

A INTELIGÊNCIA ARTIFICIAL, O DIREITO E OS VIESES

ARTIFICIAL INTELLIGENCE, LAW AND THE UNINTENTIONAL VIESES

Bruno Fediuk de Castro

Pontifícia Universidade Católica do Paraná, Curitiba, PR, Brasil. E-mail: bfc.adv@gmail.com

 <https://orcid.org/0000-0003-0947-8142>

Gilberto Bomfim

Pontifícia Universidade Católica do Paraná, Curitiba, PR, Brasil. E-mail: g_bomfim@hotmail.com

DOI: <https://doi.org/10.46550/ilustracao.v1i3.24>

Recebido em: 13.11.2020

Aceito em: 17.12.2020

Resumo: A quarta revolução industrial, caracterizada pela combinação de inovações tecnológicas e a interação entre os domínios físicos, digitais e biológicos, já impacta profundamente e de forma exponencial a sociedade, evidenciando a sua potencialidade disruptiva. Ela é caracterizada pela robótica, *big data* e, sobretudo, pela Inteligência Artificial (IA), segmento da computação que busca simular a capacidade humana de raciocinar, tomar decisões, resolver problemas, dotando softwares, por meio de algoritmos, com objetivo de automatizar vários processos. Dentre os impactos da IA, além da mudança no mercado de trabalho, identifica-se também a problemática dos vieses cognitivos. Por mais que se confie na tecnologia para lidar com as limitações cognitivas humanas, os algoritmos ainda estão mal equipados para neutralizar os vieses não intencionais aprendidos com os algoritmos. O problema que o artigo pretende apontar é como a Inteligência Artificial, por meio dos algoritmos, está equipada para neutralizar os vieses não intencionais aprendidos. O artigo utiliza o método hipotético-dedutivo para ressaltar algumas das tecnologias de IA que já se encontram disponíveis aos profissionais da área jurídica, com ênfase no COMPAS, bem como as dificuldades que precisam ser enfrentadas na evolução das ferramentas aplicadas ao Direito para eliminar os vieses tecnológicos. Conclui-se pela necessidade de contemplação dessas mudanças pela sociedade e pelo Direito, a fim de neutralizar os efeitos deletérios decorrentes do enviesamento, inicialmente pela observância de uma grande diversidade na equipe de desenvolvimento dos algoritmos de IA e de supervisão constante, inclusive viabilizando que os sistemas sejam capacitados com base em uma autoridade moral.

Palavras-chave: Direito e tecnologia. Inteligência Artificial; Algoritmos. Vieses não intencionais. Eliminação.

Abstract: The fourth industrial revolution, characterized by the combination of technological innovations and the interaction between the physical, digital and biological domains, already profoundly and exponentially impacts society, evidencing its disruptive potential. It is characterized by robotics, big data and, above all, Artificial Intelligence (AI), a segment of computing that seeks to simulate the human capacity to reason, make decisions, solve problems, easing software, through algorithms, in order to automate various processes. Among the impacts of AI, in addition to the change in the labor market, the problem of cognitive biases is identified. As much as ai rely to deal with human cognitive limitations, algorithms are still ill-equipped to neutralize the unintended biases learned from algorithms.



The problem that the article intends to point out is how Artificial Intelligence, through algorithms, is equipped to neutralize the unintentional biases learned. The article uses the hypo-thetical-deductive method to highlight some of the AI technologies that are already available to legal professionals, as an emphasis on COMPAS, as well as the difficulties that need to be faced in the evolution of the tools applied to the law to eliminate technological derivatives. It is concluded by the need to contemplate these changes by society and law, in order to neutralize the deleterious effects resulting from bias, initially by the observance of a great diversity in the team of development of AI algorithms and constant supervision, including enabling systems to be trained based on a moral authority.

Keywords: Law and technology. Artificial Intelligence; Algorithms. Unintentional biases. Elimination.

1 Introdução

As três primeiras revoluções industriais proporcionaram a produção em massa, as linhas de montagem, a eletricidade e a tecnologia da informação, elevando a renda dos trabalhadores e fazendo da competição tecnológica o objeto do desenvolvimento econômico.

A quarta revolução industrial, por sua vez, é caracterizada pela combinação das inovações tecnológicas com a interação entre os domínios físicos, digitais e biológicos, de maneira que se acredita que terá um impacto mais profundo e exponencial no mundo, o que torna a quarta revolução industrial disruptiva. Ela é evidenciada pela robótica, *big data* e, sobretudo, pela inteligência artificial, segmento da computação que busca, por meio de *softwares* e algoritmos, simular a capacidade humana de raciocinar, tomar decisões, resolver problemas, com objetivo de automatizar vários processos, desde de recomendações de filmes e livros à aprovação de créditos e contratações de pessoas.

Dentre os impactos da inteligência artificial na sociedade, há a antevisão da mudança no mercado de trabalho. Análises indicam que profissões que são muito repetitivas serão substituídas por *softwares*. E as que são preponderantemente reconhecidas pelo lado pessoal, conexas à natureza humanas, como serviços de cuidadores e de atendimento, tendem a ter seus valores pressionados para baixo em razão da robotização.

No ramo do Direito, a consciência é de que a IA promova mudanças imediatas no mercado de trabalho, tendo em vista que as soluções que utilizam programas de inteligência artificial já conseguem executar tarefas repetitivas de análise de processos e termos jurídicos com eficiência e precisão maiores do que quando as mesmas tarefas são realizadas por seres humanos. Advogados que realizam atividades que dependem de interpretações e deduções subjetivas continuarão sendo valiosos, mas assistentes jurídicos que focam na realização de trabalhos repetitivos e que não exigem trabalho de interpretação, poderão ser substituídos por soluções de inteligência artificial, as quais devem realizar as atividades com maior precisão. inclusive.

Uma das tecnologias já existentes que envolvem inteligência artificial aplicáveis ao Direito, ainda que em um estágio inicial, é o *software* denominado *Watson*, criado pela IBM, que já possui diferentes serviços como reconhecimento e análise de vídeos e imagem; interação por voz; leitura de grandes volumes de textos; criação de assistentes virtuais; entre outros. Outro sistema que usa algoritmos matemáticos que vem sendo usado nos Estados Unidos para determinar o grau de periculosidade de criminosos é o denominado *COMPAS*, projetado pela *Northpointe*, *software* criado com a intenção de tornar as decisões judiciais menos subjetivas - menos influenciáveis por

erros humanos, vieses, preconceitos ou racismo.

No Brasil, a Advocacia Geral da União (AGU) começou a desenvolver o primeiro sistema de inteligência artificial aplicável à área jurídica, denominado Sapiens, seguido do Judiciário, que vem criando o sistema denominado “*Victor*”, de iniciativa do Supremo Tribunal Federal e o “*Sócrates*”, do Superior Tribunal de Justiça, que irão aumentar a velocidade de tramitação dos processos por meio da utilização da tecnologia para auxiliar o trabalho dos advogados públicos e dos tribunais.

Nesse contexto de mudanças ocasionadas pela quarta revolução industrial, o problema que o artigo pretende apontar é como os algoritmos, utilizados nos *softwares* de Inteligência Artificial para a tomada de decisões pelas máquinas, em substituição ao profissional do Direito, estão equipados para neutralizar os preconceitos e vieses não intencionais aprendidos, muitas vezes oriundos das bases de testes com as quais os algoritmos foram treinados.

O objetivo do artigo é demonstrar algumas dificuldades que precisam ser enfrentadas pelos operadores do Direito e programadores para eliminar os vieses não intencionais dos algoritmos.

A metodologia utilizada será o método hipotético-dedutivo, com uma abordagem qualitativa, buscando-se fomentar o conhecimento a respeito da inteligência artificial e sua aplicação no Direito e na apreciação jurídica.

O trabalho encontra-se estruturado da forma a seguir apresentada. No capítulo 1, serão abordados os efeitos da quarta revolução industrial na sociedade e os conceitos de inteligência artificial, algoritmos, *machine learning* e *deep learning*. No capítulo 2, serão analisados como se formam os vieses cognitivos não intencionais nos algoritmos utilizados nos sistemas de inteligência artificial. No capítulo 3, será avaliado como os sistemas de inteligência artificial aplicáveis ao Direito, notadamente, o norte-americano denominado *COMPAS*, estão trabalhando para eliminar os vieses não intencionais de sua aplicação concreta.

2 A quarta Revolução Industrial e a Inteligência Artificial

Uma revolução industrial é o conjunto de mudanças radicais no processo produtivo, que impactam diretamente no cenário econômico e social mundial, podendo ser identificadas, a partir do século XVIII, quatro revoluções que transformaram significativamente o modelo de produção das indústrias.

A primeira revolução industrial acontece a partir de 1780, movida por tecnologias mecânicas, especialmente pela construção de ferrovias e pela invenção da máquina a vapor (SCHWAB, 2016, p. 15) máquinas a vapor, construída inicialmente por Thomas Newcomen em 1698, e aperfeiçoada por James Watt, foram utilizadas para mecanizar grande parte do processo produtivo. Assim, a produção ganhou em escala, com aumento da eficiência têxtil, aliada aos meios de transporte a vapor, que facilitaram a locomoção de insumos e produtos. É a primeira etapa do processo evolutivo da indústria (BRASIL, 2020, s.p.).

A segunda revolução industrial acontece em 1870, fincada especialmente na eletricidade e nas linhas de montagem, que permitiram outros progressos essenciais nesse período, que incluem a introdução de navios movidos a vapor, o desenvolvimento do avião e a produção em massa de bens de consumo (SCHWAB, 2016, p. 15 e 16). Com a padronização dos procedimentos e as máquinas acelerando o ritmo de trabalho e aumentando a eficiência, em 1903 teve início

um novo modelo produtivo, desenvolvido por Henry Ford, que desenvolveu um sistema de trabalho industrial para agilizar a fabricação do automóvel — o Ford T, primeiro carro do mundo produzido em série. Depois disso, veio a produção em massa de eletrodomésticos, bens de consumo, comida enlatada etc. (BRASIL, 2020, s.p.)

A terceira revolução industrial começa aproximadamente em 1960, após a Segunda Guerra Mundial, com o advento da automação, costuma também ser denominada de revolução digital ou do computador (SCHWAB, 2016, p. 15 e 16). Computadores, redes, tecnologias da informação, robôs industriais e o Controlador Lógico Programável (CLP) — computador com aplicação industrial que controla processos repetitivos de produção — permitiram a automação de diversas operações. Mais tarde, nos anos 1990, surge a *internet* e as plataformas digitais (BEZERRA, 2019, s. p.).

Acredita-se que o desenvolvimento e a incorporação de inovações tecnológicas novamente mudarão radicalmente o mundo e devem contribuir de forma significativa com o desenvolvimento da sociedade dos próximos anos. Klaus Schwab (2016, p. 11) desenvolve a ideia de que já estamos vivendo essa nova Era, que é algo diferente de tudo aquilo que já foi experimentado pela humanidade. Para ele “*estamos no início de uma revolução que irá alterar profundamente a maneira como vivemos, trabalhamos e nos relacionamos*” (SCHWAB, 2016, p. 11).

A ideia da quarta revolução industrial surge em um evento sobre automação industrial, realizado na Feira de Hannover, na Alemanha, ocorrido em 2011. Ela é baseada na enorme quantidade de informações digitalizadas e nos avanços no campo da inteligência artificial. Com isso, a indústria adquiriu autonomia na operação, combinando tecnologias cibernética, mecânica e eletrônica para promover uma sinergia entre os mundos virtual e físico, a fim de melhorar o desempenho e eficiência na produção (KELNAR, 2016, s. p.). Tecnologias habilitadoras — como *IoT*, *Cloud Computing*, *Big Data* e *Data Analytics* — conectam pessoas e dispositivos, o que permite integração de processos e operações instantâneas e guiadas por dados. Com tanta inovação, a indústria se beneficia com ganho de produtividade; aumento da segurança; eliminação de erros e desperdícios; redução de custos operacionais; transparência nos negócios e personalização em escala.

Para Luciana Pedroso Xavier e Mayara Guibor Spaler (2019, p. 543), a primeira e a segunda revoluções partiram de sistemas físicos, a terceira alcançou os sistemas cibernéticos e, finalmente, a quarta revolução viabilizou o desenvolvimento de sistemas físico-cibernéticos. Alexandre Veronese, Alessandra Silveira e Amanda Nunes Lopes Esipiñeira Lemos (2019, p. 234-235) afirmam que a quarta revolução está diretamente relacionada com a automação dos processos produtivos, no entanto, para os autores, “*a automação se revestiria de uma qualidade nova: a capacidade de aprendizado dos programas para melhorar o seu próprio desempenho*”.

A base para essa interação é a inteligência artificial (IA), que é um segmento da computação que busca simular a capacidade humana de raciocinar, tomar decisões, resolver problemas, dotando *softwares* e *hardwares*, por meio de algoritmos, de uma capacidade de automatizar vários processos.

A primeira vez que a expressão “inteligência artificial” foi utilizada foi no ano de 1956, na universidade de Dartmouth, nos Estados Unidos, quando um grupo de pesquisadores estudavam a possibilidade de aprendizado de sistemas computacionais a partir de sua experiência

própria (MICROSOFT, 2018, p. 28). Cabe aqui destacar que a expressão vem sendo utilizado de maneira indiscriminada e sem observar uma precisão conceitual pelos mais diversos veículos e autores (VERONESE, SILVEIRA, LEMOS, 2019, p 237-238) e por isso é preciso cautela com a sua utilização.

Entretanto, com a tecnologia disponível durante o período pós Segunda Guerra Mundial (1939-1945) só se conseguiram desenvolver tecnologias capazes de realizar uma única tarefa específica de forma otimizada em relação aos humanos. Por isso, naquele momento, nomearam o processo de IA Limitada (KELNAR, 2016, s. p.). A “IA” Limitada existe há décadas através de programas baseados em regras de exibições rudimentares de “inteligência” em contextos específicos. O progresso, entretanto, tem sido limitado, pois os algoritmos necessários para enfrentar os problemas de hoje são ainda muito complexos para serem programas à mão por desenvolvedores.

Os algoritmos são um passo a passo, uma sequência de instruções pré-definidas, expresso em uma linguagem matemática estilizada (KLEINBERG, 2017, s. p.) e desenvolvido com a finalidade de determinar o que um computador tem de fazer (DOMINGOS, 2015, s. p.). Em curtas linhas, todo algoritmo tem um *input* e um *output*: os dados alimentam o computador, o algoritmo segue suas instruções e daí surge o resultado esperado.

Com a evolução tecnológica, a IA atingiu um novo patamar evolutivo. O *machine learning* (aprendizado da máquina) surge como um método de análise de dados que automatiza a construção de modelos analíticos. É um ramo da inteligência artificial baseado na ideia de que sistemas podem aprender com dados, identificar padrões e tomar decisões com o mínimo de intervenção humana (CASTRO, BOMFIM, TEIDER, 2020, *passim*). O objetivo da maioria dos *machine learning* é desenvolver um mecanismo de previsão para um caso de uso específico. Um algoritmo recebe informações sobre um domínio e pesa os dados para fazer uma previsão útil. Ao disponibilizar aos computadores a “capacidade de aprender”, passando a tarefa de otimização — de pesar as variáveis nos dados disponíveis para fazer previsões precisas sobre o futuro — para o algoritmo (KELNAR, 2016, s. p.).

Algoritmos de *machine learning* aprendem por meio de treinamento. Um algoritmo recebe inicialmente exemplos cujos resultados são conhecidos, nota a diferença entre suas previsões e os resultados corretos e ajusta as ponderações das entradas para melhorar a precisão de suas previsões até que sejam otimizadas. A característica definidora dos algoritmos de *machine learning*, portanto, é que a qualidade de suas previsões melhora com a experiência. Quanto mais dados fornecemos, melhores são os mecanismos de previsão que podemos criar (KELNAR, 2016, s. p.).

Atualmente, a inteligência artificial vem evoluindo para o chamado *deep learning*, ou Computação Cognitiva. Nesta nova etapa, camadas de dados tentam imitar a conectividade de nossa rede neural biológica. Estas camadas de conexão são capazes, não apenas de aprender a como realizar uma tarefa, mas de avaliar baseando-se em grandes quantidades de dados - se uma informação (dado, imagem, etc.) tem probabilidade em ser verdadeira ou não (CHIOVATTO, 2019, p. 3).

Uma diferença fundamental do *deep learning* é que a pesquisa nesta área tenta fazer representações melhores e criar modelos para aprender a partir de dados não rotulados em grande escala. Algumas das representações são inspiradas pelos avanços da neurociência e baseadas na

interpretação do processamento de informações e padrões de comunicação em um sistema nervoso. Enquanto as “redes neurais” dos sistemas computacionais estão sendo ajustadas, elas estão constantemente produzindo respostas erradas. E por isso ainda é necessário muito trabalho manual humano: a tecnologia precisa de treino; precisa de informações para poder calibrar suas respostas e, posteriormente, acertar. Atualmente este tipo de tecnologia está em uso nas mais variadas funções, desde o reconhecimento facial do aplicativo *Facebook*, até a leitura de tumores em exames de ressonância magnética, por exemplo. Por enquanto a associação entre a análise humana e a da máquina tem porcentagem de acerto maior que cada um destes recursos utilizado sozinho (CHIOVATTO, 2019, p. 4).

Os especialistas acreditam que é possível “treinar” as máquinas para responder qualquer problema. As possibilidades seriam infinitas, bastando ensiná-las.

3 Os algoritmos e os vieses

O mundo contemporâneo proporciona uma infinidade de informações simultâneas e, buscando otimizar uma forma de absorver e interpretar a maior quantidade possível de informações, o cérebro humano filtra essas informações do ambiente e as utiliza para direcionar o modo de agir. Processar a totalidade de informação recebida de forma consciente é inviável, principalmente, porque o tempo despendido para armazenamento e interpretação seria ineficiente (LAGO, 2016, p. 1).

Para aumentar sua eficiência, o cérebro humano realiza algumas adaptações. De acordo com Thaler e Sunstein (2019, p.30) existem dois tipos de pensamento: um intuitivo e automático, que acaba sendo mais rápido e, acarreta o que se costuma associar à palavra “pensamento”, e outro denominado reflexivo e racional, e é premeditado e autoconsciente.

Para otimizar o funcionamento e a capacidade de atuação, utilizando do sistema intuitivo e automático, o cérebro humano adota uma espécie de algoritmos mentais, possibilitando a produção de julgamentos rápidos, mesmo com informação limitada, o qual também são denominados de heurísticas. Para Lago (2016, p.1). Heurísticas são “atalhos mentais para tomadas de decisões, que permitem ao ser humano ser capaz de tomar decisões e não se distrair tentando absorver todas as informações disponíveis ao seu alcance. Quando as heurísticas falham, surgem os vieses cognitivos”. Para Juarez Freitas (2013, p. 228), o sistema automático proporciona uma economia de energia, mas cobra um preço alto para tanto, principalmente ao tropeçar em questões capitais envolvendo o exercício da lógica e do discernimento.

A contemplação e o estudo dos vieses em caráter psicológico e comportamental já vem sendo objeto de estudo no campo teórico há algum tempo. Herbert Simon (1987, p. 266), desenvolvedor da teoria da racionalidade limitada, comentava acerca dos atalhos como limites cognitivos decorrentes do (limitado) conhecimento e a (restrita) capacidade computacional do ser humano.

Os vieses podem ser compreendidos como distorções cognitivas com potencial de fazer com que o intérprete cometa erros de avaliação e controle (FREITAS, 2013, p. 225). Existem vários tipos de vieses e alguns deles possuem grande potencial de influenciar negativamente as escolhas. Lucas Lago (2016, p.2) relata algumas das principais formas de vieses que nos levam a conclusões erradas em determinadas situações e podem ser potencializados pela bolha de

informação, a saber:

O viés da confirmação pode ser descrito como tendência de se lembrar, interpretar ou pesquisar informações de maneira a confirmar crenças ou hipóteses iniciais. Ou seja, temos uma tendência natural de buscar informações que reforcem o que imaginamos ser verdade. Como os filtros online nos mostram informações relacionadas ao nosso “passado”, ideias antigas tendem a ser reforçadas pelas buscas, pois resultados conflitantes são evitados pelo algoritmo que organiza essa informação.

O falso consenso pode ser explicado como a ilusão de que uma maioria concorda com um ponto de vista sobre determinado assunto, quando na verdade esse consenso não existe. As redes sociais, com seus algoritmos, tendem a afastar as pessoas que não dão os mesmos likes que você e isso pode aumentar o efeito do falso consenso, pois ao olhar na sua timeline (linha do tempo) a sua impressão será reforçada pelo viés da confirmação.

O último efeito com potencial para ser ainda mais danoso é a polarização de grupos, que acontece quando convivemos com grupos que compartilham a mesma visão em determinados temas. Deste modo, a Internet cria bolhas isoladas entre “aqueles que concordam com A” e “aqueles que concordam com B” e os dois grupos possuem poucas conexões entre si (LAGO, 2016, p. 2).

As heurísticas são naturalmente desenvolvidas, porém, não são as únicas maneiras de filtrar as informações que recebemos diariamente. Algoritmos utilizados em *softwares* de inteligência artificial e pelas empresas na *internet* auxiliam no trabalho de filtrar dados considerados irrelevantes e na tomada de decisões autônomas. A partir da inteligência artificial, as máquinas recebem milhões de dados para processar, interpretar e aprender (LAGO, 2016, p.1). Contudo, a maioria das bases de dados e classificações usadas pelos sistemas de Inteligência Artificial e *machine learning* advêm de pessoas que criaram os algoritmos, estando relacionados a pensamentos com vieses do ser humano.

A mineração de dados é um procedimento que está intimamente relacionado à análise estatística e, assim, está sujeita à análise de dados que foram ali inseridos ou captados de uma outra origem, o que tem potencial de gerar alguma forma de discriminação. Para Carolina Braga (2019, p. 681), a finalidade da *big data* é prover uma base racional em cima da qual será possível atribuir a determinado indivíduo, ou grupo de indivíduos, características específicas, possibilitando tomadas de decisões personalizadas.

Quando uma inteligência artificial é acionada, a decisão é tomada após vários cálculos matemáticos que exigem uma série de códigos de computador capazes de entender quais dados estão sendo processados. O resultado, no entanto, ainda depende das informações submetidas no processo. Assim como existem diversos vieses cognitivos no pensamento humano, também existem distorções capazes de levarem as máquinas a conclusões equivocadas/erradas em determinadas situações (CASTRO, BOMFIM, TEIDER, 2020, *passim*).

Por essas razões, o enviesamento pode possuir uma característica ou um aspecto inerente de negatividade. Em conexão com a temática de tecnologia, o enviesamento humano, que já foi considerado uma falha cognitiva (WOJCIECHOWSKI, ROSA, 2018, p. 46), possui o potencial de refletir os seus preconceitos “na informação, nos algoritmos ou no modelo de aprendizagem” (CORDEIRO; OLIVEIRA; DUARTE, 2019, s. p).

Ao abordar a discriminação nas decisões por algoritmos, Carolina Braga (2019, p.

681) afirma que a forma mais comum de discriminação que acaba sendo gerada por decisões autônomas ocorre em decorrência dos dados utilizados em seu treinamento. A autora ressalta que a mineração de dados, no período que compreende desde a coleta até a apresentação de um resultado, pode se utilizar de mecanismos que podem levar a distorções, sendo (i) definição do problema; (ii) treinamento dos dados; (iii) seleção dos dados; e (iv) utilização de *proxies* e *masking*.

Por mais que se confie na inteligência artificial para lidar com nossas frágeis limitações, os algoritmos ainda estão mal equipados para neutralizar conscientemente os vieses aprendidos com o pensamento humano. Com a evolução da inteligência artificial e sua curva de aprendizado, esses conflitos tendem a ficar mais latentes – e passíveis de correção, especialmente quando se fala de tendências ideológicas, de gênero ou raça. Assim, a IA terá o papel de guiar as decisões com maior precisão e sem os vieses de quem a programou.

Para Juarez (2013, p. 230-231), tomar ciência dos vieses é condição necessária para aprimorar a performance interpretativa, ao contrário de fingir deferência à autonomia do objeto e insistir em negar os condicionamentos. O autor destaca que caso o intérprete jurídico não esteja vigilante, ou acredite piamente em uma fantasiosa determinação do mundo pré-dado, ele acabará sendo manipulado por impulsos cegos e pré-compreensões sem freios, fazendo com que este venha a tomar decisões sob influências que não contemplam o lado racional.

Alguns exemplos de riscos e perigos de enviesamento da tecnologia (e notadamente da Inteligência Artificial) podem ser verificados, *exempli gratia*, na afirmação de que ela (a Inteligência Artificial) “*não consegue superar os vieses de seus criadores*” (UOL, 2017, s. p.); no enviesamento de ferramenta de recrutamento que demonstrou viés contra mulheres (REUTERS, 2018, s.p.); e na preocupação de que um software inteligente de segurança pública pudesse selecionar arbitrariamente os indivíduos investigados em um cenário de Direito Penal do Inimigo (CARVALHO, 2018, *passim*).

Os vieses cognitivos nasceram de uma adaptação biológica do ser humano para diminuir a quantidade de informações sobre o cérebro. Portanto, faz-se necessário tomar cuidado para “não replicar e/ou aumentar os vieses cognitivos humanos, ou pior, desenvolver vieses tecnológicos, atuais, diminuindo o risco de a tecnologia estar baseada em dados fundamentados em práticas discriminatórias passadas” (LAGO, 2016, p.1.) e que poderiam prejudicar minorias, seja, por exemplo, pela cor da sua pele, idade, sexo, ou preferência sexual.

O enviesamento da (ou na) tecnologia é uma problemática existente, latente e que já se encontra demandando interações e respostas do mundo jurídico. Nesse sentido, a partir de um caso concreto se poderá exemplificar pormenorizadamente o problema ora versado e compreender de maneira mais vertical a sua ocorrência, o seu tratamento e as suas potenciais consequências.

4 A eliminação do vieses não intencionais e o Direito

O sistema judiciário está se beneficiando dos avanços tecnológicos proporcionados pela quarta revolução industrial, sobretudo da inteligência artificial, principalmente em razão da necessidade de realizar a análise de grande quantidade de dados e tomar decisões com a maior eficiência e rapidez possível. São muitas as potencialidades do uso da Inteligência Artificial

(IA) no desenvolvimento de novas soluções para a sociedade, principalmente na área jurídica (ZIVIANI, 2020, p.11)

No Brasil, a Lei 11.419/06 (Lei do Processo Eletrônico) dispõe sobre a informatização do processo judicial, permitindo a realização de atos processuais mais adaptados à realidade, permitindo, por exemplo, a prática de atos processuais eletrônicos, inclusive por meio de videoconferência, sustentações orais e depoimentos.

Nesse caminho, a Advocacia Geral da União (AGU) começou a desenvolver, no ano de 2012, o primeiro sistema de inteligência artificial aplicável à área jurídica, denominado Sapiens, quando não havia no âmbito da administração pública nenhum *software* que fizesse uso de Inteligência Artificial. Posteriormente, o Conselho Nacional de Justiça (CNJ) instituiu, por meio da Portaria 25/19, o Laboratório de Inovação para o Processo Judicial em meio Eletrônico, com o objetivo de criar uma rede de cooperação para a construção de um ecossistema de serviços de inteligência artificial, a fim de otimizar o trabalho e maximizar os resultados.

Nesse compasso, e sobretudo com a ajuda da inteligência artificial, alguns tribunais já começaram a desenvolver ferramentas para otimizar a atuação e a prestação jurisdicional. Atualmente, existem dezenas de robôs em operação, que realizam diferentes intervenções em várias áreas, automatizando parte do trabalho até então desenvolvido manualmente pelos homens, por exemplo (i) o *Victor* no STF, que se utiliza do mecanismo de aprendizado de máquina (*machine learning*) para realizar atividades de conversão de imagens em textos no processo digital; separação de documentos, classificação das peças processuais e identificação dos temas de repercussão geral de maior incidência e (ii) o *Sócrates* no STJ, ainda em fase de testes, que pretende realizar o exame automatizado do recurso e do acórdão recorrido, disponibilizando informações relevantes, prevendo a redução de 25% do tempo entre a distribuição e a primeira decisão no REsp (RIBEIRO; MAZZOLA, 2020, S.P.).

De fato, as novas tecnologias estão revolucionando a atividade jurisdicional. Plataformas *online* de resolução de disputas, softwares jurídicos para predição de resultados (jurimetria), a utilização de robôs, decisões por algoritmo, plenário virtual, arbitragem online enfim, são muitas questões instigantes que desafiam os operadores do direito. A digitalização dos processos judiciais e a automação de seus procedimentos são uma melhoria que todos os agentes envolvidos, as partes envolvidas em um litígio e seus advogados, bem como os servidores judiciários e os juízes (CASTRO, BOMFIM, TEIDER, 2020, p.220).

Em que pese os avanços obtidos no sistema de justiça a partir da utilização da inteligência artificial, os desafios ainda são muitos, pois se não sabe ao certo como as máquinas são (e serão) alimentadas, se os algoritmos serão revelados ao público, se haverá algum tipo de participação dos operadores do direito na construção de tais ordens sequenciais e, principalmente, se os robôs irão conviver em harmonia entre si e com os homens. São questões relevantes que serão sedimentadas com o tempo.

Samuel Meira Brasil Júnior (2020, p.12) alerta que é importante observar procedimentos de conformidade e transparência, garantindo o correto uso da tecnologia e que indivíduos afetados por decisões de algoritmos tenham direito de explicação. Além disso, há que se considerar a responsabilização: “E se a IA fizer algo errado?” Outro questionamento que se faz necessário diz respeito ao enviesamento não intencional dos programas de inteligência artificial e que já se encontra demandando interações e respostas do mundo jurídico.

Muitas pesquisas mostram que, à medida que as máquinas adquirem recursos de linguagem semelhantes às humanas, elas também estão absorvendo preconceitos profundamente arraigados, ocultos nos padrões de linguagem. Exemplos práticos de enviesamento (humanos) na tecnologia se encontram presentes em várias áreas profissionais.

Cumpra mencionar o caso do *COMPAS* (sigla em inglês para “*Correcional Offender Management Profiling for Alternative Sanctions*”), que é uma ferramenta de gerenciamento de casos e apoio à decisão de juízes desenvolvida e de propriedade da *Equivant* (anteriormente *Northpointe*), utilizada por alguns tribunais norte-americanos para avaliar a probabilidade de um réu se tornar um reincidente. Ele usa um algoritmo para avaliar o risco potencial de reincidência do infrator, possuindo escalas de risco para reincidência geral e violenta e por má conduta pré-julgamento. Os resultados obtidos pelo *COMPAS* são entregues aos juízes durante sentenças criminais com objetivo de auxiliá-los na tomada de decisão.

A intenção em utilizar sistemas como o *COMPAS* é tornar as decisões judiciais menos subjetivas - menos influenciáveis por erros humanos, voluntários ou não. Afinal, seriam esses algoritmos capazes de tornar as sentenças mais justas? Estariam eles livres de preconceitos e vieses? Um estudo desenvolvido pela *ProPublica* (ANGWIN; LARSON; MATTU; KIRCHNER, 2016, s. p.) apontou que o *COMPAS* prevê que os réus afrodescendentes terão riscos mais altos de reincidência do que realmente têm, enquanto os réus tidos como brancos são classificados para ter taxas mais baixas do que realmente fazem (a *Equivant* contesta esta análise).

Por certo, um algoritmo que reflete com precisão o mundo também reflete necessariamente os preconceitos da raça humana. Nesse sentido, o computador é pior que o humano, eis que não se limita a repetir de volta para nós os nossos próprios preconceitos (tais como as heurísticas e os vieses), mas exacerba-os (LIPTAK, 2019, s. p.). Ainda que possa superar eventuais conflitos de ordem processual e procedimental quanto a legalidade de sua aplicação, sistemas como o *COMPAS* sempre estarão sob os olhares de sua efetividade e capacidade de apresentar um resultado justo. O problema em questão é que, ao menos por enquanto, os algoritmos sobre os quais são projetados os sistemas normalmente carregam em si os vieses de seus desenvolvedores.

Sistemas que empregam aprendizado de máquina, por envolverem modelos matemáticos com parâmetros abertos, possuem uma dimensão de opacidade. A escala de dados (*big data*) e os modelos empregados muitas vezes tornam difícil a explicação do resultado de uma forma compreensível para o homem, com premissas, critérios acessíveis, argumentos e conclusões. Além disso, os programas podem sofrer vieses (discriminação) e falhas, advindas do design do algoritmo, da forma de treinamento do programa, da base de dados ou da execução da programação (CASTRO, BOMFIM, TEIDER, 2020, *passim*).

É importante considerar “se a Justiça está na utilização de um algoritmo que não tenha nenhum viés cognitivo (ou seja, se um algoritmo for considerado justo, o resultado será justo, porque o meio utilizado foi justo), ou se a Justiça da decisão está no resultado”. Ele defende que a primeira indagação a ser feita quando se está trabalhando com algoritmos decisórios é: “Como vamos construí-los?”. E afirma que o ideal é que fosse um misto dos dois casos, em que o meio fosse estabelecido sem um viés cognitivo e que também o resultado não apresentasse nenhum viés cognitivo” (JÚNIOR, 2020, p.12).

Com base nisso, pesquisadores têm buscado formas de mitigar erros no processo de decisão dos algoritmos, com o intuito de eliminar os vieses não intencionais.

De acordo com Samuel Meira Brasil Júnior (2020, p.12): os “*padrões aprendidos pelas máquinas não podem ser construídos com base em vieses cognitivos. Elas precisam estar aptas a fazer distinção de casos sem viés*”. O autor sugere, ainda, técnicas e métodos em evolução e que podem ser úteis para a redução ou eliminação dos vieses, tais como o *Deep Learning*, *web* semânticas (ontologias), mineração de texto, reconhecimento de padrão e *Data Analytics*.

A *machine learning* e a Inteligência Artificial são duas tecnologias que tomam decisões baseadas em dados. São milhares deles para processar, interpretar e aprender. Apesar disso, os algoritmos responsáveis por essas tarefas são programados por pessoas. Portanto, “sem pensar” ou sem intenção, os preconceitos raciais e de gênero dominaram o processo de aprendizado da IA/ML. Os sistemas não são capazes de pensar por si só, portanto, são tão tendenciosos quanto os seres humanos que os construíram - pelo menos por enquanto.

As pesquisas demonstram que essas pessoas que desenvolveram os algoritmos eram, até pouco tempo, majoritariamente homens brancos e de classe média. Ou seja, essa falta de diversidade se reflete também na maneira como essas tecnologias vão decidir.

A necessidade de diversas equipes de desenvolvimento e conjuntos de dados verdadeiramente representativos para evitar distorções nos algoritmos é uma das recomendações principais em um longo relatório do Comitê que analisa as implicações econômicas, éticas e sociais da inteligência artificial, e publicado pela Câmara Alta do parlamento do Reino Unido (UK PARLIAMENT, 2017, s. p.).

Ainda que talvez não seja possível afastar por completo os vieses humanos replicados nos algoritmos, a diversidade dos programadores se apresenta como uma forma de ao menos mitigar os potenciais impactos que a sua falta pode acarretar.

5 Considerações finais

O mundo passa por uma transição tecnológica disruptiva. O avanço exponencial da tecnologia proporcionada pela quarta revolução industrial acelera as transformações em escala global e está cada vez mais presente nos âmbitos pessoal e profissional.

Um relatório da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) apresenta uma análise sobre os postos de trabalho que podem ser automatizados em 32 países. O resultado é que 14% dos trabalhadores, cerca de 66 milhões de pessoas, correm sérios riscos de perderem emprego para as máquinas.

Por outro lado, a Deloitte já divulgou estudos que comprovam a máxima de que a tecnologia cria mais empregos do que os extingue. Uma nova pesquisa da PwC corroborou essa conclusão. De acordo com o estudo, nos próximos 20 anos, embora possa retirar cerca de 7 milhões de empregos no Reino Unido, a IA também deve criar 7,2 milhões de vagas. Ou seja, toda essa revolução no mercado de trabalho, extinção de postos, adaptações, novas formações e novos modelos de negócio podem gerar um aumento real de 200 mil empregos.

No ramo do Direito, não é diferente. A previsão é de que a IA atinja o mercado de trabalho já a partir de 2020, tendo em vista que as soluções que utilizam inteligência artificial (*Watson*, *COMPAS* e *Victor*) já conseguem realizar tarefas com eficiência e precisão muito maiores do que quando as mesmas tarefas são realizadas por profissionais do direito.

Contudo, os sistemas de inteligência artificial não são capazes de pensar por si só,

necessitando ao menos inicialmente de um algoritmo que indique o caminho das informações sobre um domínio e a forma de sopesar os dados para fazer uma previsão útil. Assim, mesmo os sistemas de IA podem ser tão tendenciosos quanto os seres humanos que os construíram – pelo menos por enquanto.

Na medida em que os seres humanos apresentam heurísticas e vieses (que podem determinar os comportamentos dos indivíduos), pesquisas mostram que à medida que as máquinas adquirem recursos de linguagem semelhantes às humanas, elas também estão absorvendo preconceitos profundamente arraigados, ocultos nos padrões de linguagem. Estudos recentes demonstram, por meio de testes, vieses psicológicos humanos nos sistemas de inteligência artificial e *machine learning*.

Os resultados das pesquisas sugerem que os algoritmos herdaram explicitamente os mesmos preconceitos sociais que as pessoas que os programaram. Acredita-se que, embora seja uma tarefa complicada, é possível que os sistemas de IA/ML possam ser melhorados para lidar com esse viés. Hoje, essa correção já está ocorrendo em empresas como a Google e os mecanismos de pesquisa da Amazon na *web*.

A maioria das grandes empresas está começando a analisar os vieses e a tentar buscar uma solução para o problema. Um dos caminhos que vem sendo apresentado é que a equipe de desenvolvimento dos algoritmos tenha uma grande diversidade e conte com supervisão constante. Também é indicado criar um órgão de supervisão e conformidade, permitindo que os sistemas sejam capacitados com base em uma autoridade moral.

Em que pese os grandes avanços proporcionados pela inteligência artificial, a humanidade ainda deve enfrentar os desafios de desenvolver algoritmos que não sejam influenciados por vieses humanos.

Referências

ALZAMORA, Geane Carvalho; SALGADO, Tiago Barcelos Pereira; MIRANDA, Emmanuelle C. Dias. Estranhar os algoritmos: Stranger Things e os públicos de Netflix. *Revista GEMInIS*, São Carlos, UFSCar, v. 8, n. 1, p. 38-59, jan./abr. 2017.

ANGWIN, Julia; LARSON, Jeff; MATTU, Surya; KIRCHNER, Lauren. *Machine Bias*. In: ProPublica. Disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Acesso em 10 dez. 2019.

BEZERRA, Juliana. *O que foi a Terceira Revolução Industrial*. Disponível em: <https://www.todamateria.com.br/terceira-revolucao-industrial/>. Acesso em 10 dez. 2019.

BRAGA, Carolina. A discriminação nas decisões por algoritmos: Polícia predativa. *in Inteligência Artificial e direito. ética, regulação e responsabilidade*. Coord. Ana Frazão e Caitlin Mulholland. São Paulo. Editora Thomson Reuters Brasil. 2019. p. 671-697.

BRASIL. MINISTÉRIO DA INDÚSTRIA, COMÉRCIO E SERVIÇOS. *A agenda brasileira para a Indústria 4.0: O Brasil preparado para os desafios do futuro*. Disponível em: <http://www.industria40.gov.br/>. Acesso em: 20 nov. 2020.

BRITANNICA, Escola. *Revolução Industrial*. Web, 2019. Disponível em: <https://escola.britannica.com.br/artigo/Revolucao-Industrial/481567>. Acesso em 10 dez. 2019.

-
- CARVALHO, Claudia da Costa Bonard de. *A inteligência artificial na Justiça dos EUA e o Direito Penal brasileiro*, 2018. In: Consultor Jurídico (ConJur). Disponível em: <https://www.conjur.com.br/2018-jun-10/claudia-bonard-inteligencia-artificialdireito-penal-brasil>. Acesso em 10 dez. 2019.
- CASTRO, Bruno Fediuk de; BOMFIM, Gilberto; TEIDER, Lucas Hinckel. A inteligência artificial aplicada ao direito e o problema dos vieses dos algoritmos. in **Direito, tecnologia e inovação: reflexões interdisciplinares**. Org. Camila Salgueiro da Purificação Marques e Miriam Olivia Knopik Ferraz. Belo Horizonte: Editora Senso, 2020. p. 207-229.
- CHIOVATTO, Milene. Watson, uso de Inteligência Artificial (AI) e processos educativos em museus. *Revista Docência e Cibercultura*. Rio de Janeiro, v. 3, n. 2, p. 217, mai./ago. 2019 - ISSN 2594-9004. Disponível em: <https://www.e-publicacoes.uerj.br/index.php/re-doc/article/view/40293>. Acesso em 10 dez. 2019.
- CORDEIRO, António Menezes; OLIVEIRA, Ana Perestrelo de; DUARTE, Diogo Pereira [coords]. *Fintech: novos estudos sobre tecnologia financeira*. Coimbra (Portugal): Almedina, 2019.
- DOMINGOS, Pedro. *The master algorithm: how the quest for the ultimate machine learning will remake our world*. Nova Iorque: Basic Books, 2015.
- ENGELMANN, Wilson; WENER, Deivid Augusto. Inteligência Artificial e Direito. in *Inteligência Artificial e direito. ética, regulação e responsabilidade*. Coord. Ana Frazão e Caitlin Mulholland. São Paulo. Editora Thomson Reuters Brasil. 2019. p. 149-179.
- FESTINGER, Leon. **A Theory of Cognitive Dissonance**. Stanford (Estados Unidos da América): Stanford University Press, 1957.
- FREITAS, Juarez. A hermenêutica jurídica e a ciência do cérebro: como lidar com os automatismos mentais. *Revista da AJURIS*. v. 40, n. 130, junho de 2013. p. 223-244.
- HARVARD LAW REVIEW, *Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing*. Recent Case: 881 N.W.2d 749 (Wis. 2016). MAR 10, 2017. Disponível em: <https://harvardlawreview.org/2017/03/state-v-loomis/> Acesso em 10 dez. 2019.
- JÚNIOR, Samuel Meira Brasil. Sem vieses cognitivos: como usar da Justiça? *Revista digital Expojud*. Edição 1. Ano 1. Ago. 2020. Disponível em: https://d335luupugsy2.cloudfront.net/cms/files/148371/1597712577RevistaExpojud_Ago2020_EDICAO_DE_LANCAMENTO.pdf. Acesso em: 20 nov. 2020.
- KEHL, Danielle; GUO, Priscilla; KESSLER, Samuel. *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, 2017. Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School. Disponível em https://dash.harvard.edu/bitstream/handle/1/33746041/201707_responsivecommunities_2.pdf. Acesso em 10 dez. 2019.
- KELNAR, David. *The fourth industrial revolution: a primer on Artificial Intelligence (AI)*. MMC writes, 2016. Disponível em: <https://medium.com/mmc-writes/the-fourth-industrial-revolution-a-primer-on-artificial-intelligence-ai-ff5e7ffcae1>. Acesso em: 06 dez. 2019

KLEINBERG, Jon. *The mathematics of algorithm design*. Cornell University. Disponível em: <https://www.cs.cornell.edu/home/kleinber/pcm.pdf>. Acesso em 10 dez. 2019.

KUNDA, Ziva. The Case for Motivated Reasoning. *Psychological Bulletin*, Washington (Estados Unidos da América), vol. 108, n. 3, p. 480-498, 1990.

LAGO, Lucas. Heurísticas, redes sociais e algoritmos. *Centro de Estudos Sociedade e Tecnologia da Universidade de São Paulo*. Volume 1, Número 6, Junho/2016. Disponível em: http://www.hu.usp.br/wp-content/uploads/sites/26/2017/03/V1N6pt_heuristica-final.pdf. Acesso em: 20 nov. 2020.

LIPTAK, Adam. *Sent to Prison by a Software Program's Secret Algorithms*. **The** New York Times. Disponível em: https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html?smid=fb-nytimes&smtyp=cur&_r=0. Acesso em 10 dez. 2019.

MICROSOFT. *The future computed: artificial intelligence and its role in society*. Redmond: Microsoft Corporation, 2018.

MORIN, Christophe. Neuromarketing: The New Science of Consumer Behavior. *Society*, vol. 48, issue 2, p. 131-125. Disponível em: <https://link.springer.com/article/10.1007/s12115-010-9408-1#citeas>. Acesso em 10 dez. 2019.

NORTHPOINTESUITS. Disponível em: http://www.northpointeinc.com/files/downloads/Northpointe_Suite.pdf. Acesso em 10 dez. 2019.

PARCHEN, Charles Emmanuel; FREITAS, Cinthia Obladen de Almendra; MEIRELES, Jussara Maria Leal de. As técnicas de neuromarketing nos contratos eletrônicos e o vício do consentimento na era digital. *Revistas Novos Estudos Jurídicos – Eletrônica*, vol. 23, n. 2, maio-ago 2018, p. 521-548.

RIBEIRO, Darci G. MAZZOLA, Marcelo. Processo e novas tecnologias: desafios e perspectivas. *Migalhas*, sexta-feira, 20 de novembro de 2020. Disponível em: <https://migalhas.uol.com.br/depeso/316523/processo-e-novas-tecnologias--desafios-e-perspectivas>. Acesso em: 20 nov. 2020.

REUTERS. *Amazon scraps secret AI recruiting tool that showed bias against women*, 2018. Disponível em: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showedbias-against-women-idUSKCN1MK08G>. Acesso em 10 dez. 2019.

SANTOS, Marco Aurélio da Silva. “*Inteligência Artificial*”; Brasil Escola. Disponível em: <https://brasilecola.uol.com.br/informatica/inteligencia-artificial.htm>. Acesso em 10 dez. 2019.

SCHWAB, Klaus. *A Quarta Revolução Industrial*. Tradução Daniel Moreira Miranda. São Paulo: Edipro, 2016.

SIMON, Hebert Alexander. Bounded Rationality. P. 266. In: EATWELL, John [et. al.]. *The New Palgrave Dictionary of Economics*. Vol. 1. Londres (Reino Unido): MacMillan Press, 1987.

TALEB, Nassim Nicholas. *A lógica do Cisne Negro: O impacto do altamente improvável*. 10 ed. Rio de Janeiro: BestBusiness, 2016.

TVERSKY, Amos; KAHNEMAN, Daniel. Judgment under Uncertainty: Heuristics and Biases. *Science*, New Series, Estados Unidos da América, vol. 185, n. 4157, p. 1124-1131, 27 set. 1974.

UOL: GIZMODO BRASIL. *Pesquisadora da Microsoft detalha perigos do viés algorítmico no mundo real*, 2017. In: Sítio oficial do UOL. Disponível em: <https://gizmodo.uol.com.br/perigos-vies-algoritmo/>. Acesso em 10 dez. 2019.

VERONESE, Alexandre, SILVEIRA, Alessandra, LEMOS, Amanda Nunes Lopes Espiñeira. Inteligência Artificial, mercado único digital e a postulação de um direito às inferências justas e razoáveis: uma questão jurídica entre a ética e a técnica. in *Inteligência Artificial e direito. ética, regulação e responsabilidade*. Coord. Ana Frazão e Caitlin Mulholland. São Paulo. Editora Thomson Reuters Brasil. 2019. p. 233-260.

XAVIER, Luciana Pedroso, SPALER, Mayara Guibor. Patrimônio de afetação: uma possível solução para os danos causados por sistemas de inteligência artificial. in *Inteligência Artificial e direito. ética, regulação e responsabilidade*. Coord. Ana Frazão e Caitlin Mulholland. São Paulo. Editora Thomson Reuters Brasil. 2019. p. 541-560.

WOJCIECHOWSKI, Paola Bianchi; ROSA, Alexandre Morais da. *Vieses da justiça: como as heurísticas e vieses operam nas decisões penais e a atuação contraintuitiva*. Florianópolis: Emodara, 2018.

UK PARLIAMENT. *Artificial Intelligence Committee - AI in the UK: ready, willing and able?* 2017. Disponível em: <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10002.htm>. Acesso em 10 dez. 2019.

ZIVIANI, Nivio. “Estado da arte”: como a IA pode impulsionar a Justiça?. *Revista digital Expojud*. Edição 1. Ano 1. Ago. 2020. Disponível em: https://d335luupugsy2.cloudfront.net/cms/files/148371/1597712577RevistaExpojud_Ago2020_EDICAO_DE_LANCAMENTO.pdf. Acesso em: 20 nov. 2020.