, dependem de *proxies*. Uma variável *proxy* descreve algo que provavelmente

14 O termo se encaixa na ideia de Marcelo Neves a respeito da ausência de cidadania diante de exclusão social generalizada (ou ausência de acesso aos direitos disponíveis na ordem jurídica). A exclusão social não permite que pessoas estejam em condições de igualdade legal, isto é, ou estão acima da lei, como os sobreintegrados (conhecem os benefícios, mas não cumprem seus deveres legais) e os subintegrados (conhecem só a esfera punitiva da lei, mas não os seus direitos fundamentais). Muito embora as terminologias usadas não guardem perfeita correspondência, elas convergem na ausência de critérios justos e adequados; no caso da representação de dados em um *dataset training*, as razões pelas quais determinados grupos são mal representados decorrem de uma construção desigual, levada a cabo ao longo dos anos. Essa mesma construção social determina e cria as classificações apresentadas por Neves, isto é, grupos privilegiados que se desigualam por receberem os benefícios da lei, mas desconhecem os deveres, e grupos que só conhecem os deveres (às vezes, aplicados de maneira desproporcional), mas não são destinatários dos benefícios. Vide Neves, 1994.

não é, por si só, muito interessante para o sistema; entretanto, a partir dela, pode se chegar a uma variável relevante e necessária para o algoritmo (Calders; Žliobaitė, 2013, p. 52-53). Como explica O’Neil (2016), às vezes o algoritmo não terá disponível determinado dado que seja de extrema relevância para o aprendizado de máquina; todavia, esses dados podem ser substituídos pelas variáveis *proxies*, as quais estabelecem correlações estatísticas com as variáveis de interesse.

Dessa forma, por mais que dados relativos aos critérios de discriminação proibidos, como cor, gênero, orientação sexual, dentre outros, sejam “ocultados” do sistema, isto é, não sejam utilizados como atributos para o treinamento do modelo, o algoritmo, exatamente em razão dessas variáveis, poderá inferir estatisticamente que um determinado atributo mantém forte correlação com outro ao ponto de tal correlação ser crucial para o desempenho do sistema. Esse problema deriva daquilo que se chama de “codificações redundantes”, que se refere ao caso em que a associação a um grupo minoritário passa a ser codificada em outros dados (Barocas; Selbst, 2016).

Se, por exemplo, um determinado critério de discriminação proibido, como cor ou gênero, possuir grande correlação com algum atributo, com alguma característica que se pretende buscar ou prever, é uma consequência natural que membros daqueles grupos raciais ou de gênero sejam afetados de maneira desproporcional. Utilize-se, para exemplificar esse ponto, o caso em que a contratação de pessoas é feita por algoritmos de aprendizado de máquina e um dos critérios utilizados para selecionar o melhor candidato seja o nível de escolaridade e a instituição de ensino que o candidato frequentou. Caso as melhores instituições (assim consideradas pelo desenvolvedor do sistema) possuam um elevado número de estudantes brancos e homens, então ocorrerá uma codificação redundante, ou seja, a cor e o gênero serão codificados dentro do critério escolaridade.

Percebe-se que os algoritmos e seus dados carregam não só preconceitos, mas também normatividades sociais, o que inclui o próprio direito. Os preconceitos “encapsulados” nos dados estão na dinâmica cultural de um grupo social. O direito também está na cultura. Para se entender o lugar do direito nessa (tensa) relação com os dados enviesados, precisa-se compreender o direito para além da sua noção institucionalizada: o direito como semântica social ou como cultura.

DIREITO COMO CULTURA

O uso de um programa executado por algoritmos produz uma relação colaborativa entre programação e usuários, o que permite a adaptação e melhora da experiência de uso do aplicativo ou da plataforma. Mesmo com a programação para o controle automático, ainda há espaço de participação das comunidades, que é importante para melhorar a adequação da plataforma: avaliação e reputação dos usuários e aplicativos, por exemplo, que contribuem para o melhor controle da programação e – logo – das regras estabelecidas na plataforma¹⁵. A comunidade serve como uma espécie de “*Gedächtnisorganisation*”¹⁶, em que mudanças aparecem por meio de sugestões ou reclamações, consolidação de padrões técnicos e regras sociais de condutas naquele aplicativo.

Há uma linha tênue entre o aspecto binário e técnico de um algoritmo e as normas sociais. Essa linha é determinada pela programação que trabalha com códigos técnicos, que não se reduzem somente a estabilizar expectativas sociais, mas também garantem sua variabilidade e flexibilidade de conexões entre diferentes tecnologias (Vesting, 2004, p. 664). A resposta para problemas, assim, não é meramente por um processo regulatório de estabilização de expectativas normativas, mas adaptativa pelas indagações de usuários.

A programação passa, portanto, por uma influência da dimensão cultural do Direito: o código técnico não é isolado ou neutro, pois interage com outras estruturas normativas. O olhar da comunidade de usuários e do programador são fundamentais na construção de uma normatividade que está em torno da função de uso do programa. A observação do direito não parte de uma exclusiva e imediata incidência de alguma ordem jurídica localizada. Há, ao contrário, um controle próprio que incorpora diversos conteúdos jurídicos, observando diversas ordens jurídicas e as interações de sua comunidade de usuários. Tal controle passa por uma percepção própria do direito que não se encontra na diferenciação funcional e de papéis e na normatividade sistêmica. Mais que isso: o direito faz parte de uma semân-

15 Sobre comunidades *online*, veja AURAY, Nicolas. Online communities and governance mechanisms. In: BROUSSEAU, Eric; MARZOUKI, Meryem; MÉADEL, Cécile (Hrsg.). *Governance, regulations and powers on the internet*. Cambridge University Press: Cambridge, 2012, ss. 211-231.

16 Em sentido semelhante, LANGE, Andreas. Die Gaming-Community als Pionier der digitalen Bewahrung. In: KLIMPEL, Paul; KEIPER, Jürgen (Hrsg.). *Was bleibt? Nachhaltigkeit der Kultur in der digitalen Welt*. Berlin: iRights.Media, 2013, ss. 110-117.

tica social dominada também por atores externos de suas práticas formais (Rosen, 2006, p. 6-7).

Os direitos tradicionais, antes concentrados na normatividade das ordens jurídicas estatais, não se encontram somente diluídos em outras ordens jurídicas (Neves, 2009; Teubner, 2003), mas também como algo que está na semântica de práticas sociais, isto é, na cultura, que é incorporada por determinadas práticas técnicas. Portanto, percebe-se que a presença de conteúdos jurídicos normativos, cuja lógica estrutura e organiza relações mapeadas (Ladeur, 2016, p. 284), apresenta-se como elemento da cultura¹⁷.

A cultura é um sistema de imunidade ou identidade de grupos, que querem se orientar em valores comuns, experiências, expectativas e interpretações, que forma o sentido simbólico, a imagem do mundo de uma sociedade (Assmann, 1992, p. 140). Segundo Vesting (2015, p. 21), “[a] cultura então também inclui a ordem simbólica e imaginária que determina como as pessoas percebem, sentem, falam, pensam e agem no mundo e dão sentido e significado à vida”¹⁸. Mesmo com especialização intensa, o direito é também elemento cultural, pois não é estrutura isolada da sociedade (Rosen, 2006, p. 4-5)¹⁹. O direito faz parte da imaginação social, que ajuda a compreender o mundo (Rosen, 2006, p. 12). Enquanto cultura, nota-se que o direito é conhecimento local – incluindo plataformas, aplicativos e programas mediados por algoritmos em geral –, como “caracterização vernacular” conectado a “convencimentos vernaculares”, que se caracterizariam como “sensibilidade legal” (Geertz, 1983, p. 215). Se, em torno do uso dos programas mediados por algoritmos, existem regras, comunidade de usuário e programador, pode-se entender que o programa é um “lugar”, pois isso possibilita visualizar a presença do direito como cultura em comunidades que não possuem um lugar concreto ou um longo histórico de existência.

17 O trabalho pretende colocar o Direito também como prática social. Há trabalhos diferentes (cf. NADER, Laura (Ed.). *Law in culture and society: with a new preface*. Univ. of California Press, 1997) que colocam as mudanças conceituais da noção de Direito em diferentes contextos. Essa noção de cultura e Direito não impede um enfoque diferente: ao invés de Direito *na* cultura, tratamos de Direito *como* elemento da cultura, de modo que ele pode ser operacionalizado fora das instâncias tradicionais e dos papéis sociais diferenciados (juiz, promotor, advogados etc.). Assim, a percepção social de quem pode decidir, crendo normatizar com valor jurídico, torna-se importante, tal como um usuário, programador ou controlador de dados ao interpretar “liberdade de expressão”, por exemplo.

18 No mesmo sentido, GEERTZ, C. *Dichte Beschreibung*, 1983, 46; K.-H. Kohl, *Ethnologie*, 2000, 167.

19 Em sentido semelhante, GEERTZ, Clifford. *The interpretation of cultures*. New York: Basic Books, 1973, s. 5-6.

A SEMÂNTICA JURÍDICA NA COMUNIDADE DE USUÁRIOS E NO PROGRAMADOR: UMA RESPOSTA PARA O PROBLEMA

As conexões entre saber técnico e normatividade jurídica não seguem uma lógica hierárquica, mas se constroem numa rede heterárquica, que permite novas conexões (Ladeur, 2016, p. 25). Assim, não só as ordens jurídicas por meio de suas normas recebem o saber técnico: este também absorve a normatividade do direito. No contexto de transformação digital, verifica-se “um recuo do direito estabelecido pelo Estado como meio de estruturar situações da vida” e, ao mesmo tempo, “o deslocamento da responsabilidade para portadores privados nas esferas determinadas pela digitalização” (Hoffmann-Riem, 2019, p. 18). Isso permite que apareçam novos atores cognitivos, que agem como marcadores do saber e completam as formas de autocoordenação espontânea (Ladeur, 2016, p. 38). Normas e práticas não estatais coordenam cada vez mais atores da rede, sendo o direito e suas normas dissolvidas sob a descrição da elaboração de padrões técnicos (Ladeur, 2016, p. 40). O direito na amplitude de sua textura aberta permite a capacidade de adequação sob as condições de transformações contínuas da sociedade, que pode observar a ligação das instituições e dos institutos jurídicos e suas formas sociais de coordenação (Ladeur, 2016, p. 55). Desse modo, mais do que um senso de direitos e deveres, o direito se torna uma condicionante e estimulante para que algo novo possa ser produzido na prática (Ladeur, 2016, p. 56).

Uma capacidade de aprendizado e flexibilidade – tanto de ordens jurídicas quanto de práticas técnicas (inclusive a programação de algoritmos) – é influenciada por conexões horizontais e de compatibilização de expectativas (Ladeur, 1990, p. 142-143). Não é por outro motivo que se precisa admitir a coexistência do direito com outros “códigos” na internet: “No ciberespaço, devemos entender como um ‘código’ diferente regula: como o *software* e o *hardware* (ou seja, o ‘código’ do ciberespaço) que faz do ciberespaço o que ele é, também regula o ciberespaço como ele é” (Lessig, 2006, p. 5 – tradução livre). Contudo, esses códigos técnicos vão observar ao seu modo o normativo do direito. Conteúdos como “pornografia infantil”, “desrespeito à privacidade” e “discursos de ódio” são, em vários momentos, incorporados na programação como uma observação geral de várias ordens jurídicas, como se a própria programação antecipasse problemas e os evitasse simultaneamente em uma pré-seletividade. Não há uma negação, portanto, da normatividade jurídica, mas um entrelaçamento

entre esta e os códigos de programação, a partir do que a programação lê do conteúdo normativo.

Neste trabalho, o sentido de norma é entendido em sua extensão contrafactual, isto é, como demarcação positiva de uma possibilidade (Möllers, 2015, p. 125). Toda norma possui uma significação autoral que a demarca como símbolo de algo ou alguém (autoridade). Ela não pode se identificar com a realidade, pois assim perderia a condição de possibilidade, já que sempre seria realizada. Não pode também não possuir uma dimensão completamente irrealizável, dado que não a confirmaria também como possibilidade. Ou seja, o cumprimento e o desvio compõem a noção de normatividade. “A norma pode ter sua dimensão textual, representando a autorização normal [e a formalização]” (Möllers, 2015, p. 286), que está associada à aplicação. Esta é a realização concreta da própria norma (no caso em espécie), criando outra norma (Möllers, 2015, p. 182-183). Tanto na dimensão textual quanto na aplicação, há a normatividade, ao contrário da execução, que é “uma causa que permite apenas um efeito” (Möllers, 2015, p. 183), não tendo a possibilidade do desvio.

No caso da normatividade do uso de algoritmos, os termos de uso e códigos de conduta de um programa têm uma função de *pré-seletividade* na programação de algoritmos. Pode ser feito um paralelo com o texto legal: a pré-seletividade diminui o campo de possibilidades de ação, mas que pode ou não ser cumprido pelo uso do programa. Alguém que, mesmo sabendo que é proibido publicar vídeo com nudez em uma plataforma, pode utilizar de meios não captáveis pelo algoritmo programado e desviar dos termos de uso daquela plataforma. O uso se aproxima da ideia de aplicação da norma, pois, ainda que não perceba, o usuário faz um uso – em termos concretos – do programa carregado de sua pré-seletividade. Ambos, assim, são demarcados positivamente por uma possibilidade.

Por outro lado, a execução dos algoritmos em si tem a função de apenas de obedecer a um comando ou uso dado pelo usuário, que já é filtrado pelas normas de conduta e pela programação feita pelo responsável, não havendo, assim, a possibilidade de desvio ou frustração do caminho normativo. Não há normatividade, portanto, exclusivamente no funcionamento do algoritmo, pois não resta para ele a possibilidade de cumprir ou descumprir, mas apenas seguir a instrução dada.

Por essa estrutura de descrição normativa, o algoritmo pode discriminar, mas também pode incluir. Entendendo que a ponte para uma constru-

ção inclusiva passa pelo uso do programa pela comunidade de usuários e pelo programador, fica mais fácil entender o lugar de como o direito na sua função semântica pode atuar. Os usuários servem como atores de atualização da programação algorítmica: ao notar um resultado discriminatório, os usuários podem contextualizá-lo, denunciando à administração do programa para corrigir o filtro dos dados. Exemplo disso é o caso recente, ocorrido em setembro de 2020, na rede social Twitter, em que se percebeu um comportamento discriminatório na apresentação de imagens. O algoritmo dessa rede parecia “ter uma preferência” por pessoas brancas em detrimento de pessoas negras, uma vez que, independentemente do posicionamento das pessoas numa foto, o algoritmo recortava e apresentava, como pré-exibição da imagem, a parte em que estava a pessoa branca. Em nota, o porta-voz da rede social informou que iria continuar com mais análises para remover do algoritmo vieses raciais, além de abrir o código-fonte para outras pessoas revisá-lo e replicá-lo²⁰ (Hern, 2020, *online*).

Isso evidencia uma exigência que a programação observe processos de autorregulação, os quais precisam ser claros de modo que possam ser controlados por seus usuários. Essa limitação parece ser influenciada por uma dimensão cultural do direito, uma vez que a codificação técnica não se encontra isolada do meio em que se processa; ao contrário, mantém com ele uma imbricada relação, sobretudo com a construção dos algoritmos analisados neste trabalho, em razão da relevância dos dados externos e, além disso, na utilização desses algoritmos na execução de atividades sociais.

No que tange ao programador, há duas considerações: a formação dos programadores deve possuir um ensino também de ordem social, que leve em consideração a complexidade social e a responsabilidade de atuação dos programadores. Mais do que ele entender de direito e da ordem jurídica a que está vinculado, é entender que a programação é uma atividade que carrega normatividades que devem ser inclusivas. Outra consideração passa pelo entendimento de que, por meio de um programa, não se criam somente tendências de uso, mas também de normas da comunidade de usuários por meio de seus usos. Observar como essa normatividade pode

20 Nas palavras do porta-voz: “Our team did test for bias before shipping the model and did not find evidence of racial or gender bias in our testing. But it’s clear from these examples that we’ve got more analysis to do. We’ll continue to share what we learn, what actions we take, and will open source our analysis so others can review and replicate”.

ser antecipatória não só de problemas entre usuários ou com outras ordens jurídicas, mas também de problemas de inclusão ou exclusão de usuários, isto é, de problemas discriminatórios. Se, na construção da cidadania, o direito como sistema social deve incluir, o direito como dimensão cultural também não deve escapar disso: o programador e a comunidade de usuários também devem incluir por meio da semântica social do direito.

Dito isso, é possível ensinar tais algoritmos, por meio da regulação dos processos ao longo da programação, a não discriminar, acoplando, em sua programação, os valores incorporados na normatividade jurídica. Essa conclusão é salientada por Kleinberg, Ludwig, Mullainathan e Sustain (2018, p. 114), segundo os quais a proibição de discriminação pode ser implementada pela regulação do processo por meio do qual os algoritmos são construídos.

Ademais, as equipes técnicas, quando conscientes das responsabilidades decorrentes das construções algorítmicas que realizam, podem mitigar a ocorrência da discriminação por meio do reconhecimento de atributos mais capazes de conduzir a tal efeito. Ademais, a criação de times de desenvolvedores mais diversificados pode contribuir para o desenvolvimento de algoritmos mais justos, já que a maioria dos funcionários das grandes empresas de tecnologia são do gênero masculino e brancos (Hao, 2019). Karen Hao aponta que a falta de diversidade nos times de tecnologia se relaciona com os vieses que os algoritmos de IA perpetuam. Esse mesmo problema é apontado no relatório *Discriminating systems: gender, race and power AI*, elaborado em março de 2019 por Sarah M. West, Meredith Whittaker e Kate Crawford. No mesmo sentido, Fei-Fei Li, ex-chefe de IA da Google e professora em Stanford, declara que há muito mais chances de os vieses presentes na sociedade serem transportados para os algoritmos se os times não forem compostos por grupos diversos de engenheiros (Hempel, 2018).

Verificou-se, até aqui, que a discriminação é algo condenável por diversas ordens normativas estatais e até privadas (vide o caso do Twitter exposto acima), mas, a despeito disso, continua a ser um problema persistente. Há grandes problemas em aferir e determinar que uma conduta desse tipo, sobretudo aquelas relacionadas à oferta de emprego ou à demissão por pelo menos duas razões: a primeira, e talvez a mais óbvia, as pessoas costumam mentir acerca das razões que motivaram suas decisões, especialmente quando tais motivações são passíveis de repreensão jurídica e social; em segundo lugar, em certos casos, as pessoas sequer possuem consciência das motivações de seus atos (Kleinberg *et al.*, 2018).

O momento apresenta uma ideia de que os algoritmos já estão enviesados. Todavia, ele pode justamente servir como forma de desviesar as decisões humanas e promover inclusão. Por exemplo, dada a alta subjetividade na tomada de decisão na contratação de uma pessoa, o algoritmo pode já ser programado para ignorar variáveis que não são preditivas de resultado (sobrenome, cor de pele, sexo, orientação política etc.) (Kleinberg *et al.*, 2018). Assim, toma-se como exemplo uma vendedora que – por ser, em uma sociedade estruturalmente patriarcal, potencialmente mais mal avaliada que um vendedor – pode ser protegida pelo critério em que o algoritmo não leva em conta o seu gênero, o que socialmente já estaria possivelmente no cálculo (Kleinberg *et al.*, 2018). É bem verdade que essas medidas não solucionam integralmente o problema, em razão das “codificações redundantes”, apresentadas anteriormente, que conduzem o algoritmo a tomar decisões discriminatórias, a despeito de certas variáveis terem sido eliminadas do *dataset training*. Todavia, diante de um comportamento discriminatório exibido pelo algoritmo – verificado tanto por meio de processos de automonitoramento quanto de supervisão por terceiros –, é possível auditá-lo²¹, reestruturá-lo ou simplesmente abandoná-lo e construir novos modelos, buscando compreender como a discriminação foi incutida naquele sistema, o que não seria possível se a decisão fosse exclusivamente humana.

Em um artigo intitulado *An FDA for algorithms*, Tutt (2016, p. 103) afirma que os seres humanos, assim como os algoritmos de aprendizado de máquina, são um exemplo de sistemas imprevisíveis e inexplicáveis. Embora se criem diversas instituições, dentre as quais o direito, para mudar o comportamento humano e adaptá-lo à vida em sociedade, não há como ter plena certeza, isto é, não é possível prever se o comportamento humano realmente seguirá essas instituições e convenções sociais e, quando não, quais os reais motivos que conduziram àquela fuga das normas estabelecidas. Os algoritmos de aprendizado de máquina, por sua vez, embora também se apresentem como sistemas imprevisíveis e inexplicáveis em certas medidas, podem, como argumentam Kleinberg e outros (2018, p. 116), a partir da estruturação de um modelo de observação, identificação e busca de soluções adequado, pautado em determinados princípios técnicos, ser

21 Algumas diretrizes para auditoria desse tipo de sistema são sugeridas pelos autores Pasquale e Citron (2014) para aperfeiçoar a fiscalização do FTC, as quais se referem basicamente à transparência dos sistemas para facilitar a realização de testes, elaboração de relatórios de impacto e análises de risco, dentre outras. De modo similar, as organizações FAT-ML (*Fairness, Accountability and Transparency in Machine Learning Organization*) e ACM (*Association for Computing Machinery*) estabelecem princípios e diretrizes para criação e implementação de algoritmos livres de discriminação.

auditados. Isto é, é possível fazer ao algoritmo perguntas que não podem ser feitas de maneira significativa aos humanos. Em relação às decisões tomadas por algoritmos, é possível aferir, por meio dos mecanismos referidos anteriormente, quais dados foram usados para treinar o modelo algorítmico e quais não foram, além das razões que levaram o sistema àquela decisão de modo que se torna mais fácil para o sistema jurídico descobrir as motivações do sistema (Kleinberg *et al.*, 2018).

Portanto, a criação e implantação de um modelo de observação, identificação e busca de soluções adequado, baseado no diálogo entre os diversos setores da sociedade interessados na regulação dos algoritmos, não “limita simplesmente a possibilidade de discriminação de algoritmos, mas também tem o potencial de transformar os algoritmos em um poderoso contrapeso à discriminação humana e uma força positiva para o bem social” (Kleinberg *et al.*, 2018, p. 115). Talvez, na medida em que os algoritmos de aprendizado de máquina desvelam problemas profundos de discriminação, também sejam eles capazes de mitigar (para não dizer solucionar) esses mesmos problemas, por meio da disponibilização de ferramentas que outrora não existiam e que permitem uma melhor compreensão do fenômeno discriminatório.

CONCLUSÃO

O presente trabalho buscou não somente identificar os problemas de discriminação por meio de algoritmos e sua relação com o princípio da igualdade, mas também verificar respostas que o direito – para além de sua dimensão sistêmica – pode dar para mediar o problema por outras camadas sociais. Por essa razão, é importante entender que a normatividade do direito está também na comunidade de usuários e nas práticas dos programadores, o que revela a importância de impregnar a semântica jurídica da igualdade na seleção e formação de programadores.

A compreensão do direito como prática cultural, impregnada nas diversas relações sociais, além do entendimento do funcionamento dos algoritmos aprendizes nas dinâmicas sociais, possibilita, senão uma resposta para o problema da regulação algorítmica no que concerne às práticas discriminatórias levadas a cabo por esses sistemas, uma via para construção dessa solução.

Os desenvolvedores de sistemas, bem como a comunidade de usuários desses sistemas, organizadas por meio da internet, podem estabelecer

os marcadores adequados para a normatividade da programação, imbuídos de uma semântica pautada e influenciada pelas normas jurídicas voltadas à efetivação de direitos fundamentais, em especial o direito à não discriminação. Isto é, não só o Estado se apresenta como importante pilar da regulação algorítmica, mediante a formalização e institucionalização do direito, mas também, e sobretudo, outros atores diretamente envolvidos na criação e propagação de mecanismos decisórios autônomos, capazes de estruturar a vida em sociedade através da atuação em rede. A compreensão do direito enquanto expressão cultural, nesse sentido, é essencial para que se construam as pontes necessárias entre as normatividades técnicas e jurídicas visando à efetivação de direitos fundamentais.

O trabalho se encerra, mas não evita outras perguntas: que tipo de aprendizados jurídicos os programadores devem ter em sua formação? Considerando que tais atores possuem notória relevância na incorporação de normatividades na programação, em quais outros aspectos específicos e relacionados ao campo jurídico deve ser pautada a formação desses profissionais e como estabelecer essas diretrizes? Além disso, como o Estado pode regular sobre o tema sem retirar a dinâmica de inovação de atores privados? Como a autorregulação privada de plataformas de internet devem se ocupar sobre o tema? As respostas para essas perguntas exigem outras agendas de pesquisa e novas investigações, razão pela qual este trabalho deve ser visto não como o ponto final que encerra a discussão, mas como um meio que, ao mesmo tempo em que apresenta (possibilidades de) soluções para determinado problema, revela e direciona novas perguntas sobre outros aspectos relevantes.

REFERÊNCIAS

- ALI, Muhammad; SAPIEZYNSKI, Piotr; BOGEN, Miranda; KOROLOVA, Aleksandra; MISLOVE, Alan; RIEKE, Aaron. *Discrimination through optimization. Proceedings of the ACM on human-computer interaction*. [s.l.], v. 3, n. , p. 1-30, 7 nov. 2019. Association for Computing Machinery (ACM). Disponível em: <http://dx.doi.org/10.1145/3359301>.
- ALPAYDIN, Ethem. *Introduction to machine learning*. 3. ed. Massachusetts: MIT Press, 2014.
- ARTOSI, Alberto. Technical normativity. In: *Italian Philosophy of Technology*. Springer, Cham, p. 149-160, 2021.
- ASSMANN, Jan. *Das kulturelle Gedächtnis: Schrift, Erinnerung und politische Identität in frühen Hochkulturen*. München: Beck, 1992.

- BAROCAS, Solon; SELBST, Andrew D. Big data's disparate impact. *Calif. L. Rev.*, v. 104, p. 671, 2016.
- BUOLAMWINI, Joy. How I'm fighting bias in algorithm. *TEDx BeaconStreet*, 2016, TED Talks. Disponível em: https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms. Acesso em: 10 set. 2020.
- CALDERS, Toon; ŽLIOBAITĖ, Indrė. Why unbiased computational processes can lead to discriminative decision procedures. In: CUSTERS, Bart et al (Ed.). *Discrimination and privacy in the information society: data mining and profiling in large databases*. Berlin: Springer-Verlag, p. 3-26, 2013. (Studies in Applied Philosophy, Epistemology and Rational Ethics 3).
- CALO, Ryan. *Artificial intelligence policy: a primer and roadmap*. 8 ago. 2017. Disponível em: <https://ssrn.com/abstract=3015350>. Acesso em: 17 jul. 2020.
- CITRON, Danielle Keats; PASQUALE, Frank A. The scored society: due process for automated predictions. *Washington Law Review*, Vol. 89, 2014, p. 1-, U of Maryland Legal Studies Research Paper No. 2014-8, 2014.
- CORTIZ, Diogo. *Curso de inteligência artificial para todos (Parte 1)*. Diogo Cortiz, 2020. YouTube. Disponível em: <https://youtu.be/Ze-Q6ZNWpco>. Acesso em: 24 mar. 2020.
- CUSTERS, Bart. Data dilemmas in the information society: introduction and overview. In: CUSTERS, Bart et al (Ed.). *Discrimination and privacy in the information society: data mining and profiling in large databases*. Berlin: Springer-Verlag, p. 3-26, 2013. (Studies in Applied Philosophy, Epistemology and Rational Ethics 3).
- DIAKOPOULOS, Nicholas et al. *Principles for accountable algorithms and a social impact statement for algorithms*. Fairness, accountability, and transparency in machine learning. Disponível em: <https://www.fatml.org/resources/principles-for-accountable-algorithms>. Acesso em: 1º out. 2020.
- DOMINGOS, Pedro. *O algoritmo mestre*. São Paulo: Novatec, 2017.
- ERNST, Christian. *Algorithmische Entscheidungsfindung und personenbezogene Daten Privatdozent*, Juristen Zeitung 72(21), 2017.
- ESPOSITO, Elena. *Artificial communication? The production of contingency by algorithms*. In: 46, 4 Zeitschrift für Soziologie, 2017, p. 249-265.
- FLORIDI, Luciano. *The Onlife Manifesto: being human in a hyperconnected era*. New York: Springer Nature, 2015.
- GEERTZ, Clifford. *Local knowledge: further essays in interpretative anthropology*. New York: Basic Books, 1983.
- GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. *Data mining: um guia prático*. Rio de Janeiro: Elsevier, 2005.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep learning*. MIT Press. 2016. Disponível em: <http://www.deeplearningbook.org>. Acesso em: 18 ago. 2020.

HOFFMANN-RIEM, Wolfgang. Verhaltenssteuerung durch Algorithmen – Eine Herausforderung für das Recht. In: *Archiv des oeffentlichen Rechts*, Volume 142, Number 1, January 2017 Mohr Siebeck: Tübingen, 2017.

_____. Inteligência artificial como oportunidade para a regulação jurídica. *Direito Público*, [s.l.], v. 16, n. 90, dez. 2019. ISSN 2236-1766. Disponível em: <https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3756>. Acesso em: 11 ago. 2020.

JUST, Natascha. Governing online platforms: competition policy in times of platformization. In: *Telecommunications Policy, Telecommunications Policy*, Volume 42, Issue 5, June 2018, Pages 386-394. Disponível em: <https://doi.org/10.1016/j.telpol.2018.02.006>.

_____; LATZER, Michael. Governance by algorithms: reality construction by algorithmic selection on the internet. Accepted manuscript forthcoming in *Media Culture & Society*, 2016, s. 2. Disponível em: http://mediachange.ch/media/pdf/publications/Just_Latzer2016_Governance_by_Algorithms_Reality_Construction.pdf.

KAPLAN, Andreas; HAENLEIN, Michael. Siri, Siri, in my hand: who's the fairest in the land? On interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, [s.l.], v. 62, n. 1, p. 15-25, jan. 2019. Elsevier BV. <http://dx.doi.org/10.1016/j.bushor.2018.08.004>.

LADEUR, Karl-Heinz. Lernfähigkeit des Rechts und Lernfähigkeit durch Recht, Erwiderung auf J. Nocke. In: GÖRLITZ, Axel; VOIGT, Rüdiger (Hrsg.). *Postinterventionistisches Recht*, Pfaffenweiler: Centaurus 1990 (Jahresschrift für Rechtspolitologie, Bd. 4), p. 141 ss.

_____. *Die Textualität des Rechts: Zur poststrukturalistischen Kritik des Rechts*. Velbrück Wissenschaft. Weilerwist, 2016.

_____. *Recht – Wissen – Kultur*. Die fragmentierte Ordnung. Berlin Duncker & Humblot, 2016.

LERMAN, Jonas. Big data and its exclusions. *Stan. L. Rev. Online*, v. 66, p. 55, 2013. Disponível em: https://review.law.stanford.edu/wp-content/uploads/sites/3/2016/08/66_stanlrevonline_55_lerman.pdf. Acesso em: 10 jul. 2020.

LESSIG, Lawrence. *Code: Version 2.0*, New York: Basic Books, 2006.

MAINI, Vishal; SABRI, Samer. *Machine learning for humans*. Disponível em: <https://everythingcomputerscience.com/books/Machine%20Learning%20for%20Humans.pdf>. Acesso em: 17 ago. 2020.

MARR, Bernard. How much data do we create every day? The mind-blowing stats everyone should read. *Forbes*, 18 maio 2018. Disponível em: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#7574ae7d60ba>. Acesso em: 1º abr. 2020.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. *Big Data: a revolution that will transform how we live, work and think*. New York: 2013.

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

MITCHELL, Tom M. *Machine learning*. McGraw-Hill Science, 1997.

NEGÓCIO, Ramon de Vasconcelos. *Vom Fremddruck zur Selbstbeschränkung: Das Problem der Verarbeitung juridischer Normativität durch Internet-Intermediäre*. Baden-Baden: Nomos, 2020.

NEVES, Marcelo. Entre subintegração e sobreintegração: a cidadania inexistente. *Dados – Revista de Ciências Sociais*, Rio de Janeiro, v. 37, n. 2, p. 253-275, 1994.

_____. *Transconstitucionalismo*. São Paulo: WMF Martins Fontes, 2009.

O'NEIL, Cathy. *Weapons of math destruction: how big data increases inequality and threatens democracy*. New York: Crown Publishers, 2016.

PAGER, Devah; SHEPHERD, Hana. The sociology of discrimination: racial discrimination in employment, housing, credit, and consumer markets. *Annual Review Of Sociology*, [s.l.], v. 34, n. 1, p. 181-209, ago. 2008. Annual Reviews. Disponível em: <http://dx.doi.org/10.1146/annurev.soc.33.040406.131740>.

PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. *Inteligência artificial e Direito*. Curitiba: Alteridade, 2019.

PIOVESAN, Flávia. Ações afirmativas no Brasil: desafios e perspectivas. *Revista Estudos Feministas*, Florianópolis, 16(3): 424, set./dez. 2008.

RAWLS, John. *Uma teoria da Justiça*. Trad. Almiro Pisetta e Lenita M. R. Esteves. São Paulo: Martins Fontes, 2008.

ROSEN, Lawrence. *Law as culture: an invitation*. New Jersey: Princeton University Press, 2006.

SCHULZ, W.; DANKERT, K. *Die Macht der Informationsintermediäre – Erscheinungsformen, Strukturen, Regulierungsoptionen*. Bonn: Friedrich-Ebert-Stiftung, 2016.

SCOTT, Joan W. O enigma da igualdade. *Revista Estudos Feministas*, [s.l.], v. 13, n. 1, p. 11-30, abr. 2005. FapUNIFESP (SciELO). Disponível em: <http://dx.doi.org/10.1590/s0104-026x2005000100002>.

TISCHBIREK, Alexander. Artificial intelligence and discrimination: discriminating against discriminatory systems. In: *Regulating Artificial Intelligence*, ed. Thomas Wischmeyer e Timo Rademacher, 2019.

TURING, Alan. Computing machinery and intelligence. *Mind*, New Series, v. 59, n. 236, p. 433-460, out. 1950.

TUTT, Andrew. An FDA for algorithms. *Administrative Law Review*, 83 (2017). Disponível em: <https://ssrn.com/abstract=2747994>. Acesso em: 7 nov. 2020.

VESTING, Thomas. The autonomy of law and the formation of network standards. *German Law Journal*, Vol. 05, No. 06, p. 639-668, 2004.

_____. *Die Medien des Rechts: Computernetzwerke*. Weilerswist: Velbrück Wissenschaft, 2015.

WEST, S. M.; WHITTAKER, M.; CRAWFORD, K. *Discriminating Systems: Gender, Race and Power in AI*. AI Now Institute, 2019. Disponível em: <https://ainowinstitute.org/discriminatingystems.html>. Acesso em: 1º nov. 2020.

Sobre os autores:

Alan Duarte | E-mail: duarttacademic@gmail.com

Mestrando em Direito Constitucional pela Universidade Federal do Ceará. Pós-Graduado em Direito, Tecnologia e Inovação (Instituto New Law). Advogado. Graduado em Direito pelo Centro Universitário 7 de Setembro (2020).

Ramon de Vasconcelos Negócio | E-mail: negocioramon@gmail.com

Doutor em Teoria do Direito (2019) pela Goethe-Universität (Frankfurt am Main). Mestre em Direito do Estado, na subárea de Direito Constitucional (2011), pela Pontifícia Universidade Católica de São Paulo. Graduado em Direito pela Universidade de Fortaleza (2007).

Data de submissão: 14 de agosto de 2021.

Data de aceite: 10 de janeiro de 2022.