

OS VIESES ALGORÍTMICOS NA FUNÇÃO DECISÓRIA DOS SISTEMAS DE INTELIGÊNCIA ARTIFICIAL

ALGORITHMIC BIAS IN THE DECISION-MAKING FUNCTION OF ARTIFICIAL INTELLIGENCE SYSTEMS

Gustavo Noronha de Ávila¹

Professor Permanente do Programa de Pós-Graduação, Mestrado e Doutorado em
Ciência Jurídica (UNICESUMAR, Maringá/PR, Brasil)

Thais Aline Mazetto Corazza²

Doutoranda e Pesquisadora Capes no Centro Universitário de Maringá
(UNICESUMAR, Maringá/PR, Brasil)

ÁREA(S): direito público; direito penal.

RESUMO: O tema do presente artigo são os riscos dos vieses algorítmicos na tomada de decisões por máquinas. O método utilizado para a pesquisa será hipotético-dedutivo, partindo-se de premissas gerais que possam ser aplicadas a situações concretas. Discorre-se sobre o funcionamento das ferramentas de inteligência artificial, com ênfase nos algoritmos, e o aprendizado

de máquina, para se demonstrar que até mesmo as pretensas ferramentas matemáticas imparciais possuem a subjetividade de seus criadores, sendo, portanto, afetadas diretamente pela qualidade dos dados fornecidos. A abordagem de pesquisa será dedutiva e a técnica será bibliográfica. Como resultado, verificou-se a necessidade do reconhecimento da existência dos vieses algorítmicos, uma vez que há a possibilidade de serem lesados os

¹ Graduado em Direito pela Pontifícia Universidade Católica do Rio Grande do Sul (2004), Mestrado (2006) e Doutorado (2012) em Ciências Criminais pela Pontifícia Universidade Católica do Rio Grande do Sul. Realizou Estágio de Pós-Doutoramento, sob a supervisão da Profa. Dra. Lilian Milnitsky Stein, no Programa de Pós-Graduação em Psicologia da PUCRS (2018). Professor da Pontifícia Universidade Católica do Paraná (*Campus* Maringá). Professor de Especialização em Ciências Penais da Universidade Estadual de Maringá, ABDConst, Universidade Ceuma, PUCPR, Univel, Universidade Feevale e Instituto Paranaense de Ensino. Brasil. *E-mail:* gusnavila@gmail.com. Currículo: <http://lattes.cnpq.br/4220998164028087>. Orcid: <http://orcid.org/0000-0002-7239-1456>.

² Mestre e Graduada em Ciências Jurídicas pelo Centro Universitário de Maringá (UNICESUMAR, Maringá/PR, Brasil). *E-mail:* thaiscorazza@hotmail.com. Currículo: <http://lattes.cnpq.br/3674899072994320>. Orcid: <http://orcid.org/0000-0002-5133-2238>.

princípios constituidores do Estado Democrático de Direito, como o acesso à justiça e o devido processo constitucional, bem como os direitos da personalidade, para que, a partir daí, seja possível estabelecer maneiras de contorná-los, por meio de transparência algorítmica e do *machine learning*.

ABSTRACT: *The theme of this article is the risks of algorithmic bias in decision making by machines. The method used for the research will be hypothetical-deductive, based on general premises that can be applied to specific situations. It discusses the functioning of artificial intelligence tools, with an emphasis on algorithms, and machine learning, to demonstrate that even the supposed impartial mathematical tools have the subjectivity of their creators, being, therefore, directly affected by quality of the data provided. The research approach will be deductive and the technique will be bibliographic. As a result, it was found the need to recognize the existence of algorithmic biases, since there is the possibility of harming the constitutive principles of the Democratic State of Law, such as access to justice and due constitutional process, as well as personality rights, for what from there it is possible to establish ways to circumvent them, through algorithmic transparency and machine learning.*

PALAVRAS-CHAVE: inteligência artificial; direito processual; princípio constitucionais; vieses algorítmicos; vieses cognitivos.

KEYWORDS: *artificial intelligence; procedural law; constitutional principles; algorithmic biases; cognitive biases.*

SUMÁRIO: Introdução; 1 Os vieses algorítmicos; 2 Formas de controle do envio das máquinas: transparência algorítmica e *machine learning*; Conclusão; Referências.

SUMMARY: *Introduction; 1 Algorithmic biases; 2 Ways to control machine bias: algorithmic transparency and machine learning; Conclusion; References.*

INTRODUÇÃO

O presente artigo trata da inteligência artificial e dos riscos que ela apresenta para a correção e legitimidade de sistemas de tomada de decisão, tendo em vista a perspectiva do devido processo constitucional.

O método utilizado para a pesquisa será hipotético-dedutivo, já que partiremos de premissas gerais que possam ser aplicadas a situações concretas. Será analisado o funcionamento das ferramentas de inteligência artificial, com ênfase nos algoritmos, e o aprendizado de máquina (*machine learning*), para se

demonstrar que mesmo pretensas ferramentas matemáticas imparciais possuem a subjetividade de seus criadores, sendo, portanto, afetadas diretamente pela qualidade dos dados fornecidos. A abordagem de pesquisa será dedutiva e a técnica de pesquisa será bibliográfica.

É crescente a utilização dos sistemas de inteligência artificial (IA) nos mais diversos ramos, e isso decorre do incremento da eficiência e da precisão dos serviços que eles proporcionam. Esse fenômeno também ocorre no direito, por exemplo, com o uso das soluções das *lawtechs* que otimizam os serviços, em especial no que se refere à litigância de massa.

Em países como os Estados Unidos, alguns sistemas de inteligência artificial, a exemplo do *Ross* e do *Watson*, já são usados por escritórios de advocacia nas pesquisas jurídicas, na análise documental, na redação dos contratos, entre outros. E as vantagens na utilização de tais tecnologias, como maior agilidade, precisão e qualidade na realização de trabalhos repetitivos, faz os escritórios investirem cada vez mais em seu uso.

Esse fenômeno também é verificado no Brasil, ainda que em menor escala. Plataformas são incrementadas nos escritórios de advocacia para a automatização de serviços repetitivos, aumentando a média de acertos em relação ao preenchimento de dados³. Os sistemas de inteligência artificial também são usados pelos escritórios para analisar a tendência de determinados juízes no julgamento de temas específicos (BAETA, 2018), possibilitando maior eficiência na defesa. Em 2014, a Advocacia Geral da União (AGU) começou a implantar seu Sistema AGU de Inteligência Jurídica (Sapiens), o qual objetiva “facilitar o trabalho do procurador, tornando mais rápida e simplificada a produção de peças, automatizando e eliminando a necessidade de registro manual da produção jurídica” (KAMAYURÁ, 2013). Trata-se de ferramenta que ajuda até mesmo na tomada de decisões, apresentando sugestões de teses jurídicas que se enquadram ao caso em específico.

O Poder Judiciário brasileiro também criou iniciativas nesse sentido, a exemplo do TJMG, que criou um sistema para indexação automática de

³ Sobre o assunto, ver: SATURNO, A. Inteligência artificial da IBM está ajudando escritório de advocacia brasileiro. *CanalTech*, 18 de janeiro de 2018. Disponível em: <https://canaltech.com.br/inteligencia-artificial/inteligencia-artificial-da-ibm-esta-ajudando-escritorio-de-advocacia-brasileiro-106622/>. Acesso em: 24 maio 2021.

processos, com intuito de identificar, com maior agilidade, a existência de demandas repetitivas. Trata-se da ferramenta ágil, que analisa, de maneira constante, o banco de dados do tribunal, informando os juízes acerca de distorções na distribuição (ASCOM, 2017). Igualmente, o TST, em parceria com a Universidade de Brasília (UnB), desenvolveu um *software* para realizar a triagem automática de processos e o processamento de julgados que envolvem a questão jurídica para a sugestão de proposta de voto (RACANICCI, 2018).

O Supremo Tribunal Federal também esteve desenvolvendo e implementando um programa de inteligência artificial, chamado de Victor, igualmente em parceria com a UnB. O sistema busca aplicar novos conceitos e técnicas de inteligência artificial e aprendizado de máquina para necessidades relevantes em termos de processamento, classificação de peças e classificação de temas na gestão da repercussão geral no STF. Os objetivos são o aumento da celeridade de processamento, o aumento da precisão e a acurácia nas etapas envolvidas, apoiando os recursos humanos envolvidos nas atividades judiciais (INAZAWA, 2019, p. 02). Os citados processos possuem relação íntima com o desenvolvimento dos sistemas de inteligência artificial, onde as máquinas são delineadas para a execução de funções que demandariam empenho e raciocínio humano.

Não é fácil definir inteligência artificial. Trata-se de um programa multidisciplinar, e, diante de um campo tão vasto, não se pode ficar restrito a uma área específica de pesquisa. Se se almejava imitar os processos cognitivos do ser humano, atualmente seus objetivos são desenvolver autômatos que resolvam alguns problemas de maneira melhor que os humanos, através de todos os meios disponíveis. Dessa forma, a inteligência artificial acaba cruzando várias disciplinas, como a ciência da computação, a matemática (análise, probabilidades, álgebra linear, lógica, otimização), a ciência cognitiva e o conhecimento especializado dos campos aos quais se busca aplicá-la. E os algoritmos que sustentam esse conhecimento se baseiam igualmente em abordagens diversas, como redes neurais, análise semântica, aprendizagem estatística ou exploratória, representação simbólica, entre outros (VILLANI, 2018).

A recente expansão da inteligência artificial decorre de significativos avanços no aprendizado de máquinas. As técnicas de aprendizagem são uma verdadeira revolução das abordagens históricas da inteligência artificial, já que, ao invés de programar as regras (que geralmente são mais complexas do que se

imagina) que administram e conduzem uma tarefa, agora há a possibilidade de deixar a máquina descobrir por si só (VILLANI, 2018).

Hoje a inteligência artificial e a tecnologia deixaram de ser exclusividade das grandes indústrias e passaram a ser produtos disponíveis à grande parte da população, muitas vezes, inclusive, até sem custos diretos de aquisição, como é o caso das redes sociais como o *Facebook* e o *Instagram*, sendo cada vez maior o uso dessas ferramentas digitais de automatização para auxiliar a execução de tarefas que eram até então feitas por um agente humano.

O funcionamento da inteligência artificial se dá a partir de sistemas de dados (chamados de algoritmos) que são programados para responder de acordo com a base de dados disponível. Primeiramente, faz-se necessário estabelecer o mecanismo de entrada de dados (*input*), sendo que um algoritmo deve possuir um ou mais meios para recepção dos dados a serem analisados. Em um computador, a informação deve ingressar por meio digital (*bits*). Da mesma forma, é preciso ter mecanismos para a saída/retorno dos dados trabalhados (*output*), de modo que um algoritmo também deve ter um ou mais meios para retorno dos dados, os quais devem estar relacionados especificamente com o *input*. Exemplificando: caso um algoritmo de uma calculadora receba as informações para realizar a soma de $2 + 2$ (*input*), ele retornará como resultado o número 4 (*output*). O *output* é decorrente do *input*, sendo papel do algoritmo prover o regresso dos dados certos a partir dos dados de entrada (VALENTINI, 2017, p. 42-43).

Levando-se em conta que o algoritmo é isento de juízo de valor para além de sua programação, é preciso que a relação de “correção” entre o *input* e o *output* seja determinada de forma precisa, deixando de lado ambiguidades. Por esse motivo, os algoritmos necessitam ter cada passo de suas operações definido de maneira cuidadosa. Dessa forma, cada etapa da tarefa computacional deve seguir um plano de tarefas predeterminado, e o programa (computação dos dados) deverá findar após o cumprimento do roteiro. Ou seja, o algoritmo tem que ser finito, entregando algum retorno (*output*) após cumprir os passos estabelecidos. Assim, para cumprir a tarefa de forma adequada, cada operação que o algoritmo precisar realizar deve ser simples o suficiente para que possa ser feita, de modo exato e em tempo razoável e finito, por um ser humano usando caneta e papel. Conclui-se, portanto, que o algoritmo é um plano de ação predefinido que o computador deve seguir, de forma que a realização contínua

de pequenas tarefas simples possibilite a realização da tarefa solicitada sem novo dispêndio de trabalho humano (VALENTINI, 2017, p. 42-43).

Na área jurídica, a implementação de tecnologias disruptivas do seu curso normal vem se tornando cada vez mais recorrente. De forma individual, esses sistemas emergentes existentes desafiarão e transformarão a maneira como determinados serviços jurídicos são fornecidos. Já coletivamente, eles irão transformar todo o panorama das leis. Argumenta Susskind que, ao se referir à ruptura, quer falar em destruição causada pelo lado da oferta do mercado jurídico, ou seja, pelos escritórios de advocacia e demais prestadores de serviços jurídicos. Salienta que, para o consumidor dos citados serviços, essa interrupção é uma boa notícia, já que a perturbação de uma pessoa, algumas vezes, é a salvação de outra (SUSSKIND, 2013, p. 32 em diante).

As tecnologias legais disruptivas (que causam a interrupção do curso normal) são a automação documental, a conexão constante via internet, os mercados legais eletrônicos (leilões de serviços, medidores *online* de reputação e comparativos de preços), a consultoria legal *online*, o ensino *online*, as plataformas jurídicas abertas, as comunidades *online* colaborativas fechadas, a automatização de trabalhos repetitivos e de projetos, a previsão de resultados de processos e respostas automáticas a dúvidas legais em linguagem natural, o conhecimento jurídico incorporado, a resolução *online* de conflitos (*online dispute resolutions* – ODR), a análise de documentos automatizada, entre outros (SUSSKIND, 2013, p. 32 em diante).

Nesse sentido, é notório que os sistemas de inteligência artificial trazem benefícios à prática do direito, em especial atenção à já citada automatização de atividades repetitivas, proporcionando maior precisão e agilidade em sua realização. Contudo, todo esse movimento de virada tecnológica no direito vem se impondo sem que haja a preocupação adequada por parte dos juristas, pois estão encantados com os ganhos de eficiência e produtividade nas atividades a serem realizadas. As virtudes dos programas também são apresentadas e vendidas por fornecedores dos produtos, que evitam divulgar os riscos no uso dessas tecnologias.

O presente artigo visa, a partir do processualismo constitucional democrático, problematizar um dos riscos na utilização das inteligências artificiais na prolação de decisões judiciais, propondo a necessidade de se analisar o conteúdo da cláusula do devido processo constitucional como garantidora

da transparência algorítmica. No entanto, defende-se a impossibilidade de se delegar a atividade das decisões judiciais a algoritmos de inteligência artificial em razão da opacidade e da falta de controle dos vieses peculiares das decisões.

Por viés entendem-se os fenômenos de julgamento falho que levam ao que se consideram análises falsas. Podem ser cognitivos (realizado por seres humanos) ou algorítmicos (realizado por máquinas (FULTON, 2019)). O estudo dos vieses cognitivos é essencial para entendermos como são construídas as percepções e as tomadas decisões, o que tem impacto em todas as áreas da atividade humana, inclusive com influência direta nos vieses algorítmicos.

Em que pese as vantagens ressaltadas, no atual momento da tecnologia, verifica-se que os mecanismos de inteligência artificial, no âmbito do direito, devem manter apenas as funções organizacionais, consultivas e de análise dos litígios, já que é perigoso o deslocamento da função decisória para as máquinas e dificilmente atenderão aos imperativos de *accountability*⁴, típicos do devido processo, e da necessidade de um controle participativo da formação decisória, em especial em razão da ausência de transparência dos algoritmos que orientam a inteligência artificial.

Nesse contexto é que se busca averiguar alguns riscos trazidos pela utilização da inteligência artificial com função decisória no âmbito DO Direito, partindo da análise da doutrina e de sistemas que já foram implantados.

Verificou-se, como resultado, a necessidade de se reconhecer a existência de vieses algorítmicos, já que é possível serem lesados os princípios constituidores do Estado Democrático de Direito, como o acesso à justiça e o devido processo constitucional, bem como os direitos da personalidade, para que, então, a partir daí se possam estabelecer maneiras de contorná-los, utilizando-se da transparência algorítmica e do *machine learning*.

1 OS VIESES ALGORÍTMICOS

Primeiramente, vale ressaltar que os mecanismos de inteligência artificial dependem de modelos, consistentes em representações abstratas de determinado processo, que, por sua própria natureza, trata-se de simplificações

⁴ O termo *accountability* pode ser traduzido como controle, fiscalização, responsabilização, ou, ainda, prestação de contas (SPINOZA, 2012). No presente artigo, o termo será utilizado para se referir à “responsabilidade e transparência dos algoritmos”.

do mundo real e complexo. Na criação de um modelo, os programadores selecionam as informações que serão fornecidas e disponibilizadas ao sistema de inteligência artificial e que também serão usadas na resolução de questões futuras.

Sobre os modelos e as escolhas dos programadores, Cathy O’Neil nos ensina, ao abordar as chamadas *weapons of math destruction* (“armas de destruição matemáticas”), que:

Para criar um modelo, então, fazemos escolhas sobre o que é importante o suficiente para incluir, simplificando o mundo em uma versão de brinquedo que pode ser facilmente compreendida e a partir da qual podemos inferir fatos e ações importantes. Esperamos que ele execute apenas um trabalho e aceitamos que ocasionalmente funcionará como uma máquina sem noção, com enormes pontos cegos. [...] Os pontos cegos de um modelo refletem os julgamentos e as prioridades de seus criadores. Embora as opções no Google Maps e no *software* aviônico pareçam improvisadas, outras são muito mais problemáticas. [...] Isto é extremamente simples, sacrificando a precisão e o *insight* pela eficiência. (Tradução livre) (O’NEIL, 2016, p. 27).

Portanto, verifica-se que essas escolhas fazem sempre haver pontos cegos nos algoritmos, os quais refletem as prioridades, os objetivos e as concepções de quem os criou, sendo os modelos, a todo o tempo, permeados de subjetividade do sujeito criador.

Esses pontos cegos podem não ter relevância para os resultados pretendidos pelos modelos. Por outro lado, informações importantes podem ser ignoradas para a correta análise da situação, de forma a influenciar, de maneira negativa, as respostas dadas pelo sistema. Alerta Cathy O’Neil:

Algumas vezes esses pontos cegos não importam. Quando perguntamos ao *Google Maps* por direções, ele modela o mundo como uma série de estradas, túneis e pontes. Ele ignora os prédios, porque não são relevantes para sua tarefa. [...] outros (pontos cegos) são muito

mais problemáticos. O modelo aplicado nas escolas de Washington, retornando para aquele exemplo, avalia os professores em grande parte com base nas notas de estudantes nos testes, mas ignora o quanto os professores engajam os estudantes, trabalham com habilidades específicas, lidam com a gestão da sala de aula ou ajudam seus alunos com problemas pessoais e familiares. O modelo é muito simples, sacrificando sua exatidão e diferentes percepções em prol da eficiência. No entanto, do ponto de vista dos administradores ele fornece uma ferramenta efetiva para investigar centenas de professores aparentemente com um baixo desempenho, mesmo que se corra o risco de interpretar incorretamente alguns deles. (Tradução livre) (O'NEIL, 2016, p. 28).

Após a elaboração do modelo, são fornecidos dados para o sistema, de modo a possibilitar o *machine learning* (aprendizado de máquina), pelo qual a máquina analisará as informações fornecidas, seguindo as instruções estabelecidas pelo algoritmo, para encontrar padrões e, então, conseguir prever resultados.

Também impactará nos resultados a qualidade dos dados fornecidos aos sistemas de inteligência artificial. Os dados são colhidos da sociedade, que é repleta de discriminações, desigualdades e exclusões. Um estudo realizado por pesquisadores da Universidade de Oxford concluiu que o aprendizado de máquina pode confirmar padrões discriminatórios, caso eles sejam encontrados no banco de dados. Então, um sistema de classificação exato irá reproduzi-los. Dessa forma, as decisões tendenciosas são também apresentadas como resultado de um suposto “algoritmo objetivo” (GOODMAN; FLAXMAN, 2017, p. 50-57).

Existe ainda o problema de erros nos dados em massa, devido a sua baixa confiabilidade quando extraídos da internet, somado ao fato de conterem lacunas decorrentes de perdas e de interrupções. Em relação à questão dos problemas dos dados obtidos via internet:

Grandes conjuntos de dados de fontes da internet geralmente não são confiáveis, estão sujeitos a

interrupções e perdas, e esses erros e lacunas são ampliados quando vários conjuntos de dados são usados juntos. Cientistas sociais têm uma longa história de fazer perguntas críticas sobre a coleta de dados e tentar explicar quaisquer vieses em seus dados (CAIN & FINCH, 1981; CLIFFORD & MARCUS, 1986). Isso requer a compreensão das propriedades e limites de um conjunto de dados, independentemente de seu tamanho. Um conjunto de dados pode ter muitos milhões de dados, mas isso não significa que seja aleatório ou representativo. Para fazer afirmações estatísticas sobre um conjunto de dados, precisamos saber de onde os dados estão vindo; é igualmente importante conhecer e contabilizar as deficiências desses dados (BOYD, 2011, p. 05).

Nesse sentido, há de se ressaltar que nem sempre um volume grande de dados inseridos nos sistemas de inteligência artificial refletirá em decisões melhores, tendo em vista que o fator mais relevante é a qualidade dos dados, e não a sua quantidade.

Veja-se que na própria constituição dos sistemas de inteligência artificial as escolhas feitas refletem também as opiniões e prioridades dos criadores, influenciando diretamente as decisões do sistema. Não se pode olvidar a impossibilidade de uma isenção completa quando se fala de sistemas como estes que recebem tratamento de universais e “desenviesados”, vez que o ponto de partida é sempre uma atividade humana de seleção de informações e dados, os quais refletem o contexto social de quem os produziu. Essa preocupação possui relação direta com os estudos dos vieses cognitivos (*cognitive biases*). Vieses de cognição

[...] são fenômenos da (ir)racionalidade humana, estudados pelos psicólogos cognitivos e comportamentais, e representam os desvios cognitivos decorrentes de equívocos em simplificações (heurísticas) realizadas pela mente humana diante de questões que necessitariam de um raciocínio complexo para serem respondidas. Tais simplificações (heurísticas do

pensamento) são um atalho cognitivo de que se vale a mente para facilitar uma série de atividades do dia a dia, inclusive no tocante à tomada de decisão. Nesse sentido, é possível afirmar que as heurísticas sejam mesmo automatismos inconscientes (sistema 1: rápido e instintivo) decorrentes da base de experiências e conhecimentos acumulados ao longo da vida, que permitem que as pessoas amarrem seus sapatos, dirijam seus veículos, bebam um copo d'água ou realizem uma caminhada sem despender grande esforço mental em torno de tais atividades. Entretanto, há situações, sobretudo as que envolvem um raciocínio mais complexo (sistema 2: lento, deliberativo e oneroso), em que as heurísticas do pensamento (automatismos mentais) podem gerar distorções cognitivas (vieses), levando a resultados subótimos. São inúmeros os vieses de cognição identificados e catalogados pela doutrina (NUNES; LUD; PEDRON, 2018).

Para promover a identificação dos desvios cognitivos, em consonância com a abordagem de Daniel Kahneman, recorro à ficção de dois sistemas de pensamento existentes no campo da interpretação jurídica: “o sistema I (pensamento automático) e o sistema II (controle racional)” (KAHNEMAN, 2012, p. 13). O primeiro sistema é aquele que age de forma automática e rápida, tomando a maior parte das decisões por impulso, sem um senso de controle voluntário. Já o segundo sistema se refere àquelas áreas do cérebro mais novas, que são responsáveis pela concentração e pelo esforço de calcular, pelo controle e monitoramento das sugestões formuladas pelo sistema I. Ou seja, o sistema II é responsável pela intencional atenção regulatória, em que pese se revelar confinado à lei do menor esforço (KAHNEMAN, 2012).

Assim, por viés cognitivo entende-se a distorção sistemática do pensamento que acontece durante o processamento e a interpretação das informações e que afetam as decisões e julgamentos da pessoa. O cérebro humano, apesar de um poderoso processador de informações, está sujeito a limitações, e os vieses cognitivos resultam geralmente da tentativa do cérebro de tentar simplificar o processamento de informações, uma vez que o tempo disponível e a capacidade

de processamento são escassos. Estão relacionados geralmente à memória (a forma como se lembra de um evento) ou à atenção (sobre o que se presta atenção ao redor), podendo levar a tomada de decisões tendenciosas e a influenciar na forma de o indivíduo interpretar o mundo. Como exemplo, Cass Sunstein e Richard Thaler (2009) ressaltam o viés do enquadramento (que é a predisposição de interpretar à dependência do modo pelo qual a questão é enquadrada) e da aversão à perda (que é a predisposição de valorizar mais as perdas do que os ganhos).

Em razão de os vieses se mostrarem como uma característica intrínseca do pensar humano, é possível concluir que um algoritmo criado por seres humanos enviesados provavelmente sofrerá do mesmo “mal”, não propositalmente, mas em decorrência das informações que o sistema forneceu. Surgem, assim, os chamados vieses algorítmicos, que ocorrem quando o comportamento das máquinas reflete os valores humanos implícitos envolvidos na programação. Nesse sentido, se o julgador se atentar aos vieses e hábitos mentais, será capaz de reunir elementos e forças para resistir a tais condicionamentos (internos e contextuais), preconceitos que frequentemente advêm de áreas cerebrais primitivas, contrárias às justificações sensatas (SUNSTEIN, 1993).

O fato de os algoritmos serem compostos por informações escolhidas, por si só, não se trata de um problema. Porém, em sendo um dado ignorado, e, aliado à ausência de transparência dos algoritmos e à sua possibilidade de crescimento exponencial, pode constituir um mecanismo perigoso de erro ou segregação, protegido pela pretensa e suposta imparcialidade da matemática (O’NEIL, 2016, p. 28).

Em relação à questão da opacidade dos algoritmos, O’Neil afirma que a falta de transparência dos modelos torna seu funcionamento invisível para todos, com exceção dos matemáticos e dos cientistas computacionais. Assim, ainda quando errônea a decisão dos algoritmos, ela acaba se tornando imune às reclamações e discordâncias, espalhando desigualdades e contribuindo para o seu aumento por meio do *feedback*. A esse respeito, afirma que o componente-chave disso é o pernicioso ciclo do *feedback*, pois “os modelos de sentença que traçam o perfil de uma pessoa por suas circunstâncias ajudam a criar o ambiente que justifica suas suposições. Este ciclo destrutivo gira e gira e, no processo, o modelo se torna cada vez mais injusto” (O’NEIL, 2016, p. 14).

Coloca-se em pauta aqui em que medida a falta de *accountability* (entendida aqui como responsabilidade e transparência dos algoritmos) mancha (ou não) o devido processo constitucional, bem como a necessidade de uma adequada fundamentação das respostas que os algoritmos oferecem.

Como exemplo de sistema de inteligência artificial que apresenta resultados discriminatórios, pode-se citar um mecanismo utilizado nos EUA para avaliar o risco de reincidência dos acusados no país chamado COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*). Alguns Estados utilizam os dados obtidos por esse sistema para fixar a sentença do réu e quanto maior o índice de reincidência, maior será o tempo de enclausuramento do indivíduo. No entanto, uma pesquisa realizada pela ProPublica identificou que o algoritmo utilizado possui a tendência de classificar erroneamente acusados negros como possíveis reincidentes, enquadrando assim de forma errada também os acusados brancos como sujeitos com risco baixo de reincidência (LARSON; MATTU; KIRCHNER; ANGWIN, 2016).

A pesquisa cita que “é mais provável que o algoritmo classifique erroneamente um réu negro como de maior risco do que um réu branco. Os réus negros que não reincidiram tinham quase o dobro de probabilidade de serem classificados pelo COMPAS como de maior risco em comparação com seus colegas brancos (45 por cento contra 23 por cento)”. Porém, ressalta que “os réus negros que pontuaram mais alto reincidiram um pouco mais frequentemente do que os réus brancos (63 por cento contra 59 por cento)”. Afirma então que o teste demonstrou a tendência de “cometer o erro oposto com os brancos, o que significa que era mais provável prever erroneamente que os brancos não cometeriam crimes adicionais se fossem soltos em comparação com os réus negros”. Concluiu-se na pesquisa que o sistema COMPAS “subclassificou os reincidentes brancos como de baixo risco 70,5% mais frequentemente do que os negros reincidentes (48% contra 28%). A razão de probabilidade para réus brancos era ligeiramente maior 2,23 do que para réus negros 1,61” (LARSON; MATTU; KIRCHNER; ANGWIN, 2016).

Veja-se que a empresa responsável pelo *software*, a Northpointe (NORTHPOINTE, 2012), não deixa disponível ao público o algoritmo no qual se fundamenta o índice de reincidência do acusado, de maneira que o réu não conhece o motivo pelo qual o sistema o rotulou com um alto ou baixo indicador, tampouco de que forma suas respostas influenciam no resultado final. Por

enquanto, “as únicas pessoas com visibilidade do funcionamento do COMPAS são seus programadores, que estão em muitos aspectos menos equipados do que os juízes para fazer justiça”. Isso porque “os juízes têm formação jurídica, são regidos por juramentos éticos e devem prestar contas não apenas por suas decisões, mas também por seu raciocínio em pareceres publicados. Os programadores carecem de cada uma dessas proteções” (ISRANI, 2017).

É certo que cientificamente não se pode precisar quais características e indicadores fazem um indivíduo ser mais ou menos predisposto à reincidência. Para que o algoritmo chegue a esse resultado, são utilizadas correlações entre dados, como, por exemplo, a existência de parentes ou vizinhos condenados, a convivência com usuários de drogas e reincidentes, o desempenho escolar, entre outros, e a probabilidade de reincidência, dados esses que carecem de confirmação científica e acaba culminando em resultados discriminatórios. Tudo o que os computadores sabem é o que os indivíduos lhes ensinam, incluindo o preconceito, e a melhora disso só pode ocorrer com transparência e ação corretiva por parte dos humanos.

A ausência de transparência do algoritmo causa muitas críticas. Como se defender de um “índice” sem saber o método de seu cálculo? Como submeter o “índice” ao controle do devido processo constitucional? Ainda que sejam expostas as perguntas realizadas, não há como saber de que maneiras as respostas influenciam no resultado final (*output*). Assim, torna-se inviável e impossível uma ampla defesa do acusado em razão de dados matemáticos opacos e algorítmicamente enviesados e camuflados por uma segurança da matemática, que seria supostamente justa, impessoal e imparcial.

Além disso, ao menos desde Gadamer, não é possível crer na intuição positivista (científica), que afirma que a verdade é uma construção universal, imutável e única, metodologicamente controlável, ou seja, que a verdade somente pode ser desvelada a partir de uma perspectiva controlada por uma experiência científica que seja garantida por um método e que possa ser falseado (GADAMER, 2005). Isso porque, no campo processual, inúmeras foram as hipóteses em que o procedimento judicial legitimou pelo método o barbarismo, bastando lembrar o ocorrido durante o nazismo em relação às esterilizações em massa ou até mesmo como ocorre no Brasil, onde uma moradora de rua teve esterilização determinada sem direito de defesa. Para Oscar Vieira:

Esse caso, ainda que possa ser considerado uma aberração jurídica, oferece uma amostra do impacto perverso que a profunda e persistente desigualdade causa sobre o reconhecimento das pessoas como sujeitos de direitos. Embora o princípio da dignidade determine que todos devam ser tratados com igual respeito e consideração, a miséria e a marginalização parecem tornar largas parcelas de nossa sociedade moralmente invisíveis no dia a dia, perdendo, na realidade, sua condição de sujeitos de direitos (VIEIRA, 2018).

É possível verificar vieses algorítmicos em diversos outros sistemas, como é o caso do sistema de concessão de crédito europeu e norte-americano utilizado por diversas empresas para analisar o risco de empréstimo. Diversos modelos usam até mesmo dados das redes sociais do solicitante para o cálculo do *credit score*, fundamentando-se, assim, nas conexões sociais do indivíduo. Dessa maneira, o resultado se vincula diretamente ao grupo social no qual o solicitante está inserido (HARRIS, 2018).

Outro exemplo de modelo enviesado é o sistema de reconhecimento facial criado pela *Google* que identificou pessoas negras como gorilas (KASPERKEVIC, 2015). Também o sistema de busca de contatos do aplicativo *LinkedIn* demonstrou preferência por nomes de pessoas do sexo masculino (DAY, 2016). O mecanismo de inteligência artificial da *Microsoft* criado para interagir com usuários do *Twitter*, chamado TAY, também passou a reproduzir mensagens xenofóbicas, antissemitas e racistas e antissemitas, chegando a *twittar* que “Bush fez o 11 de setembro e Hitler teria feito um trabalho melhor do que o macaco que temos agora. Donald Trump é a única esperança que temos” (HUNT, 2016). O aplicativo de *chat* SimSimi, que usa a inteligência artificial para conversar com os usuários, ficou suspenso no Brasil por reproduzir conversas, palavras e ameaças com teor sexual (SOUZA, 2018).

Portanto, o que se nota é que o aprendizado da máquina (*machine learning*) trata-se de uma atividade complexa que exige cuidados dos programadores. Já há conhecimento de que não é importante a quantidade de informações enviadas para alimentar os sistemas de inteligência artificial, e sim a qualidade delas, tendo em vista que dados enviesados ensinarão a

máquina a desempenhar suas funções de maneira enviesada também, fazendo perpetuar as desigualdades na sociedade, os erros e outras mazelas, de maneira automatizada.

2 FORMAS DE CONTROLE DO ENVIESAMENTO DAS MÁQUINAS: TRANSPARÊNCIA ALGORÍTMICA E MACHINE LEARNING

Todo o quadro exposto vem ocorrendo durante as iniciativas, no campo do direito, ao se usar algoritmos de inteligência artificial para auxiliar o trabalho dos profissionais jurídicos e tribunais, todas elas já apresentadas no início do presente trabalho. Diante da busca pela eficiência do neoliberalismo processual e diante do próprio desconhecimento da virada tecnológica na área jurídica, os sujeitos são induzidos a um encantamento com as potenciais simplificações de atividades jurídicas, sem que se tenham fixados os critérios de respeito a pressupostos jurídicos essenciais, como os inerentes ao devido processo constitucional, a fim de controlar a utilização inadequada dessas novas ferramentas.

Com transparência e responsabilidade, os algoritmos do sistema de justiça criminal têm potencial para o bem. Como exemplo, pode-se citar o uso de um programa de avaliação de risco em New Jersey, conhecido como Avaliação da Segurança Pública, para reformar seu sistema de fiança, que levou a uma redução de 16% na população carcerária antes do julgamento. O mesmo algoritmo auxiliou o condado de Lucas, em Ohio, a dobrar o número de liberações prejudgmento sem fiança e cortar o prejudgmento de crimes pela metade. O diferencial é que o funcionamento desse programa foi detalhado em um relatório publicado, permitindo que aqueles com experiência no assunto confirmem que variáveis moralmente problemáticas (e constitucionalmente inadmissíveis), como raça, gênero e variáveis, que poderiam representar os dois, não estavam sendo considerados na análise feita pelo algoritmo (ISRANI, 2017).

Portanto, é fundamental que se tenha um elevado grau de transparência algorítmica, a fim de permitir que os afetados pelo modelo compreendam o que determina o resultado alcançado pelo sistema de inteligência artificial. O Parlamento Europeu, atento a essa questão, em resolução de 16 de fevereiro de 2017, “considerando que a aprendizagem automática oferece enormes benefícios econômicos e inovadores à sociedade ao melhorar significativamente a capacidade de análise de dados, suscitando simultaneamente desafios no que

respeita à garantia da não discriminação, ao processo equitativo, à transparência e à inteligibilidade dos processos decisórios”, postulou o princípio ético de transparência para orientar a regulação da robótica:

Realça o princípio da transparência, nomeadamente o facto de que deve ser sempre possível fundamentar qualquer decisão tomada com recurso a inteligência artificial que possa ter um impacto substancial sobre a vida de uma ou mais pessoas; considera que deve ser sempre possível reduzir a computação realizada por sistemas de IA a uma forma compreensível para os seres humanos; considera que os robôs avançados deveriam ser dotados de uma “caixa preta” com dados sobre todas as operações realizadas pela máquina, incluindo os passos da lógica que conduziu à formulação das suas decisões (PARLAMENTO EUROPEU, 16 de fevereiro de 2017).

Veja-se que tornar acessíveis os algoritmos para a população em geral, e não apenas para os especialistas, é essencial para que se possa ter o controle e a fiscalização deles, já que se deve ter em mente a subjetividade que permeia o processo de tomada de decisão das máquinas e sua capacidade de atingir, de maneira drástica, os direitos dos indivíduos.

Há algum tempo a questão dos vieses vem sendo discutida no âmbito do direito, em especial quando se trata de decisões judiciais. Vieses cognitivos são atributos inerentes ao ser humano, já que o cérebro apresenta recursos cognitivos limitados e, em razão disso, acaba criando alguns atalhos para a tomada de decisões, de maneira a utilizá-lo de forma mais eficiente:

Ao decidir, pessoas se valem de heurísticas, ou seja, de atalhos cognitivos a partir dos quais lidam com as limitações de tempo e com a incompletude dos dados disponíveis: certas informações são privilegiadas, certas lacunas são inconscientemente completadas, certas percepções desencadeiam reações emocionais intensas, certas projeções de significado são rejeitadas pelos incômodos que nos provocam. [...] Como nossa capacidade de processamento de informação é

limitada, temos de selecionar com cuidado os objetos que receberão esse tipo de atenção, extremamente custosa em termos de tempo e energia. Necessitamos de heurísticas para viabilizar uma tomada de decisão eficiente, em que seja alcançado um nível de precisão aceitável (embora não ótimo) com um processamento cerebral que não comprometa demasiadamente a celeridade (HORTA; COSTA, 2017, p. 18).

Entretanto, quando comprometem a tomada de decisões por parte dos juízes, os vieses são assaz danosos, uma vez que são levados em consideração, ainda que inconscientemente, fatores que não deveriam intervir no julgamento.

Um dos principais vieses cognitivos identificados pela doutrina é o viés de confirmação (*confirmation bias*), que se caracteriza pela tendência de buscar e interpretar informações que ratifiquem as concepções já previamente constituídas pelo sujeito (PEER; GAMLIEL, p. 114-115). O impacto desse viés no processo de tomada de decisão é relevante, uma vez que o julgador tenderá a beneficiar evidências que confirmem sua hipótese, rejeitando as que apontem para solução diversa (NUNES; LUD; PEDRON, 2018).

Há inúmeros outros vieses que interferem no modo de pensar e decidir dos indivíduos. Pelo viés social, formam-se as impressões sobre as pessoas com base no grupo social em que estão inseridas. O efeito de ancoragem (*anchoring effect*) representa a dificuldade de se distanciar de uma primeira impressão ou de um primeiro dado que é fornecido (STAFFORD, 2017). Ainda, tem-se o viés de conformidade, segundo o qual as pessoas de um grupo possuem a tendência de evitar o dissenso e aderir à opinião da maioria (NUNES; LUD; PEDRON, 2018).

Em que pese a existência de diversos estudos a respeito dos vieses cognitivos, existe muita dificuldade ao tratar do tema, já que muitos julgadores ainda se consideram imparciais, e, portanto, não desenvolvem técnicas apropriadas para a superação do enviesamento – as técnicas de “desenviesamento” ou *debiasing* (NUNES; LUD; PEDRON, 2018). Questionando a imparcialidade dos juízos, Doyle aponta que “vários estudos sugeriram que os juízes podem ser tendenciosos a favor de evidências que confirmam suas hipóteses anteriores e podem desconsiderar evidências que não correspondem às suas suposições anteriores” (DOYLE, 2016, p. 39). O mesmo ocorre com as

ferramentas de inteligência artificial que muitos consideram como isentas de parcialidade.

Importante salientar que a tomada de consciência da existência dos vieses e da impossibilidade de imparcialidade total é fundamental para que o problema seja enfrentado:

Não estamos, contudo, condenados a repetir inconscientemente nossos vieses. Estudos contemporâneos apontam que essas respostas intuitivas e automáticas podem ser corrigidas, e os vieses, reduzidos, com o uso intensivo do raciocínio, mediante esforço e controle deliberado [...] compreender as nossas limitações cognitivas é fundamental para entender como os seres humanos tomam decisões e, com isso, interferir conscientemente nesses mecanismos (HORTA; COSTA, 2017, p. 20).

Contudo, existe uma agravante. As decisões tomadas por seres humanos podem ser impugnadas, já que se pode delimitar os fatores que levaram a determinada resposta, sendo que a própria pessoa que proferiu a decisão deve demonstrar o que o induziu a tal resposta (art. 93, IX, da CF/1988 e art. 489 do CPC). Por outro lado, os algoritmos utilizados nas ferramentas de inteligência artificial são obscuros para a maioria das pessoas, inclusive para seus programadores, tornando-os, de certo modo, inatacáveis. Isso porque, apesar de os programadores conhecerem o algoritmo no qual o modelo se baseia, nem sempre conseguem decifrar o que levou a máquina a chegar a um determinado resultado. A exemplo, há o sistema de inteligência artificial criado pelo *Facebook* para negociação, onde os robôs desenvolveram uma linguagem própria, indecifrável para os programadores, quando foram colocados para negociar entre si (FIELD, 2017). Por tais motivos, no âmbito do direito, considera-se problemática a atribuição de função decisória aos sistemas de inteligência artificial.

Importante destacar aqui que mesmo a estruturação de bases de dados de julgados e precedentes, após a adoção de força normativa aos últimos, caso ela seja realizada de modo enviesado e acrítica, pode levar à indução de resultados distorcidos nos julgamentos, culminando na perpetuação de erros ou reduzindo a possibilidade de diferenciação (*distinguishing*) de casos, o que

se considera impensável num sistema jurídico que busca sofisticar o uso da jurisprudência, como é o caso do Brasil após o Código de Processo Civil de 2015. Sobre a análise crítica dos precedentes no Brasil, vide Viana (2018).

Como já exposto inicialmente, é inegável que o uso de máquinas traz inúmeros benefícios à prática jurídica. O implemento de sistemas de inteligência artificial para as atividades de pesquisa, organização e classificação de informações, vinculação de casos a precedentes e elaboração de contratos tem efetividade prática, proporcionando precisão e celeridade. Porém, atribuir função decisória às máquinas equivalente a um juiz ampliará as desigualdades que existem no sistema judiciário, tudo com o respaldo de um decisionismo tecnológico. Isso porque, ainda que as decisões proferidas pelos juízes sejam enviesadas, existe certo grau de acesso aos motivos (mesmo enviesados, subjetivos ou errados) que os levaram a tomar determinada posição, já que devem fundamentar suas decisões. Dessa forma, os afetados podem discuti-las e impugná-las.

Já os resultados obtidos por sistemas de inteligência artificial são indiscutíveis por aqueles que não conhecem o algoritmo e seu funcionamento, como é o caso de advogados, juízes e membros do Ministério Público. Sobre essa questão, o instituto de pesquisa *AI Now*, da Universidade de Nova York, fez a seguinte recomendação em seu relatório apresentado no ano de 2017:

1. Agências públicas centrais, como as responsáveis pela justiça criminal, saúde, educação e assistência social, não devem mais utilizar IA e sistemas algorítmicos incompreensíveis (“caixa preta”). Isso inclui a utilização de modelos pré-treinados sem revisão e validação, sistemas de IA autorizados por fornecedores externos e processos algorítmicos criados internamente em empresas privadas. O uso de tais sistemas por agências públicas fomenta sérias preocupações quanto ao devido processo e, no mínimo, deveria ser possível realizar audiências públicas, testes e revisões, bem como respeitar padrões de *accountability* (CAMPOLO; SANFILIPPO; WHITTAKER; CRAWFORD, 2017).

Não obstante o alerta, a Suprema Corte de Wisconsin debateu, no caso *State v. Loomis*, a utilização do sistema COMPAS (algoritmo para análise de

risco dos acusados). A respeito do caso que originou o julgamento em questão, merece destacar que:

Em preparação para a sentença, um oficial do Departamento de Correções de Wisconsin produziu um PSI que incluía uma avaliação de risco COMPAS. [...] Como a metodologia do COMPAS é um segredo comercial, apenas as estimativas de risco de reincidência são reportadas ao tribunal. Na audiência de condenação de Loomis, o tribunal de primeira instância referiu-se à avaliação do COMPAS em sua determinação de condenação e, com base em parte nesta avaliação, condenou Loomis a seis anos de prisão e cinco anos de supervisão estendida. Loomis entrou com um recurso pós-condenação no tribunal de primeira instância, argumentando que a confiança do tribunal no COMPAS violava seus direitos ao devido processo. Como os relatórios do COMPAS fornecem dados relevantes apenas para grupos específicos e porque a metodologia usada para fazer os relatórios é um segredo comercial, Loomis afirmou que o uso da avaliação do COMPAS pelo tribunal infringia seu direito a uma sentença individualizada e seu direito de ser condenado em informação precisa (CRIMINAL LAW, 2016).

Em que pese o recurso, a Suprema Corte entendeu que a utilização desse sistema não violava o direito do acusado ao devido processo, em que pese a metodologia utilizada para o cálculo do índice ser indecifrável para os juízes e para o réu.

No sistema processual brasileiro, o uso de mecanismos ocultos para as partes na tomada de decisões viola as garantias processuais constitucionais, em especial a ampla defesa e o contraditório⁵, que somente serão efetivadas com o conhecimento dos pressupostos usados no julgamento. Assim, merece especial atenção a crescente implantação de sistemas de inteligência artificial no âmbito

⁵ A respeito do contraditório como comparticipação e garantia de influência das partes, ver: THEODORO JÚNIOR; NUNES, 2009, p. 109; e NUNES, 2004.

do Poder Judiciário brasileiro, especialmente em razão da tendência gradual de delegação de mais atividades para as máquinas, diante da sobrecarga judiciária e do quadro de litigância de massa. Veja-se que o próprio Supremo Tribunal Federal, ao proclamar a implementação da ferramenta Victor (PEIXOTO, 2020), asseverou que o sistema não irá se limitar a seu objetivo inicial, pois,

Como toda tecnologia, seu crescimento pode se tornar exponencial e já foram colocadas em discussão diversas ideias para a ampliação de suas habilidades. O objetivo inicial é aumentar a velocidade de tramitação dos processos por meio da utilização da tecnologia para auxiliar o trabalho do Supremo Tribunal (BRASIL, 2018).

É fato que, com o desenvolvimento rápido das tecnologias, as ferramentas de inteligência artificial terão ampliadas as suas funções. Contudo, a implementação desses sistemas sem a correta reflexão e análise sobre seus impactos e seu modo de funcionamento traz grandes riscos para o direito do ponto de vista sistêmico, tendo em vista que tornará indiscutíveis ou reduzirá drasticamente a possibilidade de discussão das decisões enviesadas com base em argumentos puramente quantitativos.

Nesse sentido foi o alerta feito pelo Professor de Direito da Universidade do Colorado, ao afirmar:

Sistemas tecnológicos que utilizam inteligência artificial são cada vez mais utilizados na aplicação do direito. Estes sistemas podem conter valores sutilmente embutidos no seu design tecnológico. Essa observação se torna particularmente importante no contexto do direito, tendo em vista a relevância das questões em jogo, incluindo a perda de liberdade, propriedade ou direitos. Sistemas tecnológicos legais que empregam inteligência artificial demandam cuidados e conhecimentos especiais em seu desenvolvimento, uma vez que o uso da inteligência artificial pode levantar problemas específicos em relação aos valores embutidos, os quais podem ser relevantes, mas difíceis de observar (SURDEN, 2017).

O Parlamento Europeu, em sua Resolução de 14 de março de 2017, igualmente afirmou que

[...] os dados e /ou os procedimentos de baixa qualidade em que se baseiam os processos de tomada de decisão e os instrumentos analíticos podem traduzir-se em algoritmos parciais, correlações ilegítimas, erros, numa subestimação das implicações jurídicas, sociais e éticas, no risco de utilização de dados para fins discriminatórios ou fraudulentos e na marginalização do papel dos seres humanos nestes processos, podendo resultar em processos imperfeitos de tomada de decisão, com um impacto nocivo nas vidas e nas oportunidades dos cidadãos, mormente nos grupos marginalizados, bem como em consequências negativas para as sociedades e as empresas (PARLAMENTO EUROPEU, 14 de março de 2017).

Dessa maneira, deve ser cuidadosa a implementação dessas ferramentas e cercada por um amplo debate sobre as melhores formas de se realizar o *machine learning*, a fim de se reduzir, ao máximo possível, o enviesamento das máquinas. Também se faz necessária a implementação de mecanismos para assegurar a transparência algorítmica, essencial para a compreensão do processo de tomada de decisões dos sistemas de inteligência artificial.

Ao se verificar a virada tecnológica no direito, torna-se necessária a rediscussão do próprio conteúdo dos seus princípios fundamentais, em especial do princípio do devido processo constitucional, a fim de adotar ações corretivas para controlar os poderes de decisão, que eram cognitivos e agora são tecnológicos/algorítmicos, de forma a reduzir sua opacidade e aumentar sua transparência (*accountability*).

O advento da Lei Geral de Proteção de Dados (Lei nº 13.709/2018) trouxe, em seu art. 6º, os princípios que as atividades de tratamento de dados pessoais deverão observar, além da boa-fé, que auxiliam na solução dos possíveis problemas enfrentados no presente estudo. Esses princípios são o da finalidade, da adequação, do livre acesso, da necessidade, da qualidade dos dados, da transparência, da segurança, da prevenção da não discriminação e da responsabilização e da prestação de contas.

Ademais, diante de uma decisão automatizada que não admitiu recurso para um tribunal superior aplicando precedente inadequado, o art. 20 da mesma lei possibilita, por exemplo, via recurso, requerer a revisão da citada decisão por um sujeito humano, ao dispor que o titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade, devendo o controlador fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial.

Da mesma forma que ocorre com os vieses cognitivos humanos, nos vieses algorítmicos deve-se primeiro reconhecer o problema a fim de se buscar soluções capazes de reduzir ou eliminar o enviesamento das máquinas. É certo que o uso irracional de sistemas de inteligência artificial no direito, com objetivos mercantis dos fornecedores de serviços ou em busca de rapidez e eficiência diante das demandas massificadas, poderá lesionar os princípios constituidores do Estado Democrático de Direito, como o acesso à justiça, a ampla defesa e o devido processo constitucional.

Assim como ocorreu no passado com a Revolução Industrial, deve-se perceber que os sistemas de inteligência artificial decorrentes das novas tecnologias não predeterminam um único resultado maléfico ou benéfico. Faz-se necessário que o direito e seus estudiosos utilizem essas tecnologias disruptivas e fixem pressupostos que aspirem à melhoria e à correção do sistema jurídico.

CONCLUSÃO

Não se pode frear o movimento de inserção de mecanismos tecnológicos no direito. É inegável os diversos benefícios que as máquinas podem trazer ao sistema. Todavia, é imprescindível que se tenha cuidado na sua implementação, já que, apesar de serem pretensamente objetivas, as ferramentas de inteligência artificial também possuem subjetividades que emergem tanto no momento de elaboração dos algoritmos quanto no momento de fornecimento de dados para o *machine learning*.

Sendo assim, necessário se faz o reconhecimento da existência dos vieses algorítmicos, tendo em vista que as máquinas, muitas vezes, comportam-se

de maneira a refletir os valores humanos que estão implicitamente envolvidos na programação. Esse fator, somado à opacidade dos algoritmos (que são indecifráveis para a maior parte da população), faz desabrochar os riscos que tais mecanismos podem acarretar para o devido processo constitucional, em razão de impossibilitar o exercício da garantia da ampla defesa e do contraditório, e por violar também o acesso à justiça. Como vimos, há também implicações bastante significativas aos direitos da personalidade, especialmente a privacidade.

Fatos assim já vêm ocorrendo em sistemas implantados em outros países, como o citado COMPAS, no qual o índice de reincidência dos acusados exibe resultados discriminatórios e, ao mesmo tempo, não impugnáveis para os sujeitos que foram afetados pela decisão, em razão de sua opacidade. A menção de outros exemplos neste artigo esclarece o risco sistêmico da implementação irracional de sistemas de inteligência artificial no âmbito do Poder Judiciário, com objetivos predominantemente quantitativos e sem a fixação de pressupostos jurídicos essenciais para o domínio das novas ferramentas.

Esse risco mencionado acontece pela falta de conhecimento dos estudiosos e profissionais do direito acerca das potencialidades das tecnologias de inteligência artificial, bem como em decorrência do modelo neoliberal de processo e a incessante busca por rapidez e simplificação dos procedimentos em massa, visando ao aumento da eficiência do sistema sob a ótica da produtividade em detrimento de critérios qualitativos. Ou seja, perde-se a qualidade em busca da quantidade.

Portanto, assim como somente a partir do reconhecimento, por parte da doutrina e dos julgadores, de que o ser humano não é naturalmente imparcial foi possível traçar soluções para os vieses cognitivos dos sujeitos tomadores de decisões, faz-se necessário também o reconhecimento e a ampliação dos estudos sobre os vieses algorítmicos para que se possa pensar em mecanismos para contorná-los.

O incremento de programas de aprendizagem de máquina e de transparência algorítmica são exemplos de instrumentos que podem ajudar nesse objetivo, de maneira a evitar a perpetuação das desigualdades, dos erros e das demais mazelas sociais que são decorrentes da ilusão da perfeita imparcialidade da matemática. Também a Lei Geral de proteção de Dados (Lei

nº 13.709/2018) trouxe, no bojo dos arts. 6º e 20, grande potencial para orientar possíveis soluções para os problemas narrados na pesquisa, a exemplo da possibilidade de revisão da decisão automatizada, que não admitiu recurso para um tribunal superior, aplicando precedente inadequado por um sujeito humano.

Por fim, faz-se necessário analisar e redefinir o conteúdo do devido processo constitucional como garantidor dessa transparência algorítmica e também como pressuposto da apreciação e interpretação das novas ferramentas que causarão impacto na cognição e na prática dos atos processuais.

REFERÊNCIAS

ASCOM - Assessoria de Comunicação Institucional. Gestão de precedentes é tema de encontro no TJM: Ministros da Comissão Gestora de Precedentes do STJ compartilharam experiência com o Tribunal mineiro. *Tribunal de Justiça de Minas Gerais - TJMG*, 7 de agosto de 2017. Disponível em: <http://www.tjmg.jus.br/portal-tjmg/noticias/gestao-de-precedentes-e-tema-de-encontro-no-tjmg.htm#.WvMAfogyv00>. Acesso em: 24 maio 2021.

BAETA, Z. Escritório já usam robôs que ajudam na escolha de estratégias nos tribunais. *Valor Econômico*, São Paulo, 16 de janeiro de 2018. Disponível em: <http://www.valor.com.br/legislacao/5259801/escritorios-ja-usam-robos-que-ajudamna-escolha-de-estrategias-nos-tribunais>. Acesso em: 24 maio 2021.

BOYD, D.; CRAWFORD, K. Six provocations for big data. A decade in internet time: Symposium on the Dynamics of the Internet and Society. Paper presented at *Oxford Internet Institute's*, 2011. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431. Acesso em: 1º mar. 2021.

BRASIL. *Inteligência artificial vai agilizar a tramitação de processos no STF*. Supremo Tribunal Federal: Brasília, 30 de maio de 2018. Disponível em: <http://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=380038>. Acesso em: 6 mar. 2021.

CAMPOLO, A.; SANFILIPPO, M.; WHITTAKER, M.; CRAWFORD, K. AI NOW 2017 Report. *AI Now Institute*, New York (2017). Disponível em: https://ainowinstitute.org/AI_Now_2017_Report.pdf. Acesso em: 23 abr. 2021.

CRIMINAL LAW. State v. Loomis: Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing. *Harvard Law Review*, 881 N.W. 2d 749, 2016. Disponível em: <https://harvardlawreview.org/2017/03/state-v-loomis/>. Acesso em: 23 mar. 2021.

DAY, M. How LinkedIn's search engine may reflect a gender bias. *The Seattle Times*, 31 de agosto de 2016. Disponível em: <https://www.seattletimes.com/business/microsoft/how-linkedins-search-engine-may-reflect-a-bias>. Acesso em: 6 maio 2021.

DOYLE, N. J. Confirmation bias and the due process of inter partes review. *IDEA: The Journal of the Franklin Pierce Center for Intellectual Property*, v. 57, p. 29-70, 2016.

FIELD, M. Facebook shuts down robots after they invent their own language. *The Telegraph*, 1º de agosto de 2017. Disponível em: <https://www.telegraph.co.uk/technology/2017/08/01/facebook-shuts-robots-invent-language>. Acesso em: 2 mar. 2021.

FULTON, S. What is bias in AI really, and why can't AI neutralize it? *ZDNet*, Nova York, 17 jul. 2019. Disponível em: <https://www.zdnet.com/article/what-is-bias-in-ai-really-and-why-cant-ai-neutralize-it/>. Acesso em: 25 maio 2021.

GADAMER, H.-G. *Verdade e método: fundamentos de uma hermenêutica filosófica*. 7. ed. Tradução: Enio Paulo Giachini. Petrópolis: Vozes, 2005.

GOODMAN, B.; FLAXMAN, S. R.. European Union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine*, v. 38, n. 03, p. 50-57, 2017.

HARRIS, J. The tyranny of algorithms is part of our lives: soon they could rate everything we do. *The Guardian*, 5 de março de 2018. Disponível em: <https://www.theguardian.com/commentisfree/2018/mar/05/algorithms-rate-credit-scores-finances-data>. Acesso em: 6 maio 2021.

HORTA, R. de L.; COSTA, A. A. Vieses na decisão judicial e desenho institucional: uma discussão necessária na era da pós-verdade. *Cadernos Adenauer 1*, p. 11-34, 2017.

HUNT, E. Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter. *The Guardian*, 24 de março de 2016. Disponível em: <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter>. Acesso em: 6 maio 2021.

INAZAWA, P. et al. Projeto Victor. *Especial Machine Learning*, p. 19-46, 2019. Disponível em: https://cic.unb.br/~teodecampos/ViP/inazawa_etal_compBrasil2019.pdf. Acesso em: 24 maio 2021.

ISRANI, E. T. When an algorithm helps send you to prison. *The New York Times*, Nova York, 26 out. 2017. Disponível em: <https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html>. Acesso em: 25 maio 2021.

KAHNEMAN, D. *Thinking, fast and slow*. London: Penguin Books, 2012.

KAMAYURÁ, U. Advocacia-Geral aposta em inteligência artificial e automação de processos para agilizar trabalhos jurídicos. *AGU*, 2013. Disponível em: <https://>

agu.jusbrasil.com.br/noticias/100362832/agu-aposta-em-inteligencia-artificial-e-automacao-de-processos-para-agilizar-trabalhos-juridicos. Acesso em: 24 maio 2021.

KASPERKEVIC, J. Google says sorry for racist auto-tag in photo app. *The Guardian*, 1º de julho de 2015. Disponível em: <https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app>. Acesso em: 25 maio 2021.

LARSON, J.; MATTU, S.; KIRCHNER, L.; ANGWIN, J. How we analyzed the COMPAS. Recidivism algorithm. *ProPublica*, 23 de maio de 2016. Disponível em: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Acesso em: 12 maio 2021.

NORTHPOINTE. *Practitioners Guide to COMPAS*. 17 de agosto de 2012. Disponível em: http://www.northpointeinc.com/files/technical_documents/FieldGuide2_081412.pdf. Acesso em: 14 maio 2021.

NUNES, D. J. C. O princípio do contraditório. *Revista Síntese de Direito Civil e Processual Civil*, Porto Alegre, n. 29, maio/jun. 2004.

NUNES, D.; LUD, N.; PEDRON, F. *Desconfiando da (im)parcialidade dos sujeitos processuais: um estudo sobre os vieses cognitivos, a mitigação de seus efeitos e o debiasing*. Salvador: JusPodivm, 2018.

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

PARLAMENTO EUROPEU. *Disposições de direito civil sobre robótica*. Resolução do Parlamento Europeu, de 16 de fevereiro de 2017, que contém recomendações à Comissão sobre disposições de Direito Civil sobre Robótica. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_PT.pdf. Acesso em: 25 maio 2021.

PARLAMENTO EUROPEU. *Resolução do Parlamento Europeu, de 14 de março de 2017*. Dispõe sobre as implicações dos grandes volumes de dados nos direitos fundamentais: privacidade, proteção de dados, não discriminação, segurança e aplicação da lei. Disponível em: <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0076+0+DOC+XML+V0//PT>. Acesso em: 15 abr. 2021.

PEER, E.; GAMLIEL, E. Heuristics and biases in judicial decisions. *Court Review*, v. 49, p. 114-118. Disponível em: <http://aja.ncsc.dni.us/publications/courtrv/cr49-2/CR49-2Peer.pdf>. Acesso em: 18 maio 2021.

PEIXOTO, F. H. Projeto Victor: relato do desenvolvimento da inteligência artificial na repercussão geral do Supremo Tribunal Federal. *Revista Brasileira de Inteligência Artificial e Direito - RBIAD*, v. 1, n. 1, p. 1-22, 2020. Disponível em: <https://rbiad.com.br/index.php/rbiad/article/view/4>. Acesso em: 20 maio 2021.

RACANICCI, J. Judiciário desenvolve tecnologia de voto assistido por máquinas. *Jota*, 2018. Disponível em: <https://www.jota.info/justica/judiciario-desenvolve-tecnologia-de-voto-assistido-por-maquinas-08012018>. Acesso em: 24 maio 2021.

SATURNO, A. Inteligência artificial da IBM está ajudando escritório de advocacia brasileiro. *CanalTech*, 18 de janeiro de 2018. Disponível em: <https://canaltech.com.br/inteligencia-artificial/inteligencia-artificial-da-ibm-esta-ajudando-escritorio-de-advocacia-brasileiro-106622/>. Acesso em: 24 maio 2021.

SOUZA, E. de. Simsimi é suspenso no Brasil; entenda caso do app de chat online. *TechTudo*, 18 de abril de 2018. Disponível em: <https://www.techtudo.com.br/noticias/2018/04/simsimi-e-suspenso-no-brasil-entenda-caso-do-app-de-chat-online.ghtml>. Acesso em: 6 maio 2021.

SPINOZA, R. M. Accountability. In: CASTRO, C. L. F.; GONTIJO, C. R. B.; AMABILE, A. E. N. (org.). *Dicionário de políticas públicas*. Barbacena: EdUEMG, 2012.

STAFFORD, T. Biases in decision making. *Tribunals*, 2017. Disponível em: <https://www.judiciary.uk/wp-content/uploads/2018/02/stafford-biases-in-decision-making-winter-2017.pdf>. Acesso em: 29 abr. 2021.

SUNSTEIN, C. *The partial Constitution*. Cambridge: Harvard University Press, 1993.

SUNSTEIN, C.; THALER, R. *Nudge*. Rio: Elsevier, 2009.

SURDEN, H. Values embedded in legal artificial intelligence. *U of Colorado Law Legal Studies Research Paper*, n. 17-17, 15 de março de 2017. Disponível em: <https://ssrn.com/abstract=2932333>. Acesso em: 1º fev. 2021.

SUSSKIND, R. *Tomorrow lawyers: an introduction to your future*. Oxford: Oxford University Press, 2013.

THEODORO JÚNIOR, H.; NUNES, D. J. C. Uma dimensão que urge reconhecer ao contraditório no direito brasileiro: sua aplicação como garantia de influência, de não surpresa e de aproveitamento da atividade processual. *Revista de Processo*, n. 168, p. 107-141, 2009.

VALENTINI, R. S. Julgamento por computadores? As novas possibilidades da juscibernética no século XXI e suas implicações para o futuro do direito e do trabalho dos juristas. *Tese (Doutorado em direito)* – Faculdade de Direito, Universidade Federal de Minas Gerais. Belo Horizonte, 2017. Disponível em: https://repositorio.ufmg.br/bitstream/1843/BUOS-B5DPSA/1/vers_o_completa_tese_romulo_soares_valentini.pdf. Acesso em: 20 maio 2021.

VIANA, A.; NUNES, D. *Precedentes: a mutação do ônus argumentativo*. Rio de Janeiro: GEN Forense, 2018.

VIEIRA, O. V. *Justiça, ainda que tardia*: moradora de rua teve esterilização determinada sem direito de defesa. 9 de junho de 2018. Disponível em: <https://goo.gl/2DrLRh>. Acesso em: 20 abr. 2021.

VILLANI, C. Donner uns sens à lí'intelligence artificielle: pour une stratégie nationale et européenne. *AI for Humanity*, 2018. Disponível em: <https://www.aiforhumanity.fr>. Acesso em: 24 maio 2021.

Submissão em: 24.08.2021

Avaliado em: 10.11.2021 (Avaliador B)

Avaliado em: 07.01.2022 (Avaliador D)

Aceito em: 16.09.2022